Disfluencies affect language comprehension: evidence from event-related potentials and recognition memory

Lucy J. MacGregor

A thesis submitted in fulfilment of requirements for the degree of

Doctor of Philosophy

to

School of Philosophy, Psychology and Language Sciences

University of Edinburgh

2008

Declaration

I hereby declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledgement is made in the text.

Lucy J. MacGregor

Abstract

Everyday speech is littered with disfluencies such as filled pauses, silent pauses, repetitions and repairs which reflect a speaker's language production difficulties. But what are the effects on language comprehension?

This thesis took a novel approach to the study of disfluencies by combining an investigation of the immediate effects on language processing with an investigation of the longer-term effects for the representation of language in memory. A series of experiments is reported which reflects the first attempt at a systematic investigation of the effects of different types of disfluencies on language comprehension.

The experiments focused on the effects of three types of disfluencies—*ers*, silent pauses, and repetitions—on the comprehension of subsequent words. Critical words were either straightforward continuations of the pre-interrupted speech or a repair word which corrected the pre-interrupted speech. In addition, the effects that occur when *er*, repetition, and repair disfluencies themselves are processed, were assessed.

ERPs showed that the N400 effect elicited in response to contextually unpredictable compared to predictable words was attenuated by the presence of a pre-target *er* reflecting a reduction in the standard difference where unpredictable words are more difficult to integrate into their contexts. This finding suggests that *ers* may reduce the extent to which listeners make predictions about upcoming words. In addition, words preceded by an *er* were more likely to be correctly recognised in a subsequent memory test. These findings demonstrate a longer-term consequence for representation which may reflect heightened attention during processing. Silent pauses did not affect the N400 but there was some indication of an effect on recognition memory. Repetition disfluencies did not affect the N400 or recognition memory. These findings demonstrate the importance of the nature of the disruption to speech. For all types of disfluent utterances, unpredictable words elicited a Late Positive Complex (LPC), possibly reflecting processes associated with memory retrieval and control as listeners attempted to resume structural fluency after any interruption.

Ers themselves elicited standard attention-related ERP effects: the Mismatch Negativity (MMN) and P300 effects, supporting the possibility that *ers* heighten attention. Repetition disfluencies elicited a right posterior positivity, reflecting detection of the disfluency and possibly syntactic reanalysis. Repair disfluencies elicited an early frontal negativity, possibly related to the detection of a word category violation, and a P600 effect, reflecting syntactic reanalysis. The presence of an *er* preceding the repair eliminated the early negativity, but had no effect on the P600 suggesting that *ers* may prepare listeners for the possibility of an upcoming repair, but that they do not reduce the difficulty associated with reanalysis.

Taken together, the results from the studies reported in the thesis support an account of disfluency processing which incorporates both prediction and attention.

Acknowledgements

I would like to acknowledge a number of funding sources. The research was funded primarily by an ESRC (1+3) Postgraduate Award (PTA-030-2003-01430). The ESRC also funded an Overseas Institutional Visit, which allowed me to spend three months at the Learning, Research and Development Center, University of Pittsburgh collaborating on research outside of my PhD topic and developing my research skills. The School of Philosophy, Psychology and Language Sciences, University of Edinburgh provided additional funding for travel to Stirling University, participant payment and conference attendance; a Grindley grant funded my attendance at the Architectures and Mechanisms of Language Processing Conference 2007; a University of Edinburgh Staff Research Development Grant awarded to Dr. Martin Corley partially funded Experiment 2 and my attendance at the Cognitive Neuroscience Society Conference 2005. The research presented in this thesis benefited from the feedback I received from researchers at these conferences.

I would like to thank Dr. Martin Corley for encouraging my initial interest in psycholinguistics and introducing me to disfluency research. His supervision, continual enthusiasm and support at all stages of the project have been invaluable.

All Event-Related Potential (ERP) data was collected in the Psychological Imaging Laboratory (PIL) at Stirling University. Through the PIL I have acquired technical skills and had the opportunity to develop my critical thinking within a supportive environment. For this, I am indebted to Professor David Donaldson.

I am grateful to all the members of the PIL for sharing their knowledge of ERPs and for exposing me to the field of memory research. Thanks in particular to Catriona Bruce for help with data collection (Experiments 3 and 4), and to Ric Sharp for the use of his AUTOERP script which enabled automatisation of much of the ERP data processing.

Thanks to Jos van Berkum for guidance on running ERP studies on spoken language comprehension, and for the opportunity to attend the Brain Mechanisms and Cognitive Processes of Discourse Workshop, Leiden, March 2007.

I thank members of the Edinburgh Disfluency Group, particularly Phil Collard, Dr. Robin Lickley, and Michael Schnadt for discussions about disfluencies.

Thanks to Corey McMillan and Suzy Moat for help with LaTeX.

Thanks to Dr. Manon Jones and Matt McGeehan for providing the voices of the experiments.

Matt McGeehan and Evelyn Newell helped with editing, and to them I am especially grateful.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Spontaneous speech is characterised by the presence of disfluencies. These are phenomena that interrupt the flow of speech and do not add propositional content to an utterance (Fox Tree, 1995, p.709): filled pauses (*er* and *um*), silent pauses, prolongations, repetitions and repairs. Disfluencies, which are the norm not the exception in everyday conversational speech, are the topic of this thesis.[1]

This thesis examines disfluent speech comprehension. To date, our understanding relies predominantly on a small number of experimental investigations using behavioural measures to assess comprehension. The studies suggest that disfluencies can affect both language processing and the resulting representation of a message, but the methodologies limit the conclusions that can be drawn. Further, no studies have attempted to systematically compare the effects of different types of disfluencies on language comprehension, nor have they used the same stimuli and participants in a combined investigation of the online effects during processing and the longer-term consequences for representation.

[1]Disfluencies associated with pathological speech behaviours such as stuttering, usually termed *dysfluencies*, are not considered here.

In the last twenty years there has been a wealth of research using Event-Related Potentials (ERPs) to investigate language comprehension, but almost no ERP studies investigating the comprehension of disfluent speech. ERPs are derived from Electroencephalograms (EEGs)—scalp-recorded voltages which reflect the neuronal activity of the brain—and are time-locked to events of interest. They can be used to selectively track specific aspects of cognitive processing with high temporal resolution, and therefore offer the potential to enhance our understanding of how disfluencies affect language comprehension.

This thesis presents a series of experiments designed to investigate the effects of disfluencies in speech on langugage comprehension. The experiments focused on the effects of three types of interruptions—*ers*, silent pauses, and repetitions—on the comprehension of subsequent words. In the majority of the experiments the critical words were straightforward continuations of the pre-interrupted speech; in one experiment the critical word was a repair word which corrected the pre-interrupted speech. Immediate effects on processing were assessed using response times in a Lexical Decision Task (LDT) and using ERPs. Longer-term consequences for language representation were assessed using a surprise recognition memory test. In addition, the effects that occur when *er*, repetition, and repair disfluencies themselves are processed, were assessed.

1.1 Thesis overview

Chapters 2–4 provide the background for the experimental work. In Chapter 2 I review the disfluency literature. I introduce the relevant terminology and concepts, and describe the conditions under which disfluencies are typically produced. The focus of the chapter is on how disfluent speech affects language comprehension and I discuss the existing evidence and evaluate the claims which have been made. Chapter 3 introduces ERPs. I explain how ERPs can be used as an experimental tool and discuss the potential and limitations of ERP data and methodology for investigating cognition. Chapter 4 introduces the ERP effects that have been most commonly observed during language comprehension. I discuss the evidence supporting an association between these effects and aspects of semantic and syntactic processing.

Chapters 5–9 comprise the experimental work and form the main body of the thesis. In Chapter 5, I describe the experimental designs and discuss design decisions. I then describe details about the methods. In Chapter 6 I describe two experiments which investigated the effects of the filled-pause *er* on the comprehension of subsequent words. *Ers* interrupt speech, delay the onset of subsequent information, and are a vocalisation with an unambiguously disfluent phonological form. The results demonstrate an impact of *ers* on language processing and representation and suggest that *ers* affect attentional processes, semantic integrative processes, and processes related to resuming structural fluency after a disruption. The following two chapters consider other types of disfluencies: silent pauses and repetitions. In Chapter 7 I describe an experiment which investigated the effects of silent pauses on language comprehension. Silent pauses interrupt speech, delay the onset of subsequent information, but in contrast to *ers* are not vocalised. The results show some similarities and some differences between the effects of silent pauses and *ers*. Chapter 8 presents an experiment which investigated the effects of repetitions on language comprehension. Repetitions also interrupt speech and delay the onset of subsequent new information but they are vocalisations of a lexical item and rendered disfluent only by the context in which they are encountered. Results show some similarities and some differences between the effects of repetitions and those of *ers* and silent pauses. In addition, effects of repetitions themselves are observed. Chapter 9 returns to *ers*, but in the context of repair disfluencies. I describe an experiment which investigated whether the effects of repairs on language processing are modified by the presence of an *er*: results suggest that they do. In addition, effects of *ers* themselves are observed. Chapter 10 draws together the main findings from the experimental work. I suggest what features of disfluencies may have driven the observed effects and describe a possible mechanism to account for the findings, which incorporates both prediction and attention.

Chapter 2

Disfluencies in speech

2.1 Introduction

To produce an utterance, speakers must first conceptualise the message they wish to convey. Translating this pre-verbal message into a linguistic form requires syntactic planning, retrieval of the appropriate words in the correct grammatical form, and specification of the articulatory and phonetic details. The resulting phonological form can then be articulated. Accompanying these three stages is a monitoring process: speakers monitor both their internal (covert) speech before it has been articulated and their overt speech (for explanatory purposes, I assume the basic structure of a Leveltian model of speech production: Levelt, 1983, 1989).

It is widely accepted that language production is an incremental process—speakers conceptualise and formulate their utterances online, while they are speaking (Marslen-Wilson & Tyler, 1980). It is therefore unsurprising that difficulties sometimes arise. Such difficulties, which can occur at any stage of the production process, result in disfluencies.

This chapter reviews the disfluency literature. The aim is to provide the theoretical background for the experimental work presented in the thesis and the review is necessarily selective. I introduce the various types of disfluencies and the terms by which they are known. I explain the conditions under which they are produced and their possible causes, and I briefly discuss how disfluencies affect listeners' metalinguistic judgements about speakers. The focus of the chapter is on how disfluencies affect listeners' language comprehension. There are few empirical studies investigating the comprehension of disfluent speech, but importantly they show that effects can be observed, at least within constrained situations. I discuss the main findings and highlight the limitations of the conclusions that can be drawn.

2.2 What are disfluencies?

Disfluencies are phenomena that interrupt the flow of speech and do not add propositional content to an utterance (Fox Tree, 1995, p.709). Disfluencies are an integral feature of spoken language affecting around six in every hundred words of spontaneous speech (this count excludes silent pauses: Bortfeld, Leon, Bloom, Schober, & Brennan, 2001; Fox Tree, 1995; Lickley, 1995; Shriberg, 2001). Because disfluent speech is the norm not the exception, the effects of disfluencies on language comprehension are of considerable relevance to psycholinguistics.

Disfluencies have been classified in various ways and there is some variability in the terminology, in part as a consequence of the diversity of research orientations (social psychological, linguistic, computer linguistic, psycholinguistic). The aim here is not to provide an exhaustive review but to capture the essence of the phenomena and to define the disfluency types and terms used throughout the thesis.

There are broadly six types of disfluent phenomena which may be produced during production difficulties:

1. **Filled pauses**: all types of hesitation devices such as *uh, er, ah, um, em, erm, mm.* Filled pauses are sometimes referred to as fillers.
2. **Silent pauses**: periods of unusually long silence.
3. **Repetitions**: repeated phonemes, words or phrases.
4. **Repairs**: changes to phonemes, words or phrases. Repairs involve the backtracking of speech to correct a speech error, qualify a part of speech, or change the meaning. Repairs are sometimes referred to as corrections, substitutions, replacements or revisions. If a repair restarts with no retracting of the pre-interrupted speech it is sometimes referred to as a deletion.
5. **Lexical fillers**: conventional words which are semantically redundant in the utterance, such as *I mean, y'know, basically, like.* Lexical fillers are sometimes classified as discourse markers or interjections, rather than disfluencies, and treated as having a different function.
6. **Prolongations**: syllabic lengthening, such as "the" pronounced *thee*, or "to" pronounced *tooo*.

The experimental work presented in the thesis focused on the filled pause *er*, silent pauses, repetitions, and repairs, and these are discussed in more detail below. Silent pauses can be fluent, reflecting the natural prosody of an utterance, or disfluent (Duez, 1985; Ferreira, 2007; Zellner, 1994) and because of the difficulty in distinguishing the two, particularly when they occur between clauses, silent pauses are often excluded from disfluency studies (Bortfeld et al., 2001) or conflated with filled pauses (e.g., Hawkins, 1971). Because of the scarcity of silent pause studies, silent pauses are discussed alongside those relating to filled pauses.

2.2.1 Filled and silent pause disfluencies

Filled pauses involve an interruption to speech and a delay followed by continuation of the pre-disfluency utterance, with no backtracking of the original utterances.

Filled pauses often occur with other disfluencies. For example, they may be preceded by a prolongation and followed by a silent pause (see 1).

(1) Go from thee ***uh*** right until you get to the end.

The phonological form of filled pauses depends, in part, on a speaker's accent. For example, the production of an *uh* in American English is equivalent to the *er* used by many British English speakers. In this thesis I make no distinction between *uh* and *er* and in discussing the literature refer to the term used by the relevant author.

2.2.2 Repetition disfluencies

Repetitions involve an interruption to speech which is followed by a repetition of some part of the pre-interruption utterance, usually a function word (e.g., Maclay & Osgood, 1959). The message is unaltered (see 2).

(2) Go from the **the** right until you get to the end.

Heike (1981) distinguishes between retrospective and prospective repetitions. This distinction, originally based on assumptions about the differing functions of the repetitions, is also supported by differences in durational and fundamental frequency properties between the two types (Shriberg, 1995; Plauche & Shriberg, 1999). Retrospective repetitions are considered to be a bridging device to re-establish fluency in speech after its disruption due to another disfluency such as a repair or a filled pause. In studies of disfluency effects they are often considered alongside and compared to repairs since both effectively resume fluency after an interruption. Prospective repetitions however, are considered to function like filled pauses, introducing a delay associated with speech production difficulties (for details of disfluency production, see section 2.3). Studies of the production and comprehension of disfluency do not typically distinguish between the two types, but unless otherwise stated, the term "repetition" in the thesis refers to prospective repetitions.

2.2.3 Repair disfluencies

Repairs are the most complex type of disfluency. They involve an interruption to speech and sometimes a pause, followed by new information which replaces some part of the pre-interruption speech and may alter the pre-interruption message. Levelt (1983) proposed a description of the structure of repairs (see also Nakatani & Hirschberg, 1994). Three major parts of a repair utterance are identified: the *reparandum*, the *edit interval* and the *repair* (see Figure 2.1). The reparandum is the original speech produced in error which will be corrected, the edit interval (sometimes called the disfluency interval) is the period immediately after the point where fluent speech is interrupted (interruption point) and may include a filled pause. The repair itself is the information which replaces the reparandum. In more complex repairs, the post-interruption continuation may include some retracing (repetition) of the original utterance in addition to the repair.

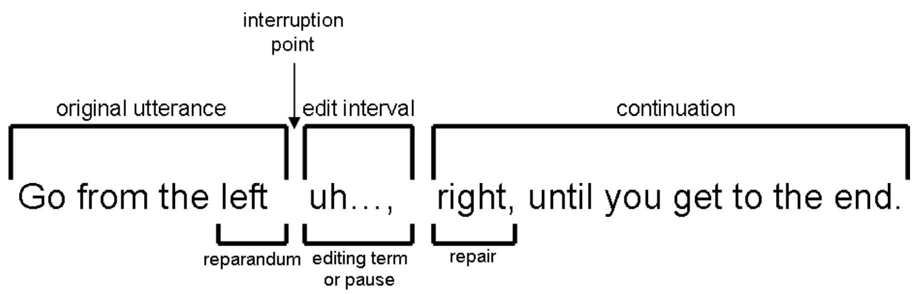

Figure 2.1: Structure of a simple repair with no retracing of the original utterance in the repair section. The reparandum is a single lexical item, "left", which is replaced by the repair word "right". The edit interval between the reparandum and the repair includes the filler *uh*. Adapted from Levelt (1983, p.43).

Levelt (1983) distinguished between overt and covert repairs. Covert repairs are changes to inner speech and are made by the speaker before articulation. They are a result of self-monitoring (see section 2.3). An editing term such as *uh* or a repetition may be indicative of a covert repair and introduces time for replanning (Blackmer & Mitton, 1991).

Overt repairs are corrections to mistakes in articulated speech. Unless otherwise stated, the term “repair” in the thesis refers to overt repairs. Further divisions of overt repairs have been made (Levelt, 1983), based, for example, on whether the repair changes the message to something *different* (D-repairs), corrects an *error* (E-repairs), or makes the speech more *appropriate* (A-repairs). For a more detailed review of repair subtypes, see Lickley (1994).

2.3 When are disfluencies produced?

Filled pauses, silent pauses, and repetitions typically occur when there is no appropriate phonological plan ready for execution, which may be due to conceptualisation or formulation difficulties. Conceptualisation difficulties occur if a speaker is uncertain about the message they wish to convey, perhaps because they lack the required semantic knowledge. Formulation difficulties may occur if there are multiple ways in which the information can be conveyed, for example if options exist at a word level (e.g., synonyms) or at a structural level. Difficulties may also arise if the specific word is hard to access (e.g., if it is of low frequency or low contextual probability) or if the syntactic structure of upcoming speech is complex and hard to construct.

Speakers monitor their speech both before and after it has been articulated (Levelt, 1983, 1989) which enables the detection and correction of mistakes. Overt repairs occur after the execution of an inappropriate phonological plan and indicate misconceptualisation or mis-formulation of an utterance which was not detected during monitoring of inner speech, or indicate mis-articulation (cf. Ferreira, Lau, & Bailey, 2004). If a mistake is detected in the speech plan before it has been executed, filled pauses and repetitions may be articulated as a by-product.

The possibility that different disfluencies reflect different cognitive processes has not been widely investigated and there is little research suggesting why a particular disfluency might be produced at a given time. It is also important to note that a single disfluency may reflect more than one process and serve more than one function simultaneously (Bortfeld et al., 2001). Repairs differ from filled pauses, silent pauses and repetitions because they correct an error, but it is less clear whether there are differences between the causes of other disfluencies.

Furthermore, the use of disfluencies varies between individuals. For example, slow speakers tend to use filled and silent pauses and repetitions and faster speakers tend to produce more repairs (Ferreira et al., 2004). Therefore differences between speakers may reflect individual differences or preferences as well as differences in production processes (Bortfeld et al., 2001).

Below I summarise the evidence showing in what circumstances speakers may produce filled pauses, repetitions, and repairs: I consider each disfluency in turn. I briefly consider the possibility that disfluencies are devices used explicitly by speakers (much like regular words) rather than merely symptoms of production processes.

2.3.1 Production of filled and silent pauses

Conceptualisation difficulties

Filled and silent pauses have been observed in situations where conceptualisation difficulties arise because speakers lack the semantic knowledge required to produce an utterance. Smith and Clark (1993) showed that speakers' answers to general knowledge questions incorrect when they were preceded by filled pauses than when they were fluent. Furthermore, using a subsequent Feeling Of Knowing (FOK) measure (Hart, 1965), speakers rated themselves as less likely to know the correct answer when they had been disfluent.

Filled pauses have also been observed in situations where conceptualisation difficulties arise because of ambiguity regarding the message to be conveyed. For example, in the Thematic Apperception Test where speakers describe cartoon pictures, those which had elicited a wider variety of themed responses in an independent study (and were therefore deemed more ambiguous) were associated with a higher rate of pauses (Siegman & Pope, 1966). Even within a constrained speech production task, message uncertainty can arise. In a maze description task, as the number of potential routes within the maze increased, filled pause rate increased (Christenfeld, 1994).

Formulation difficulties

Schachter, Rauscher, Christenfeld, and Crone (1994) used vocabulary size as an index of lexical options, and demonstrated a relationship between the numbers of different words used by academics in different fields and the numbers of filled pauses produced when lecturing (see also Schachter, Christenfeld, Ravina, & Bilous, 1991).

Evidence for an association between syntactic options and pauses comes from observations that filled and silent pauses tend to occur at clause boundaries (Boomer, 1965; Ford, 1982; Hawkins, 1971; Holmes, 1988), or before more complex syntactic structures (Maclay & Osgood, 1959). Because new clauses tend to occur with new "idea units" (Butterworth, 1975), it is assumed that planning typically takes place at clause boundaries. Thus the typical occurrence of pauses at clause boundaries is interpreted as evidence for their association with syntactic planning difficulties (although for data showing no differences in the numbers of filled pauses produced between-clauses and within-clauses, see Bortfeld et al., 2001).

Oomen and Postma (2001a) manipulated formulation difficulty experimentally using a dual-task paradigm. Speakers performing a picture storytelling task produced more filled pauses when they had to simultaneously perform a tactile-form recognition task assumed to reduce the resources available for speech formulation. It is worth noting that in addition to increasing planning difficulties the divided-attention introduced by the dual task presumably also reduced the availability of resources for self-monitoring. Given the claims that reduced self-monitoring reduces filler rate (section 2.3.1) it must be assumed that in the task described, any reduction in self-monitoring when attention was divided was a less important influence on filled pause production than the concurrent increase in planning difficulties.

Although pauses commonly occur between clauses, they also occur within clauses, in response to lexical retrieval difficulties. For example, pauses tend to occur before information considered to be of relatively low accessibility for the speaker, and of high informational content for listeners. Studies have found more pauses before low frequency words (Levelt, 1983), before content words (Goldman-Eisler, 1958a, 1958b; Maclay & Osgood, 1959; Martin, 1967), before discourse-new items (Arnold, Wasow, Losongco, & Ginstrom, 2000; Barr, 2001), before unfamiliar items (Arnold, Hudson Kam, & Tanenhaus, 2007), and before words with low contextual probability (Beattie & Butterworth, 1979; Cook, 1969; Goldman-Eisler, 1958a, 1958b; Tannenbaum, Williams, & Hillier, 1965) than before the relevant control words.

It is possible that within-clause and between-clause pauses are functionally distinct and reflect different aspects of the language production process. This proposal is supported by a fMRI study (Kircher, Brammer, Levelt, Bartels, & McGuire, 2004). Six participants were scanned whilst they described Rorschach inkblots (21 minutes of speech each). Relative to an articulation baseline, within-clause silent pauses were associated with an increased BOLD (Blood Oxygenation Level Dependent) response at regions previously implicated in lexical retrieval (superior and middle temporal gyri bilaterally) whereas between-clause pauses were not.

Self-monitoring

A number of studies have related the production of filled pauses to successful self-monitoring—the ability of a speaker to detect and correct mistakes made in the speech plan or in articulated speech. Three studies have attempted to manipulate self-monitoring experimentally and have observed effects on filled pause rate (Christenfeld, 1996; Postma, Kolk, & Povel, 1990; Postma & Kolk, 1992). In Christenfeld's (1996) study, speakers were required to synchronise their speech to a metronome. This was assumed to cause speakers to direct their attentional resources to the rhythm rather than the content of their speech and to reduce self-monitoring. The results showed a decrease in filled pause rate. In Postma et al.'s (1990) study, the importance of speech accuracy was emphasised to speakers. This was assumed to cause speakers to increase self-monitoring and the assumption was supported by a decrease in the speech errors that were made. There were no differences in the numbers of filled pauses, which meant there was an increase in the ratio of filled pauses to errors. In Postma and Kolk's (1992) study, speakers' auditory feedback was masked by noise and this was assumed to cause a reduction in the effectiveness of self-monitoring. Fewer filled pauses were produced.

Observational support for a relationship between filled pause production and self-monitoring comes from Christenfeld and Creager (1996), who observed the filled pause rates of speakers who had consumed varying quantities of alchohol. Alchohol decreases general self-awareness and the ability to attend to more than one task and it was assumed that it also decreases the extent or effectiveness of self-monitoring. Results showed a small, but significant, negative correlation between (approximate) alchohol intoxication and filled pause rate.

A relationship between filled pause production and successful self-monitoring can account for the inconsistent findings relating anxiety to filled pauses (Christenfeld, 1995). Anxiety manipulations can typically, but not reliably, increase a speaker's self-consciousness about their speech, which is assumed to increase self-monitoring. Christenfeld and Creager (1996) showed that when speakers' attention to their own speech was increased independently of anxiety (listeners heard an amplified version of their own speech whilst they were talking), the rate of filled and silent pauses to errors increased. When anxiety was increased, there was no increase in filled pause rate, although there was a (marginal) increase in the rate of repairs .

2.3.2 Production of repetitions

Formulation difficulties

Repetitions, like filled pauses, have been associated with formulation difficulties. A higher repetition rate has been observed at between-clause rather than within-clause boundaries (Holmes, 1988) and before noun phrases of greater complexity (Clark & Wasow, 1998), suggesting that syntactic construction or planning is an underlying cause (see also Oomen & Postma, 2001a). A higher repetition rate has also been observed for function than for content words, which may be because function words tend to occur at the beginnings of utterances and because they are high frequency and often short, making them easier to access and repeat (Clark & Wasow, 1998).

It has also been suggested that repetitions function similarly to filled pauses, for example providing speakers with more time with which to resolve production difficulties (Clark & Wasow, 1998; Maclay & Osgood, 1959). The possibility that repetitions can be used interchangeably with pauses, and thus may reflect similar processes, is compatible with studies showing that when speakers were instructed to avoid silent pauses but maintain speech rate, they produced more repetitions (Beattie & Bradbury, 1979), particularly of function words (Howell & Sackin, 2001). It is of course possible that these results reflect task-related processes involved with avoiding silent pauses which do not occur in everyday speech.

2.3.3 Production of repairs

Self-monitoring

Repairs are a clear indication of self-monitoring: they occur when a speaker has detected the production of an error and corrects it. Postma and Kolk (1992) reduced speakers' ability to self-monitor by masking the auditory feedback of their speech with noise, resulting in a reduction in the number of overt speech errors which were repaired.

2.3.4 Disfluencies: symptoms or signals?

A topic of some controversy within the disfluency literature relates to the function of filled pauses and repetitions: are they merely *symptoms* of production difficulties or intentional *signals* by a speaker to listeners used to convey information about their production state and to perform specific functions (Clark, 1994; Clark & Fox Tree, 2002; Fox Tree & Clark, 1997). Filled pauses and repetitions may be used, for example, as a device to hold the floor, informing listeners of the speaker's trouble, but indicating that they will resume speaking shortly and should not be interrupted (Maclay & Osgood, 1959). Smith and Clark (1993) suggested a difference between speakers' use of *ums* and *uhs*, with *ums* signalling a greater production difficulty and hence a longer delay, than *uhs*. It has been argued that silent pauses may reflect an unplanned discontinuity in speech; when a speakers knows of upcoming production difficulty, they can produce a filled pause (Wingate, 1984) or a repetition (Clark & Wasow, 1998) which maintains vocalisation. Filled pauses and repetitions may also be planned devices used to help restore continuity after an interruption due to a (covert) repair (Clark & Wasow, 1998). Alternatively, filled pauses and repetitions may be symptomatic of the increase in time required for re-formulation after a covert repair (Blackmer & Mitton, 1991).

A problem with the "disfluency-as-signal" view is that although disfluencies are produced more frequently in dialogue when there is a threat of interruption (Siegman & Pope, 1966), they do still occur in monologue. It can be argued that even in monologue, speakers have an audience in mind, but it is unlikely that filled pauses are used solely as a signal. The debate is interesting but disfluencies may of course convey information to listeners in the absence of any intentionality on the part of the speaker. The issue of intentionality is not considered further in this thesis.

2.3.5 Summary: the production of disfluencies

The distribution of disfluencies in speech is not arbitrary, but reflects an association between disfluencies and production difficulties. Filled pauses, silent pauses and repetitions typically occur when there are difficulties with the conceptualisation or formulation of speech and these disfluencies introduce a delay during which the difficulties can be resolved. Repairs occur when errors are detected in the pre-articulated speech plan (covert repairs) and these repairs may be marked by a filled pause. Repairs may also occur after articulation (overt repairs). In the remainder of the chapter I consider disfluencies from the perspective of listeners, focusing on the effects of disfluencies on language comprehension.

2.4 What are the effects of disfluencies on language comprehension?

Disfluencies might be expected to have no effects on language comprehension, because a number of studies show that listeners do not perform well in disfluency detection tasks (Bailey & Ferreira, 2003). For example, in transcription tasks, listeners fail to reproduce many filled pauses (e.g., Lindsay & O'Connell, 1995; Martin, 1967) and tend to displace within-clause filled pauses to between clauses (e.g., Duez, 1985; Martin, 1967; Martin & Strange, 1968). Performance is even worse for repetitions and repairs (e.g., Lickley, 1995). More details of disfluency detection can be found in Lickley (1994) and a description of a strategy which may be used by listeners to distinguish between repetition and repair disfluencies can be found in Levelt (1983). However, there is no reason to suppose that explicit detection of disfluencies is needed for disfluencies to have an effect on listeners.

When questioned, listeners cite production difficulties as the most likely cause of disfluences in speech (Fox Tree, 2002) and a number of studies have demonstrated effects of disfluencies on listeners' metalinguistic judgements. For example, speakers have been rated as less honest (Fox Tree, 2002), less intelligent (Christenfeld, 1995), less knowledgeable about their topic (Brennan & Williams, 1995) and to be having greater production difficulties (Fox Tree, 2002), when they produce filled and unfilled pauses than when their speech is fluent. Brennan and Schober (2001) showed that ratings of knowledge were lower and ratings of production difficulties were higher as the duration of disfluent speakers' pauses increased. Furthermore, ratings of knowledge were lower, and ratings of production difficulties higher, for filled pauses than for silent pauses of the same duration. The effects of disfluencies on listeners' metalinguistic representation of speech indicate listeners' sensitivity to disfluencies. But what are the effects of disfluencies on language comprehension?

Disfluencies might be expected to affect language comprehension because they interrupt listeners' perception of the flow of speech, introduce a delay, and usually cause ungrammaticality which may be a burden on syntactic and semantic processing. At what stages of language comprehension might effects be observed? If effects of disfluencies can be observed on listeners' final representation of an utterance then it can be assumed that disfluencies also affected the processing of the utterance, as it was heard. However, although effects of disfluencies on representation reflect probable differences in the way in which disfluent utterances were processed, the reverse is not true: if effects of disfluencies can be observed during listeners' processing of an utterance, this does not necessarily mean that there will be any longer-term consequences for the representation of the message.

I review the evidence from studies that investigated the effects of disfluencies on language comprehension. To date, most of these have focused on *er* disfluencies, and I therefore consider these disfluencies first. I then consider the effects of repetition and repair disfluencies. For each type of disfluency, I first discuss any studies that assessed listeners' final representation of disfluent speech and I then discuss the effects on processing.

2.4.1 Effects of filled and silent pauses

Filled pauses introduce an unambiguously disfluent delay into the speech signal which may be why most studies of disfluencies have focused on *ers*.

Effects of filled and silent pauses on representation

Bailey and Ferreira (2003) investigated the effects of filled pauses on listeners' comprehension of garden path sentences containing a subordinate-main ambiguity which resolved to the nonpreferred subordinate interpretation (see 3a–3c; the ambiguous head noun is indicated in bold). An offline grammaticality judgment task was used to assess listeners' final representation of the utterances, and to make an inference about the way the disfluent utterances had been processed.

(3a) While the man hunted the **deer** ran into the wood
(3b) While the man hunted the *uh uh* **deer** ran into the wood
(3c) While the man hunted the **deer** *uh uh* ran into the wood

Utterances were rated grammatical more often when an *uh uh* pause was placed before the ambiguous head noun (3b: 85%) than when it was placed after the ambiguous head noun (3c: 60%) and as often as the baseline fluent condition (3a:

83%). The basic pattern was replicated when the *uh uh* pause was replaced by a non-speech interruption (e.g., a cat meow, a doorbell ring, or a cough).

The similar results for filled pauses and environmental noises suggest that the effects may be caused by the time which the interruptions add to the signal: there is a greater delay before the disambiguation and hence a greater time during which the incorrect parse is held. Bailey and Ferreira (2003) also suggest that listeners may be sensitive to the distribution of filled pauses in speech, which tend to occur at the initiation of a major constituent such as a phrase boundary or complex noun phrase (Clark & Wasow, 1998; Ford, 1982; Hawkins, 1971). They suggest that an *uh uh* pause may provide a cue to the upcoming structure which effectively alters listeners' predictions.

Bailey and Ferreira's (2003) findings demonstrate that filled pauses can affect listeners' representation of an utterance. Although these effects reflect probable differences in the way in which disfluent utterances were processed, offline grammaticality judgements do not measure processing as it occurs.

Effects of filled and silent pauses on language processing

Fox Tree (2001) assessed the effects of filled pauses on immediate language processing using a word monitoring task (Marslen-Wilson & Tyler, 1980) where listeners must press a button in response to specific target words in spoken utterances. Response times to correctly identify targets were used as an index of the ease with which information can be integrated (semantically and syntactically) into the context at that point (Fox Tree, 1995; Fox Tree & Schrock, 1999; Marslen-Wilson & Tyler, 1980). Both English and Dutch listeners were faster to identify target words when they were preceded by an *uh* disfluency than when they occured in control utterances where the filled pause had been excised; by contrast, no differences were shown for *ums*.

Fox Tree (2001) suggests (see also Fox Tree, 2002) that the processing benefit that is demonstrated for words following an *uh* arises because *uhs* heighten listeners' attention to upcoming speech (for a similar account of *oh*, see Fox Tree & Schrock, 1999). She argues further that the longer delay created by *ums* than *uhs* and the longer pause which typically follows (Clark & Fox Tree, 2002; Smith & Clark, 1993), makes heightening attention following *um* not a useful strategy for listeners preparing to process upcoming words (although note that post-filler pause differences were only observed between *uh* and *um* for the Dutch stimuli). Fox Tree (2002) assumes filled pauses are used and planned for by the speaker like regular words, but her arguments have relevance too for accounts which do not make such assumptions.

A concern about Fox Tree's (2001) findings is that response times to correctly identify targets following an *uh* were compared to response times to targets in utterances where the *uhs* had been excised. Because of periods of silence that had occurred before and after the *uh*, the control condition included a silent pause (of 704ms in the English sentences and 380ms in the Dutch sentences) before the target and was therefore not a fluent control. The faster response times to post-*uh* targets may therefore reflect either a processing benefit for these words or a processing hindrance for post-silence words and would be better compared to response times to words in a fluent condition.

Further experiments that suggest that filled pauses affect attentional processes have been carried out by Brennan and Schober (2001). They assessed language processing using a referential communication task. Listeners followed instructions which were spontaneous utterances selected from those produced by a male native English speaker during a different task (using a procedure adapted from Van Wijk & Kempen, 1987). Participants were required to press a button to select a referent from a set of geometric objects displayed on a computer screen. Response times and accuracy were used as a general index of the ease with which the target word (targets were repairs) could be processed. Response times were faster to targets that were preceded by an *uh* (4a) than to those without (4b) and this was not at the expense of accuracy. In fact, there were fewer errors to targets preceded by an *uh* than to those without.

(4a) move to the yel- *uh* **purple square**

(4b) move to the yel- **purple square**

(4c) move to the yel- . . . **purple square**

Unlike Fox Tree (2001), Brennan and Schober (2001) controlled for both the phonological form and the duration of the filled pause by including conditions where the *uh* had been excised to form an utterances without a pause and where it was replaced by a silent pause of the same length (4c). Response times and error rates were unaffected by type of pause suggesting that the disfluency effects were driven by the time which the filled pause added to the signal, not by the phonological form.

These studies demonstrate observable effects of filled and silent pauses on processing, and suggest that pauses may affect attentional processes. However, it can also be argued that word monitoring and button-pressing tasks do not require comprehension of the utterance for successful performance (Bailey, 2004), and that the results are task specific. Furthermore, language comprehension is incremental—words are semantically and syntactically integrated into their context as they are heard (Altmann & Kamide, 1999; Marslen-Wilson & Welsh, 1978; Sedivy, Tanenhaus, Chambers, & Carlson, 1999; Traxler, Bybee, & Pickering, 1997), but these tasks do not assess the comprehension of an utterance as each new word is heard.

Work by Arnold and colleagues (Arnold, Tanenhaus, Altmann, & Fagnano, 2004) addresses these issues in an attempt to refine an account of how listeners respond to disfluency in real time. Using a visual world paradigm (first demonstrated by

Cooper, 1974) which was developed by Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995), participants' eye movements to depictions of four objects on a computer screen were monitored as they responded to auditory instructions to move the objects with a computer mouse. Fixations are thought to reflect lexical access and can therefore be used to track the time course of the processing of continuous speech.

The presence of the filled pause disfluency (*thee uh*) before the mention of the target object increased the probability of an initial eye movement to an object which had not been previously mentioned (discourse-new); in contrast, when the instructions were fluent, participants were more likely to look first at a previously mentioned (discourse-old) object. There is increasing evidence that listeners make online predictions during language comprehension (e.g., Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003; Kamide, Scheepers, & Altmann, 2003; Knoeferle, Crocker, Scheepers, & Pickering, 2005; Pickering & Garrod, 2007; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005), and Arnold et al.'s (2004) findings suggest that filled pauses may lead listeners to update their predictions about upcoming words. Specifically, listeners appear to predict that discourse-new items are more likely to be mentioned following a disfluency. One possibility is that listeners attribute the disfluency to the speaker's production difficulties associated with the upcoming mention of a hard-to-access (discourse-new) object. Alternatively, listeners may be sensitive to the distributional occurrence of disfluency in natural speech: speakers are more likely to be disfluent before mentioning a discourse-new object (Arnold, Fagnano, & Tanenhaus, 2003; Arnold & Tanenhaus, 2007).

In a similar study, listeners' assumption about a speaker's production difficulty was manipulated through object familiarity (Arnold et al., 2007). Participants' eye movements were monitored while they viewed pictures of familiar objects with well-known and easily accessible names (e.g., an icecream cone) and unfamiliar, difficult-to-describe squiggly shapes. When the instructions were disfluent, listeners looked to the unfamiliar object more frequently than to the familiar object (for further evidence that filled pauses direct listeners' attention towards unfamiliar referents, see Barr, 2001). When speakers were fluent, listeners looked to the familiar and unfamiliar objects with equal frequency which does not replicate the preference for the more accessible object in Arnold et al. (2004). Moreover, because the frequency of looks to familiar objects when speakers were fluent was similar to the frequency of looks to unfamiliar objects when the speakers were disfluent, it may be that the disfluency *reduced* expectation for the mention of the more accessible object, rather than *increased* the expectation for the mention of the less accessible object.

Arnold et al. (2007) also investigated whether listeners' attribution of the cause of the disfluency could alter the effects of disfluency. Attribution was assessed in a post-experiment questionnaire. When listeners were told that the speaker was agnosic and thus had difficulty in naming objects, they attributed disfluency to the speaker's agnosia. The results showed that listeners no longer used disfluency in their predictions, suggesting that listeners had made the inference that disfluency was not a reliable cue for the upcoming mention of the unfamiliar object. In an experiment where listeners attributed a speaker's disfluency to distraction by construction noises and beeps (present when the speaker was being recorded) rather than to language production difficulties, the results suggested that disfluency continued to affect listeners' predictions.

The results from these two visual world studies (Arnold et al., 2004, 2007) demonstrate an effect of filled pauses on linguistic prediction. Because the visual world paradigm provides a continuous measure of lexical activation, it is a useful way of assessing the cognitive processes underlying language processing. However, the paradigm has limitations. Listeners are presented *a priori* with a restricted set of candidate images: in Arnold et al. (2004, 2007), these (four) items provided the sentence completions. Such constraints are rare in natural situations. It is possible that the changes to predictive processes observed by Arnold et al. (2004, 2007) are a result, at least in part, of the availability of a potential referent. In other words, although listeners can predict specific completions from a limited choice, it is unclear whether changes to predictive processes would also be observed in the absence of a constrained set of referents, as in natural speech. One possibility is that disfluency would no longer affect prediction, and any standard predictions would be unaffected. Alternatively, disfluency would affect predictive processes, and lead to a reduction in the extent to which any standard specific predictions are made based on the available context. For example, following a highly constrained utterance, listeners may no longer expect the most predictable word.

2.4.2 Effects of repetitions

There are only two studies that have investigated the effects of repetition disfluencies on language comprehension: one that investigated the effects on the processing of subsequent words (Fox Tree, 1995) and one that investigated the effects that occur when the repetition disfluency itself is processed (McAllister, Cato-Symonds, & Johnson, 2001). No study has investigated whether effects of repetitions can be observed on the final representation of an utterance.

Effects of repetitions on language processing

Fox Tree (1995) investigated the effects of repetitions (typically of multiple words) using a word monitoring task (see section 2.4.1). Response times to correctly identify target words were unaffected by the presence of a preceding repetition, relative to utterances where the repetitions had been excised or excised and replaced by a pause. There is therefore no evidence for an effect of repetitions on processing, which is in contrast to the faster response times to targets following an *er* which were observed in a similar study (Fox Tree, 2001, section 2.4.1) showing that *er* led to faster response times.

Using Event-Related Potentials (ERPs), McAllister et al. (2001) showed effects of repetitions themselves. ERPs provide an online measure of processing (see chapter 3) and have been widely used to investigate aspects of language comprehension (see Chapter 4), but as yet, have not been fully exploited as a tool for investigating disfluency processing. In McAllister et al.'s (2001) study, participants listened to recordings of speech which were immediately followed by the presentation of a tangram picture, and had to decide whether or not the picture matched the description. Some of the descriptions included repairs and some included repetitions. Relative to the first occurrence of the word, the repeated words elicited a negativity in the time period 300–500ms after the word onset which the authors interpret as an N400 effect associated with an increase in the difficulty of semantically integrating repeated words into their context (section 4.3.1). For the repair results, see section 2.4.3. However, statistically, the difference was only marginally significant. It should also be noted that the authors only present analyses from midline electrodes and do not show the scalp distributions of the ERP effect.

2.4.3 Effects of repairs

Repairs usually result in a syntactic violation and therefore the processing of repairs might be expected to cause syntactic processing difficulties and require additional resources to be resolved.

Effects of repairs on language representation

Lau and Ferreira (2005) demonstrated effects of the reparanda of repair utterances on the final representation of garden path sentences using offline grammaticality judgements. Participants heard sentences which were either fluent (examples 5a and

5b) or contained verb-verb repairs (examples 5c–5d). The target verb (in bold) was either an unambiguous past participle and required a reduced relative continuation (5a) or ambiguous between a favoured simple past tense which results in a garden path to the ultimately incorrect main verb continuation, and the past participle (5b). In the repair conditions, the ambiguous target (repair) verbs were preceded by reparanda verbs (in italics) which were either unambiguous and took the correct past participle form (5c) or were ambiguous and would at first be assigned the favoured main verb form (5d).

(5a) The little girl **chosen** for the role celebrated with her parents and friends

(5b) The little girl **selected** for the role celebrated with her parents and friends

(5c) The little girl *chosen*, uh, **selected** for the role celebrated with her parents and friends

(5d) The little girl *picked*, uh, **selected** for the role celebrated with her parents and friends

Grammaticality judgements were high (mean over 90% over all stimuli). As expected, for fluent utterances, grammaticality judgements were higher when the critical verb was unambiguous (5a) than ambiguous (5b). Critically, in disfluent utterances, the ambiguity of the reparanda affected the grammaticality judgement: grammaticality judgements were higher when the reparanda were unambiguous (5c) than ambiguous (5d). Thus the reparanda affected the final representation of the utterances, even though the repaired versions were the same. The results reflect probable differences in processing, but as already mentioned (section 2.4.1) grammaticality judgments do not measure processing as it occurs.

Effects of repairs on language processing

Fox Tree (1995) investigated the processing of repairs using a word monitoring task (see section 2.4.1). Repairs were effectively deletions (false starts); the reparanda, comprising multiple words, were replaced with a different message with a different syntactic structure (D-repairs). There were no editing terms such as an *uh* during the edit interval. Response times to correctly identify repairs (the first word in the continuation, which was often a function word) were slower than for control words in fluent utterances where the reparanda had been excised, or replaced by a silent pause of the same duration. This suggests that the processing of repairs incurs a cost, perhaps reflecting the additional resources or time required for semantic or syntactic integration, although response times do not enable conclusions to be drawn about the specific aspects of language processing which are affected.

It has been suggested that filled pauses, during the edit interval of repairs, aid the processing of repairs by helping listeners to detect the presence of repairs. For example, Levelt (1989, p.481) suggests that an editing expression like *er* or *uh* may "warn the addressee that the current message is to be replaced". Hindle (1983) suggests that an *er* is an "edit signal".

Brennan and Schober (2001) investigated the processing of repairs and specifically the effect of an *uh* at the edit interval, using a referential communication task (see section 2.4.1). They tested two hypotheses generated from Levelt (1989): first, that the interruption of a word is a cue that the word is to be replaced and so a mid-word repair should be easier to process than a between-word repair, and secondly, that an *uh* is a cue for a repair and so a mid-word repair with *uh* should be easier to process than mid-word repair without *uh*.

The results provide partial support for Brennan and Schober's (2001) first hypothesis that word interruptions cue repairs: listeners' responses were more accurate for mid-word repairs than for between-word repairs. However, the experiment was confounded by the limited set of referents: because the targets were repair words, the reparanda were particularly informative in reducing the number of potential referents. When the number of referents was increased from two to three and so reduced the informativeness of the reparanda as a cue, the advantage afforded by the filled pause, although still present, was attenuated. Furthermore, response times did not differ between mid-word and between-word repairs (although responses to mid-word repairs were faster numerically: $F1$ significant, $F2$ not significant). These results are compatible with a *post hoc* analysis of Fox Tree's (1995) data demonstrating no correlation between the response latencies to targets and the length (in terms of number of words and duration) of the reparandum. Also, listeners did not make more errors in response to repetition targets than to repairs, providing no evidence that interruptions cue repair disfluencies specifically.

The results also provide partial support for the second hypothesis that *uh* filled pauses cue repairs: response times were faster and accuracy higher for mid-word repairs which included an *uh* at the edit interval than for those without. However, the effects did not depend on the phonological form of the filler, but on the time introduced by the pause (section 2.4.1).

Brennan and Schober (2001) suggest that pauses and word fragments in reparanda may be cues that there is a problem with speech, and prepare listeners for the possibility of a repair. The edit interval may give listeners time during which the incorrect message can be cancelled (although because of the limited number of referents the incorrect information was probably more obvious in Brennan and Schober's (2001) task than in conversational speech). The increase in stress often observed on repair words (relative to fluent controls and words in the reparandum, e.g., Howell & Young, 1991; Levelt & Cutler, 1983), may help listeners identify the disfluency as a repair rather than a repetition (see also Levelt, 1983; Fox Tree, 1995), locate the repair onset and suppress information contained in the reparandum.

Ferreira et al. (2004) proposed that the processing of repairs may employ mechanisms similar to those involved in the recovery from garden path sentences because both appear to require structural reanalysis. This claim is supported by McAllister et al.'s (2001) study which used ERPs to investigate the processing of repairs and repetitions (see section 2.4.2). Repairs were effectively deletions (false starts); the reparanda, comprising multiple words, were replaced with a different message with a different syntactic structure (D-repairs). There were no editing terms such as an *uh* during the edit interval. Relative to fluent control words, repairs elicited a positivity over posterior scalp locations in the time period 0–800ms after the word onset, which the authors interpret as a P600 effect associated with syntactic reanalysis and repair (section 4.4). It should be noted that the authors only present analyses from midline electrodes and do not show the scalp distributions of the ERP effect, and so the findings require replication.

2.4.4 Summary: the effects of disfluencies

Few studies have investigated the effects of disfluencies on language comprehension and there has been no systematic comparison between the effects of different types of disfluencies. Effects have been observed on processing and on the final representation of the message. The majority of the research to date has focused on the effects of filled pauses. Response times suggest that filled pauses may affect attentional processes and eye gaze measures collected during visual world experiments suggest that filled pauses may affect predictive processes. It is not clear what features of filled pauses may be driving the effects. For example, the effects may occur because of the delay which filled pauses add to the signal, but the nature of the disruption may also be important. The limited evidence on repetitions suggests that they have limited (or no) effects on comprehension. Repairs may cause processing difficulties which can be alleviated by the presence of an *er*.

2.5 Conclusions

Disfluencies are an integral feature in spoken language and their distribution within speech is not arbitrary. Filled pauses, silent pauses and repetitions are observed in response to difficulties with the production of upcoming speech (conceptualisation, planning, and lexical retrieval). Repairs occur when the production process has failed and an error has already been articulated. Repairs sometimes include a filled pause during the edit interval before the production of the repair word.

Given the prevalence of disfluencies in speech, the impact of disfluencies on language comprehension is of considerable relevance to psycholinguists. There have been few experimental investigations, but behavioural studies show that effects can be observed on processing, and also on the final representation of speech. The majority of previous work has focused on the effects of filled pauses. Response times suggest that filled pauses may heighten listeners' attention to upcoming words, and eye gaze measures recorded during visual world experiments suggest that filled pauses may alter listeners' predictions about upcoming words. However, it is possible that the engagement of attentional and predictive processes suggested by these studies are a result of the tasks and the situational contexts in which the disfluencies were heard. The response time studies did not require that participants comprehended the utterances to complete the task, and it is possible that participants were strategically processing the utterances. The visual world experiments presented participants with a set of four referents from which one would always be mentioned. It is possible that in natural speech where the context does not constrain the possibilities to such an extent, filled pauses would not affect listeners' predictions. Alternatively, filled pauses may reduce the extent to which the context is used to make specific predictions. The limited evidence on other types of disfluencies suggests that repetitions may have limited (or no) effects on comprehension, but that repairs cause processing difficulties which may be alleviated by the presence of an *er*.

This thesis took a novel approach to the study of disfluencies, by combining an investigation of the effects on language processing, with an investigation of the longer-term effects on the representation of the message in memory. An attempt was made to address some of the limitations of the behavioural methodologies, by using ERPs to assess processing. ERPs provide a way to track language processing online, while listeners comprehend disfluent utterances without engaging in a secondary task. In addition, ERPs are sensitive to semantic and syntactic processes and therefore offer the potential to determine which aspects of language comprehension are affected by disfluencies. A surprise recognition memory test was used to assess whether any effects of disfluencies during processing would have longer-term consequences for the representation of the message which would be observable up to 55 minutes after initial processing. Furthermore, by comparing the effects of different types of disfluencies across similar experiments, an attempt was made to determine which features of disfluencies may be driving the effects which were observed. Before the experimental work is presented (Chapters 5–9), I describe the ERP methodology and its application to questions about language processing.

Chapter 3

Event-Related Potentials

3.1 Introduction

Event-Related Potentials (ERPs) are derived from scalp-recorded electroencephalograms (EEGs), a measure of differences in electrical potentials (voltages), Electrical potentials associated with sensory, cognitive and motor events—ERPs—are isolated by averaging the activity elicited in response to multiple examples of the event of interest (Luck, 2005).

The most often cited benefit of the ERP methodology for investigating aspects of cognition is its high temporal resolution. Data are acquired every 4−5 milliseconds which enables the continuous online tracking of cognitive processing. To exploit the methodology to its full, however, it is important to understand that the neural origins of ERP data constrain the inferences that can be drawn from their analysis.

The chapter summarises the neural source of the ERP signal, discusses the recording and processing procedures which enable the extraction of clean data, and explains the analysis techniques which permit specific inferences to be made about a cognitive process. The aim is to give an overview of the potential and the limitations of ERP methodology and data for investigating cognitive processes.

3.2 Neural origin of ERPs

The neuron doctrine, one of the fundamental tenets of modern neuroscience, is that neurons form the structural and functional basis of the nervous system (Cajal, 1909-1911). Neurons process and transmit information in the form of electrical potentials and communicate with each other via chemical and electrical synapses. Neuronal activity produces two types of electrical potentials—action potentials and postsynaptic potentials—which result from ion flow across membranes. The focus in this section is on the relationship between the generation of these electrical potentials (neuronal electrogenesis) and their detection at the scalp, and therefore only limited details about the ionic basis of neuronal activity are described (for more detail, see Kandel, Schwartz, & Jessell, 2000). Figure 3.1 shows the basic structure and features of a typical neuron.

3.2.1 Electrogenesis

Neurons maintain a negative electrical potential (of about -60mV) across their membranes created by the difference in ion composition of intracellular and extracellular fluids. This is due to the selective permeability of cell membranes to cellular ions, in particular potassium [K^+] which predominates on the inside and sodium [Na+] which predominates on the outside. An external stimulus causes [Na^+] channels to open, raising the membrane potential which depolarises the cell (making it more positive) and initiates a positive feedback process. If depolarisation raises the membrane potential above a critical threshold an action potential, also called a spike, is generated. Thus action potentials are "all or none". An action potential is a transient change of about 100mV in electrical potential across a cell membrane

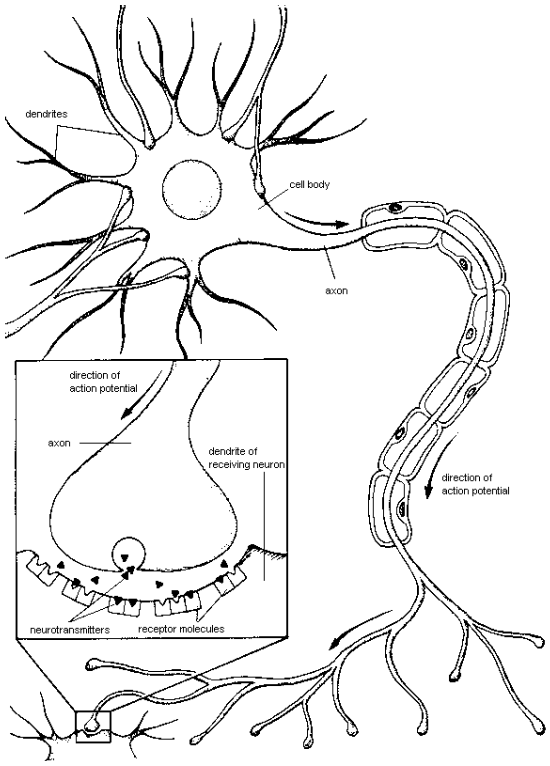

Figure 3.1: The structure of a typical neuron which comprises dendrites, a cell body (soma) and an axon. Action potentials propagate along the axon via a wave of depolarisation reulting from ion flow across the membrane. Information passes between connecting neurons at synapses. The arrival of an action potential at the synapse causes release of neurotransmitters from the presynaptic cells which bind to receptors in the postsynaptic cells. This causes opening or closing of ion channels, leading to voltage change in the post-synaptic membrane which results in (excitatory or inhibitory) postsynaptic potentials. Diagram adapted from `www.pfizer.com/brain/images/neuron\_large`.

lasting about 1ms. The depolarisation is followed by hyperpolarisation making the cell membrane more negative than the resting potential. An action potential propagates from dendrites, along the neuronal axon via the sequential opening and closing of voltage-gated ion channels and passes onto subsequent neurons via synapses, the junction connecting neurons. When an action potential reaches a synapse, ion channels open allowing an influx of calcium ions [Ca^{2+}]. This causes the release of neurotransmitters from pre-synaptic cells which bind to receptors on the post-synaptic cell membranes. The subsequent opening or closing of ion channels (depending on whether the neurotransmitter/synapse is excitatory or inhibitory) results in a graded change in the electrical potential across the membrane known as a post-synaptic potential lasting tens to hundreds of milliseconds. If the neurotransmitter is excitatory, its release causes an increase in permeability to [Na^+] at the apical dendrites. An inward flow of current depolarises the membrane and if depolarisation reaches a certain threshold, an action potential will be generated in the post-synaptic cell. Inhibitory neurotransmitters hyperpolarise the post-synaptic cell. Importantly, the negativity in the extracellular fluid of the apical dendrite and positivity in the cell body (and basal dendrites) form a dipole. Although this voltage created by a single neuron is tiny, under certain conditions, the dipoles from multiple neurons will summate to produce a voltage large enough to be recorded at the scalp, and it is these dipoles which form the basis of scalp recorded EEGs and ERPs.

Action potentials can be recorded from single electrodes ("single-unit" recordings) or from large populations of neurons ("multi-unit recordings") and are typically measured by in-vivo recordings. They are rarely detected by scalp electrodes, although an exception is when they occur in cortical structures close to the scalp (Allison, Wood, & McCarthy, 1986; Wood, 1987; Wood & Allison, 1981). This is because of the timing of action potentials and the physical arrangement of axons. As discussed, the propagation of an action potential comprises rapid transient depolarisation and hyperpolarisation along an axonal membrane. If two action potentials are propagated simultaneously along two parallel axons, their voltages will summate, which can be recorded by surface electrodes. However, if the action potentials are slightly asynchronous, which is usually the case, the depolarisation of one will coincide with the hyperpolarisation of the other and thus they cancel each other out to an extent, reducing the voltage which can be recorded at the surface.

Post-synaptic potentials can also be recorded in-vivo, but only from large populations of neurons (local field potentials). However, they can be recorded at the scalp, because their duration is longer that that of action potentials.

For the summation of post-synaptic dipoles to occur, a sufficient number of neurons must produce dipoles approximately simultaneously and be spatially aligned in an "open field", for example in parallel. This allows electrical activity to be detected outside of the neuronal population (e.g., at the scalp). Summation occurs maximally in cortical pyramidal cells, where neurons are aligned perpendicular to the surface of the cortex (Kutas & Dale, 1997). As the orientation between two dipoles increases beyond 90 degrees, the signals will cancel to an increasing extent with complete cancellation at 180 degrees where the positive dipole of one neuron is aligned with negative dipole of another neuron. This means that because of the folds of the cortex, not all synchronous dipoles of cortical neurons will summate. In a "closed field", cell bodies are clustered together with their dendrites extending outwards. Even in synchronous depolarisation of these cells, the dipoles cancel each other out producing an electrical potential of zero outside the structure (see Figure 3.2). The neurons in the thalamus and midbrain nuclei are arranged in such a way as to be undetectable by surface electrodes.

The constraints discussed above highlight the selectivity with which EEG reflect the neuronal activity of the brain. This selectivity means that failure to detect a difference in scalp recorded EEG activity between two experimental conditions does not necessarily imply identical neuronal activity (and by inference cognitive processing) in the two cases. Null results should therefore always be interpreted with caution (Rugg & Coles, 1995; Kutas & Dale, 1997; L. J. Otten & Rugg, 2005).

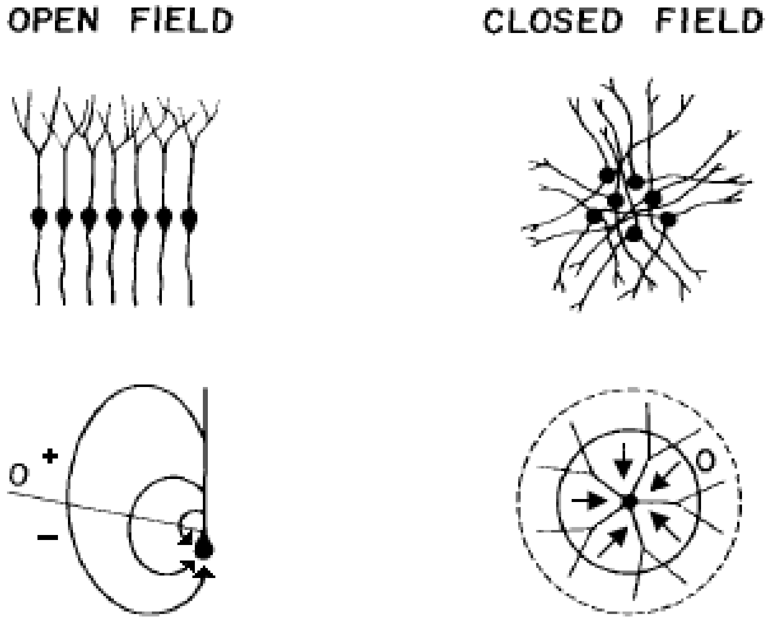

Figure 3.2: Predicted current flow and electrical potential fields produced by synchronous depolarisation of neurons aligned in parallel (open field) or in a group with cell bodies clustered in the centre and dendrites spreading radially (closed field). Diagram adapted from Allison et al. (1986)

3.2.2 Volume conduction

As discussed in the previous section, scalp-recorded EEGs reflect voltages generated primarily by the postsynaptic activity of large populations of neurons (estimated minimum of 1000–10,000, Luck, 2005; Kutas, Van Petten, & Kluender, 2006) which are synchronously active and whose dipoles are spatially aligned. Voltage generated by this subset of neuronal activity is propagated to the scalp because the brain, the skull and the scalp act as volume conductors.

Propagation of the electrical activity depends on the position and orientation of the generators and on the shape and conductivity of the brain, skull and scalp. The low conductivity of the skull attenuates the signal and also induces lateral spreading (Koles, 1998). Together with the differences in head size and shape between participants, these factors render identification of the neuronal generators, from EEG activity at the scalp (source localisation) difficult. Source localisation is an example of the "inverse problem" because for any given voltage distribution there are an infinite number of possible configurations of dipoles (Luck, 2005) which could have given rise to such a pattern. It is possible to use additional constraints and mathematically model the possible sources of activity, but the spatial resolution of ERPs remains weak. ERPs are therefore not helpful for localising cognitive functions in the brain.

The main benefit of ERP methodology is in addressing temporal questions. Because data can be acquired every 4 to 5 milliseconds, voltage changes shown in the EEGs reflect the activity of the working brain almost as it happens. This property has been exploited by psychologists interested in the moment by moment processes of various cognitive operations including language comprehension and memory retrieval.

3.3 Recording ERPs

Voltage is the difference between electrical potentials at two points and therefore the voltage observed at any scalp electrode is the difference between electrical activity at that so called active electrode (EA) and some other reference electrode (ER). However a direct subtraction of activity at these two sites would reflect activity at both the EA and the ER and include a contribution from any environmental electrical activity. Thus, a third ground electrode (EG) is placed at a different site on the scalp (or on the body) and differential amplifiers are used to amplify the difference between the voltage at pairs of electrode sites (EA−EG) minus (ER−EG). Any contribution from the ground electrode which is equal at the active and reference electrodes (and assumed to be background environmental electrical activity) will then be eliminated in the subtraction, leaving a measure of activity at the active electrode (relative to the chosen reference site). Before discussing the requirements of an appropriate reference site, details of the active electrodes and their placement on the scalp are summarised.

3.3.1 Active electrodes

Scalp electrodes are metal disks which form a circuit with the scalp via conductive gel. The metal chosen should be slow to corrode so that its conductance remains high. The experiments in this thesis use Ag/AgCl electrodes—silver electrodes coated in a thin layer of silver-chloride—but tin electrodes are a common alternative. Both Ag/AgCl and tin electrodes attenuate low fequencies and are therefore inappropriate for recording slow voltage changes. The current takes the path of least resistance, and therefore to maximise the activity recorded at the electrodes it is important that the impedance (impediment to current flow) between the scalp and the electrodes is kept to a minimum and does not vary between the ground, reference, and active electrodes. This is done by gently abrading the skin to remove the outer layer of dead skin cells. In the experiments reported in the thesis, electrode impedances were reduced and kept below 5kΩ during recording. Low impedance is important to avoid decreased common mode rejection—a decrease in the ability of an amplifier to eliminate environmental noise—and increased skin potentials (between surface and deeper layers of skin) which contaminate the signal with low frequency noise.

The placement of the active electrodes is usually based on the International 10-20 system (Jasper, 1958) which uses features of the skull—nasion, inion and the left and right pre-auricular points—to define lines of latitude and longitude and then places electrodes at regular intervals along these lines (see Figure 3.3).

Electrode names refer to the general scalp region (Fp = frontal pole; F = frontal; C = central; P = parietal, O = occipital), and to the position relative to the midline (z = midline; odd number = left hemisphere; even number = right hemisphere; larger numbers refer to greater distances from the midline). The scheme has since

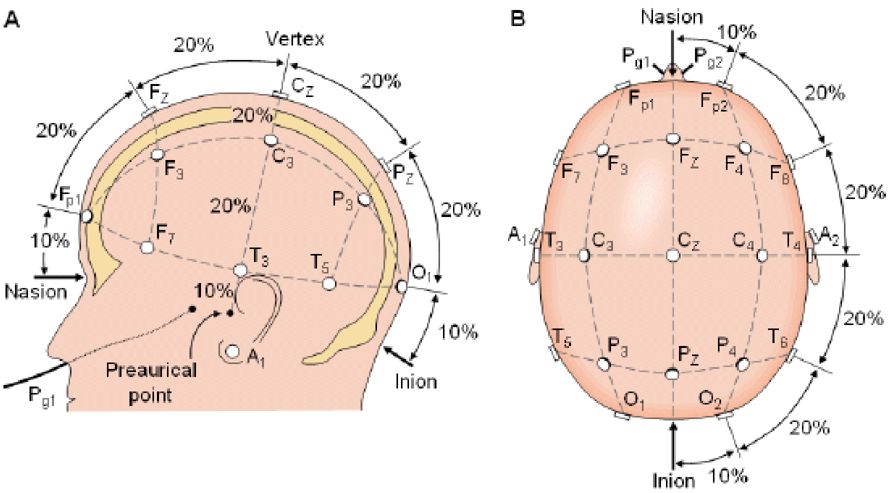

Figure 3.3: The International 10-20 system for electrode placement seen from (A) left and (B) above the head. The reference points refer to the following: A = ear lobe, C = central, Pg = nasopharyngeal, T = temporal, P = parietal, F = frontal, Fp = frontal polar, O = occipital. Adapted from Jasper (1958).

been extended to include more sites (for a more recent version see American Electroencephalographic Society, 1994, Figure 3.4) and the data reported in the thesis were recorded from electrode sites based on this extended 10-20 system.

3.3.2 Reference electrodes

The reference electrode is placed at a convenient site chosen to be as electrically insulated from brain activity as possible but sensitive to similar external sources of electrical noise as scalp electrodes and not biased towards either hemisphere. A common choice is the mastoids—the bony protrusions behind the ears. Typically, data is collected using one mastoid as the reference and the other acts as an active electrode. This enables monitoring of mastoid activity, which although relatively inert, will not be completely neutral or shielded from activity. The data are then reconstructed offline by calculating the average of both left and right mastoid recordings to eliminate any hemispheric bias in recording (see Luck, 2005, p.108 for further explanation).

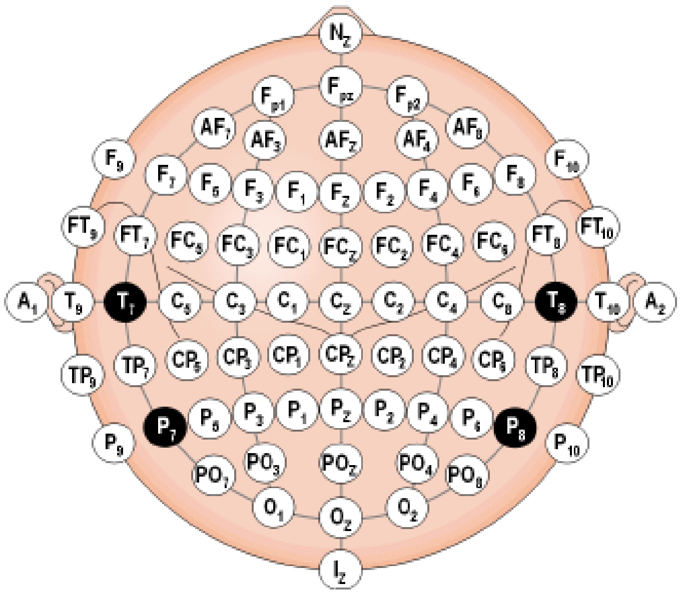

Figure 3.4: The Extended International 10-20 system for electrode placement, standardised by the American Electroencephalographic Society. Adapted from American Electroencephalographic Society (1994).

Although data could be collected using a linked mastoid reference, where wires from the left and right electrodes are physically linked, this can be problematic for two reasons. First, it creates a zero resistance path between the linked electrodes which can distort recordings at the scalp electrodes. Secondly, if the left and right mastoids have different impedances, the linked mastoid will reflect the mastoid with the lowest impedance producing a hemispheric bias (Miller, Lutzenberger, & Elbert, 1991).

Another reference that can be used is a global average: an average of all scalp recordings. Using a global average reference means that an increase in voltage at one site will always be accompanied by a decrease in voltage at another site which must be taken into account when interpreting the data. In addition, if the electrodes do not cover a large enough proportion of the scalp, use of a global average can lead to waveform distortions and subsequent misinterpretations (see Luck, 2005, p.109). Nevertheless, if voltages are recorded equally from a dense array of electrodes covering a sufficiently large proportion of the head, a close approximation of the true average can be obtained (an exact average would require covering all surfaces of the head, including face: Dien, 1998).

The position of the reference electrode affects the morphology (polarities and spatial distribution) of the waveforms observed at the active electrode. In particular, activity at active electrodes closer to the reference will be attenuated more than at those further away. In line with the majority of previous language comprehension research, the experiments in the thesis used an average of left and right mastoid recordings.

3.3.3 Amplifying, digitising and filtering

The voltages between active/reference/ground electrodes are recorded as an analogue signal and amplified before being filtered to attenuate frequencies outside of a given range and digitised (AD conversion) to convert the analogue voltage into a discrete form.

Digitisation converts the EEGs into numerical representations at a particular sampling frequency or rate (e.g., 200Hz or every 5ms). The resolution (e.g., 16 bits) specifies the number of discrete values (e.g., 65536) which can be produced over a range of voltages (e.g., +/−5 V). To capture all of the analogue signal in digital form, the sampling frequency should be at least twice the highest frequency in the signal (the *Nyquist theorem*). If the sampling rate is too low, low frequency artefacts may be introduced (aliasing). To minimise the possibility of aliasing caused by arbitrarily high frequencies, amplifiers typically include a low-pass filter which attenuates high frequencies. A high-pass filter attenuates low frequencies, which are commonly produced as a result of gradual voltage shifts caused by skin potentials. Such low frequencies are problematic because they may cause amplifier blocking. The reported experiments use a filter of 0.01–40 Hz and a sampling frequency of either 200Hz or 250Hz.

3.4 From EEGs to ERPs

Neuronal activity, reflecting the cognitive processing associated with an event, causes only very small voltage changes (e.g., 5–10μV: Kutas & Dale, 1997) which are difficult to distinguish from the background EEG (typically 50–100μV). To extract the "signal" (time-locked ERP) from the "noise" (all EEG unassociated with the event of interest), the EEG data are processed to remove or correct for contamination by artefacts such as ocular activity, amplifier saturation, voltage drift and muscular activity, and averaged. This section discusses the issues surrounding averaging and artefact reduction.

3.4.1 Averaging

To obtain ERPs, EEGs are recorded during the presentation of multiple trials of an event of interest and segmented into epochs (time windows) corresponding to the events. Data are then averaged at each sampled time point within an epoch (Coles & Rugg, 1995) to produce an ERP for a particular event. ERPs are further averaged over participants to make a grand average from which the data is interpreted. The pattern of the voltage changes time-locked to an event of interest is plotted relative to the average activity over a pre-stimulus baseline, usually a period of 100–200ms. There is therefore nothing inherently meaningful about the polarity or absolute voltage of the waveform; rather these features are used in a relative sense, to compare the effects across different processing conditions. Waveform morphology including polarity is influenced by the location and orientation of underlying neuronal generators relative to the scalp electrodes, choice of reference electrode and baseline conditions and the pre-stimulus activity. Issues surrounding data interpretation are discussed in section 3.5.

Four assumptions underlie averaging (Glaser & Ruchkin, 1976; Spencer, 2005): first, that the signal and noise are independent and sum linearly to produce the recorded waveform; secondly, that the signal is the same over trials of a particular event of interest; thirdly, that the noise varies randomly between events; fourthly, that the noise is constant (the mean and variance are similar) between events. If these assumptions are met experimentally, averaging will leave voltage patterns (ERPs) which are temporally related to the processing of that type of event.

As the number of trials in the average ERP increases, the contribution of noise will tend towards 0μV, which increases the signal-to-noise ratio (S/N). The reduction in noise is equal to the square root of the number of trials and therefore because the signal is unaffected by the averaging process, the S/N ratio increases as a function of the square root of the number of trials. This means a two-fold increase in S/N requires an average of four trials, a three-fold increase in S/N requires nine trials, a four-fold increase in S/N requires sixteen trials, and a five-fold increase requires twenty five trials. The experiments reported in the thesis use a criterion of sixteen trials as the minimum required for all conditions for each participant for inclusion of that participant in the grand average.

In practice, the assumptions of averaging are rarely met experimentally. The signal is unlikely to be constant. For example, periods of fatigue, boredom and attention lapse will introduce variance as the recording session progresses (Ruchkin, 1988), and the signal may even be absent from some trials (e.g., as a result of guessing in a memory experiment). In addition, the morphology, amplitude and latency of waveforms will vary between trials and between participants. Average waveforms may not, therefore, represent the waveforms for any individual trial or participant. While inter-trial variability in morphology and amplitude are not usually problems, as the averaged ERP waveform can be interpreted to reflect the mean amplitude and mean morphology, inter-trial variability in trial onset time (latency jitter) can distort the average waveform. As latency jitter increases, there is more variability in the latency of the peak which decreases the amplitude and increases the temporal duration of the peak. Onset (and offset) times of any divergences in an average waveform will represent the earliest onsets (and latest offsets) from the contributing waveforms. If latency jitter differs across conditions, this may lead to the false conclusion that conditions differ in the extent to which they draw upon particular processes. However, the impact of waveform variation on interpretation is minimsed by performing analyses on the mean amplitude over a time window and thereby paying less attention to peak-based measures. It is important to note that mean amplitude measures will only correct for latency jitter if the time window is chosen to capture the full duration of the effect. In practice this may be difficult due to the presence of overlapping components (Luck, 2005).

Noise tends not to be constant and does not always vary randomly. For example, muscular artefacts may vary between trials and eye movements are associated with large and quite systematic voltage patterns. Not all sources of noise will therefore be eliminated by averaging. Contamination of the signal by systematic noise can be reduced by correcting for or removing such artefacts. The main sources of contamination are discussed below.

3.4.2 Artefact correction/rejection

The most effective way to reduce the impact of noise on the signal is to record clean data (Luck, 2005). However, even with relatively clean data, it is still necessary to minimise the contribution of artefacts which will not be removed by averaging.

Ocular artefacts

Eye blinks and eye movements (electro-oculargram; EOG) are prominent sources of contamination of the EEG signal and result from changes to the electrical potential between the (positive) cornea and the (negative) retina. The movement of the eyelid across the eye during an eye blink is characterised by a monophasic deflection in the waveform of around 50–100μV which lasts around 200–400ms. This vertical EOG (VEOG) can be measured by placing electrodes above and below one of the eyes and is assumed to be representative of the activity produced by both. The VEOG is observable in the scalp recorded EEGs such that the voltage at any given electrode site is equal to the voltage at the eye electrodes multiplied by a propagation factor, plus EEG activity recorded at that electrode (Luck, 2005). The propagation factor varies depending on the site because the contribution of VEOG to scalp EEG increases in amplitude at electrode sites towards the front of the head (Lins, Picton, Berg, & Scherg, 1993). Lateral eye movements result in a positive deflection in the waveform in the direction of movement. This Horizontal EOG (HEOG) can be recorded by placing electrodes either side of the eyes.

Ocular artefacts can be reduced by asking participants to fixate on a cross in the centre of the screen and blink only when instructed. In relatively lengthy experiments, however, this can become tiring for participants. Furthermore, such instructions introduce a secondary task which may cause participants to approach the task differently and engage in differential cognitive processing (e.g., Verleger, 1991). The introduction of such systematic noise could be falsely attributed to the experimental manipulation.

All contaminated trials can be rejected but this can result in a large proportion of rejected trials and the remaining trials may not be representative of the whole data set (Gratton, 1998). It is therefore preferable to estimate the contribution of activity from systematic eye blinks to the EEGs and remove it from the data. Data containing unsystematic eye movements are rejected. There are several methods of correcting for eye blinks. Most use regression techniques and assume a linear relationship between the EOG and the EEG. They involve the calculation of the propagation factor between the EOG and the EEG at each scalp electrode and then subtraction of the corresponding proportion of the EOG from the EEG at each scalp electrode. Such estimation is made on an individual participant basis and derived from an average of their blinks in the experiment. The experiments reported in the thesis made use of such regression techniques. One limitation of this correction method is that subtraction of the EOG not only removes true ocular activity, but also the brain activity which is also recorded by the eye electrodes.

Another problem is that the propagation factors will be the same for eye blinks and eye movements. Alternative methods of ocular artefact correction which address these problems include the dipole modelling procedure and independent component analysis (ICA) (for a review see Gratton, 1998).

Amplifier saturation, voltage drift and muscular activity

In addition to ocular artefacts, there are other sources of contamination of the EEG and affected trials are typically rejected. Trials are rejected if they include any portion of saturation, where voltage exceeds the capacity of the amplifiers. As discussed (section 3.3.3) amplifier gain should be set to minimise this possibility. High frequency potentials are produced by muscular activity (electromyogram; EMG). This noise can be minimised by ensuring the comfort of participants to reduce muscle activity and attenuated by the low-pass filter of the amplifier.

Low fequency potentials or slow voltage shifts are caused by changes in skin or electrode impedances which occur if participants sweat or electrodes move. This noise can be avoided by ensuring low and stable impedances during recording. In addition, trials are rejected if voltage change over a given time reaches a certain threshold. The current experiments used a drift criterion of a change of 50μV over 2000ms.

3.5 Inferences from ERPs

Once EEG data have been processed to extract the signal from the noise, the resulting ERPs can be interpreted. ERP data, like behavioural data, are inherently correlational: the underlying neural activity is correlated with a cognitive process but is not regarded as its direct manifestation. In this section I discuss the inferences about cognitive processing which can be drawn from ERP data.

3.5.1 Components

ERPs are observable as a pattern of changing voltage over time and comprise a series of positive and negative going waves plotted relative to the average activity over a pre-stimulus baseline (usually of 100–200 ms). Observable effects reflect the summation of multiple underlying, so-called *latent* components which are potentially difficult to isolate. The term "component" is traditionally used to refer to an observable pattern of activity with specific polarity, timing and general scalp distribution which occurs with a particular cognitive function. Components may be labelled according to their typical physical characteristics, for example the N400 is a negative going wave which peaks around 400ms after word onset, or according to the conditions under which they are typically observed, for example the Mismatch Negativity is a negative going wave which is elicited in response to stimuli which mismatch with the surrounding acoustic context. Any observable waveform reflects contributions from all cognitive operations at that time and therefore cannot be assumed to be the neural correlate of any specific cognitive function. In practice, therefore, components of interest are defined as the difference between the ERP activity elicited for two experimental conditions (difference waves) averaged over the appropriate time window. Subtraction of the ERPs recorded in the two conditions is assumed to isolate the component which reflects the cognitive process that the experimental manipulation seeks to elicit.

The subtraction methodology (first conceived by Donders, 1968, translated by Koster, 1969) is predicated on two assumptions. The first is that the latency of equivalent components across conditions is identical; latency differences would produce distinct peaks in the subtraction waveforms, incorrectly implying qualitative differences in cognitive processes (Coles & Rugg, 1995). The second assumption is that cognitive processes are additive and independent ("pure insertion", see Friston, Price, Fletcher, Moore, & Frackowiak, 1996) and therefore subtraction should isolate a difference in a single cognitive process; failure of pure insertion means that the subtraction waveform might not reflect the addition of the cognitive process of interest, but rather an interaction between between an added component and existing components. Careful consideration must be given to the design and interpretation of all cognitive psychology experiments to try to fulfil the demands of the assumption of pure insertion (for alternatives to subtraction methods, see Aguirre, 2003).

There are two main types of ERP components (Fabiani, Gratton, & Coles, 2000; Luck, 2005): exogenous (or early) components and endogenous (or late) components. Exogenous components are elicited in response to physical characteristics of a stimulus and reflect sensory aspects of stimulus processing. They are relatively insensitive to a person's cognitive state. Of greater interest to cognitive psychologists are endogenous components which are elicited in response to the perceptual and cognitive processing associated with a stimulus (or the expectation that a stimulus will appear). Mesogenous components possess characteristics of the two main types (Fabiani et al., 2000).

The functional interpretations of some components are fairly well understood, at least to the extent that the component can be used as a reliable index of a particular cognitive process. For example, the N400 is thought to reflect ease of semantic integration and is often used as an index of semantic mismatch (see chapter 4). It is not always necessary to know the functional interpretation of a component to derive useful conclusions from an experiment because differences in waveforms may themselves be informative (e.g., Van Berkum et al., 2005). In such a case, the term "component" may be used simply to refer to the part of the waveform chosen for analysis. In chapter 4 I discuss the ERP components of most relevance to language comprehension.

The remainder of this section discusses the types of inferences about cognitive processing which can be drawn from ERPs. Although the most often cited experimental benefit of ERPs is their excellent temporal resolution, inferences are not limited to those of timing, but include inferences about sensitivity, functional equivalence, degree of engagement, and to an extent, functional interpretation and neuronal generators. Such inferences are drawn from quantitative temporal and amplitude differences and qualitative topographic differences in ERP data (Van Berkum, 2005).

3.5.2 Inferences from amplitude and temporal differences

Quantitative differences in amplitude are interpreted as changes in the extent to which a particular underlying cognitive process is engaged. It is important to be aware that between-condition amplitude differences may also occur for other reasons, for example, differences in latency jitter (see section 3.4.1) and the number of trials in the ERP average. Another cause of the misattribution of amplitude differences to the degree of cognitive engagement stems from variability in the proportion of trials which reflect the cognitive process of interest. For example, in memory studies responses which correctly identify previously presented stimuli as "old" may be produced by guessing rather than remembering: the associated cognitive processes will be reflected in the average ERPs. This means that an amplitude increase may reflect an increase in the probability of engaging memory processes on a single trial, rather than an increase in the degree to which memory processes are engaged per se. Quantitative differences in timing are taken to reflect differences in the temporal engagement of different processes. The onset of differences between two conditions reflects the latest time at which differences between cognitive processes emerge.

3.5.3 Inferences from topographic differences

Qualitative differences in scalp distribution (topographic differences) are attributed to qualitatively distinct cognitive processes, based on the assumption that a specific cognitive process is associated with invariant underlying neuronal activity. However, the absence of differences in scalp distributions is not strong evidence for the engagement of identical cognitive processes because the scalp-recorded voltages are compatible with an infinite number of underlying generators (section 3.2.2).

3.6 Analysis of ERPs

This section provides an overview of the types of analyses which are performed in order that inferences can be drawn (discussed above; section 3.5). Typically, the subtraction method is used to extract ERP components of interest. The components are then quantified by taking the mean amplitude over a defined time window (relative to the mean voltage over the pre-stimulus interval). Mean amplitude is preferable to peak amplitude because it is less sensitive to high frequency noise and is independent of changes to the variance which may occur because of different numbers of trials (Luck, 2005).

3.6.1 Analysis of Variance (ANOVA)

The most commonly used statistical test to analyse ERP data is the Analysis of Variance (ANOVA). The ANOVA model is based on an assumption of sphericity, which posits that the variance of the difference scores between all possible pairs of variables is equal (and therefore all pairs of variables are correlated to the same degree). ERP data typically violate this assumption because scalp recorded data are likely to be more correlated from electrodes which are closer together. The result is an increase in the probability of a Type I error (false positive) but this can be corrected for by the Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) which decreases the degrees of freedom and consequently the F-ratio, effectively making the test more conservative (and increasing the p-value). The ANOVAs reported in this thesis make use of the Greenhouse-Geisser correction where appropriate.

3.6.2 Amplitude analyses

The presence of quantitative differences are assessed using an ANOVA. Electrode data for each experimental condition are typically divided further into factors corresponding to different scalp regions, for example location [frontal, parietal], hemisphere [left, right], and site [superior, inferior]. Chapter 5 provides details about the factors used in the experiments reported in the thesis. Where amplitude differences are identified it is important to assess whether the scalp distributions of the ERPs are equivalent, because differences in the distribution of ERPs are assumed to reflect the engagement of different cognitive processes across conditions. In addition, interactions between amplitude and scalp regions must be interpreted with caution because of their ambiguous cause (see section 3.6.3) and so topographic analyses are performed.

3.6.3 Topographic analyses

The presence of distributional differences are assessed using an ANOVA. However, before being submitted to such a test, the data must be rescaled to remove amplitude differences between conditions, whilst maintaining the relative between-electrode voltage differences within conditions. Such normalisation is necessary because changes in the extent to which the underlying neuronal generators are active lead to a multiplicative effect on ERPs rather than an additive effect, as assumed by the ANOVA model. In other words, a two-fold increase in source strength leads to a two-fold increase in voltage at the corresponding electrode sites, whereas the ANOVA model assumes a constant increase in voltage (McCarthy & Wood, 1985). Without normalisation, an interaction between condition and location can arise from a change in source strength rather than from the activation of different generators. The Max/Min normalisation method rescales the amplitude of the effect (the difference between two conditions) at each electrode relative to all other electrodes. It maintains the pattern of relative differences whilst removing atypically large amplitude differences between electrodes which are likely to give rise to spurious condition by location interactions in the analysis. Normalisation prevents false positives, which would lead to an incorrect conclusion that topographies are different. The method requires finding the maximum and minimum value in each condition, subtracting the minimum from every data point, and dividing it by the difference between maximum and minimum. There is some debate as to the effectiveness of different normalisation procedures in enabling conclusions to be drawn about differences in underlying neuronal generators (see Dien & Santuzzi, 2005; Urbach & Kutas, 2002, 2006; Wilding, 2006). Topographic analyses of ERP data reported in this thesis use the Max/Min method.

3.7 Conclusions

ERPs enable the tracking of cognitive processes with high temporal resolution. The precise nature of the relationship between scalp recorded electrical activity and underlying neuronal generators remains unclear, but with appropriately designed experiments, ERPs enable inferences about cognitive processes to be made. Chapter 4 reviews how ERPs have been used to investigate the cognitive processes of language comprehension.

Chapter 4

ERPs and language

4.1 Introduction

ERPs have been widely used to investigate language processing since the demonstration by Kutas and Hillyard (1980) that sentence-final words that were semantically incongruous with their preceding sentence context elicited a negative waveform relative to congruous completions (see 1; the incongruous word is in bold and the congruous word is in parentheses).

(1) I take my coffee with cream and **dog**/(sugar)

Since the discovery of this effect, termed the N400, a wealth of research has demonstrated a relationship between ERPs and aspects of language comprehension. Observable, replicable patterns of ERP activity have been used to inform and support theories of language comprehension. To date, however, relatively few studies have used ERPs to investigate the comprehension of natural speech, and almost none have used them to investigate the effects of disfluencies.

This chapter examines the main ERP components associated with language comprehension. I briefly review effects related to the sensory processing of auditory information, and then focus on the effects related to semantic and syntactic processing. The aim is to demonstrate the potential of ERPs for selectively tracking aspects of language comprehension, and to give an overview of the ERP effects which are of particular relevance for the predictions and interpretations of the experiments reported in the thesis.

As discussed in Chapter 3, scalp recorded ERPs are not a direct manifestation of cognition, but reflect a correlation with the processes being engaged. Furthermore, because the interpretation of ERP waveforms requires a comparison between conditions the effects which I describe refer to activity relative to a baseline condition, for example, the ERPs to unpredictable relative to predictable.

4.2 Auditory sensory processing effects

Early sensory processing of auditory stimuli is associated with early (1–10ms post-stimulus onset) ERP components that originate in the brain stem. These are followed by cortical components, including the P1, N1 and P2 which reflect late sensory and early perceptual processes. Of greater interest to cognitive psychologists are the components that are sensitive to the cognitive and attentional demands imposed by processing of stimuli. These are described below.

4.2.1 Mismatch Negativity (MMN)

The mismatch negativity (MMN) is a negative wave largest over central/fronto-central sites, which peaks between 100 and 250ms after stimulus onset (Näätänen, Gaillard, & Mantysalo, 1978; Näätänen, 2001). It is elicited in response to stimuli which are acoustically deviant within their context, irrespective of task demands and occurs even when participants are not required to attend to the stimuli. It is therefore thought to reflect automatic neural processes associated with an acoustic mismatch between the stimulus and the sensory memory trace created by previous stimuli which is known as echoic memory (Näätäanen & Winkler, 1999; Schröger, 1997).

4.2.2 P300 effect

The MMN is typically followed by the P300 (P3) effect which comprises several subcomponents. The two main components are the frontally maximal P3a and the parietally maximal P3b (Squires, Squires, & Hillyard, 1975), though in the absence of a distinction, the term P3 is typically used to refer to the P3b (Luck, 2005). In the original Squires et al. (1975) paper, both the P3a and the P3b were elicited by unpredictable, infrequent changes in pitch or intensity, but the P3b was only elicited when such changes were task relevant. Other studies have suggested that the P3b is elicited by infrequent stimuli which are nonetheless expected by participants, whereas the P3a is associated with truly unexpected infrequent stimuli (Verleger, Jasowski, & Wauschkuhn, 1994) but this distinction between levels of automaticity is not without its critics (see Luck, 2005, p.42). The commonly held view (Donchin, 1981; Donchin & Coles, 1988; Polich, 2004) is that the P300 reflects the identification of (P3a), and attentional orientation to (P3b), deviant stimuli, and the subsequent induced memory updating (P3b).

4.3 Semantic processing effects

Language comprehension is fundamentally about extracting meaning. Semantic processing is associated with distinctive patterns of ERP activity, of which the N400 effect is the most studied and understood. I describe the features of the N400, the conditions under which it is observed, and the factors which influence its amplitude. I discuss how the interpretation of the N400 enables it to be used to investigate the integration of semantic information during comprehension. I also introduce two other ERP components which are sometimes observed alongside the

N400, the Phonological Mismatch Negativity (PMN) and the Late Positive Complex (LPC).

4.3.1 N400 effect

The N400 effect is a relative negativity in the ERP waveform, often peaking at approximately 400ms after stimulus onset. The effect (typically measured between 300–500ms) is broadly distributed over the scalp and maximal over the centro-parietal sites with a bilateral distribution for auditory stimuli, but with a slight right hemisphere bias for written stimuli. In response to written stimuli, the effect onsets around 200–250ms but auditory stimuli tend to elicit an earlier onsetting effect (as early as 50ms after word onset in continuous speech, Holcomb & Neville, 1991), because of co-articulatory information, for example (Connolly & Phillips, 1994; Hagoort & Brown, 1999; Van den Brink, Brown, & Hagoort, 2001), although it has been suggested that early effects may reflect a different process (see section 4.3.2). The effect usually lasts until around 600ms post-stimulus onset. Figure 4.1 shows an example ERP waveform which shows an N400 effect.

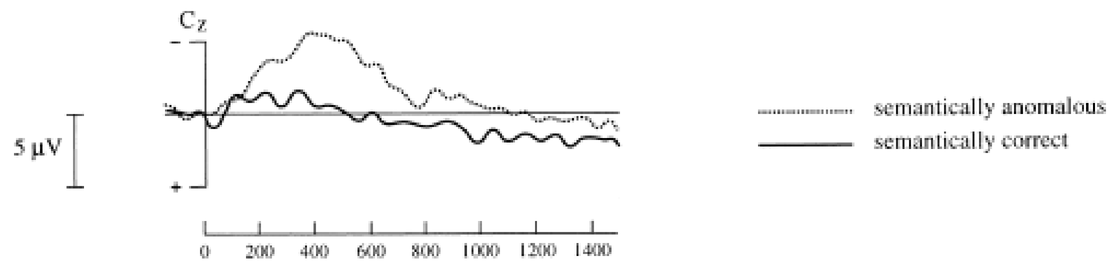

Figure 4.1: Example ERP waveform showing an N400 effect for spoken language at the Cz electrode. ERP waveform relative to the semantically correct (solid line) and semantically incongruent (dotted line) word onsets. Voltage (in μV) is plotted on the y-axis (negative up) and time (in ms) is plotted on the x-axis. Relative to semantically correct words, semantically incongruent words show a relative negativity which emerges around 100ms and peaks around 400ms. Diagram adapted from Hagoort and Brown (2000b).

Slight timing differences notwithstanding, the N400 is relatively insensitive to the manner of stimulus presentation. N400s have been observed for written stimuli presented both slowly (e.g., Kutas & Hillyard, 1980) and using the rapid serial visual presentation technique (e.g., Van Berkum, Hagoort, & Brown, 1999). N400s have been observed for auditory stimuli where the critical word was separated from its context by a silent pause (e.g., Besson, Faita, Czternasty, & Kutas, 1997; Van Petten, Coulson, Rubin, Plante, & Parks, 1999), for connected speech where the sentence-final target word was spliced into the context (e.g., McCallum, Farmer, & Pocock, 1984), and for fully connected and fluent speech (e.g., Hagoort & Brown, 2000b; Holcomb & Neville, 1991).

The N400 was initially observed in response to semantically incongruous relative to congruous words (Kutas & Hillyard, 1980), but it is not exclusive to spoken and written words. N400s are also observed in signed languages (e.g., Kutas, Neville, & Holcomb, 1987; Holcomb & Neville, 1990, 1991) and in response to non-verbal, but meaningful stimuli, such as line drawings (Federmeier & Kutas, 2001; Ganis, Kutas, & Sereno, 1996), photographs (McPherson & Holcomb, 1999), faces (Bobes, Valdes-Sosa, & Olivares, 1994) and environmental sounds (Cummings, Ceponiene, Koyama, & Saygin, 2006; Van Petten & Rheinfelder, 1995) suggesting it may not be a language specific effect. However, non-verbal N400s do show some variability in scalp topography and additional effects on other components are sometimes observed. This suggests the engagement of only partially overlapping neuronal populations and that semantic representations of verbal and non-verbal information are partially distinct (Federmeier & Kutas, 2001). Pseudowords (pronounceable non-words), but not unpronounceable non-words, elicit an N400 (e.g., Bentin, McCarthy, & Wood, 1985; Chwilla, Brown, & Hagoort, 1995; Holcomb & Neville, 1990), although only within a predominantly meaningful context (Hahne & Jescheniak, 2001).

The N400 amplitude is larger for low frequency than for high frequency words (e.g., Van Petten, 1993; Van Petten & Kutas, 1990), for open-class than for closed-class words (Van Petten & Kutas, 1991), and for concrete than for abstract words, although the scalp distribution of the effect for abstract words is more frontally focused implying the activation of additional neurons. Tasks requiring deeper semantic analysis or imagery also elicit larger effects (e.g., Kounios & Holcomb, 1994; Paller, Kutas, & Mayes, 1987; West & Holcomb, 2000). However, these stimulus-related effects (for more details, see Kutas et al., 2006) may not be stimulus-inherent, but related to the context in which they appear. For example, the effect of word frequency diminishes throughout the sentence as contextual constraint increases (Van Petten & Kutas, 1990, 1991) and the effect of concreteness is eliminated when congruency is controlled (Holcomb, Kounios, Anderson, & West, 1999). Also, the greater N400 amplitude for open-class words may be explained by an account based on contextual word predictability because closed-class words are more prevalent within language and therefore may be more expected or more easily accessed during speech production.

Research has led to a generally accepted functional interpretation of the N400, that it provides an index of the ease of retrieval or integration of the stored conceptual knowledge associated with a word (or other meaningful stimuli), which is dependent on both the stored representation itself, and on the semantic constraints imposed by the interpretative context (Kutas et al., 2006, p.10). Contextual constraints can be formed at a number of levels—word, sentence, discourse, and pragmatic. I consider the influence of different contexts on the N400 and how the different contextual constraints interact.

N400 amplitude: word- and sentence-level constraints.

A number of studies have demonstrated N400 effects to words in sentential contexts. The amplitude of the N400 is inversely related to the predictability or expectancy of a word as determined by its cloze probability (Kutas & Hillyard, 1984), defined as the percentage of people producing that word in a sentence completion task.

However, the N400 is more than a mere index of cloze probability and when word predictability is held constant, the influence of other factors can be observed. The N400 is sensitive to the semantic and associative relationships between words in the absence of a sentential context. It has been shown that target words preceded by, or presented concommitantly with, a prime word elicit a larger N400 when the prime is semantically or associatively unrelated to the target (e.g., Bentin et al., 1985; Bentin, Kutas, & Hillyard, 1993; Coulson, Federmeier, Van Petten, & Kutas, 2005; Heinze, Muente, & Kutas, 1998; Holcomb & Neville, 1990; Kutas, 1993; Rhodes & Donaldson, 2008). For evidence of that the N400 depends on lexical association rather than semantic feature overlap see Rhodes and Donaldson (2008). The lexical N400 is qualitatively identical to the N400 found in a sentential context (Kutas, 1993).

N400 amplitude: discourse-level constraints.

The N400 is not just sensitive to semantic fit at an individual word or sentential level but also to semantic fit at a higher discourse level (Federmeier & Kutas, 1999; Nieuwland & Van Berkum, 2006; St. George, Mannes, & Hoffman, 1994; Van Berkum, Hagoort, & Brown, 1999; Van Berkum, Zwisterlood, Hagoort, & Brown, 2003; Van Berkum et al., 2005). Van Berkum and colleagues presented participants with written (Van Berkum, Hagoort, & Brown, 1999) or spoken (Van Berkum et al., 2003) isolated sentences such as (2a), or following a discourse which rendered one of the completions anomalous (2b).

(2a) Jane told her brother that he was exceptionally **quick/slow**

(2b) Jane was to wake her brother at five o'clock in the morning. But the brother had already dressed. Jane told her brother that he was exceptionally **quick/slow**.

Relative to discourse compatible words ("quick"), the discourse anomalous words ("slow") elicited a standard N400 which was significantly reduced when the sentences were not preceded by the discourse. The discourse-related N400 showed no qualitative differences in terms of latency and scalp topography to a standard sentence-related N400 obtained in the same experiment (but see M. Otten & Van Berkum, 2007, for evidence of some distributional differences between word-level and discourse-level N400 effects).

A meaningful context need not be linguistic. The N400 has also been observed related to words which, in a pre-test, were judged to be semantically unrelated to a preceding musical excerpt (Koelsch et al., 2004). The results demonstrate that a musical context can provide constraints on semantic processing.

N400 amplitude: pragmatic-level constraints.

The N400 is also sensitive to contextual information extrinsic to the current linguistic information (pragmatic information) such as world knowledge contained in semantic memory (Hagoort, Hald, Bastiaansen, & Petersson, 2004; Hald, Steenbeek-Planting, & Hagoort, 2007). For example, using sentences such as (3), it has been shown that, relative to plausible and true alternatives ("yellow"), world-knowledge violations, although plausible ("white") elicited an N400 effect which was not qualitatively different to an N400 elicited to outright semantic violations ("sour") (Hagoort et al., 2004).

(3) Dutch trains are **yellow/white/sour** and very crowded

A study by Federmeier and Kutas (1999) provides further evidence for the influence of semantic memory on the N400 amplitude, and on semantic processing of discourse. Participants were presented with sentences such as (4) which ended in one of three types of targets: the most expected word ("palms"), an unexpected word which was from the same semantic category as the expected word and therefore shared features ("pines"), or an unexpected word which was from a different category ("tulips").

> (4) They wanted to make the hotel look more like a tropical resort. So along the driveway they planted rows of **palms/pines/tulips**

Both unexpected words had cloze probabilities of 0, were rated equally implausible in offline studies and elicited an N400 relative to the most expected completion. However, the N400 effect was reduced for unexpected words from the same semantic category as the expected word. Furthermore, the reduction was even greater when the sentences were highly constraining even though the words were rated less plausible. These findings suggests that N400 amplitude is not merely an index of the contextual fit of a word (neither unexpected word fitted well, in terms of plausibility and cloze probability), but is partly dependent on the semantic associations between the critical word and contextually expected word held in long-term memory.

N400 amplitude: drawing inferences about semantic processing

The temporal equivalence of the N400 in response to semantic violations at a word-, sentence-, discourse-, and pragmatic-level, offers no support to a model of language comprehension that gives temporal precedence to lexical over sentential discourse or one that gives precedence to sentential over discourse processes (Hald et al., 2007; Kutas et al., 2006; Van Berkum, Hagoort, & Brown, 1999; Van Berkum et al., 2003). It is more compatible with a mechanism that utilises and integrates semantic information from various levels within the same time frame and which does not require distinctions to be made between word meaning, discourse context, and world knowledge (e.g., Clark, 1996; Jackendoff, 2002; MacDonald, Pearlmutter, & Seidenberg, 1994).

It should be noted that different levels of constraints have been shown to elicit different effects in non-temporal aspects of EEG. Specifically, world-knowledge violations, but not sentential violations, have been shown to elicit a peak in the gamma frequency band, suggesting differences in some aspect of processing of the two types of information, to which the N400 is not sensitive (Hagoort et al., 2004).

Several studies have directly investigated the interaction of different contextual constraints on the N400, and by inference the way semantic information at different levels is utilised during sentence processing.

When sentential context is unconstraining, lexical assocation and sentence-level expectancy have an additive influence on the N400 amplitude (Van Petten, 1993; Van Petten, Weckerly, McIsaac, & Kutas, 1997). However, as sentence-level constraints become stronger, either by manipulating the syntax (Hoeks, Stowe, & Doedens, 2004) or the semantics (Coulson et al., 2005; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007) they become the greater influence on N400 amplitude, demonstrating a preference for higher-level information in semantic integration.

However, higher-level information does not always take precedence in semantic integration. A study on negation (Fischler, Bloom, Childers, Roucos, & Perry Jr., 1983) showed no evidence of an N400 in response to sentences containing world-knowledge violations such as (5a) relative to (5b).

(5a) A robin is not a **bird**

(5b) A robin is a **bird** .

In this study, the semantic relationship between sentence constituents (“robin” and “bird”), in terms of category membership, or maybe feature overlap, seems to take precedence over the sentence-level constraints which lead to a violation of world-knowledge. For similiar results with quantification, see Kounios and Holcomb (1992).

Discourse-level and pragmatic-level constraints have an interactive effect on the N400 amplitude, although neither appear to override the other. Appropriate local discourse can ease the integration of information which conflicts with world-knowledge (Hald et al., 2007; Nieuwland & Van Berkum, 2006), but correct world-knowledge can be less acceptable if the local discourse is not supportive (Hald et al., 2007).

Because the N400 reflects the ease of integrating incoming information into the widest context available, the N400 amplitude is smaller when presented information matches the predictions developed from the surrounding context. If expectations are violated, processing is easier when words share semantic features with the most expected infomation. If the context is unconstraining, processing is easier when presented information fits well semantically into the context.

4.3.2 Phonological Mismatch Negativity (PMN)/N200

The onset latency of the N400 is related to the time at which critical information becomes available to the listener, and as mentioned previously (section 4.3.1), is typically earlier for spoken than for written stimuli. Variability in the onset of negativity during auditory speech comprehension has been demonstrated by Van Petten et al. (1999) who presented participants with sentences such as (6).

> (6) It was a pleasant surprise to find that the car repair bill was only seventeen **dollars/dolphins/scholars/hospitals**.

There were four types of sentence final words: congruent (“dollars”), incongruent which shared initial phonemes with the congruent ending (“dolphins”), incongruent which rhymed with the congruent endings ("scholars"), or fully incongruent ("hospitals"). Relative to the congruent endings, incongruent endings which rhymed elicited a negative ERP effect which was similar to the fully incongruent condition, onsetting around 150ms post-stimulus. By contrast, the negative ERP effect to incongruous endings which shared initial phonemes with the expected completion, only onset at around 400ms post-stimulus. By time-locking ERPs to the isolation point (the time at which the exact lexical candidate became clear) it was shown that ERP differences between congruent words and incongruent words sharing initial phonemes began at the isolation point, the point at which listeners could realise that the word was not the most expected. When critical words did not share the initial phonemes with the most congruent completition, differences emerged around 200ms before the isolation point, indicating that comprehenders start processing linguistic perceptual input with only partial information.

Other studies have observed similar early onsetting negativities (Connolly, Stewart, & Phillips, 1990; Connolly, Phillips, Stewart, & Brake, 1992; Connolly & Phillips, 1994; Hagoort & Brown, 2000b; Van den Brink et al., 2001; Van den Brink & Hagoort, 2004) although there is some disagreement as to whether the early onsetting negativity is indistinguishable from the N400 (Van Petten et al., 1999; Van den Brink & Hagoort, 2004) or is a distinctive perceptual component with a fronto-central scalp distribution which reflects the extent to which a word's phonological form fits into the context (based on previous semantic and syntactic constraints) which has been termed a Phonogical Mismatch Negativity (PMN) (Connolly et al., 1990; Connolly & Phillips, 1994) or an N200 (Van den Brink et al., 2001). A more recent study (Diaz & Swaab, 2007) found an early onsetting negativity in response to violations of phonological expectancy, but only to words in lists (which arguably emphasised phonological processing and did not require engagement of lexical integration) and not in sentences. As Kutas et al. (2006, p. 19) point out, the important conclusion is semantic context effects can be observed with auditory stimuli before word identification is complete.

4.3.3 Late Positive Complex (LPC)

Most ERP studies of semantic processing have focused on the N400, but the effect has sometimes been followed by a Late Positive Complex (LPC), a positive deflection in the waveform observed approximately 500-900 ms after stimulus onset. The effect has a frontal focus and may be more prominent over the left hemisphere. The functional interpretation of the LPC is not clear and the term LPC has been used quite generally to refer to positivities elicited under a number of conditions: it probably does not reflect a unitary phenomenon. Note that the term LPC is sometimes used to refer to parietal positivities such as the P600 (see section 4.4.1) and the P300 (see section 4.2.2) and the functional relationship between the P600, the P300 and the LPC is not currently clear (Federmeier et al., 2007).

A frontal LPC has been observed in response to unexpected words that complete highly constraining sentences relative to unexpected words that complete more weakly constrained sentences and to expected words (Federmeier et al., 2007). It has also been observed in response to probe words which are unrelated to preceding jokes (Coulson & Wu, 2005). One suggestion is that the LPC reflects control processes associated with memory retrieval and suppression of semantic information, for example the most contextually predictable word when an unexpected word is presented (Federmeier et al., 2007). This explanation is compatible with the demonstration of an LPC in response to unexpected words following English sentence fragments, particularly in idiomatic expressions (and thus highly constraining sentences) as well as unexpected switches into Spanish (Moreno, Federmeier, & Kutas, 2002).

The frontal distribution of the LPC is similar to positivities that are observed in studies of memory and are associated with retrieval effort (Ranganath & Paller, 1999; Rugg, Allan, & Birch, 2000), or attempts to retrieve source information from memory (Senkfor & Van Petten, 1998). Furthermore, the distribution of the LPC is consistent with a generator in the left inferior prefrontal cortex (Coulson & Wu, 2005), a brain region which is activated during memory tasks (Buckner, 2003; Wagner, Pare-Blagoev, Clark, & Poldrack, 2001). Thus an interpretation related to memory control seems plausible.

The LPC is attenuated for repeated words in a sentence context relative to their first presentation. This has led to the suggestion that it reflects the processes involved with retrieving and updating working memory as words (or other stimuli) are processed (Van Petten, Kutas, Kluender, Mitchiner, & McIsaac, 1991). The demands on these processes are reduced when a word appears for a second time.

4.4 Syntactic processing effects

In addition to semantic processing, language comprehension involves the processing of syntactic information which has been associated with distinctive ERP components. This provides evidence for a psychological distinction between the processing of semantic information and of syntactic information, and enables components to be used to selectively track these processes. The most widely investigated syntactic-related ERP component is the P600 which is sometimes refered to as the Syntactic Positive Shift (SPS). I describe the effect, the conditions under which it has been observed and the cognitive processes which it is thought to reflect. I discuss how the P600 has been used to investigate the interaction between semantic and syntactic information during language comprehension. I also introduce three other ERP components which have been related to aspects of syntactic processing, the Left Anterior Negativity (LAN), Early LAN (ELAN), and the NRef.

4.4.1 P600 effect

The P600 is a relative positivity in the ERP waveform, peaking at around 600ms after stimulus onset although often appearing as a long-lasting positive shift with no clear peak (Kutas et al., 2006). The effect is typically measured between 500–800ms, although it may onset as early as 200ms (Kutas et al., 2006) and is usually maximal over centro-parietal and parietal regions. The effect sometimes has an anterior maximum, which suggests the engagement of different processes. Figure 4.2 shows an example ERP waveform which shows a P600 effect.

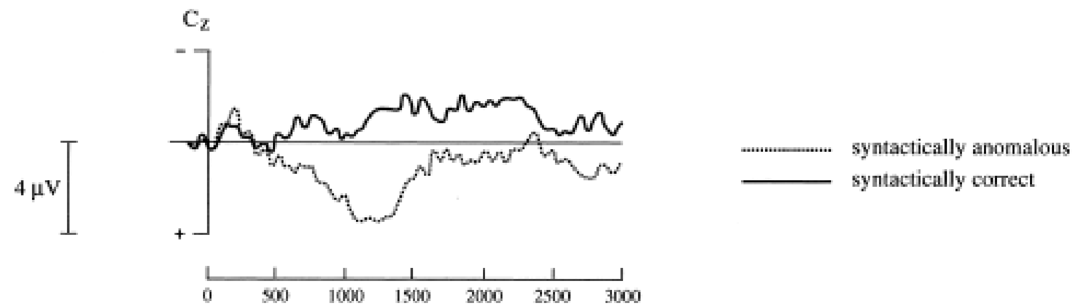

Figure 4.2: Example ERP waveform showing a P600 effect for spoken language at the Cz electrode. ERP waveform relative to the syntactically correct (solid line) and syntactically incongruent (dotted line) word onsets. Voltage (in μV) is plotted on the y-axis (negative up) and time (in ms) is plotted on the x-axis. Relative to syntactically correct words, syntactically incongruent words show a sustained relative positivity which emerges around 500ms. Diagram adapted from (Hagoort & Brown, 2000a)

One of the earliest studies identifying an ERP effect associated with syntactic processing and the one which coined the term "P600" observed the positivity response to violations of verb subcategorisation preferences (Osterhout & Holcomb, 1992), for example in response to "to" in (7a), relative to "to" following a verb (7b).

(7a) The broker persuaded **to** sell the stock was sent to jail.

(7b) The broker hoped **to** sell the stock.

It has also been observed in response to violations of phrase structure or word category (Friederici, Pfeifer, & Hahne, 1993; Friederici, Hahne, & Mecklinger, 1996; Hahne & Friederici, 2002; Osterhout & Holcomb, 1992; Neville, Nicol, Barss, Forster, & Garrett, 1991) and word order (Hagoort, Brown, & Groothusen, 1993) and in response to various types of morphosyntactic violations, including those of subject-verb number or subject-verb tense disagreement (Coulson, King, & Kutas, 1998; Osterhout, McKinnon, Bersick, & Corey, 1996; Osterhout & Mobley, 1995; Hagoort et al., 1993; Van Berkum, Brown, & Hagoort, 1999a; Vos, Gunter, Kolk, & Mulder, 2001), of article-noun number or gender disagreement (Gunter, Stowe, & Mulder, 1997; Hagoort, 2003) and of prounoun case disagreement (Coulson et al., 1998). The P600 is elicited to these violations in semantically meaningless sentences (Hahne & Jescheniak, 2001), but in sentences comprising pseudowords it is reduced (Yamada & Neville, 2007) or eliminated (Münte, Matske, & Johannes, 1997, although in this study morphosyntactic violations elicited an earlier negativity).

A P600-type effect has also been elicited in response to syntactically complex sentences (Friederici, Hahne, & Saddy, 2002; Kaan & Swaab, 2003b), or to non-preferred, but grammatical continuations (Osterhout & Holcomb, 1992; Kaan & Swaab, 2003a; Van Berkum et al., 2003). In these situations, the effect has shown an anterior focus (though see, Kaan, Harris, Gibson, & Holcomb, 2000, for an example of a posterior positivity in response to syntactically complex continuations).

The most common interpretation of the P600, is that it is a family of components which reflect similar but distinct aspects of syntactic processing (Carreiras, Salillas, & Barber, 2004; Friederici, Mecklinger, Spencer, Steinhauer, & Donchin, 2001; Friederici et al., 2002), such assyntactic repair or reanalysis following the violation of syntactic preferences (e.g., Friederici, 1995, 2002; Friederici et al., 1996; Gunter et al., 1997) or syntactic integration difficulties in general (Kaan et al., 2000; Kaan & Swaab, 2003b, 2003a). For discussion of potential functional distinctions between the anterior and posterior positivities, see Carreiras et al. (2004); Friederici et al. (2002); Kaan and Swaab (2003a); Kutas et al. (2006).

The P600, like the N400 (section 4.3.1), may not be a language-specific effect. It has been elicited in non-linguistic contexts, for example musical excerpts, in response to deviations from expected musical notes (Besson & Faita, 1995; Patel, Gibson, Ratner, Besson, & Holcomb, 1998). This has led to a domain general interpretation of the P600 as reflecting structural integration processes (Patel et al., 1998). Others have suggested that the P600 is a type of P300 component elicited by oddballs (see section 4.2.2). For discussion of this possibility, see Coulson et al. (1998); Friederici et al. (2001); Gunter et al. (1997); Osterhout and Hagoort (1999); Osterhout et al. (1996).

A recent body of work has shown that the P600 can be elicited in situations where an N400 might be expected because of semantic violations (Kim & Osterhout, 2005; Kolk, Chwilla, van Herten, & Oor, 2003; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003; Nieuwland & Van Berkum, 2005; Van Herten, Kolk, & Chwilla, 2005). For example, Kuperberg et al. (2003) demonstrated a P600 in response to “eat” in (8a) relative to (8b) despite the absence of a violation of syntactic constraints or preferences (for a review, see Kuperberg, 2007).

(8a) Every morning at breakfast the eggs would **eat** ...
(8b) Every morning at breakfast the boys would **eat** ...

These findings are consistent with a syntax-related interpretation of the P600 but also with the possibility that the P600 reflects a more general monitoring process which occurs when there is a conflict between the linguistic item that is presented and the item that was expected (Vissers, Chwilla, & Kolk, 2006).

4.4.2 Early Left Anterior Negativity (ELAN) and LAN

When sentences violate syntactic preferences, for example in garden path sentences, then a lone P600 is typically observed (Friederici, Steinhauer, & Frisch, 1999; Osterhout & Holcomb, 1992; Osterhout, Holcomb, & Swinney, 1994; Mecklinger, Schriefers, Steinhauer, & Friederici, 1995). In cases of outright syntactic violation, the P600 is sometimes preceded by an early negativity with a left anterior focus (LAN), although a bilateral distribution has been observed (Hagoort, Wassenaar, & Brown, 2003). The effect does not usually form a clear peak, but is observable between 300 and 500ms post stimulus. A negativity with similar scalp distribution but with an earlier onset of around 160ms (Early LAN: ELAN) has also been identified. Despite the topographical similarities of the LAN and the ELAN, a functional distinction between the two effects has been proposed (Gunter, Friederici, & Schriefers, 2000).

Whilst it is agreed that the ELAN and LAN, like the P600, are sensitive to syntactic aspects of processing, there is less certainty about what specific processes they reflect. ELANs are elicited by word category errors (e.g., Friederici et al., 1993, 1996; Hahne & Friederici, 1999; Neville et al., 1991) and may reflect local phrase structure building where word-category information is used to assign initial syntactic structure (Friederici, 1995; Friederici et al., 1996).

LANs, like ELANs, are elicited by word category errors, but are also observed in response to morphosyntactic errors both in word (e.g., Coulson et al., 1998; Gunter et al., 1997; Hagoort & Brown, 2000a; Osterhout & Mobley, 1995) and pseudoword (e.g., Münte et al., 1997) contexts. These findings have led to the proposal that the LAN reflects difficulties in thematic role assignment which makes use of morphosyntactic information such as verb inflections (Friederici, 2002). An alternative interpretation of the LAN is that it reflects general working memory processes.

This proposal is largely based on the observation of LANs in long-distance dependency (filler-gap) constructions (Kluender & Kutas, 1993; King & Kutas, 1995). The presentation of unexpected information which must be reconciled with earlier information may trigger a search back process whilst the current information is maintained. For example, pronouns whose antecedents are not immediately obvious may trigger some kind of search back through the memory trace (Coulson et al., 1998). See Lau, Stroud, Plesch, and Phillips (2006) for an alternative interpretation involving the prediction of syntactic categories.

4.4.3 NRef

The NRef is a frontally maximal sustained negative shift which has been elicited in response to referentially ambiguous phrases such as "the girl" following a discourse involving two girls (Van Berkum, Brown, & Hagoort, 1999b; Van Berkum et al., 2003), although there is no strong reason to suppose that the NRef is specific to referential processes (Van Berkum, Koornneef, Otten, & Nieuwland, 2007). One possibility is that it reflects control processes which are engaged during the attempts to resolve the ambiguity associated with multiple potential referents. Such control processes include making inferences or the search for cues in the episodic memory of the discourse. Another possibility is that it reflects the processing costs associated with the increase in demands on working memory that are imposed when two alternative referential interpretations are stored. This second possibility may explain the distributional similarity of the NRef to the LAN which has also been associated with increase in working memory demands (Van Berkum et al., 2007).

4.5 Conclusions

A number of ERP components have been associated with aspects of language processing and ERPs have been widely used to investigate language comprehension.

However, as yet their benefits as tool for measuring cognitive processing online (Chapter 3) have not been fully exploited to investigate disfluent speech comprehension. Four experiments presented in this thesis used ERPs to investigate the effects of disfluencies on language processing. The next chapter introduces the design and methods of the experiments.

Chapter 5

General methods

5.1 Introduction

This chapter introduces the core methods used in the five experiments presented in the thesis. I describe the experimental designs and discuss design decisions. I then describe details about the stimuli, participants, procedures, data processing, and analyses. Modifications of the methods presented here and additional information specific to each experiment are included in the relevant experimental chapters.

5.2 Experimental design

The aim of the thesis was to compare the effects of comprehending disfluent and fluent speech. This was done in five experiments. The effects on processing were assessed in Experiment 1 using a Lexical Decision Task (LDT), and in Experiments 2–5 using ERPs. Longer-term consequences for representation were assessed by a surprise recognition memory test (Experiments 1–4)[1].

[1] Experiment 5 included a memory test but participants performed below chance for all conditions. The test is not reported further in the thesis.

Each experiment employed a 2 X 2 design with factors of fluency [fluent, disfluent] and a secondary manipulation. The design meant that the impact of the secondary manipulation could be observed in fluent utterances and the resulting effect subsequently compared with the effect produced by the same manipulation in disfluent utterances. This second order comparison was chosen for three reasons. First, disfluent and fluent utterances necessarily contain inherent physical differences (e.g., the presence of an *er*) which would be expected to affect the ERP waveforms. This means that a direct comparison between waveforms for fluent and disfluent utterances may reflect physical differences between the stimuli, rather than differences in the cognitive processes of interest. In particular, the presence of a filled or silent pause before the target word resulted in a very different pre-target baseline compared to a fluent condition where the pre-target baseline was another word. Such stimuli differences are inherent to any investigation of disfluent speech and cannot be overcome through counterbalancing. Secondly, the 2 X 2 design enabled the effects of disfluencies on specific aspects of language comprehension to be assessed by using standard experimental manipulations. The standard effect of these manipulations on fluent utterances, with their agreed functional interpretations, could then be compared with the effects of the manipulation for disfluent utterances. Thirdly, establishing standard effects for fluent utterances confirmed that the dependent measures (e.g., reaction times, ERPs, recognition probability) showed the expected sensitivity with the experimental stimuli and procedures. Failure to obtain these standard effects for fluent utterances would prevent any meaningful interpretation of the disfluency related effects.

In Experiments 1–4, the secondary manipulation was of contextual word predictability. Compared to predictable words, unpredictable words are costly for semantic aspects of language comprehension: they are associated with slower LDTs and the elicitation of an N400 ERP effect. Disfluent pauses often, although not always, precede words which the speaker is having difficulty retrieving (Arnold et al., 2000, 2007; Barr, 2001; Beattie & Butterworth, 1979; Cook, 1969; Levelt, 1983; Maclay & Osgood, 1959; Martin, 1967; Tannenbaum et al., 1965) and predictability is one way of manipulating lexical retrieval difficulties. The experiments assessed whether *ers*, silent pauses and repetition disfluencies affect the semantic processing of contextually unpredictable compared to predictable words, using LDTs and ERPs. The longer-term representations of these words were assessed using recognition memory.

In Experiment 5, the secondary manipulation was the presence or not of a repair disfluency. The processing of repair disfluencies has not been widely investigated but it was hypothesised that they are costly for syntactic aspects of language comprehension and would elicit a LAN and a P600 ERP effect. Filled pauses sometimes, although not always, precede repairs (Brennan & Schober, 2001; Hindle, 1983; Levelt, 1983). The experiment assessed whether an *er* affects the syntactic processing of repairs, using ERPs.

5.2.1 Stimuli considerations

The aim of the experiments was to investigate the effects of the natural phenomena of disfluencies in speech on language processing. Stimuli were artificially created to sound as natural as possible. The use of spontaneously produced speech (e.g., from a corpus) would be in many ways preferable, ensuring, for example, natural prosody and intonation of the disfluencies and of the speech in which they were embedded. However, because of the lack of previous ERP studies in the area, it was considered extremely important to exert as much control over the stimuli as possible. Interpretation of the ERP data necessitated attribution of effects to the presence or absence of disfluencies and not to non-experimentally manipulated aspects of the stimuli, such as physical differences (see Chapter 4). The use of spontaneously produced speech would have made control of the stimuli difficult.

One possibility might have been to elicit disfluencies under controlled settings, for example using the network task (Levelt, 1983; Oomen & Postma, 2001b), or the Map task (Anderson et al., 1991) but the numbers of disfluencies produced would be lower than ideal for the purpose of stimuli generation. For example, picture-story descriptions in a divided attention task (simultaneous tactile task) designed to increase disfluency rate resulted in a mean rate of only 3 fillers per hundred words (Oomen & Postma, 2001a). As discussed previously (Chapter 3), ERP analysis requires a large number of trials to produce a sufficiently large signal-to-noise (S/N) ratio. The experiments reported in the thesis employed a minimum criterion of sixteen trials for each condition for each participant for inclusion of that participant in the grand average. This criterion was chosen partly for pragmatic reasons because doubling the S/N again would require 25 trials. Increasing the criterion would result in the exclusion of more participants' data or the need for more stimuli which would increase the duration of the experiment. In order to acquire sufficient trial numbers after artifact rejection, experiments were designed with 40 trials per condition. It would be difficult to obtain such large numbers of stimuli with controlled features using spontaneously produced speech. By contrast, creating the stimuli from scratch enabled full control over the syntax of the utterances and the placement of the disfluencies and also made it possible to create stimuli with chosen variation within a systematic framework.

5.2.2 Procedural considerations

To avoid introducing additional cognitive processes and altering the cognitive processing under investigation, participants were not given a secondary task during the comprehension part of the ERP experiment (e.g., Hagoort & Brown, 2000b; Van Berkum et al., 2005, 2007). Psycholinguistic studies of speech comprehension typically include some task (e.g., word monitoring) which provide the means for assessing comprehension. One of the benefits of ERPs is that they avoid the need for a secondary task, because the dependent measure (scalp-recorded voltage) is sensitive to the comprehension process itself. However, language comprehension studies using ERPs often still include some task, typically comprehension questions or grammaticality judgements. Although such tasks may keep the participant alert and can act as a check that they are paying attention, they are not necessary. Further, a secondary task may be distracting and therefore alter the participant's comprehension process.

5.3 Stimuli

Stimuli were 1-sentence or 2-sentence utterances formed from two stimuli banks, the first of which was used for Experiments 1–4 and the second for Experiment 5 (see Appendices B and C).

5.3.1 Stimuli for Experiments 1–4

The stimuli for Experiments 1–4 were highly constrained fluent and disfluent utterances which ended in either predictable or unpredictable words. One concern with this method is that unpredictable words may also be implausible, for example "burger" following "I'm really thirsty. Let's go to the pub for a..." Since the focus of the present work is on predictability this issue is not discussed further but future work could distinguish between these two types of difficult-to-integrate words. Utterances were constructed in pairs such that each predictable final word was an unpredictable word for a corresponding utterance. Furthermore, predictable and unpredictable targets completed fluent and disfluent utterances so that across participants (in each experiment), each target appeared in every condition. This double counterbalancing ensured that targets were perfectly controlled for grammatical class, duration, frequency, imageability and concreteness. Utterances were selected from a larger set which had been submitted to a cloze probability pre-test (see Appendix A) on the World Wide Web (`http://www.language-experiments.org`) using a minimum of 17 participants per sentence. Pre-test participants were recruited through an email advertisement with the University of Edinburgh and through links from other websites. Each participant saw 84 or 85 sentence frames, one at a time, which they were required to complete with the first single word that occurred to them as the likely end to the sentence. Utterances selected for inclusion in the stimulus set had a mean high cloze probability of 0.84 (range 0.52–1) and a mean low cloze probability of 0. Within utterance pairs, the predictable and unpredictable targets had different phonetic onsets so that the targets would be acoustically distinguishable at word onset[2]. This avoided potential smearing of ERP effect onsets (e.g., Van Petten et al., 1999, see section 4.3.2). Table 5.1 shows an example stimulus set from Experiments 1 and 2.

Table 5.1: Example stimulus set comprising two highly constraining sentence frames, crossed with two target words, which were either predictable or unpredictable in context. Target words are shown in bold. Half of the utterances were disfluent and contained a disfluency (an *er*, a silent pause, or a repetition of the pre-target word) before the target word and is indicated by *.

Predictable	Everyone's got bad habits and mine is biting my	*	**nails**
	That drink's too hot; I've just burnt my	*	**tongue**
Unpredictable	Everyone's got bad habits and mine is biting my	*	**tongue**
	That drink's too hot; I've just burnt my	*	**nails**

Stimuli were digitally recorded (16-bit, 44000Hz mono) by a female native English speaker at natural speaking rate. Fluent and disfluent contexts were recorded for each stimulus pair, with the utterance final word replaced by the word "pen", chosen because the initial plosive meant it could be easily identified in a spectrogram. Each of the target words was recorded as the utterance-final word of a non-experimental fluent carrier sentence that had similar grammatical structure to the experimental sentence. Disfluent contexts were recorded with an *er* before the target, and with

[2]One target pair inadvertently had the same acoustic onset.

other features of disfluency such as vowel lengthening where natural for the speaker. This resulted in 35% of disfluent utterances containing lengthened pre-target words and therefore in clear indications of disfluency before the hesitation itself. Following recording, utterances were edited (using Adobe Audition, `http://www.adobe.com/products/audition/`) to excise the word "pen" and splice the appropriate targets onto the sentence frames. Such splicing meant that across conditions, targets were acoustically identical. Following the editing procedure, all auditory stimuli were converted to 16-bit 22050 Hz .wav files, and the amplitudes normalised so that the acoustic volume was approximately matched across stimuli. Utterances were individually checked (acoustically and visually) to determine the acoustic onset of the targets.

Experiments 1 and 2 focused on the effects of *er* pauses and used the original stimuli. Experiment 3 focused on silent pauses and used modified versions of the stimuli: the disfluent stimuli were created by excising the pre-target *ers* and replacing them with a silence of equivalent duration. Experiment 4 focused on repetitions: disfluent stimuli were created from the fluent stimuli by copying the pre-target word (or for some stimuli several words: see section 8.2.3) and splicing them in.

The experiments all included filler utterances to mask the potentially salient features of the experimental utterances. There were 80 fillers of varying constraint, about similar topics to the experimental utterances, and of similar length, duration and grammatical structure. Forty were fluent and 40 contained disfluencies of various types (*er* pauses, silent pauses, repetitions, and repairs) in various locations. In Experiment 3, a number of the fillers which included an *er* were altered to replace the *ers* with silent pauses of identical duration. In Experiment 4, a number of the fillers which included an *er* were altered to replace the *er* with a repetition. Fillers were recorded as complete utterances, with no splicing.

5.3.2 Stimuli for Experiment 5

The stimuli for Experiment 5 were utterances that were either control utterances that did not include a repair, or repair utterances that included a repair disfluency comprising two consecutive verbs, the second of which was the repair target. Utterances were either *er*-free ($-er$) or included an *er* ($+er$) before the repair or control target word. Targets were followed by a few words to end the sentence. Utterances were constructed in sets such that each repair word was a control word for a corresponding utterance. Furthermore, control and repair targets appeared in $-er$ and $+er$ contexts so that across participants, each target appeared in every condition. This counterbalancing ensured that targets were perfectly controlled for grammatical class, duration, frequency, imageability and concreteness. Table 5.2 shows an example stimulus set from Experiment 5.

Table 5.2: Example stimulus set comprising control utterances or utterances with a repair. Repair and control word targets are shown in bold. Half of the utterances included an *er* before the target word, indicated in square brackets.

control	It was warm today until the sun	[*er*]	**hid**	behind the clouds
repair	It was warm today until the sun went	[*er*]	**hid**	behind the clouds

Stimuli were digitally recorded (16 bit, 44000Hz mono) by a male native English speaker at a natural speaking rate. Utterances were recorded for each stimulus set in each of the four conditions (-*er* control utterances, $-er$ repair utterances; $+er$ control utterances; $+er$ repair utterances). Following recording, utterances were edited (using Adobe Audition, `http://www.adobe.com/products/audition/`) to excise the target word (for -*er* conditions) or the *er* and the target word (for $+er$ conditions) to create four utterance frames. Utterance completions were formed using those from $+er$ utterances from the onset of the *er*. This string was spliced onto the end of $+er$ utterance frames to form the *er* utterances. The *er* was then excised from the utterance completion and the resulting string spliced onto the end of the -*er* utterance frames to form the -*er* utterances. The method ensured that across conditions, target words and the *ers* (in the two +*er* conditions) were acoustically identical. Following the editing procedure, all auditory stimuli were converted to 16-bit 22050Hz .wav files, and the amplitudes normalised so that the acoustic volume was approximately matched across stimuli. Utterances were individually checked (acoustially and visually) to determined the acoustic onset of the targets.

The experiment included filler utterances to mask the potentially salient features of the experimental utterances. There were 160 fillers about similar topics to the experimental utterances, and of similar duration. One hundred and twenty contained *ers* at various locations within the sentences. Fillers were recorded as complete utterances, with no splicing.

5.4 Participants

Participants were recruited from the University of Edinburgh and University of Stirling and were right-handed, native English speakers, aged 16–36 years, with normal (or corrected-to-normal) vision, who reported no hearing or reading difficulties. Participants were compensated at rate of £5 per hour for participating and (in Experiments 3–5) had the option of receiving part payment in course credits. Informed consent was obtained prior to participation and participants were fully debriefed after taking part. The experiments were approved by the University of Edinburgh (Experiment 1) or the University of Stirling (Experiments 2–5) Psychology Ethics Committees.

5.5 Procedure

Experiment 1 was a behavioural Lexical Decision Task. In Experiments 2–5, ERP data was collected while participants listened to speech for comprehension with no secondary task. Experiments were compiled in E-Prime sofware (Psychology Software Tools Inc., `www.pstnet.com`). Auditory stimuli were presented via loudspeakers at a volume comfortable for the participant. The ERP experiments were run from a computer in an adjacent room to the participant and the participants were observed by the experimenter throughout the experiment, via a video link.

There were two parts to Experiments 1–4 and one part to Experiment 5. The first part was designed to investigate the effects of disfluency on language processing and these were assessed using reaction time data from a LDT or from ERP data collected during listening; the second part was designed to investigate the longer-term effects on language representation and these were assessed using data from a recognition memory test. During the listening part of the experiment, participants were unaware that their memory would subsequently be tested. Details of the LDT are given in section 6.2. The remainder of this chapter focuses on the ERP experiments.

5.5.1 Part 1: language processing

In the ERP experiments, participants were told that they would hear a series of utterances, which were re-recorded extracts from natural conversations. Participants were further advised that because the utterances would be heard out of context, some would make more sense than others. They were instructed to listen for understanding, just as they would in a natural situation. There was no other task. It was emphasised to participants that they should relax, keep as still as possible, and fixate their eyes on a cross in the centre of the screen, since any movements would introduce artefacts into the EEG recordings. In Experiments 3–5, participants were shown the effects of their movements and tension on their continuous EEG in an attempt to discourage such actions.

Recordings were presented in blocks lasting 5 to 14 minutes, separated by a break of a few minutes, signalled to the participant by a visual instruction to take a break. In Experiments 3 and 4, participants were given the option of taking extra breaks if they felt it necessary, although no participant took this option. In Experiments 2–5, additional breaks were introduced by the experimenter if necessary (e.g., to fix an electrode). Each block started with a visual reminder for participants to relax, stay still and to keep their eyes on the cross in the centre of the screen. The start of each recording was signalled visually by the appearance (for 250ms) of a yellow fixation cross [+] on a black screen, which flashed blue once (for 250ms) and returned to yellow as the utterance began. The fixation cross remained on the screen for the duration of the utterance, after which the screen was blanked for 1500ms.

EEG recording and data processing

In Experiments 2–4, EEGs were recorded from 61 Ag/AgCl electrodes embedded in an elasticated cap, based on an extended version of the international 10−20 system (Jasper, 1958): Fz, FCz, Cz, CPz, Pz, POz, Oz, Fp1, Fp2, AF7, AF8, AF3, AF4, F7, F8, F5, F6, F3, F4, F1, F2, FT7, FT8, FC5, FC6, FC3, FC4, FC1, FC2, T7, T8, C5, C6, C3, C4, C1, C2, TP7, TP8, CP5, CP6, CP3, CP4, CP1, CP2, P7, P8, P5, P6, P3, P4, P1, P2, PO7, PO8, PO5, PO6, PO3, PO4, O1, O2. Panel (a) in Figure 5.1 shows a diagram of electrode placement. Data were recorded (Neuroscan 4.2 Acquire software, Neuromedical Supplies, `http://www.neuro.com`) using a left mastoid reference and re-referenced off-line to the average of left and right mastoid recordings. EOGs were recorded from electrodes located above and below the left eye to monitor eye blinks and from electrodes located on the outer canthi of both eyes to monitor lateral eye movements. Electrode impedances were kept below

5kΩ. The analogue recordings were amplified (Contact Precision amplifiers)(band pass filter 0.01–40Hz) and continuously digitised (16 bit) at a sampling frequency of 200Hz (5 ms per point). Before offline averaging, the continuous EEG files for each participant were segmented into 1350ms epochs starting 150ms before the target onset, baseline corrected relative to the mean amplitude over the pre-target interval, and screened for artefacts.

In Experiment 5, data were recorded using a different set of amplifiers. EEGs were recorded from 62 electrodes as described above, but including Fpz, CB1, CB2 and omitting AF7 and AF8 (see Figure 5.1 panel b). Data were recorded (NeuroScan 4.3 Acquire sofware, Neuromedical Supplies, `http://www.neuro.com`) using a central scalp reference electrode located between Cz and CPz, but again re-referenced offline to the average of left and right mastoid recordings. The digital recordings were amplified (band pass filter 0.01–40 Hz) and continuously digitised (16 bit) at a sampling frequency of 250Hz (4 ms per point). Before offline averaging, the continuous EEG files for each participant were segmented into 2150ms epochs starting 152ms before the target onset, baseline corrected relative to the mean amplitude over the pre-target interval, and screened for artefacts.

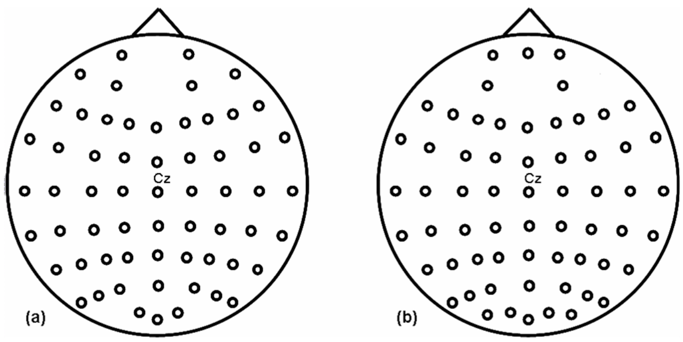

Figure 5.1: Schematic maps of the electrode sites from which ERP data were recorded. Panel (a) shows the 61 electrodes employed in Experiments 2–4, and panel (b) shows the 62 electrodes used in Experiment 5. Electrode Cz is labelled for reference.

Data were processed (NeuroScan 4.3 Edit software, Neuromedical Supplies, `http://www.neuro.com`) and epochs were excluded when any channel became saturated (exceeding 495μV), when baseline drift (absolute difference in amplitude between the first and last data point of each individual epoch) was greater than 50μV per 2000ms on any EEG channel (i.e., 33.75μV for Experiments 2–4 and 53.8μV for Experiment 5), or when amplitude on any channel (excluding VEOG) was greater than 75μV. A minimum of 16 artefact-free trials was required from each participant in each condition to ensure an acceptable signal-to-noise ratio. The effect of eye-blink artefacts was minimised by estimating and correcting their contribution to the ERP waveforms using a regression procedure (Neuroscan Ocular Artifact Reduction). This blink correction involves calculating an average blink from a minimum of 32 blinks for each participant, and removing the contribution of the blink from all other channels on a point-by-point basis (see Chapter 3).

After artefact rejection/correction, the resulting ERPs were smoothed over 5 points so that each sampling point represented the average over the two previous and two subsequent points. ERPs were averaged over the events of interest and then over multiple participants to form the grand average ERPs.

5.5.2 Part 2: recognition memory

In Experiments 1–4, following the first part of the experiments, participants took part in a surprise recognition memory test for the target ("old") words. These "old" words had been either contextually predictable or unpredictable and had appeared in either fluent or disfluent contexts. Words were presented visually (white courier font, 18 point, on a black background) and interspersed with frequency-matched "new" words, which had not been heard at any point during the first part of the experiment. Participants discriminated between old and new words as accurately as possible by pressing one of two response keys (counterbalanced across participants) on a PST Serial Response box (Psychology Tools Inc., `http://www.pstnet.com`). After a blank screen (250ms), the start of each presentation of a target word was signalled by the appearance of a white fixation cross (400ms), which was replaced by the stimulus (750ms), after which the screen was blanked (1750ms). Responses made later than this were not recorded.

5.6 ERP analyses

ERPs were quantified by measuring the mean amplitude (relative to the pre-stimulus baseline) over the time windows of interest, chosen based on previous literature and by inspection of the waveforms.

The data were analysed using Analysis of Variance (ANOVA). As discussed previously in (section 3.5.2), the ANOVA model assumes sphericity—homogenity of variance among levels of each factor—which is typically violated by ERP data. The Greenhouse-Geisser correction was therefore used to adjust the degrees of freedom for non-sphericity; corrected F-ratios and p values are reported where appropriate. Statistical significance was assessed with an alpha level of 0.05; no Bonferroni-correction was applied for planned subsidiary analyses. Only significant effects (and marginally significant effects where considered important for the argument) are reported.

5.6.1 Amplitude analyses

Fluent and disfluent conditions were analysed over the time windows of interest to assess for the presence of significant effects. In Experiments 2–4 this meant comparison of ERPs in response to predictable and unpredictable words. In Experiment 5, this meant comparison of ERPs in response to repair words and fluent controls.

Two ANOVAs were used to compare the experimental manipulations at electrode locations distributed over the scalp, using a “hemispheric analysis” and a “midline analysis". Figure 5.2 shows the electrode sites used in Experiments 2–4. In Experiments 2–4, the first "hemispheric analysis" incorporated factors of predictability [predictable, unpredictable], location [F, FC, C, CP, P], hemisphere [left, right], and site [superior: electrode 1/2, medial: electrode 3/4, inferior: electrode 5/6] and thus assessed for the presence of any hemispheric and site differences.

In the absence of hemispheric differences, or if the hemispheric analysis suggested a larger effect at midline sites, the second "midline analysis" was performed on only midline electrodes and incorporated a factor of location [F, FC, C, CP, P, PO]). Only significant results relevant to the experimental manipulation are reported. Depending on the distribution of the effect under investigation, subsequent analyses were performed at more localised scalp regions. Results from subsequent analyses are only reported when relevant to comparisons between fluent and disfluent conditions.

Because the pre-target baselines for fluent and disfluent materials were recorded from different points in the utterances (disfluent baselines are typically obtained mid-pause), direct comparisons for targets in fluent and disfluent conditions could not be made: instead interaction analyses were used to compare the size of the effects of interest across conditions.

To establish that this comparison was meaningful, it was important to first ensure that there was no distributional difference between the effects obtained in fluent and disfluent conditions (see section 5.6.2).

When significant effects were found for both fluent and disfluent conditions, and there was no evidence for distributional differences, analyses were performed to assess for quantitative differences in the magnitude of the effects between fluent and disfluent conditions which incorporated a factor of fluency [fluent, disfluent] into the "hemispheric" and "midline" analyses. Similarly, when appropriate, analyses were performed to assess for quantitative differences between the effects over time, for fluent and disfluent conditions separately.

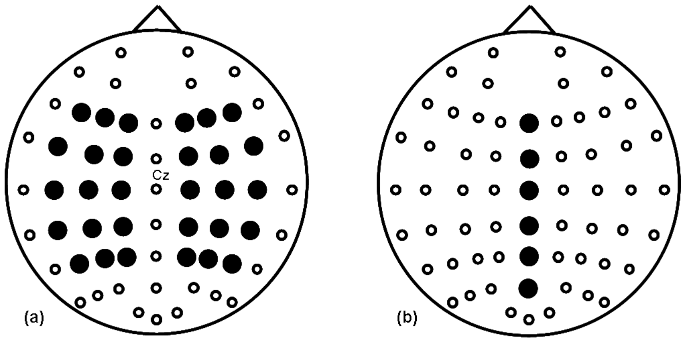

Figure 5.2: Schematic maps of the 61 electrode sites with highlighted sites from which ERPs were analysed in Experiments 2–4. Panel (a) shows the electrodes used in the "hemispheric analyses" which incorporated factors of predictability [predictable, unpredictable], location [F, FC, C, CP, P, PO], hemisphere [L, R], and site [superior: electrode 1/2, medial: electrode 3/4, inferior: electrode 5/6], and panel (b) shows the electrodes used in the "midline analyses" which incorporated factors of predictability [predictable, unpredictable] and location [F, FC, C, CP, P, POz]. Electrode Cz is labelled for reference.

5.6.2 *Topographic analyses*

Topographic analyses were performed on mean voltage difference scores (e.g., between ERPs for unpredictable and predictable targets) which were calculated for each electrode, in both fluent and disfluent conditions, and normalised for amplitude differences using the Max/Min method (see section 3.5.3, McCarthy & Wood, 1985). Topographic analyses assessed whether there were qualitative differences in the distribution of ERP effects between fluent and disfluent utterances, for each of the time windows.

Three types of ANOVAs were performed on the normalised differences. The first used data from all of the electrodes and incorporated factors of fluency [fluent, disfluent] and site and assessed for a distributional difference between conditions over the whole scalp. The other two used data from the electrodes from the "hemispheric" and "midline" amplitude analyses (section 5.6.1).

Where the effects for the time windows were significant, comparisons were also made between successive time windows for fluent and disfluent conditions separately, to assess for distributional changes in the effects over time. These incorporated a factor of epoch into the analyses.

5.7 Memory analyses

Experiments 1–4 included a memory test for the utterance final words. Memory performance was quantified as the probability of correctly identifying "old" (previously heard) words as a function of fluency and predictability. Null responses were excluded from the analyses. To control for differences in individual memory performance, stimulus identity was treated as a random factor. Although adjustments for individual error-rates are traditionally made using measures such as d', they are inappropriate here; the properties of "old" stimuli are determined by their context of occurrence and hence there are no comparable categories of "new" stimuli. Using stimulus identity as a random factor ensures that per-participant biases to respond "old" or "new" are controlled for, across each experiment. Twelve target words were inadvertently repeated in the stimuli for Experiments 1–4, resulting in 148 distinct targets. Analyses with data from the repeated targets removed did not affect the outcome unless reported.

Differences in the memorability of predictable and unpredictable words from fluent and disfluent utterances were assessed using an ANOVA with factors of predictability [predictable, unpredictable] and fluency [fluent, disfluent]. The main ANOVA was followed by planned t-tests to assess for predictability effects for fluent and disfluent conditions separately, and to assess for fluency effects for predictable and unpredictable words separately.

Chapter 6

The effects of *er* filled pauses on language comprehension

Parts of this chapter, including the data presented in Experiment 2 (section 6.3), are reported by Corley et al. (2007). The publication is reproduced in Appendix D.

6.1 Introduction

Approximately six in every hundred words of speech are affected by disfluencies such as the filled pauses, *um* and *er* (Fox Tree, 1995). Moreover, the distribution of disfluency is not arbitrary. For example, filled pauses tend to occur before low frequency (Levelt, 1983) and unpredictable words (Beattie & Butterworth, 1979), in circumstances where the speaker is faced with multiple semantic (Schachter et al., 1994) or syntactic (Holmes, 1988) possibilities, as well as in cases where other types of uncertainty occur (Smith & Clark, 1993). But what are the effects of filled pauses on listeners and on language comprehension?

In this chapter I describe two experiments that investigated the effects of the filled pause *er* on language comprehension, using different methodologies.

6.1.1 Background literature

Although the majority of psycholinguistic research on speech comprehension has been conducted using idealised, fluent utterances, behavioural studies suggest that filled pauses can affect listeners.

Speakers have been rated as less likely to know answers to general knowledge questions when their answers were preceded by disfluent pauses (Brennan & Williams, 1995), suggesting that listeners are sensitive to the uncertainty conveyed by pauses at a metacognitive level. Offline questionnaire studies additionally reveal that filled pauses can influence grammaticality ratings for garden path sentences, showing effects on listeners' final representation of utterances which reflect probable differences in the ways in which they have been processed (Bailey & Ferreira, 2003).

Filled pauses also have effects which can be observed on processing (see section 2.4.1). Using a word monitoring task (Marslen-Wilson & Tyler, 1980), Fox Tree (2001) showed that listeners were faster to identify target words when they were preceded by an *uh*. From this it has been argued that *uhs* heighten listeners' immediate attention to upcoming speech. Attention can also account for Brennan and Schober's (2001) findings that participants were quicker to identify target objects in a visual array when instructions included a hesitation before the target. Two studies by Arnold and colleagues (Arnold et al., 2004, 2007) investigated how listeners respond to disfluency in real time. Using a visual world paradigm (first demonstrated by Cooper, 1974) which was developed by Tanenhaus et al. (1995), participants' eye movements to depictions of objects on a computer screen were monitored as they responded to auditory instructions to move the objects with a computer mouse. The presence of a disfluency (*thee uh*) before the target object increased the probability of an initial eye movement to an object that was considered to be of low accessibility for the speaker and was new in the discourse (Arnold et al., 2004) or unfamiliar (Arnold et al., 2007). In contrast, when the instructions were fluent, participants were more likely to look first at an object that was considered to be of high accessibility for the speaker and had been previously mentioned (Arnold et al., 2004) or familiar (Arnold et al., 2007). Arnold et al.'s (2004, 2007) findings suggest that filled pauses may lead listeners to update their predictions about upcoming words. Specifically, listeners appear sensitive to the fact that speakers find it more difficult to retrieve the names of discourse-new (Arnold et al., 2000; Arnold & Tanenhaus, 2007) and unfamiliar (Arnold et al., 2007) objects and can predict that these items are more likely to be mentioned following disfluency. Thus there is some evidence that disfluent pauses may affect both predictive (Arnold et al., 2004, 2007) and attentional (Brennan & Schober, 2001; Fox Tree, 2001) processes. The methodologies which have been used to date to investigate disfluency processing, however, impose constraints on the interpretations which can be drawn from the data.

The visual world paradigm, which enables the online tracking of eye movements, provides a continuous measure of lexical activation. The methodology is therefore useful for assessing the cognitive processes that underly incremental language processing. However, the paradigm is limited. Listeners are presented *a priori* with a restricted set of candidate images: in Arnold et al. (2004, 2007), these (four) items provided the sentence completions. Such constraints are rare in natural situations. It is possible therefore, that changes to predictive processes observed by Arnold et al. (2004, 2007) are a result, at least in part, of the availability of a potential referent. In other words, although listeners can predict specific completions when information is available, it is unclear whether changes to predictive processes would also be observed in the absence of a constrained set of referents. In addition, the Arnold et al. (2004, 2007) study does not incorporate any assessment of the longer-term consequences of disfluency.

The brief review of the literature above highlights (at least) two outstanding questions about the impact of disfluent pauses on language processing. First, are changes to linguistic processes such as prediction observed in the absence of a constrained set of referents? Secondly, are any online changes observed associated with longer-term consequences? Two experiments were designed to address these issues and investigate the claims that filled pauses affect predictive and attentional processes, using different methodologies to those used previously.

Filled pause disfluencies are associated with retrieval difficulties and therefore tend to occur before less accessible words. Previous studies have manipulated accessibility through discourse-mention of an object (Arnold et al., 2004) and through object familiarity (Arnold et al., 2007). Another way of manipulating accessibility is via contextual word predictability and this is used in the two experiments reported in this chapter. The filled pause *er* (the British English version of *uh*) was used, rather than *um*, for consistency with the majority of studies on filled pauses. Further, it has also been suggested that *uhs* and *ums* may differ in distribution and reflect different production difficulites. For example, in a Dutch corpus, Swerts (1998) observed that *uhs* occur more frequently at phrase-internal than phrase-initial locations which led to the suggestion that *uhs* reflect lexical retrieval difficulties (Shriberg, 1994). *Ums* were more likely to occur at phrase-initial locations leading to the suggestion that they reflect planning difficulties associated with the subsequent production of new, long, or complex clauses.

Because *ers* tend to precede less predictable items in speech (Beattie & Butterworth, 1979) listeners may interpret *ers* as a signal that the following words may not follow from the preceding context. If this is the case, the presence of filled pauses before target words should reduce the standard benefit in the ease of recognising and integrating predictable compared to unpredictable words. Changes to the immediate access and integration of words indicating differences in the processing of the input may result in changes to the representation of the message, particularly of the words immediately following the disfluency. An effect in memory for these words would provide evidence for this, as well as a longer-term correlate of any effects observed at the time the utterances were heard. The effects on language processing were investigated using a Lexical Decision Task (Experiment 1), and ERPs (Experiment 2). Longer-term effects for language representation were assessed using a surprise recognition memory test following the first part of each experiment.

6.2 Experiment 1. Investigating the effects of *ers* on language comprehension using a Lexical Decision Task

In a Lexical Decision Task (LDT), participants discriminate between words and non-words. Response times to correctly identify letter strings as words reflect the ease with which that word is activated from the lexicon (lexical access). LDTs involve responses to words in single or multiple word contexts. Access is easier, and response times quicker, for high frequency words (e.g., Whaley, 1978), for words with fewer syllables (e.g., Eriksen, Pollack, & Montague, 1970), for words with many orthographic neighbours (although clear effects are only observed for low frequency words, e.g., Grainger, 1990) and for words preceded by a semantically related prime (Meyer & Schvaneveldt, 1971; Neely, 1991), which could be a word, a sentence or even a picture. Semantic priming demonstrates the impact of pre-target contextual information on lexical processing; contextually relevant information can aid access. A number of studies show that following the presentation of incomplete sentence frames, words are identified more quickly in a LDT when they are more predictable based on the previous context (Blank & Foss, 1978; Fischler & Bloom, 1979; Kleinman, 1980; Schuberth & Eimas, 1977; Schwanenflugel & Shoben, 1985).

The way in which contextual information exerts an impact on lexical access has been widely investigated using the cross-modal priming technique (e.g., Swinney, 1979; Tabossi, 1988; Zwiterslood, 1989), a variant of the LDT. In a cross-modal priming paradigm, participants listen to speech and make a lexical decision to words presented visually on a computer screen. The relationship between the target word and the spoken context is manipulated. There is still some debate as to whether context affects the initial access of a word or the post-access selection, but either mechanism may come about because context imposes constraints on the semantic aspects of the upcoming word (Schwanenflugel & LaCount, 1988; Tabossi, 1988). The relevant point for the purpose of the current experiment is that predictions generated from contextual information can affect response times to words in a LDT.

6.2.1 Experimental rationale

The aim of the first experiment was to investigate whether *ers* would affect language comprehension, as measured by lexical decision times to subsequent predictable or unpredictable targets in a cross modal-type paradigm, and by a subsequent recognition memory test for the targets. Participants listened to highly constrained fluent or disfluent utterance frames and made a lexical decision to visually presented targets which completed the utterances. These utterance-final targets were either predictable or unpredictable words, or non-word controls.

It was predicted that following the presentation of fluent utterances, times to respond correctly would be faster for contextually predictable compared to unpredictable words. If *ers* heighten listeners' attention to upcoming information (e.g., Brennan & Schober, 2001; Fox Tree, 2001) times to respond correctly should be faster for all targets following an *er* compared to those in fluent utterances. Additionally (or alternatively), if *ers* alter listeners' predictions about upcoming information, namely reducing the extent to which predictable (highly accessible) words are expected (e.g., Arnold et al., 2004), the faster responses to predictable compared to unpredictable words observed in fluent utterances should be reduced by the presence of a pre-target hesitation. If an effect is observed during processing, a longer-term effect in memory might also be observed.

6.2.2 Methods

Stimuli

The stimuli (see section 5.3 and Appendix B) were highly constrained fluent and disfluent utterances coupled with predictable (cloze probability 0.84, range 0.52–1) or unpredictable (cloze probability 0) written target words which completed the utterances. Disfluent utterances included an *er* before the target word and other features of disfluency, such as vowel lengthening where natural for the speaker. For each predictable and unpredictable word, a non-word counterpart was formed with similar numbers of syllables and letters. Table 6.1 shows an example stimulus set.

Table 6.1: Example stimulus set comprising two highly constraining auditorily presented sentence frames, crossed with two visually presented target words which were either predictable or unpredictable in context. Target words are shown in bold. Half of the utterances were disfluent and contained an *er* filled pause before the target, indicated in square brackets.

Predictable	Everyone's got bad habits and mine is biting my	[*er*]	**nails**
	That drink's too hot; I've just burnt my	[*er*]	**tongue**
Unpredictable	Everyone's got bad habits and mine is biting my	[*er*]	**tongue**
	That drink's too hot; I've just burnt my	[*er*]	**nails**

Participants

Thirty two right-handed native English speakers (13 male; mean age 21 years; range 18-41 years) took part in the experiment.

Procedure

There were two parts to the experiment (section 5.5). The first part was designed to investigate the effects of *ers* on processing and these were assessed using reaction time data from a LDT; the second part was designed to investigate the longer-term effects on representation and these were assessed using data from a recognition memory test. During the listening part of the experiment, participants were unaware that their memory would subsequently be tested.

Participants were told that they would hear a series of utterances which were re-recorded extracts from natural conversations. Participants were further advised that because the utterances would be heard out of context, some would make more sense than others. For each utterance, at the point when the final word should be heard, the recording ended and a string of letters appeared on the screen which was a word or a non-word. Participants were instructed to listen for understanding, just as they would in a natural situation. As soon as the letters appeared on the screen, participants had to discriminate between words and non-words as quickly and accurately as possible by pressing one of two response keys using the index fingers of their left and right hands (counterbalanced across participants).

One hundred and sixty experimental stimuli (20 each of fluent predictable, fluent unpredictable, disfluent predictable, disfluent unpredictable, and 40 fluent non-word and disfluent non-word) were presented auditorily until the point at which the utterance-final word should have been heard. At this point the target appeared visually on the screen. Experimental utterances were interspersed with 80 filler utterances which were also interrupted before visual presentation of the final word or non-word (40 of each). Utterances were presented in two blocks lasting approximately 15 minutes each, separated by a break of a few minutes.

The start of each utterance was signalled visually by the appearance (for 250ms) of a yellow fixation cross [+] on a black screen, which flashed blue once (for 250ms) and returned to yellow as the utterance began. At the point where the final word should have been heard, the cross was replaced by the target word. The screen went blank as soon as participants had made their response, or if no response was made, after 2500ms. Following the completion of each utterance, the screen was blanked for 750ms.

Following the first part of the experiment, participants took part in a surprise recognition memory test for the utterance-final ("old") words. Words were presented visually and interspersed with frequency-matched "new" words which had not been seen or heard at any point during the first part of the experiment. Participants discriminated between old and new words as accurately as possible by pressing one of two response keys (counterbalanced across participants).

6.2.3 Lexical decision results

Lexical decision times were quantified as the mean response times to correctly identify letter strings as words by fluency and predictability. Null responses were excluded from the analyses. Twelve target words were inadvertently repeated resulting in 148 distinct targets. Analyses with data from the repeated targets removed did not affect the outcome.

Participants responded correctly to 97.2% of the words (0.31% null responses) and 98.0% of the non-words (0.53% null responses). Results are reported only for words. Figure 6.1 shows the mean response times to correctly identify the utterance-final letter strings as words, by fluency and predictability.

Differences in the response times to correctly identify predictable and unpredictable words from fluent and disfluent utterances were assessed using an ANOVA with factors of predictability [predictable, unpredictable] and fluency [fluent, disfluent]. The main ANOVA was followed by planned t-tests to assess for predictability effects for fluent and disfluent conditions separately, and to assess for fluency effects for predictable and unpredictable words separately.

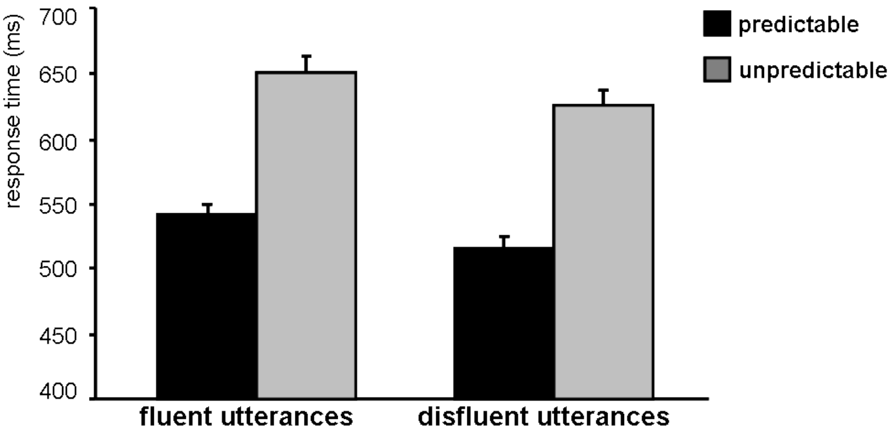

Figure 6.1: Lexical Decision Task response times to correctly identify utterance-final words which were predictable (black) or unpredictable (grey) in their contexts, for fluent and disfluent conditions. Error bars represent one standard error of the mean. Response times were faster for predictable words for both fluent and disfluent conditions. Response times were faster following a disfluency, for both predictable and unpredictable words.

Response times were faster for predictable than for unpredictable words [528ms vs. 639ms: $F1(1,31) = 29.902$, $\eta_p^2 = .491$, $p < .001$, $F2(1,79) = 127.394$, $\eta_p^2 = .617$, $p < .001$], and faster for words following an *er* than for words in fluent utterances [572ms vs. 596ms: $F1(1,31) = 22.409$, $\eta_p^2 = .420$, $p < .001$, $F2(1,79) = 6.562$, $\eta_p^2 = .077$, $p < .012$]. There was no interaction between fluency and predictability [$F < 1$] and therefore no evidence that the presence of an *er* differentially affected the response to predictable compared to unpredictable words.

6.2.4 Recognition memory results

Memory performance was quantified as the probability of correctly identifying "old" words by fluency and predictability. Overall, 66% of the old words were correctly recognised. Figure 6.2 shows the recognition probability of utterance-final words by fluency and predictability.

Memory analyses were performed following the strategy described in the General Methods Chapter (section 5.7). Words that had been unpredictable in their context were more likely to be correctly recognised than words which had been predictable

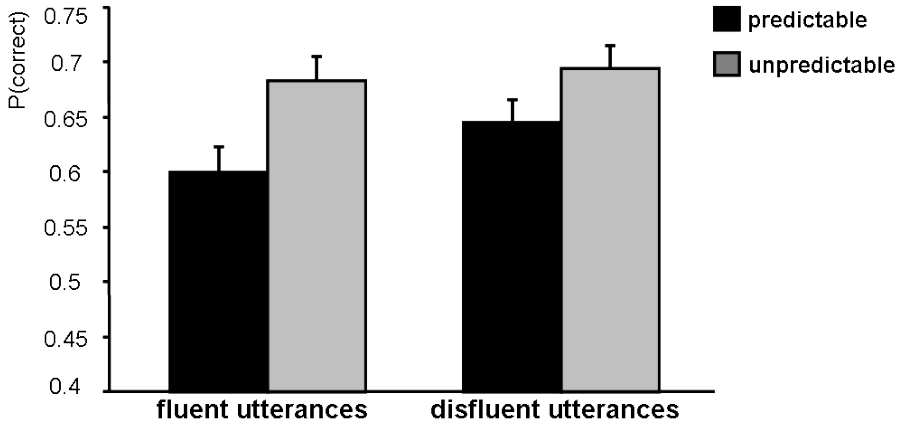

Figure 6.2: Probability of correctly recognising words that were originally predictable (black) or unpredictable (grey) in their contexts, for fluent and disfluent conditions. Error bars represent one standard error of the mean. Unpredictable words were more likely to be correctly recognised than predictable words for both fluent and disfluent conditions. Predictable words were more likely to be correctly recognised for disfluent than for fluent conditions.

[69% vs. 62%: $F(1, 147) = 14.221$, $\eta_p^2 = .088$, $p < .001$]. Planned one-tailed t-tests showed that unpredictable words were more likely to be remembered than predictable words, for both fluent utterances [68% vs. 60%: $t(147) = 2.073, p = .020$] and for disfluent utterances [70% vs. 65%: $t(147) = 3.385, p = .001$]. Numerically, words were more likely to be correctly recognised when they were preceded by an *er*, although this was not significant [67% vs. 64%: $F(1, 147) = 2.505$, $\eta_p^2 = .017$, $p = .116$]. Predictable words were more likely to be correctly recognised if they had occurred following an *er* [65% vs. 60%: $t(147) = 1.805, p = .036$] but there was no difference for unpredictable words [$t < 1$]. There was no interaction between fluency and predictability [$F(1, 147) = 1.006$, $\eta_p^2 = .007$, $p = .318$].

6.2.5 Discussion

The aim of the experiment was to investigate whether the filled pause *er* affects language comprehension. Effects on processing were assessed using an LDT and longer-term consequences for language representation were assessed using a recognition memory test: participants listened for understanding to a series of short utterances and responded to predictable and unpredictable target words which were presented visually interspersed with non-word controls. Utterances were either fluent or disfluent, containing an *er* before the target.

Summary and interpretation

As expected, response times were faster for predictable than for unpredictable targets. As discussed earlier (section 6.2.1) there are two plausible explanations of this finding. First, that context had a pre-lexical influence on word recognition and that at the point of presentation, the most predictable word was already activated (to an extent) by the preceding context. Alternatively, it may be that context had a post-lexical influence on word recognition and that upon presentation of the most predictable word, the preceding context facilitated its access. The relevant point is that lexical decision times were sensitive to contextually generated predictions.

There was also an effect of disfluency: targets preceded by an *er* were responded to more quickly. This finding is compatible with the suggestion that *ers* heighten listeners' attention to upcoming information. There was no interaction between predictability and fluency: *ers* did not reduce the response time difference between predictable and unpredictable words. The LDT therefore failed to find support for the hypothesis that filled pauses alter listeners' predictions about upcoming information. However, this result is at odds with the recognition memory data. As expected, words which had been unpredictable in their original context were more likely to be subsequently recognised than those which had been predictable. More interestingly, predictable words were more likely to be correctly recognised when they followed an *er* (results were only significant in planned t-tests). The result demonstrates a longer-term consequence of disfluency for language representation.

Furthermore, the effect suggests that filled pauses perhaps *do* affect immediate linguistic processes such as prediction, but that the LDT was not sensitive to these processes. Alternatively, perhaps the effects of filled pauses on prediction which have been previously demonstrated using the visual world paradigm (Arnold et al., 2004, 2007) were a result of the constrained set of referents to which the utterances referred. Although the utterances in the present experiment were highly constraining with potentially predictable endings, in the absence of supporting visual information it may not have made sense for listeners to make explicit predictions.

Lexical decisions have been criticised for not being a true measure of the automatic process of lexical access. This is because response times do not just reflect the time associated with access, but also the time associated with non-automatic attentional processes which LDTs typically encourage (e.g., Neely, Keefe, & Ross, 1989; Seidenberg, Waters, Sanders, & Langer, 1984; Shelton & Martin, 1992). Such processes include strategic generation of expectancies and *post-access* checking. Participants were most likely aware that sometimes the target was the most predictable utterance completion. By generating this word in advance, when it was the target they were able to make the "word" response more quickly than for an alternative target. This may have been particularly likely for the disfluent utterances: the *er* may have been an explicit "get ready" signal which cued the generation of a subsequent predictable word and perhaps enabled some degree of motoric preparation in advance of the target appearing. Such an explanation can account for the faster response times for words following disfluency, which was of comparable magnitude for predictable and unpredictable words. The explanation is additionally supported to an extent, by faster responses for the non-words following a disfluency [618ms vs. 643ms: $F1(1, 31) = 11.464$, $\eta_p^2 = .270$, $p = .002$, $F2(1, 79) = 7.992$, $\eta_p^2 = .092$, $p = .006$]. Participants may have also used information after initial lexical access to aid their decision. For example, the existence of a meaningful relationship between the target and the context meant that the target must be a word and participants could respond on the basis of such a semantic-relationship estimate and not on the ease of accessing the target from the lexicon.

Given the contradiction between the lexical decision data and the recognition memory data, it is likely that the LDT is insensitive to predictive processes on which disfluencies exert an immediate influence. In addition, LDTs do not provide insight into the continual processes at work during processing itself.

6.3 Experiment 2. Investigating the effects of *ers* on language comprehension using ERPs

Experiment 2 investigated the effects of *ers* on language comprehension using ERPs (rather than LDTs) as an online measure of processing.

ERPs—neural activity recorded at the scalp, time-locked to the onset of a cognitive event of interest and averaged over multiple events—are ideal for investigating the functional and neural basis of spoken language comprehension. There is no need for a contextually relevant visual presentation with its attendant constraints (e.g., in contrast to the visual world paradigm), and participants need not perform any other task other than listen to the experimental stimuli (e.g., in contrast to LDTs). This means that ERPs provide an ideal means to investigate how listeners process disfluent speech in a situation which is a close analogue to everyday language comprehension (see Chapter 3).

6.3.1 Experimental rationale

The aim of the experiment was to investigate the effects of *ers* on language comprehension, but to address the methodological concerns of the LDT. ERPs were used as an online measure of the processing of post-disfluency words. The likelihood of strategic processing was reduced by the absence of a secondary task. Longer-term effects on representation were again assessed using a recognition memory test.

Participants listened to fluent and disfluent utterances, for understanding, with no secondary task imposed, while EEGs were recorded. ERPs were formed relative to utterance-final targets which were either predictable or unpredictable words. The focus was on the N400, an ERP component associated with the meaningful processing of language (Kutas & Hillyard, 1980, 1984, see section 4.3). During processing, each word must be integrated with its linguistic context, from which it can often be predicted. Where integration is difficult, for example because a word is not predictable, a negative change in voltage is observed at the scalp, relative to more easily integrated words.

If filled pauses alter listeners' predictions about upcoming information, namely reducing the extent to which predictable (highly accessible) words are expected (e.g., Arnold et al., 2004, 2007), it is hypothesised that the large N400 for unpredictable compared to predictable words observed in fluent utterances, will be reduced by the presence of a pre-target *er*. If the ERP data show immediate effects of *ers* on language processing, then as for Experiment 1, an *er* may enhance the recognition probability of subsequent words, at least of those which are contextually predictable.

6.3.2 Methods

Stimuli

The stimuli were edited from those in Experiment 1 (see section 5.3 and Appendix B), and were highly constrained fluent and disfluent utterances ending in predictable (cloze probability 0.84, range 0.52–1) or unpredictable (cloze probability 0) target words. Table 6.2 shows an example stimulus set.

Table 6.2: Example stimulus set comprising two highly constraining sentence frames, crossed with two target words which were either predictable or unpredictable in context. Target words are shown in bold. Half of the utterances were disfluent and contained an *er* before the target word, indicated in square brackets.

Predictable	Everyone's got bad habits and mine is biting my	[*er*]	**nails**
	That drink's too hot; I've just burnt my	[*er*]	**tongue**
Unpredictable	Everyone's got bad habits and mine is biting my	[*er*]	**tongue**
	That drink's too hot; I've just burnt my	[*er*]	**nails**

Participants

Twelve right-handed native English speakers (6 male; mean age 21 years; range 16-35 years) took part in the experiment.

Procedure

The procedure followed the procedure described in the General Methods chapter (see section 5.5). There were two parts to the experiment. The first part was designed to investigate the effects of *ers* on online processing and these were assessed using ERPs formed from EEGs collected during natural listening; the second part was designed to investigate the longer-term effects on representation and these were assessed using data from a recognition memory test. During the listening part of the experiment, participants were unaware that their memory would subsequently be tested.

One hundred and sixty experimental stimuli (40 each of fluent predictable, fluent unpredictable, disfluent predictable, disfluent unpredictable) were presented auditorily. Experimental utterances were interspersed with 80 filler utterances. Utterances were presented in two blocks lasting approximately 15 minutes each, separated by a break of a few minutes. EEG was recorded from 61 scalp electrodes using a left mastoid reference and re-referenced offline to the average of left and right mastoid recordings.

Following the first part of the experiment, participants took part in a surprise recognition memory test for the utterance-final ("old") words. Words were presented visually interspersed with frequency-matched "new" words which had not been heard at any point during the first part of the experiment. Participants discriminated between old and new words as accurately as possible by pressing one of two response keys (counterbalanced across participants).

Before offline averaging, the continuous EEG files for each participant were segmented into 1350ms epochs starting 150ms before the target onset and screened for artefacts. Artefact rejection resulted in the exclusion of 25% of the trials. Grand average ERPs were formed time-locked to the onsets of the utterance-final predictable and unpredictable words from fluent and disfluent utterances making four conditions: fluent predictable, fluent unpredictable, disfluent predictable, disfluent unpredictable, with mean trial numbers of 26, 27, 27, and 25 respectively.

6.3.3 ERP results

Figures 6.3 and 6.4 show the grand average ERPs time-locked to the utterance-final word onsets for fluent and disfluent utterances respectively.

For both fluent (Figure 6.3) and disfluent (Figure 6.4) utterances, relative to predictable words, unpredictable words show a negativity over 300–500ms, which is broadly distributed over the scalp, but appears larger over central and midline sites. The negativity closely resembles N400 effects shown in previous studies (e.g., Kutas & Hillyard, 1980, 1984; Van Berkum, Brown, & Hagoort, 1999a; Van Petten et al., 1999). Following the negativity, differences between fluent and disfluent utterances emerge. For disfluent utterances, unpredictable words show a relative positivity, over left parietal and frontal midline sites. There is no sign of a positivity in fluent utterances.

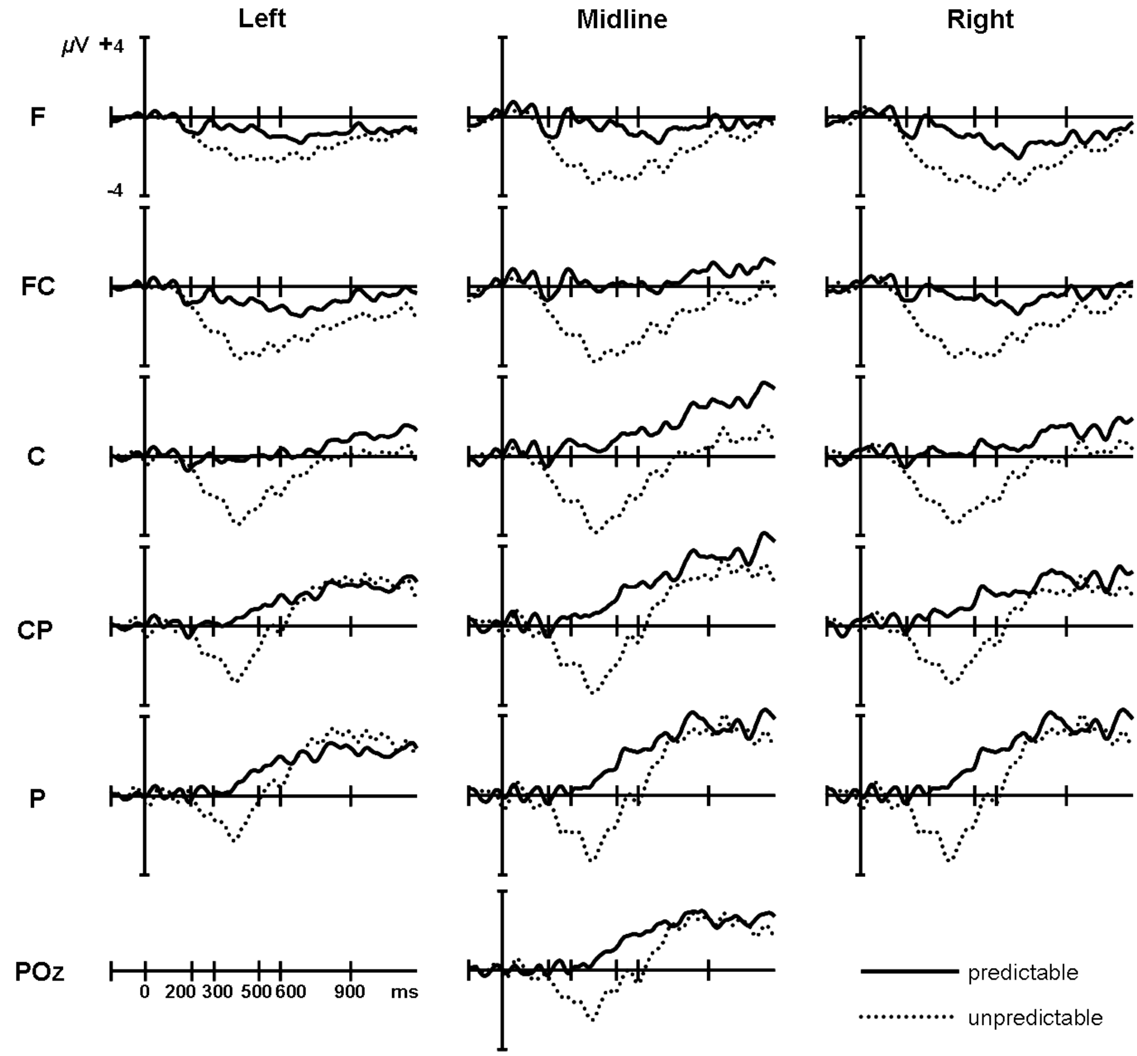

Figure 6.3: Grand average ERPs (n=12) for *fluent* utterances relative to predictable (solid lines) or unpredictable (dotted lines) target word onsets at frontal (F), fronto-central (FC), central (C), centro-parietal (CP), parietal (P), and occipito-parietal (PO) locations, for electrodes at the midline and grouped over left (electrodes 1, 3, 5) and right (electrodes 2, 4, 6) hemispheres. Unpredictable words show a relative negativity, which is broadly distributed over the scalp, but appears larger over central/centro-parietal/parietal and midline sites. The effect emerges around 200ms and is larger in the standard N400 time window (300–500ms).

ERPs were quantified by measuring the mean amplitude of the ERP differences between unpredictable and predictable words, for fluent and disfluent utterances over two time windows: the standard N400 time window (300–500ms) and a later (600–900ms) time window. The topographic distributions of the effects for fluent and disfluent utterances in these time windows are shown in Figure 6.5.

Magnitude and topographic analyses were performed following the strategy described in the General Methods Chapter (sections 5.6.1 and 5.6.2).

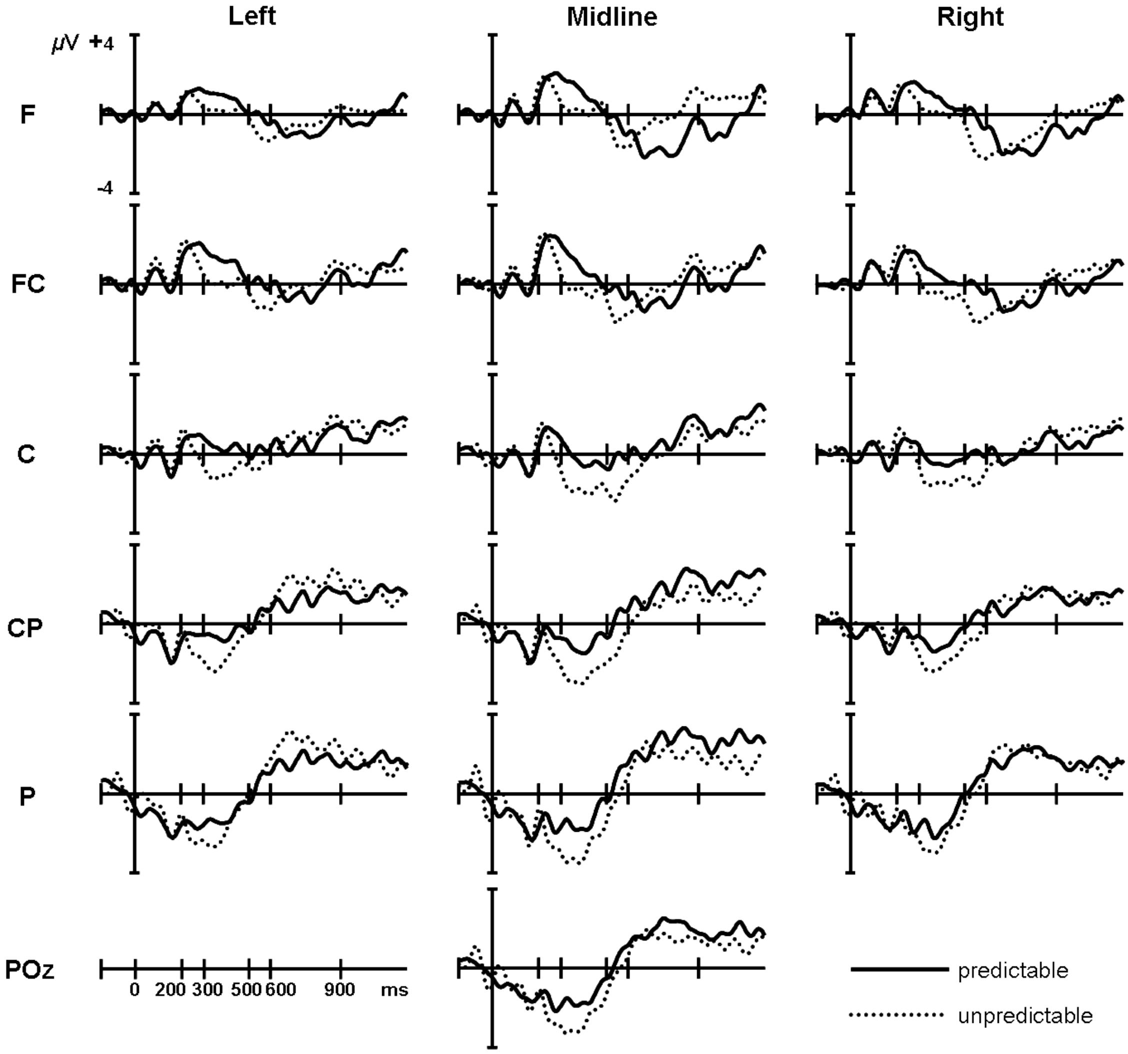

Figure 6.4: Grand average ERPs (n=12) for *disfluent* utterances relative to predictable (solid lines) or unpredictable (dotted lines) target word onsets at frontal (F), fronto-central (FC), central (C), centro-parietal (CP), parietal (P), and occipito-parietal (PO) locations, for electrodes at the midline and grouped over left (electrodes 1, 3, 5) and right (electrodes 2, 4, 6) hemispheres. Unpredictable words show a relative negativity, which is broadly distributed over the scalp, but appears larger over centro-parietal/parietal and midline sites. The effect emerges around 200ms and is larger in the standard N400 time window (300–500ms). Following the negativity unpredictable words show a relative positivity over the frontal midline sites and over left centro-parietal/parietal sites.

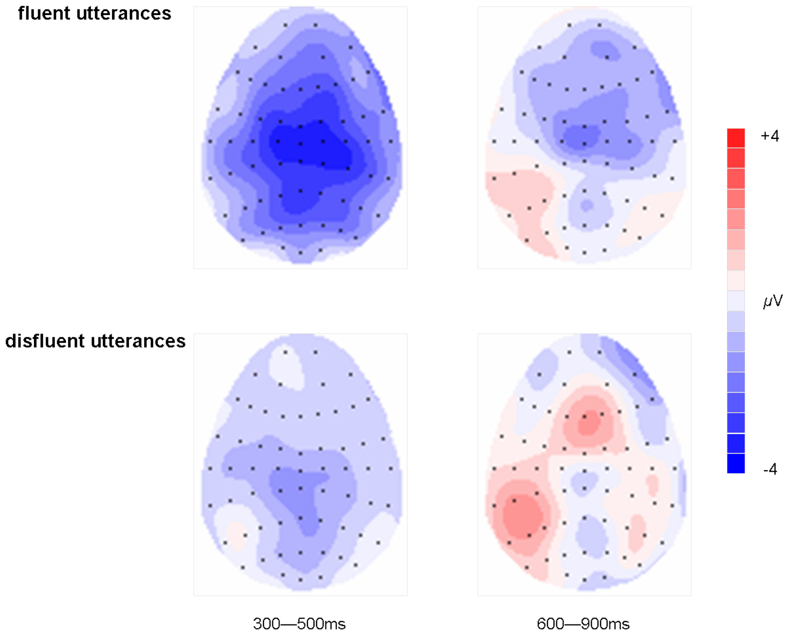

Figure 6.5: Scalp topographies (n=16) showing the predictablility effects over two time windows: 300–500ms and 600-900ms, for *fluent* and *disfluent* utterances. Both fluent and disfluent utterances show a negative effect over the 300–500ms time window, which is larger over centro-parietal/parietal and midline sites. The effect appears larger and more broadly distributed over the scalp for fluent utterances. Over the 600–900ms time window fluent utterances show some indication of a weaker negativity with a more central/fronto-central focus. Disfluent utterances show a positive effect over frontal midline sites and over left centro-parietal/parietal sites.

300–500ms

Over the standard N400 time window of 300–500ms, a topographic analysis incorporating all 61 electrodes provided no evidence of a distributional difference between the effects for fluent and disfluent conditions [for site: $F(60, 660) = 1.70$, $\eta_p^2 = .134$, $p = .191$; other Fs < 1; see Figure 6.5]. Thus there was no reason to suppose that different neural generators are responsible for the recorded effects of predictability and no reason to suppose the engagement of different cognitive processes during this time period. The effects for fluent and disfluent conditions were therefore analysed separately and then quantitatively compared.

Hemispheric analyses established that the distributions of the fluent and disfluent N400 effects were not lateralised. For fluent utterances, there was a main effect of predictability [$F(1, 11) = 43.93$, $\eta_p^2 = .800$, $p < .001$] and an interaction between predictability and site [$F(2, 22) = 8.95$, $\eta_p^2 = .448$, $p = .010$], reflecting a relative negativity which was greater over sites closer to the midline. No other effect was significant. For disfluent utterances, no effect reached significance.

Since no effects involving hemisphere were found for either fluent or disfluent conditions, further analyses concentrated on midline electrodes. For fluent utterances, the midline analysis showed a main effect of predictability [$F(1, 11) = 28.184$, $\eta_p^2 = .719$, $p < .001$], reflecting a relative negativity. For disfluent utterances there was a marginally significant effect of predictability [$F(1, 11) = 3.874$, $\eta_p^2 = .260$, $p < .079$], reflecting a very small relative negativity for unpredictable words.

The quantitative comparison of the effects between fluent and disfluent conditions showed a main effect of predictability [$F(1, 11) = 18.68$, $\eta_p^2 = .629$, $p = .001$], reflecting the relative negativity for unpredictable words and an interaction of fluency with location [$F(5, 55) = 13.03$, $\eta_P^2 = .542$, $p = .002$], reflecting general frontal positivity relative to the baseline in the disfluent conditions. Importantly, fluency interacted with predictability [$F(1, 11) = 7.78$, $\eta_p^2 = .414$, $p = .018$], establishing that the N400 effect for fluent utterances [$2.99\mu V$] was reduced for disfluent utterances [$1.18\mu V$].

As a final check, a topographic analysis was performed for the midline N400 effects to examine whether there was a distributional difference between the N400 effects for fluent and disfluent conditions. There were no observable difference [for location: $F(5, 55) = 1.18$, $\eta_p^2 = .10$, $p = .309$; other Fs < 1].

600–900ms

Over 600–900ms, a topographic analysis incorporating all 61 electrodes provided no evidence of a distributional difference between the effects for fluent and disfluent conditions. Hemispheric and midline topographic analyses also failed to find distributional differences. Thus despite the apparent difference in scalp topographies between the fluent and disfluent conditions (see Figure 6.5) there was no reason to suppose that different neural generators are responsible for the recorded effects of predictability. However, there was no *a priori* reason to quantitatively compare the effects and therefore the effects are only reported for fluent and disfluent conditions separately.

For fluent utterances, the hemispheric analysis showed a marginal effect of predictability [$F(1, 11) = 4.392$, $\eta_p^2 = .285$, $p = .060$], reflecting a weak relative negativity for unpredictable words. The midline analysis showed no significant effects.

For disfluent utterances, the hemispheric analysis showed a three-way interaction between predictability, location and site [$F(8, 88) = 4.344$, $\eta_p^2 = .286$, $p = .026$], reflecting a relative positivity for unpredictable words over the frontal sites close to the midline. The midline analysis showed no significant effects.

Effects over time

For fluent utterances, the topographic analysis incorporating all 61 electrodes, and the hemispheric and midline topographic analyses, provided no evidence of a distributional difference between the effects over 300–500ms and 600–900ms (see Figure 6.5). Thus the effects were quantitatively compared. Although no effects involving hemisphere were found for either the 300–500ms or 600–900ms time windows, the quantitative comparison used the hemispheric analysis because no significant effects had emerged for the midline analysis in the 600–900ms time window. There was a main effect of predictability [$F(1, 11) = 21.205$, $\eta_p^2 = .658$, $p = .001$], and interactions between epoch and predictability [$F(1, 11) = 7.869$, $\eta_p^2 = .417$, $p = .017$], and between epoch, predictability and site [$F(2, 22) = 4.321$, $\eta_p^2 = .282$, $p = .048$], reflecting an N400 which became smaller and more focused over central sites over time.

For disfluent utterances, the topographic analysis incorporating all 61 electrodes, and the hemispheric and midline topographic analyses, provided no evidence of a distributional difference between the effects over 300–500ms and over 600–900ms. Thus the effects were quantitatively compared. Although no effects involving hemisphere were found for either the 300–500ms or 600–900ms time windows, the quantitative comparison used the hemispheric analysis because no significant effects had emerged for the midline analysis in the 600–900ms time window. There was an interaction between epoch and predictability [$F(1, 11) = 6.821$, $\eta_p^2 = .383$, $p = .024$], reflecting the negativity in the early time window and the positivity in the later time window. There was also an interaction between predictability, location and site [$F(8, 88) = 5.177$, $\eta_p^2 = .320$, $p = .016$], reflecting relative negativity for unpredictable words at posterior midline sites.

6.3.4 Recognition memory results

Memory performance was quantified as the probability of correctly identifying "old" words by fluency and predictability. Overall, 64% of the old words were correctly recognised (24% false alarms). Figure 6.6 shows the recognition probability of utterance-final words by fluency and predictability.

Memory analyses were performed following the strategy described in the General Methods Chapter (section 5.7) and used in Experiment 1. Words that had been unpredictable in their context were more likely to be correctly recognised than words that had been predictable [69% vs. 58%: $F(1, 147) = 23.48$, $\eta_p^2 = .138$, $p < .001$]. This was the case for words from fluent utterances [68% vs. 55%: $t(147) = 4.229$, $p < .001$] and for words from disfluent utterances [70% vs. 62%: $t(147) = 2.489$, $p = .014$]. Importantly, disfluency also had a long-term effect: words which were

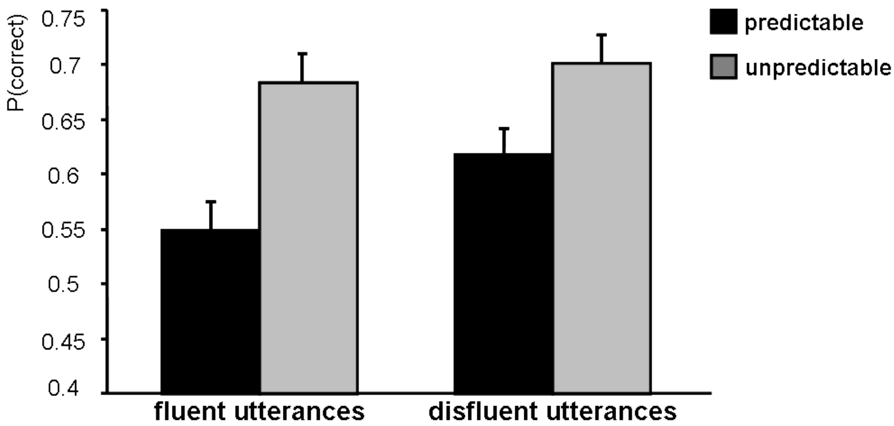

Figure 6.6: Probability of correctly recognising words that were originally predictable (black) or unpredictable (grey) in their contexts, for fluent and disfluent utterances. Error bars represent one standard error of the mean. Unpredictable words were more likely to be correctly recognised than predictable words. Words from disfluent utterances were more likely to be correctly recognised than those from fluent utterances, and this was primarily for predictable words.

preceded by an *er* were more likely to be recognised than those from fluent utterances [66% vs. 62%: $F(1, 147) = 4.31$, $\eta_p^2 = .029$, $p = .040$], primarily predictable words [62% vs. 55%: $t(147) = 2.175$, $p = .031$; for unpredictable words $t < 1$]. There was no interaction between fluency and predictability [$F(1, 147) = 1.127$, $\eta_p^2 = .008$, $p = .290$].

6.3.5 Discussion

The aim of the experiment was to investigate the effects of the filled pause *er* on language comprehension: participants listened for understanding to a series of short utterances which ended in either predictable or unpredictable target words. Utterances were either fluent or disfluent, containing an *er* before the target words.

Summary and interpretation

ERPs showed that when target words were preceded by an *er*, the N400 effect, traditionally associated with the semantic processing of less compared to more predictable words, was substantially reduced. Disfluency also had a longer-term effect:

words following an *er* were more likely to be recognised in a subsequent memory test. These findings suggest that words were processed differently as a consequence of the *er*. Since the N400 differences correspond to differences in memory performance, we can additionally conclude that the ERP differences are not due to contamination of the N400 waveform by spillover effects from the processing of the *er*.

One possible account of these findings is based on linguistic prediction, or expectancy. There is increasing evidence that listeners make online predictions during language comprehension (e.g., Altmann & Kamide, 1999; DeLong, Urbach, & Kutas, 2005; Van Berkum et al., 2005; Wicha, Bates, Moreno, & Kutas, 2003; Wicha, Moreno, & Kutas, 2003, 2004). Furthermore, as discussed earlier (section 6.1) eye-tracking evidence suggests that disfluency marked by filled pauses such as *uh* may lead listeners to update their predictions about upcoming words (Arnold et al., 2004, 2007). Specifically, Arnold et al. (2004, 2007) showed that following a filled pause, listeners were more likely to predict the upcoming mention of an object whose name was considered less accessible for the speaker (a discourse-new or unfamiliar object), albeit from a limited set of candidate referents. In the absence of sufficient (e.g., visual) information, an *er* may cause a reduction in extent to which specific predictions are made, which increases integration difficulty of the predictable word, leading to the N400 attenuation observed here.

Alternatively, it is possible that the *er* affects post-lexical factors, which operate once the target has been heard. Previous research has shown that the N400 is sensitive to differences in the semantic fit of words that do not differ in terms of predictability (Van Berkum et al., 2003). Evidence from the speech production literature (e.g., Levelt, 1989) shows that fillers such as *er* often co-occur with other disfluent phenomena such as repairs. Repairs are hypothesised to be more difficult to integrate syntactically and semantically, because some kind of revision must take place (see Experiment 5, section 9.2). A similar process could be responsible for post-*er* integration in the current experiment: disfluency could add to the difficulty with which both predictable and unpredictable words are integrated.

Both of these views predict that words following disfluency will be better remembered, as demonstrated in the present study. As in Experiment 1, an *er* increased the probability that words were correctly recognised. These data are compatible with that suggestion that *ers* affect attentional processes (Brennan & Schober, 2001; Fox Tree, 1995). The possibility that *ers* affect both prediction and attention is discussed further in section 10.3.

The late positivity

Following the attenuated N400 for disfluent utterances, there was some indication of a relative positivity (600–900ms) for unpredictable relative to predictable words, which was not observed in fluent utterances (although there was no indication of a distributional difference between the effects for fluent and disfluent conditions in this time window, nor between the effects between this time window and the N400 time window for the disfluent conditions or the fluent conditions). One possibility is that the effect reflects spillover from the attenuated N400 for disfluent conditions. Alternatively, the timing and distribution of the effect are compatible with its identification as an (albeit weak) Late Positive Complex (LPC), which is sometimes observed following an N400 and is thought to reflect aspects of memory retrieval and control (Federmeier et al., 2007; Van Petten et al., 1991, see section 4.3.3). These processes may be engaged as listeners maintain the unpredictable word in working memory as they activate the pre-interruption context to resume structural fluency of the message after the interruption. The LPC is discussed further in section 10.3.

Whatever the detailed mechanism, and whatever the interpretation of the later positivity, disfluency clearly affects the processing of language. But what is it about *er* that causes a processing change? One view is that there is nothing intrinsic to *er* that allows it to be understood as a disfluent signal. Instead, the N400 attenuation and subsequent effects on memory might be attributed to timing differences in the fluent and disfluent utterances: in the disfluent utterances, the *er* necessarily introduces more time between the context and the (predictable or unpredictable) target word. Among competing possibilities, listeners could be sensitive to the disfluent phonological form "*er*". The question surrounding the nature of the signal is addressed in subsequent experiments reported in the thesis (see Chapters 7 and 8), but it is secondary to the primary motivation for the current study, which is to demonstrate that filled pauses affect language comprehension and that these effects can be observed using ERPs and a recognition memory test.

The effect of *ers* demonstrated in this experiment is profound: differences in the processing of words in an utterance are visible immediately after the disfluency is encountered, and after a substantial delay (of up to 55 minutes after the first few utterances are heard) participants are more likely to recognise words which have been preceded by disfluency. Using a combined ERP and memory approach, it is clear that the electrophysiological differences observed following *ers* are not merely epiphenomena, but reflect differences in online processing which have longer-term effects on representation.

6.4 Conclusions

The chapter reported two experiments that investigated the effects of the filled pause *er* on language comprehension. *Ers* resulted in faster responses to correctly identify subsequent words in a Lexical Decision Task (Experiment 1), suggesting easier activation of subsequent words, possibly due to heightened attention. *Ers* also resulted in a reduction in the standard N400 effect for subsequent words (Experiment 2), suggesting changes to the relative ease of integration of subsequent predictable compared to unpredictable words, possibly due to a reduction in the extent to which listeners made specific predictions about upcoming words. In addition, unpredictable words elicited a relative positivity when they were preceded by an *er*, but not in fluent utterances. The interpretation of this effect is not clear, but it may reflect memory retrieval and control processes which were engaged as listeners attempt to resume structural fluency after the interruption. Words preceded by *ers* were also more likely to be subsequently recognised (Experiments 1 and 2), demonstrating longer-lasting consequences for representation. The data raise a number of questions, one of which is related to the nature of the disfluent signal. Experiments 3 (Chapter 7) and 4 (Chapter 8) investigate whether the observed effects are specific to the filled pause *er*, or whether the delay introduced by the pause, or the nature of the disfluency, are also important.

Chapter 7

The effects of silent pauses on language comprehension

7.1 Introduction

Experiments 1 and 2 demonstrated that the filled pause *er* affects language comprehension. The results raised a number of interesting questions, one of which concerns the nature of the disfluent signal: what is it about the disfluent utterances which led to the observed effects? One possibility is that the effects are due to the phonological form of the filled pause *er* (e.g., Fox Tree, 2001). Another possibility is that the effects are due to timing differences between the fluent and disfluent utterances (e.g., Bailey & Ferreira, 2003; Brennan & Schober, 2001): in the disfluent utterances, the *er* necessarily introduced more time between the context and the (predictable or unpredictable) target word. Disfluent silent pauses, like filled pauses, interrupt the flow of speech and delay the onset of subsequent new information. Unlike filled pauses they are not vocalised and not always a clear sign of disfluency. In this chapter I describe an experiment that investigated the effects of disfluent silent pauses on language comprehension.

7.2 Experiment 3. Investigating the effects of silent pauses on language comprehension

Experiment 3 investigated the effects of disfluent silent pauses on language comprehension using a modification of Experiment 2.

7.2.1 Background literature

Disfluent silent pauses can be difficult to distinguish from fluent pauses that reflect the natural prosody of an utterance (Duez, 1985; Ferreira, 2007; Zellner, 1994), particularly when they occur between clauses. For this reason, they are often excluded from disfluency studies (Bortfeld et al., 2001) or conflated with filled pauses (e.g., Hawkins, 1971). Within-clause silent pauses, like within-clause filled pauses, are associated with lexical retrieval difficulties (e.g., Goldman-Eisler, 1958a, 1958b; Kircher et al., 2004; Maclay & Osgood, 1959; Martin, 1967) although silent pauses may reflect a less planned interruption than filled pauses (e.g., Clark & Wasow, 1998; Wingate, 1984, see section 2.3.4). However, similar distributions and similar causes of filled and silent pauses do not imply similar effects on langugage comprehension.

Previous studies have not directly investigated the effects of disfluent silent pauses on language comprehension and there is little evidence as to whether the effects of filled pauses are a result of the phonological form of the filler or of the time which they add to the speech signal.

Fox Tree (2001) investigated the effects of filled pauses on language processing using a word monitoring task (see section 2.4.1). Response times to correctly identify target words following an *uh* were compared to response times to target words in utterances where the *uhs* had been excised but several hundred milliseconds of silence preceding the target remained. The results showed faster responses to words preceded by an *uh* than to words preceded by a silence, suggesting that the phonological form *uh* may be important.

The faster response times to post-*uh* targets may reflect a processing benefit for these words or a processing hindrance for post-silence words and so the results would be better compared to response times to words in a fluent condition. Further, the *uhs* and silent pauses were of differing durations and so results may reflect these timing differences.

Brennan and Schober (2001) also investigated the effects of an *uh* on language processing but, unlike Fox Tree (2001) controlled for both the phonological form and the duration of the filled pause by including a condition where the *uhs* had been excised to form fluent utterances and a condition where the *uhs* were replaced by silent pauses of the same durations. Participants listened to instructions and were required to press a button to select a target from a set of geometric objects displayed on a computer screen (see section 2.4.1). Response times and error rates were unaffected by type of pause suggesting that the disfluency effects were driven by the time which the filled pause added to the signal, not by the phonological form.

Bailey and Ferreira (2003) demonstrated effects of filled pauses on listeners' language comprehension using offline grammaticality judgements of garden path sentences which assessed listeners' final representation of the utterance (see section 2.4.1). Although the study did not include a silent pause condition, similar results were observed for filled pauses and environmental noises which led Bailey and Ferreira (2003) to suggest that the effects of filled pauses may be driven by the time which they add to the signal.

Besson et al. (1997) used ERPs to investigate the effects of an unexpected delay on listeners' processing of connected speech. Although the study was not concerned with disfluency, the results are relevant to the current investigation of the effects of disfluent silent pauses. Participants heard utterances which were either highly constrained proverbs which ended in predictable words, or unconstrained and hence ended in unpredictable words. Sometimes the utterance-final critical word was delayed unexpectedly by 600ms. Following the pause, which was associated with the elicitation of a negative-positive complex (discussed further in section 9.2), unpredictable words elicited an N400 relative to predictable words which onset around 250ms. In comparison, the N400 for standard utterances onset earlier, around 150ms. The authors suggest that the later onset of the N400 following an unexpected pause may be because of the absence of co-articulatory cues, or because of the surprise of not hearing a word when it was expected. The authors do not report any differences in amplitude of the N400 between the standard utterances and those which included a delay.

The evidence for the importance of the phonological form of the filler on filled-pause effects is limited and inconsistent. Experiment 3 was designed to investigate whether the ERP and memory effects that were observed in Experiment 2 would extend to silent pause disfluencies.

7.2.2 Experimental rationale

The aim of the experiment was to investigate whether silent pauses would lead to effects on language comprehension that were similar to those observed with *ers*, in Experiment 2 (see section 6.3). Such findings would suggest that the effects are not driven by the phonological form of the filled pause and do not depend on a vocalisation, and would implicate temporal delay as a causal factor. The online processing of post-disfluency words was again assessed using ERPs: specifically, the N400 effect was used as an index of semantic integration processes, although effects in other time windows were analysed. Longer-term effects on representation were assessed using a recognition memory test.

7.2.3 Methods

Stimuli

The stimuli were edited from those in Experiment 2 (see section 5.3 and Appendix B), and were highly constrained fluent and disfluent utterances ending in predictable (cloze probability 0.84, range 0.52–1) or unpredictable (cloze probability 0) target words. Disfluent utterances were edited from the fluent utterances by excising the *er* before the utterance-final target and replacing it with a silence of identical duration, to form a silent pause before the utterance-final word. Table 7.1 shows an example stimulus set.

Table 7.1: Example stimulus set comprising two highly constraining sentence frames, crossed with two target words which were either predictable or unpredictable in context. Target words are shown in bold. Half of the utterances were disfluent, and contained a silent pause before the target word, indicated in square brackets.

Predictable	Everyone's got bad habits and mine is biting my	[...]	**nails**
	That drink's too hot; I've just burnt my	[...]	**tongue**
Unpredictable	Everyone's got bad habits and mine is biting my	[...]	**tongue**
	That drink's too hot; I've just burnt my	[...]	**nails**

Participants

Sixteen right-handed native English speakers (6 male; mean age 22 years; range 17–29 years) took part in the experiment.

Procedure

The procedure followed the procedure described in the General Methods Chapter (section 5.5) and which was followed in Experiments 2. There were two parts to the experiment. The first part was designed to investigate the effects of silent pause disfluencies on online processing and these were assessed using ERP data collecting during natural listening; the second part was designed to investigate the longer-term effects on representation and these were assessed using data from a recognition memory test. During the listening part of the experiment, participants were unaware that their memory would subsequently be tested.

One hundred and sixty experimental stimuli (40 each of fluent predictable, fluent unpredictable, disfluent predictable, disfluent unpredictable) were presented auditorily. Experimental utterances were interspersed with 80 filler utterances. Utterances were presented in four blocks lasting approximately 7 minutes each (12 participants) or two blocks (4 participants) lasting approximately 15 minutes each, separated by a break of a few minutes. EEG was recorded from 61 scalp electrodes using a left mastoid reference and re-referenced offline to the average of left and right mastoid recordings.

Following the first part of the experiment, participants took part in a surprise recognition memory test for the utterance-final ("old") words. Words were presented visually and interspersed with frequency-matched "new" words, which had not been heard at any point during the first part of the experiment. Participants discriminated between old and new words as accurately as possible by pressing one of two response keys (counterbalanced across participants).

Before offline averaging, the continuous EEG files for each participant were segmented into 1350ms epochs starting 150ms before the target onset and screened for artefacts. Artefact rejection resulted in the exclusion of 30% of the trials. Grand average ERPs were formed, time-locked to the onsets of the utterance-final predictable and unpredictable words from fluent and disfluent utterances making four conditions: fluent predictable, fluent unpredictable, disfluent predictable, disfluent unpredictable, with mean trial numbers of 24, 23, 24, and 25 respectively.

7.2.4 ERP results

Figures 7.1 and 7.2 show the grand average ERPs time-locked to the utterance-final word onsets for fluent and disfluent utterances respectively.

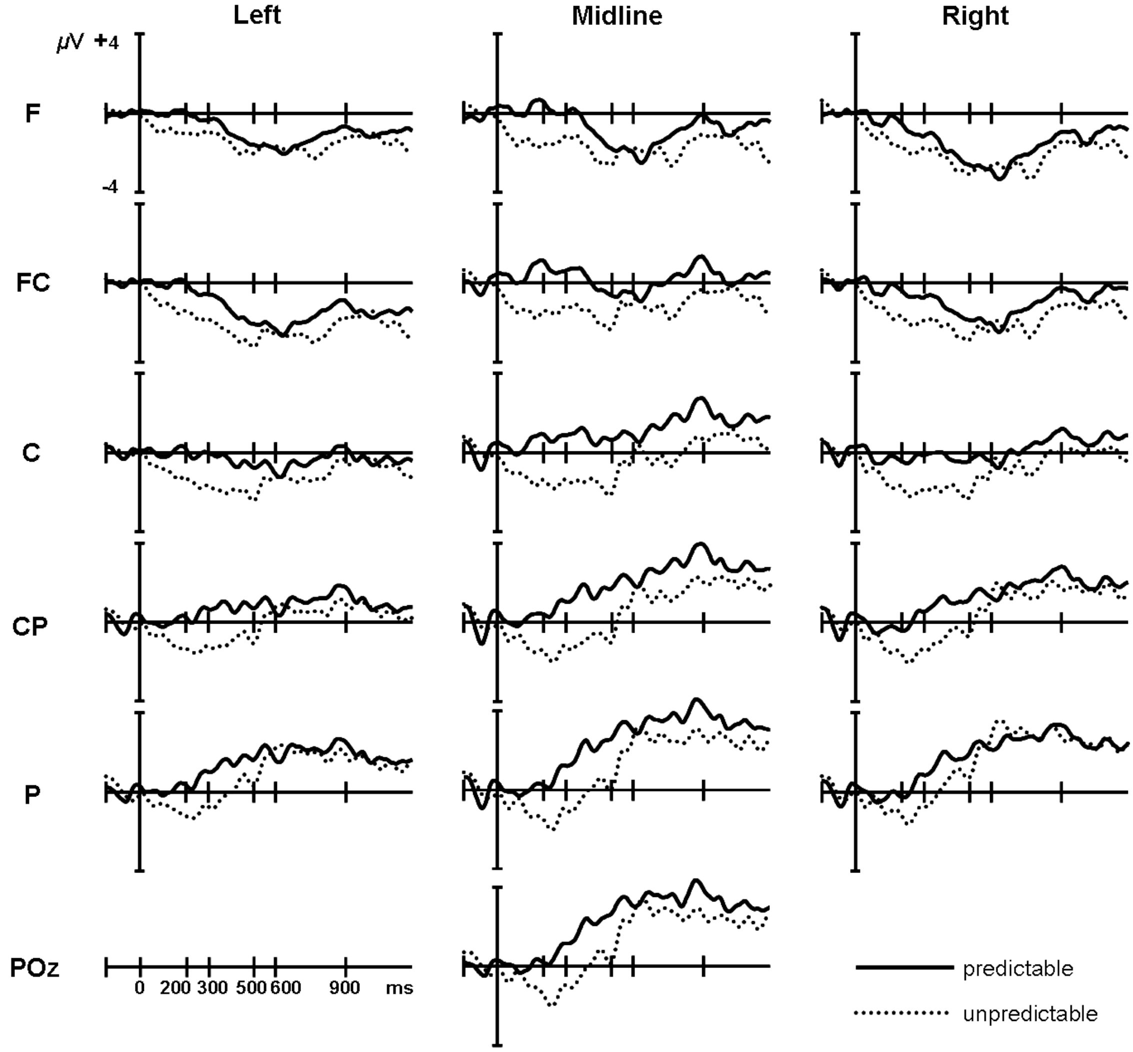

Figure 7.1: Grand average ERPs (n=16) for *fluent* utterances relative to predictable (solid lines) or unpredictable (dotted lines) target word onsets at frontal (F), fronto-central (FC), central (C), centro-parietal (CP), parietal (P), and occipito-parietal (PO) locations, for electrodes at the midline and grouped over left (electrodes 1, 3, 5) and right (electrodes 2, 4, 6) hemispheres. Unpredictable words show a relative negativity which is broadly distributed over the scalp. The effect emerges very soon after word onset, particularly over frontal and over left hemisphere sites and is present over the standard N400 time window (300–500ms).

For both fluent utterances (Figure 7.1) and disfluent utterances (Figure 7.2), relative to predictable words, unpredictable words show a negativity over the 300–500ms

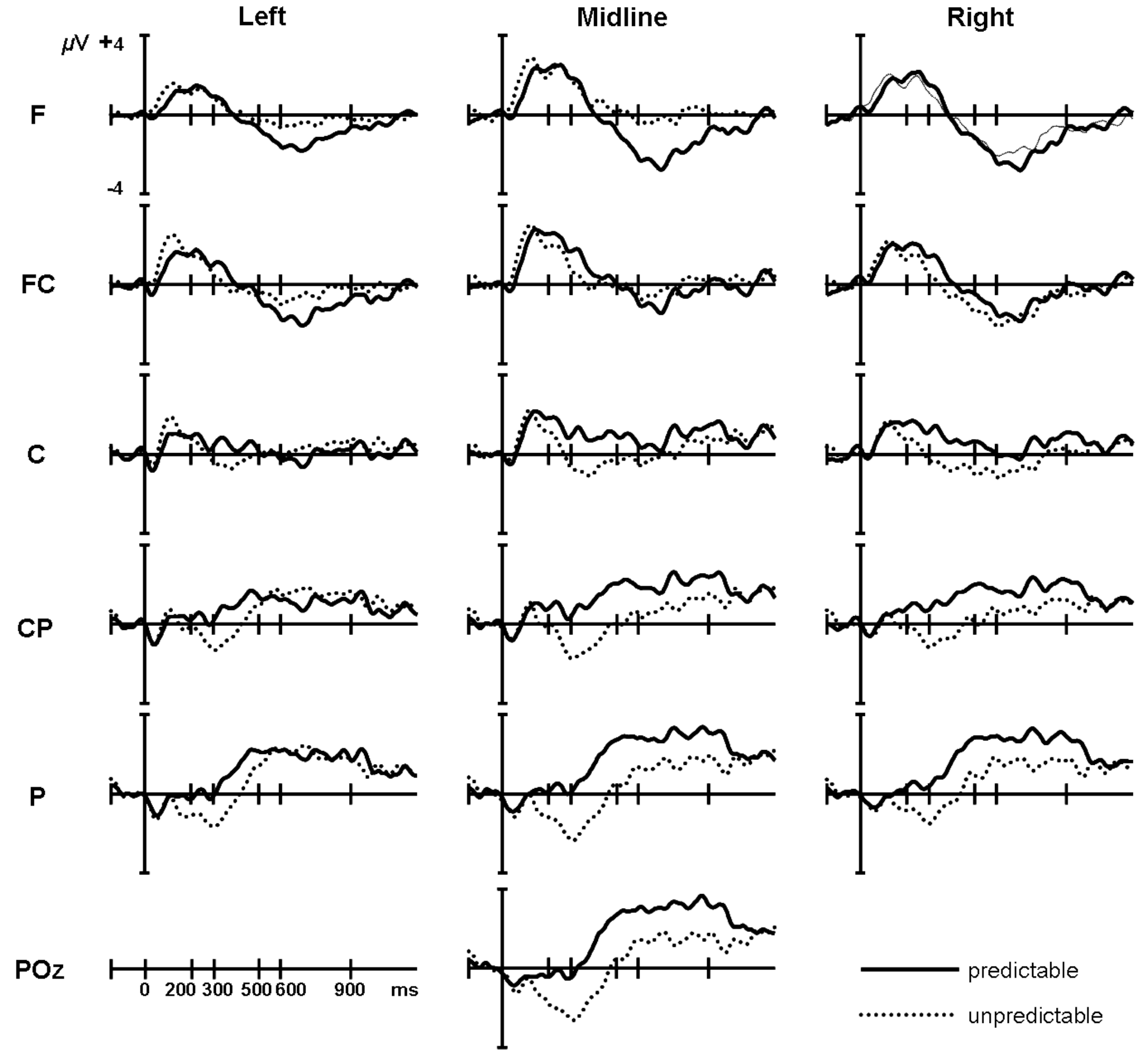

Figure 7.2: Grand average ERPs (n=16) for *disfluent* utterances relative to predictable (solid lines) or unpredictable (dotted lines) target word onsets at frontal (F), fronto-central (FC), central (C), centro-parietal (CP), parietal (P), and occipito-parietal (PO) locations, for electrodes at the midline and grouped over left (electrodes 1, 3, 5) and right (electrodes 2, 4, 6) hemispheres. Unpredictable words show a relative negativity over central/centro-parietal sites, particularly at the midline. The effect emerges around 100ms and is larger in the standard N400 time window (300–500ms). Following the negativity unpredictable words show a relative positivity over frontal/fronto-central sites with a left hemisphere bias.

time window, which is broadly distributed over the scalp, but appears larger over central/centro-parietal and midline sites. These features are compatible with its identification as an N400 effect. Following the N400 effect, differences emerge between fluent and disfluent utterances. For fluent utterances, the relative negativity continues, but appears smaller and more focused at central sites. For disfluent utterances, unpredictable words show a relative positivity over bilateral frontal and fronto-central sites and over the left hemisphere. Differences between fluent and disfluent utterances are also apparent preceding the N400 effect. For fluent utterances unpredictable words show a relative negativity which is more focused over frontal sites than a typical N400 effect. For disfluent utterances there is no indication of such an early onsetting negativity.

ERPs were quantified by measuring the mean amplitude of the ERP difference for predictable and unpredictable words, for fluent and disfluent utterances over three time windows: the standard N400 time window (300–500ms), plus an early (0–200ms) and a later (600–900ms) time window. Topographic distributions of the effects (difference between ERPs for unpredictable and predictable words) for fluent and disfluent utterances over these time windows are shown in Figure 7.3.

Magnitude and topographic analyses were performed following the strategy described in the General Methods Chapter (sections 5.6.1 and 5.6.2) and used in Experiment 2.

300–500ms

Over the standard N400 time window, 300–500ms, a topographic analysis incorporating all 61 electrodes provided no evidence of distributional differences between the effects for fluent and disfluent utterances [for site: $F(60, 900) = 5.244$, $\eta_p^2 = .259$, $p = .004$; other Fs < 1; see Figure 7.3]. Thus there was no reason to suppose that different neural generators are responsible for the recorded effects of predictability and no reason to suppose the engagement of different cognitive processes during this time period. The effects for fluent and disfluent utterances were therefore analysed separately and then quantitatively compared.

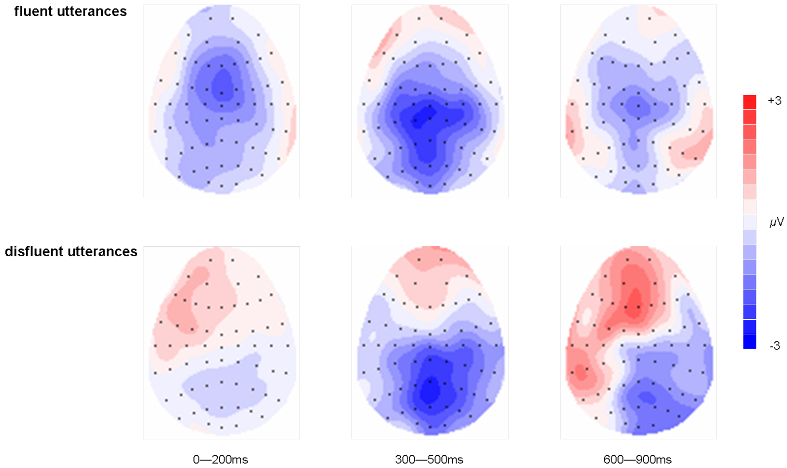

Figure 7.3: Scalp topographies (n=16) showing the predictability effects over three time windows: 0–200ms, 300–500ms and 600-900ms, for *fluent* and for *disfluent* utterances. Fluent utterances show two negative effects with a broad scalp distribution. Over the 0–200ms time window the negativity is focused over frontal/fronto-central sites; over 300–500ms and 600–900ms the negativity is focused over central/centro-parietal sites, although weaker in the later time window. Disfluent utterances also show a negative effect with a broad scalp distribution that is focused over central/centro-parietal sites over 300–500ms. Over 0–200ms there is some indication of a similar, but weaker negativity. Over 600–900ms there is a positivity over frontal/fronto-central sites with a slight left hemisphere bias and over left centro-parietal sites. The positivity is accompanied by posterior negativity.

Hemispheric analyses established that the distributions of both the fluent and disfluent N400 effects were not lateralised. For fluent utterances, there was a main effect of predictability [$F(1, 15) = 4.756$, $\eta_p^2 = .241$, $p = .046$] and an interaction between predictability and site [$F(2, 30) = 9.740$, $\eta_p^2 = .394$, $p = .004$], reflecting a relative negativity for unpredictable words, which was greater towards the midline. No other effect was significant. For disfluent utterances, the main effect of predictability did not reach significance [$F(1, 15) = 2.344$, $\eta_p^2 = .135$, $p = .147$], but there was a three-way interaction between predictability, location and site [$F(8, 120) = 5.385$, $\eta_p^2 = .264$, $p = .006$] reflecting a focused negativitity for unpredictable words over posterior midline sites, typical of an N400 effect.

Since no effects involving hemisphere were found for either fluent or disfluent conditions, further analyses concentrated on midline electrodes. For fluent conditions the midline analysis showed a main effect of predictability [$F(1,15) = 7.150$, $\eta_p^2 = .323$, $p = .017$], reflecting a relative negativity for unpredictable words. For disfluent utterances, there was a marginally significant main effect of predictability [$F(1,15) = 4.024$, $\eta_p^2 = .212$, $p = .063$] and an interaction between predictability and location [$F(5,75) = 5.215$, $\eta_p^2 = .258$, $p = .027$], reflecting a relative negativity which was greater at more posterior locations.

The quantitative comparison of the effects between fluent and disfluent utterances showed a main effect of predictability [$F(1,15) = 10.334$, $\eta_p^2 = .408$, $p = .006$], reflecting the greater negativity for unpredictable words. There was no interaction between fluency and predictability [$F < 1$] and therefore no evidence of a difference between the amplitude of the N400 effect for fluent utterances [2.035μV] and for disfluent utterances that included a silent pause [1.502μV].

0–200ms

Over 0–200ms, a topographic analysis incorporating all 61 electrodes showed a marginally significant interaction between fluency and site [$F(60,900) = 2.6$, $\eta_p^2 = .148$, $p = .058$; see Figure 7.3], suggesting distributional differences between the effects for fluent and disfluent utterances. This finding was supported by a hemispheric analysis which revealed interactions betweens fluency and location [$F(4,60) = 4.170$, $\eta_p^2 = .218$, $p = .042$] reflecting greater negativity at frontal locations for fluent utterances, and between fluency and hemisphere [$F(1,15) = 7.243$, $\eta_p^2 = .326$, $p = .017$], reflecting a slight left lateralised frontal positivity for disfluent utterances. Distributional differences were further supported by a midline analysis which showed an interaction between fluency and location [$F(5,75) = 4.623$, $\eta_p^2 = .236$, $p = .033$], reflecting the negativity at frontal locations for disfluent utterances. The distributional differences suggested that (at least partially) different neural generators are responsible for the recorded effects of predictability and that different cognitive processes were engaged during this time period. The effects for fluent and disfluent utterances were therefore analysed separately and not quantitatively compared.

For fluent utterances, the hemispheric analysis provided no evidence of any laterality differences. There was however, a main effect of predictability [$F(5, 75) = 11.127$, $\eta_p^2 = .426$, $p = .005$], reflecting relative negativity for unpredictable words, and interactions between predictability and site [$F(2, 30) = 20.825$, $\eta_p^2 = .581$, $p < .001$] and between predictability, location and site [$F(8, 120) = 6.296$, $\eta_p^2 = .295$, $p < .001$], reflecting greater negativity at frontal locations towards the midline. The midline analysis revealed a main effect of predictability [$F(1, 15) = 15.990$, $\eta_p^2 = .516$, $p = .001$], reflecting a relative negativity.

For disfluent utterances, the hemispheric analysis suggested laterality differences. There was no main effect of predictability, but interactions between predictability and hemisphere [$F(1, 15) = 5.944$, $\eta_p^2 = .284$, $p = .028$] and between predictability, hemisphere, and site [$F(2, 30) = 6.179$, $\eta_p^2 = .292$, $p = .016$] reflecting a small relative positivity over the left hemisphere, at inferior sites, but a small relative negativity. The midline analysis revealed no significant effects.

600–900ms

Over 600–900ms, a topographic analysis incorporating all 61 electrodes showed an interaction between fluency and site [$F(60, 900) = 2.877$, $\eta_p^2 = .161$, $p = .030$; see Figure 7.3], suggesting distributional differences between the effects for fluent and disfluent utterances. These findings were supported by interactions between fluency and location for a hemispheric analysis [$F(4, 60) = 4.802$, $\eta_p^2 = .242$, $p = .025$] and for a midline analysis [$F(5, 75) = 4.499$, $\eta_p^2 = .231$, $p = .016$], reflecting frontal positivity for the disfluent utterances and centro-parietal/parietal positivity for fluent utterances. Given their distributional differences, the effects for fluent and disfluent utterances were analysed separately and not quantitatively compared.

For fluent utterances, the hemispheric analysis provided no evidence for any laterality differences. There was an interaction between predictability and site [$F(2, 30) = 4.710$, $\eta_p^2 = .239$, $p = .041$] and an interaction between predictability, location and site [$F(8, 120) = 7.971$, $\eta_p^2 = .347$, $p < .001$], reflecting negativity at central locations, particularly midline sites. The midline analysis revealed no significant effects reflecting the very focused distribution and small amplitude of the negativity.

For disfluent utterances, the hemispheric analysis suggested laterality differences. There was an interaction between predictability and location [$F(4, 60) = 9.181$, $\eta_p^2 = .380$, $p = .001$], and interactions between predictability and hemisphere [$F(1, 15) = 6.298$, $\eta_p^2 = .296$, $p = .024$], between predictability, location and site [$F(8, 120) = 17.247$, $\eta_p^2 = .535$, $p < .001$], between predictability, hemisphere and site [$F(2, 30) = 11.810$, $\eta_p^2 = .441$, $p = .001$] and between predictability, location, hemisphere and site [$F(8, 120) = 2.494$, $\eta_p^2 = .143$, $p = .043$]. These effects reflect the relative positivity at frontal locations and over the left hemisphere, which at frontal locations was greater at more superior sites, and over the left hemisphere was greater at more inferior sites particularly at posterior locations. The midline analysis revealed an interaction between predictability and location [$F(5, 75) = 17.500$, $\eta_p^2 = .538$, $p < .001$], reflecting greater positivity at the frontal location.

Effects over time

For fluent utterances, the topographic analysis incorporating all 61 electrodes provided no evidence for a distributional difference between the effects over 0–200ms and 300–500ms. However, both hemispheric and midline topographic analyses suggested a distributional difference (see Figure 7.3). The midline topographic analysis showed an interaction between epoch and location [$F(5, 75) = 6.010$, $\eta_p^2 = .286$, $p = .012$], reflecting the frontal focus of the negativity over the earlier time window and the more posterior focus over the later time window. The hemispheric topographic analysis revealed that this was particularly the case for sites closer to the midline: there was an interaction between epoch, location and site [$F(8, 120) = 7.671$, $\eta_p^2 = .338$, $p < .001$] reflecting the greater negativity at the frontal/posterior locations for sites closer to the midline. Given their distributional differences, the effects for the two epochs were not quantitatively compared.

The topographic analysis incorporating all 61 electrodes provided no evidence for distributional differences between the fluent effects over 300–500ms and 600–900ms, however the hemispheric topographic analysis did suggest a distributional difference. There was an interaction between epoch and location [$F(4, 60) = 4.099$, $\eta_p^2 = .215$, $p = .047$] and between epoch, location and site [$F(8, 120) = 4.509$, $\eta_p^2 = .231$, $p = .012$], reflecting a broader distribution of negativity over scalp locations in the earlier time window, and a more focused negativity in the later time window, particularly over central midline sites. Given their distributional difference, the effects for the two epochs were not compared.

For disfluent utterances, the topographic analysis incorporating all 61 electrodes provided no evidence of distributional differences between the effects over 0–200ms and 300–500ms. Given no evidence for distributional differences, the effects for the two time windows were quantitatively compared. Since effects involving hemisphere were found for the 0–200ms time window (though not for the 300–500ms time window), the quantitative comparison of the effects focused on the hemispheric electrodes. The hemispheric analysis showed a four-way interaction between epoch, predictability, location and site [$F(8, 120) = 5.938$, $\eta_p^2 = .284$, $p = .004$], reflecting a greater negativity in the later epoch, particularly at posterior locations and superior sites. There were also interactions between predictability, location and site [$F(8, 120) = 3.184$, $\eta_p^2 = .175$, $p = .038$], reflecting greater negativity at posterior locations and superior sites and between predictability, hemisphere and site

[$F(5, 75) = 5.942$, $\eta_p^2 = .284$, $p = .021$], which reflects greater positivity over the left inferior sites in the early time window.

As a final check, a topographic analysis was performed to examine whether there were any distributional differences between the two time windows at these electrodes: the hemispheric topographic analysis did suggest a distributional difference, although only over sites that were not deemed directly relevant to the current investigation. There was a four-way interaction between epoch, location, hemisphere and site [$F(8, 120) = 3.147$, $\eta_p^2 = .173$, $p = .022$], reflecting the effect over left inferior sites, which was positive in the earlier time window, but negative in the later time window.

The topographic analysis incorporating all 61 electrodes provided no evidence of distributional differences between the disfluent effects over 300–500ms and 600–900ms. However, the hemispheric topographic analysis did suggest a distributional difference. There was a weak interaction between epoch and hemisphere [$F(1, 15) = 4.317$, $\eta_p^2 = .223$, $p = .055$] and between epoch, hemisphere and site [$F(2, 30) = 5.122$, $\eta_p^2 = .255$, $p = .022$], reflecting the left lateralised positivity in the later time window, particularly over inferior sites. Given their distributional differences, the effects for the two time windows were not quantitatively compared.

7.2.5 Recognition memory results

Memory performance was quantified as the probability of correctly identifying "old" words, by fluency and predictability. Overall, 52% of the old words were correctly recognised[1] (20% false alarms). Figure 7.4 shows the recognition probability of utterance-final words by fluency and predictability.

[1] One item was corrupted and excluded from the analyses. In addition, 5% of the items were not responded to by participants (within the allocated time) resulting in no data for these items. These items are excluded from the analyses.

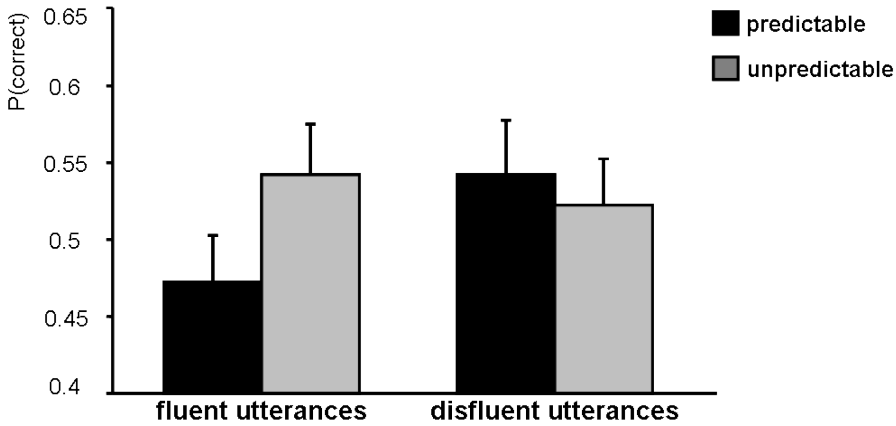

Figure 7.4: Probability of correctly recognising words that were originally predictable (black) or unpredictable (grey) in their contexts, for fluent and disfluent utterances. Error bars represent one standard error of the mean. Predictable words were marginally more likely to be correctly recognised when utterances were disfluent than when they were fluent, but there were no differences for unpredictable words.

Memory analyses were performed following the strategy described in the General Methods Chapter (section 5.7) and used in Experiments 1 and 2.

There were no main effects of predictability or fluency [Fs < 1]. However, there was a weak interaction between fluency and predictability [$F(1, 113) = 3.226$, $\eta_p^2 = .028$, $p < .075$], reflecting the tendency for words that were predictable utterance endings to be more likely to be recognised when utterances were disfluent than fluent [54% vs. 47%: $t(130) = 1.406$, $p = .081$] which was not the case for words which were unpredictable [52% vs. 54%: $t < 1$]. The marginally significant result for predictable words follows the same pattern as the results in Experiments 1 and 2, although performance is somewhat lower in all conditions.

7.2.6 Discussion

The aim of the experiment was to investigate whether the effects of the filled pause *er* on language comprehension that were observed in Experiment 2 (see section 6.3) are driven by phonological form of the filler, or could be (at least partly) due to the delay which they introduce into the signal. The effects of disfluent silent pauses on language comprehension were investigated using a modification of Experiment 2: participants listened for understanding to a series of short utterances that ended in either predictable or unpredictable target words. Utterances were either fluent or disfluent containing a silent pause before the target words.

Summary and interpretation

ERPs showed the presence of an N400 effect for both fluent and disfluent utterances. In contrast to the results for *ers* (Experiment 6.3), the amplitude of the N400 was unaffected by the presence of a silent pause. There is therefore no evidence that silent pauses affect the difference between the ease of integrating predictable and unpredictable words and no support for the possibility that the effects of *ers* on the N400 observed in Experiment 6.3 are due to the increase in time which *ers* add to the speech signal. The data are compatible with the possibility that the phonological form of the filler *er* is important. If as suggested previously (section 6.3.5), the attenuation of the N400 for targets preceded by an *er* reflects a reduction in the extent to which listeners make predictions following such disfluencies, the lack of an attenuation for targets preceded by a silent pause suggests that the silent pauses in the present experiment did not affect predictive processes. Prediction is discussed further in section 10.3.1.

However, analyses of the ERPs after the N400 effect provided evidence for processing differences between fluent and disfluent utterances. For fluent utterances, the N400 effect continued beyond the 300–500ms epoch and was still significant (although weaker) in the 600–900ms time window. By contrast, for disfluent utterances unpredictable words elicited a positivity at mid-frontal and left posterior sites. The effect is similar to the positivity that was observed in response to unpredictable words following an *er* (Experiment 2; Figure 6.5): the timing and distribution of the effect are compatible with its identification as a Late Positive Complex (LPC) which may reflect memory retrieval and control processes (Federmeier et al., 2007; Van Petten & Kutas, 1991, see section 4.3.3). Previously it was suggested (section 6.3.5) that these processes may be engaged as listeners attempt to resume structural fluency after the disruption and the presence of the effect following a silent pause suggests it may be due to the increase in time which disfluency adds to the speech signal. Its presence for disfluent but not for fluent utterances, in the absence of differences in the N400 effect between fluent and disfluent utterances, suggests an independence between the the LPC and the N400 effect and between the cognitive processes which these two effects reflect. The relationship between these processes, and their engagement following different types of disfluencies, is discussed further in Chapter 10.

Differences in processing were also indicated in an early time window (0–200ms) preceding the N400 effect. For fluent utterances, unpredictable words elicited a frontal/fronto-central negativity that was topographically distinct from the N400 effect. The distribution of the effect is compatible with its identification as a Phonological Mismatch Negativity (PMN: Connolly et al., 1990; Connolly & Phillips, 1994, see section 4.3.2). The PMN has been previously observed in N400 paradigms and is thought to reflect detection of a mismatch between the phonological onset of the presented word with a listener's expectations. Although there is some debate as to whether negativities which sometimes precede the N400 effect are distinguishable from the N400 effect itself, the distributional difference between the two negative effects here implies the activation of different populations of neurons and therefore the engagement of different cognitive processes. The presence of a PMN would be unsurprising since all unpredictable words had phonological onsets which differed from those of the predictable words. It is less clear, however, why it would be present in the current data but not in response to the same fluent stimuli when they were presented in Experiment 2. The only difference between the two experiments was the nature of the pause (silence or *er*) in the disfluent utterances, and therefore the explanation probably relies on the impact of the disfluent utterances on the processing of the fluent utterances.

If the early negativity is a PMN, its absence for disfluent utterances suggests that the unpredictable words were no longer perceived to mismatch phonologically with expectations. There are at least two possible reasons for this. First, the delay may render the mismatch less salient. Second, the presence of a pause may disrupt predictive processes: following a silent pause, listeners may reduce the extent to which they make predictions about upcoming words. If this account is true, it is perhaps surprising that there were no observable effects on the N400 effect, which is also sensitive to extent to which predictions have been made.

Besson et al. (1997) observed an N400 which onset around 150ms in response to utterance-final words presented in fluent connected speech and was delayed by around 100ms when words were preceded by an unexpected silent pause. Given the controversy surrounding the existence of the PMN (see section 4.3.2), it is possible that the early onsetting negativity in fluent utterances in the present study reflects an earlier onsetting N400. Whether the early negativity is a PMN or an early manifestation of the N400, its onset is concerning because it appears to emerge at the target onset. Given that time is required to perceive a stimulus, very early onsetting effects suggest differences that precede the target, either inherent to the stimuli, or in processing. It is not clear how such differences could have come about since the experiment employed a fully counterbalanced design (all targets appeared as predictable and unpredictable words in fluent and disfluent utterances) and targets were acoustically identical across all conditions. One possiblity is that the target splicing employed to achieve such a design resulted in differences in the naturalness of the stimuli across conditions. The acoustic token may have been perceived as unnatural (e.g., due to volume or pitch) when it was an unpredictable word completing a fluent utterance, leading to early onsetting differences in this condition only.

Such an explanation seems unlikely because following the creation of the utterances, stimuli were normalised for volume and individually checked. Furthermore, there were no early onsetting differences to fluent stimuli in Experiment 2.

In addition to the effects on processing, there was some indication of longer-term effect of disfluency: predictable words were marginally more likely to be recognised in a subsequent memory test when they had originally been heard following a silent pause. This finding patterns with the effects of the filled pause *er* on recognition memory (Experiment 2; section 6.3) and is compatible with the possibility that silent pauses heightened listeners' attention in a similar way to *ers*.

Comparing silence with ers

The data do not provide strong support for claims that the effects of *ers* on all aspects of language comprehension are driven by the time which these disfluent pauses add to the speech signal (e.g., Brennan & Schober, 2001; Bailey & Ferreira, 2003); the phonological form of the filler *er* may play some role. Unlike silence, *ers* are vocalisations and unambiguously disfluent. These features may be important in terms of the effects on comprehension. However, the possibility that time is an important feature of disfluency cannot be excluded because although there was no effect of silent pauses on the N400, the effects on the memorability of predictable words and on the LPC were similar to those observed with *ers*.

The absence of an effect on the N400 may be because of the nature of the stimuli. Disfluent stimuli were created by excising *ers* and replacing them with silence of identical duration. This method enabled the use of the same stimuli as Experiment 2 and ensured that the durations of the disfluent pauses for each stimulus in the present experiment were similar to those in Experiment 2. However, the method may have introduced unnatural co-articulation into the pre-disfluent word which meant the stimuli with silent pauses were not perceived to be naturally disfluent.

In addition to the time that disfluencies add to the speech signal, another important feature of disfluencies is the nature of the disruption itself. Repetition disfluencies, like disfluent pauses, delay the onset of subsequent new information and like filled pauses are a vocalisation. However, in contrast to both filled and unfilled pauses, they are lexicalised (although for the suggestion that *ers* are also lexicalised, see Clark & Fox Tree, 2002) which may result in different effects on comprehension. The nature of the disfluent signal is addressed further in Experiment 4 (Chapter 8).

7.3 Conclusions

The chapter reported an experiment that investigated the effects of silent pause disfluencies on language comprehension. There was no evidence that silent pauses reduced the standard N400 effect and therefore no evidence that the effect of *ers* on the relative ease of integrating subsequent predictable compared to unpredictable words is due to the increase in time which *ers* add to the signal. However, unpredictable words following a silent pause elicited a Late Positive Complex (LPC) which was not present for fluent utterances, and there was some indication that predictable words were more likely to be subsequently recognised suggesting longer-lasting consequences for representation.

Although the data do not provide strong evidence to rule out the possibility that the effects of the filled pause *er* observed in Experiment 2 can be attributed to the phonological form of the filler, there is some indication that the effects on some aspects of comprehension may be driven by the increase in time which any disfluent pause adds to the signal. In addition to delay, the nature of the disfluent disruption may be an important feature in disfluency processing. Repetition disfluencies delay the onset of subsequent new information (like *ers* and silent pauses), are vocalisations (like *ers*), but contain lexical information. Experiment 4 (Chapter

8) investigates whether the observed effects of *er* and silent pauses are elicited in response to repetitions.

Chapter 8

The effects of repetitions on language comprehension

8.1 Introduction

Experiment 3 demonstrated that silent pauses have some similar effects to the filled pause *er* on language comprehension. Although the results do not rule out the possibility that the effects of *ers* can be attributed to the phonological form of the filler or the presence of a vocalised disfluency, the possibility remains that the effects are driven by the increase in time which an interruption adds to the signal. In addition to time, another important feature of disfluencies which may be of importance in terms of their effects on language comprehension is the nature of the interruption itself. For example, the effects of vocalised interruptions on language comprehension may depend, in part, on the type of information contained in the vocalisation. Repetition disfluencies, like filled and silent pauses, delay the onset of subsequent new information, and like filled pauses are vocalisations. However, they are lexicalised which means they are not inherently disfluent, but are rendered disfluent only by the context in which they are heard, and are of the same type of vocalisation as the surrounding speech. In this chapter I describe an experiment that investigated the effects of disfluent repetitions on language comprehension.

8.2 Experiment 4. Investigating the effects of repetitions on language comprehension

Experiment 4 used a modification of Experiment 2.

8.2.1 Background literature

Repetition disfluencies have been observed in similar situations to disfluent pauses (Beattie & Bradbury, 1979; Howell & Sackin, 2001) and it has been suggested that they have similar functions for the speaker, for example introducing more time during which production difficulties can be resolved (Blackmer & Mitton, 1991; Clark & Wasow, 1998; Maclay & Osgood, 1959). However, it is not clear whether disfluent pauses and disfluent repetitions have similar effects on language comprehension.

Only two studies have investigated the comprehension of repetition disfluencies: one investigated the effects on the processing of subsequent words (Fox Tree, 1995) and another investigated the effects that occur when the repetition disfluency itself is processed (McAllister et al., 2001). No study has investigated the effects of repetitions on language representation.

Fox Tree (1995) investigated the effects of repetitions (which were predominantly of multiple words) using a word monitoring task (see section 2.4.2). Response times to correctly identify target words were unaffected by the presence of a preceding repetition, providing no evidence for an effect on processing. These results contrast with those from a similar word monitoring study, which showed faster response times to targets following an *er* (Fox Tree, 2001). Extending Fox Tree's (2001) interpretations that the faster response times reflected heightened attention (see also Brennan & Schober, 2001), the lack of an effect for repetition may reflect the failure of repetitions to heighten listeners' attention to upcoming speech, perhaps because of the resources required to process the lexical information in the repetition. Although observable effects of repetitions on language comprehension have not been demonstrated using reaction times, this does not exclude the possibility that effects could be shown using ERPs.

Using ERPs, McAllister et al. (2001) showed effects of the processing of repetitions themselves (see section 2.4.2). Participants listened to recordings of speech which were immediately followed by the presentation of a tangram picture, and had to decide whether or not the picture matched the description. Some of the descriptions included repetitions. Relative to the first occurrence of the word, the repeated words elicited a negativity in the time period 300–500ms after the word onset, which the authors interpret as an N400 effect. However, statistically, the difference was only marginally significant. Furthermore, the authors only present analyses from midline electrodes and do not show the scalp distributions of the ERP effect. If the results are truly significant, they suggest that listeners are sensitive to the presence of repetition disfluencies. More specifically, the negativity suggests that semantic integration difficulty is greater for words when they are repeated, than when they first occur.

A number of studies have used ERPs to investigate the effects of processing repeated words that are not in the context of disfluent speech. Relative to the first occurrence of a word, repeated words in lists and sentences are commonly associated with an attenuation of the N400 (Besson, Kutas, & Van Petten, 1992; Besson & Kutas, 1993; Ledoux, Traxler, & Swaab, 2007; Okita & Jibu, 1998; Rugg, 1985), particularly if the repeated word is presented immediately after its first occurrence (Nagy & Rugg, 1989), indicating easier integration (in contrast to McAllister et al., 2001).

It is important to note that the reduction in the N400 is sometimes reported as an increase in positivity. However, the presence of a discourse context which renders repeated words unpredictable or unnatural, can reverse this effect. For example investigations of co-referential processing where two expressions refer to the same entity (see 1) have shown an increase in the N400 for repeated words ("Matt"), relative to pronoun controls ("he"), which were more predictable (Swaab, Camblin, & Gordon, 2004).

(1) Matt went swimming after **Matt**/(he) had dinner

Experiment 4 was designed to investigate whether the ERP and memory effects that were observed in Experiments 2 and 3 would extend to repetition disfluencies.

8.2.2 Experimental rationale

The aim of the experiment was to investigate whether repetitions would lead to effects on language comprehension that were similar to those observed with *ers* in Experiment 2 (see section 6.3). Such findings would implicate temporal delay as a causal factor; failure to observe similar results would demonstrate that the nature of the disruption to speech is important. The online processing of post-disfluency words was assessed using ERPs: specifically, the N400 effect was used as an index of semantic integration processes, although effects in other time windows were also analysed. Longer-term effects on representation were assessed using a recognition memory test.

Additionally, an exploratory investigation of the ERP effects associated with the processing of a disfluent repetition itself was performed. The ERPs for repeated words were compared with those for identical control words from fluent utterances. Given the exploratory nature of the investigation and based on the limited evidence on the effects of repeated words, no strong predictions were made.

8.2.3 Methods

Stimuli

The stimuli were edited from those in Experiment 2 (see section 5.3 and Appendix B), and were highly constrained fluent and disfluent utterances ending in predictable (cloze probability 0.84, range 0.52–1) or unpredictable (cloze probability 0) target words. Disfluent utterances were created from the fluent utterances by copying the pre-target word and splicing it into the speech stream, after the original, to form a repetition before the utterance-final target word.

The disfluent utterances differed from those in previous experiments in that they were identical to the fluent utterances up until the point of the repetition disfluency, and did not include any other features of disfluency. This design was chosen for two reasons. First, in the disfluent utterances used in the previous experiments, the pre-target words were sometimes lengthened and phonologically altered (e.g., from "thuh" to "thee"). In natural speech, when originally altered words are repeated, the repetition usually involves a return to the normal phonological form. To create utterances like this would have involved splicing in a different acoustic token from another part of the utterance, or a different utterance altogether. By using the fluent utterances, it was easier to create natural sounding pre-target repetitions which had normal phonological forms (e.g., "thuh"). Secondly, using fluent and disfluent utterances which were identical until the point of the repetition made it possible to compare the effect of processing the repetition itself, with the processing of an acoustically identical fluent control word. A pause of 200ms was inserted between the two tokens of the repetition. The duration chosen was based on the pauses which have been observed during naturally occuring repetitions (Fox Tree, 1995) and was shortened or lengthened where this resulted in more natural sounding speech. The repetition was typically of a single function word but occasionally, two or three words were repeated where it resulted in more natural sounding speech. Table 8.1 shows an example stimulus set.

Table 8.1: Example stimulus set comprising two highly constraining sentence frames, crossed with two target words which were either predictable or unpredictable in context. Target words are shown in bold. Half the utterances were disfluent and contained a repetition before the target word, indicated in square brackets.

Predictable	Everyone's got bad habits and mine is biting	my	[*my*]	**nails**
	That drink's too hot; I've just burnt	my	[*my*]	**tongue**
Unpredictable	Everyone's got bad habits and mine is biting	my	[*my*]	**tongue**
	That drink's too hot; I've just burnt	my	[*my*]	**nails**

Participants

Sixteen right-handed native English speakers (7 male; mean age 22 years; range 19-35 years) took part in the experiment.

Procedure

The procedure followed that described in the General Methods Chapter (section 5.5) and which was followed in Experiments 2 and 3. There were two parts to the experiment. The first part was designed to investigate the effects of repetition disfluencies on online processing and these were assessed using ERP data collected during natural listening; the second part was designed to investigate the longer-term effects on representation and these were assessed using data from a recognition memory test. During the listening part of the experiment, participants were unaware that their memory would subsequently be tested.

One hundred and sixty experimental stimuli (40 each of fluent predictable, fluent unpredictable, disfluent predictable, disfluent unpredictable) were presented auditorily. Experimental utterances were interspersed with 80 filler utterances. Recordings were presented in two blocks lasting approximately 15 minutes each, separated by a break of a few minutes. EEGs were recorded from 61 scalp electrodes using a left mastoid reference and re-referenced offline to the average of left and right mastoid recordings.

Following the first part of the experiment, participants took part in a surprise recognition memory test for the utterance-final ("old") words. Words were presented visually and interspersed with frequency-matched "new" words, which had not been heard at any point during the first part of the experiment. Participants discriminated between old and new words as accurately as possible by pressing one of two response keys (counterbalanced across participants).

Before offline averaging, the continuous EEG files for each participant were segmented into 1350ms epochs starting 150ms before the target onset and screened for artefacts. Artefact rejection resulted in the exclusion of 36% of the trials. Grand average ERPs were formed, time-locked to the onsets of the utterance-final predictable and unpredictable words ("word-locked") from fluent and disfluent utterances making four conditions: fluent predictable, fluent unpredictable, disfluent predictable, and disfluent unpredictable, with mean trial numbers of 26, 26, 25, and 26 respectively.

Grand average ERPs were also formed, time-locked to the onsets of repetitions ("repetition-locked") from the disfluent utterances, and to corresponding control words which preceded the utterance-final words. No differences were observed between the effects of repetitions relative to their control between utterances that ended in predictable words and utterances that ended in unpredictable words. This was expected because the counterbalancing ensured that the predictable-ending utterances and the unpredictable-ending utterances were identical until the repeated word (in repetition utterances) or until the predictable/unpredictable utterance-final word (in control utterances). Therefore, the data for the repeated and control words are presented collapsed over the predictable/unpredictable conditions. This resulted in two conditions: repetition, and control, with mean trial numbers of 26 in each condition.

8.2.4 ERP results: word-locked

Figures 8.1 and 8.2 show the grand average ERPs time-locked to the utterance-final word onsets for fluent and disfluent utterances respectively.

For both fluent utterances (Figure 8.1) and disfluent utterances (Figure 8.2), relative to predictable words, unpredictable words show a negativity over the 300–500ms time window, which is broadly distributed over the scalp, but appears larger over centro-parietal/parietal and midline sites. These features are compatible with its identification as an N400 effect. Following the negativity, differences emerge between fluent and disfluent utterances. For fluent utterances, the relative negativity continues, but appears smaller and more focused at central sites. For disfluent utterances, unpredictable words show a relative positivity over frontal and fronto-central sites bilaterally, and over left centro-parietal/parietal sites.

ERPs were quantified by measuring the mean amplitude of the ERP difference between predictable and unpredictable words, for fluent and disfluent utterances over two time windows: the standard N400 time window (300–500ms) and a later (600–900ms) time window. Topographic distributions of the effects (difference between ERPs for unpredictable and predictable words) for fluent and disfluent utterances over these time windows are shown in Figure 8.3.

Magnitude and topographic analyses were performed following the strategy described in the General Methods Chapter (sections 5.6.1 and 5.6.2) and used in Experiments 2 and 3.

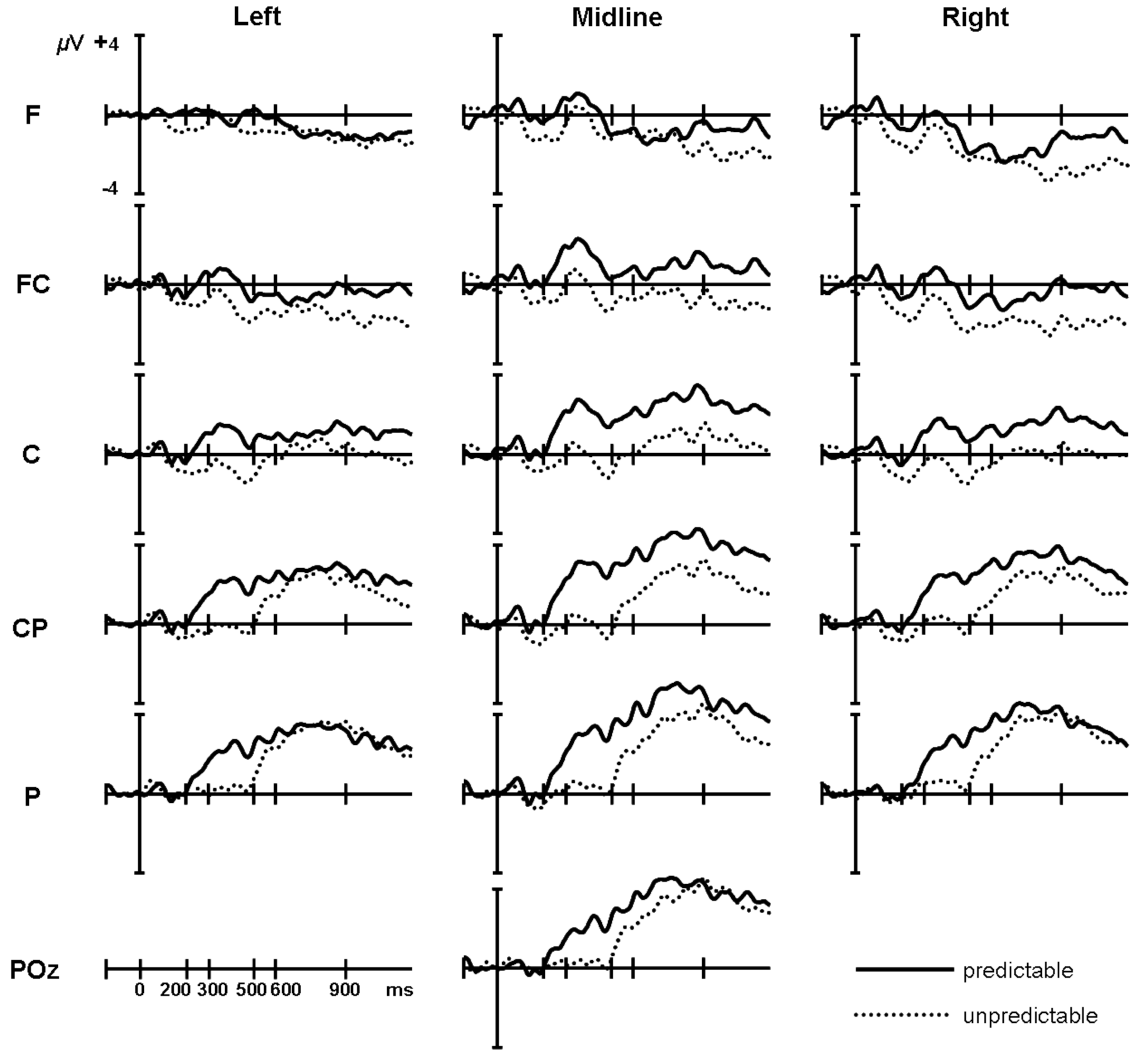

Figure 8.1: Grand average ERPs (n=16) for *fluent* utterances relative to predictable (solid lines) or unpredictable (dotted lines) target word onsets at frontal (F), fronto-central (FC), central (C), centro-parietal (CP), parietal (P), and occipito-parietal (PO) locations, for electrodes at the midline and grouped over left (electrodes 1, 3, 5) and right (electrodes 2, 4, 6) hemispheres. Unpredictable words show a relative negativity, which is broadly distributed over the scalp, but appears larger over central/centro-parietal/parietal and midline sites. The effect emerges around 200ms and is larger over the standard N400 time window (300–500ms).

300–500ms

Over the standard N400 time window of 300–500ms, a topographic analysis incorporating all 61 electrodes provided no evidence of a distributional difference between the effects for fluent and disfluent conditions [for site: $F(60, 900) = 4.844$, $\eta_p^2 = .244$, $p = .007$; other Fs < 1; see Figure 8.3]. Thus there was no reason to suppose that

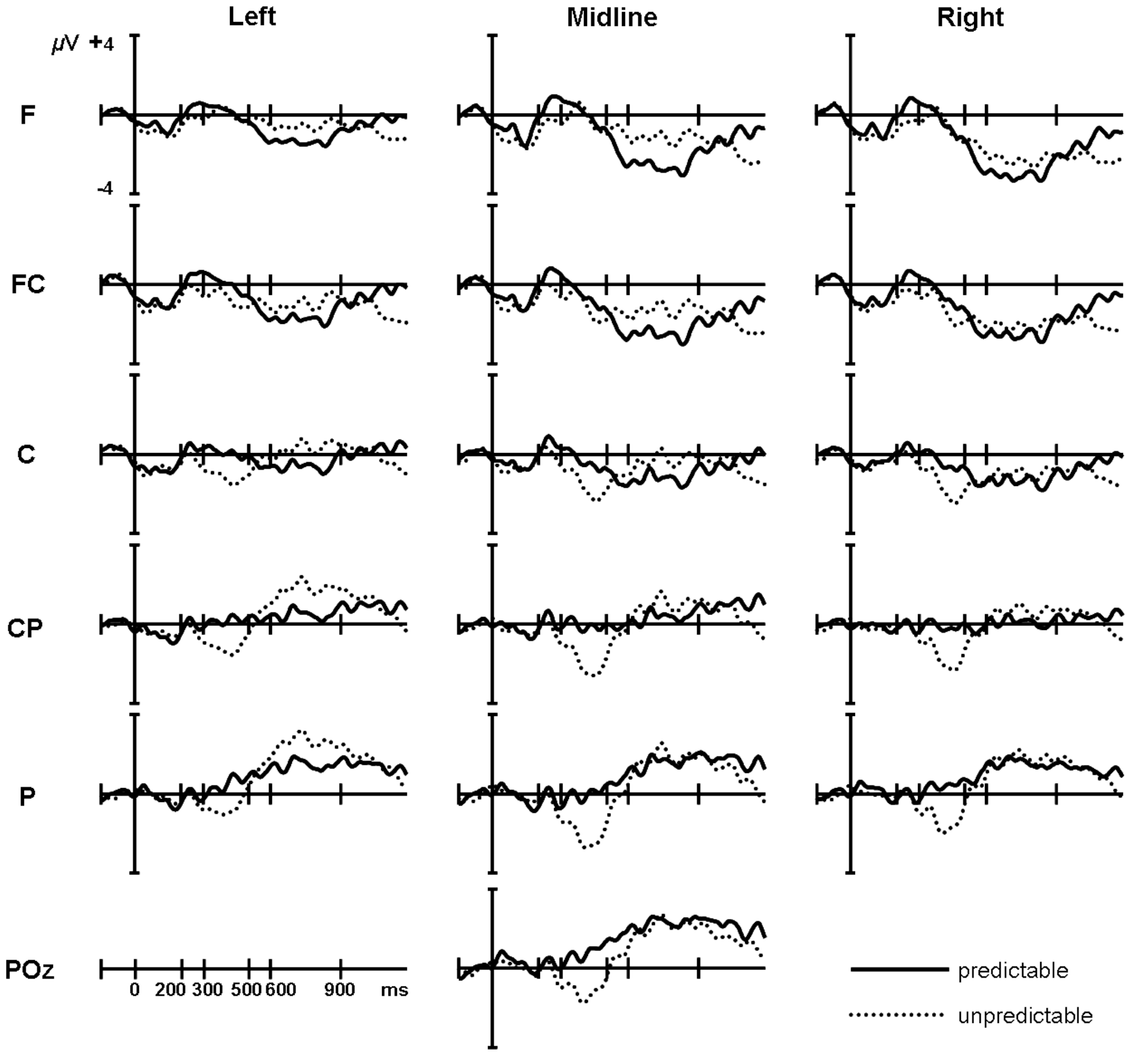

Figure 8.2: Grand average ERPs (n=16) for *disfluent* utterances relative to predictable (solid lines) or unpredictable (dotted lines) target word onsets at frontal (F), fronto-central (FC), central (C), centro-parietal (CP), parietal (P), and occipto-parietal (PO) locations, for electrodes at the midline and grouped over left (electrodes 1, 3, 5) and right (electrodes 2, 4, 6) hemispheres. Unpredictable words show a relative negativity, which is broadly distributed over the scalp, but appears larger over centro-parietal/parietal and midline sites over the standard N400 time window (300–500ms). Following the negativity, unpredictable words show a relative positivity over frontal/fronto-central sites bilaterally, and over left centro-parietal/parietal sites.

different neural generators are responsible for the recorded effects of predictability and no reason to suppose the engagement of different cognitive processes during this time period. The effects for fluent and disfluent conditions were therefore analysed separately and then quantitatively compared.

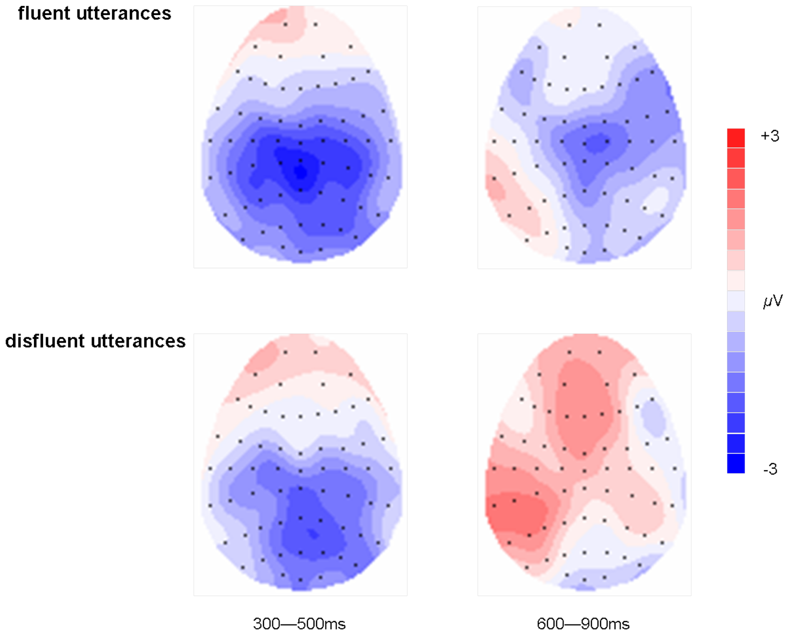

Figure 8.3: Scalp topographies (n=16) showing the predictability effects over two time windows: 300–500ms and 600-900ms, for *fluent* and *disfluent* utterances. Both fluent and disfluent utterances show a negative effect over the 300–500ms time window, which is broadly distributed over the scalp, but appears larger over centro-parietal/parietal and midline sites. Over the 600–900ms time window fluent utterances show some indication of a similar, but weaker negativity. Disfluent utterances show a positive effect over frontal and front-central sites bilaterally and over left centro-parietal/parietal sites.

Hemispheric analyses established that the distributions of the fluent and disfluent N400 effects were not lateralised. For fluent utterances, there was a main effect of predictability [$F(1, 15) = 17.76$, $\eta_p^2 = .542$, $p = .001$] and an interaction between predictability and site [$F(2, 22) = 5.26$, $\eta_p^2 = .259$, $p = .025$], reflecting a relative negativity for unpredictable words which was greater towards the midline. For disfluent utterances, there was also a main effect of predictability [$F(1, 15) = 6.65$, $\eta_p^2 = .307$, $p = .021$], reflecting a relative negativity for unpredictable words, but no other significant effects.

Since no effects involving hemisphere were found for either fluent or disfluent conditions, further analyses concentrated on midline electrodes. For fluent utterances, the midline analysis showed a main effect of predictability [$F(1, 15) = 16.222$, $\eta_p^2 = .520$,

$p = .001$] and an interaction between predictability and location [$F(5, 75) = 4.052$, $\eta_p^2 = .213$, $p = .032$], reflecting a relative negativity which was larger at more posterior locations. Disfluent utterances showed a main effect of predictability [$F(1, 15) = 5.635$, $\eta_p^2 = .273$, $p = .031$] and a marginal interaction between predictability and location [$F(5, 75) = 3.392$, $\eta_p^2 = .184$, $p = .067$], again reflecting a relative negativity and indicating a tendency towards it being larger at more posterior locations.

The quantitative comparison of the effects between fluent and disfluent conditions showed a main effect of predictability [$F(1, 15) = 16.36$, $\eta_p^2 = .522$, $p = .001$], reflecting the relative negativity for unpredictable words. There was also a main effect of fluency [$F(1, 15) = 10.15$, $\eta_p^2 = .404$, $p = .006$], reflecting the greater overall positivity of the ERPs for fluent utterances. Importantly, there was no interaction between fluency and predictability [$F(1, 15) = 1.23$, $\eta_p^2 = .076$, $p = .285$], and therefore no evidence of a difference between the amplitude of the N400 effect for fluent utterances [1.93μV] and for disfluent utterances that included a repetition [1.35μV].

As a final check, a topographic analysis was performed for the midline N400 effects to examine whether there was a distributional difference between the N400 effects for fluent and disfluent conditions. There was no observable difference [for location: $F(5, 75) = 7.127$, $\eta_p^2 = .322$, $p = .004$; other Fs < 1].

600–900ms

Over 600–900ms, a topographic analysis incorporating all 61 electrodes provided no evidence of distributional differences between the effects for fluent and disfluent conditions [for site: $F(60, 900) = 2.261$, $\eta_p^2 = .131$, $p = .007$; other Fs < 1]. Hemispheric and midline topographic analyses also failed to find distributional differences. Thus despite the apparent difference in scalp topographies between the fluent and disfluent conditions (see Figure 8.3) there is no reason to suppose that different neural generators were responsible for the recorded effects of predictability. However, there was no *a priori* reason to quantitatively compare the effects and therefore the effects are only reported for fluent and disfluent conditions separately.

For fluent utterances, the hemispheric analysis provided no evidence for laterality differences. There were interactions between predictability and site [$F(2, 30) = 5.692$, $\eta_p^2 = .275$, $p = .021$] and between predictability, location and site [$F(8, 120) = 7.548$, $\eta_p^2 = .335$, $p = .004$], reflecting a greater negativity towards midline sites, and more so at central locations. The midline analysis showed a main effect of predictability [$F(1, 15) = 10.635$, $\eta_p^2 = .415$, $p = .005$].

For disfluent utterances the hemispheric analysis provided no evidence for laterality differences. There was an interaction between predictability, location and site [$F(8, 120) = 7.714$, $\eta_p^2 = .340$, $p = .002$], reflecting a relative positivity at the frontal location which was larger for superior sites. The midline analysis showed no significant effects.

Effects over time

For fluent utterances, the topographic analysis incorporating all 61 electrodes, and the hemispheric and midline topographic analyses provided no evidence of a distributional difference between the effects over 300–500ms and 600–900ms (see Figure 8.3). Thus the effects for the two time windows were quantitatively compared. Since no effects involving hemisphere were found for either the 300–500ms or 600–900ms time windows, the quantitative comparison of the effects focused on the midline electrodes. The midline analysis revealed a main effect of predictability [$F(1, 15) = 14.557$, $\eta_p^2 = .493$, $p = .002$] and an interaction between epoch and predictability [$F(1, 15) = 5.851$, $\eta_p^2 = .281$, $p = .029$], reflecting the larger negativity in the earlier time window.

For disfluent utterances, the topographic analysis incorporating all 61 electrodes showed an interaction between epoch and site [$F(60,600) = 2.952$, $\eta_p^2 = .164$, $p = .032$], suggesting a distributional difference between the effects over the 300–500ms and 600–900ms time windows. Given their distributional differences which suggest the presence of two distinct effects and the engagement of different populations of neurons, the effects for the two epochs were not quantitatively compared.

8.2.5 ERP results: repetition-locked

Figure 8.4 shows the grand average ERPs time-locked to the repetition and control word onsets.

Relative to fluent control words, repeated words show a positivity which is broadly distributed over the scalp, but appears larger and longer-lasting over central sites with a slight right hemisphere bias. The effect onsets around 50ms, is larger over 100–400ms and continues until around 600ms. It is possible that the positivity encompasses two overlapping broadly-distributed positive effects: an early effect over frontal sites, and a later effect over central sites (see section 8.2.7).

ERPs were quantified by measuring the mean amplitude of the ERP difference between repetition and control words over 100–400ms. Effects were also analysed over shorter time windows of 50–150ms and 150–400ms, but no differences between these time windows were observed, and the results are not reported. The effects were not analysed after 400ms because of the overlap with the word-locked effects. The topographic distribution of the effect over the 100–400ms time window is shown in Figure 8.5.

Analyses were performed for the single time window and incorporated a factor of repetition [control, repetition] rather than of predictability.

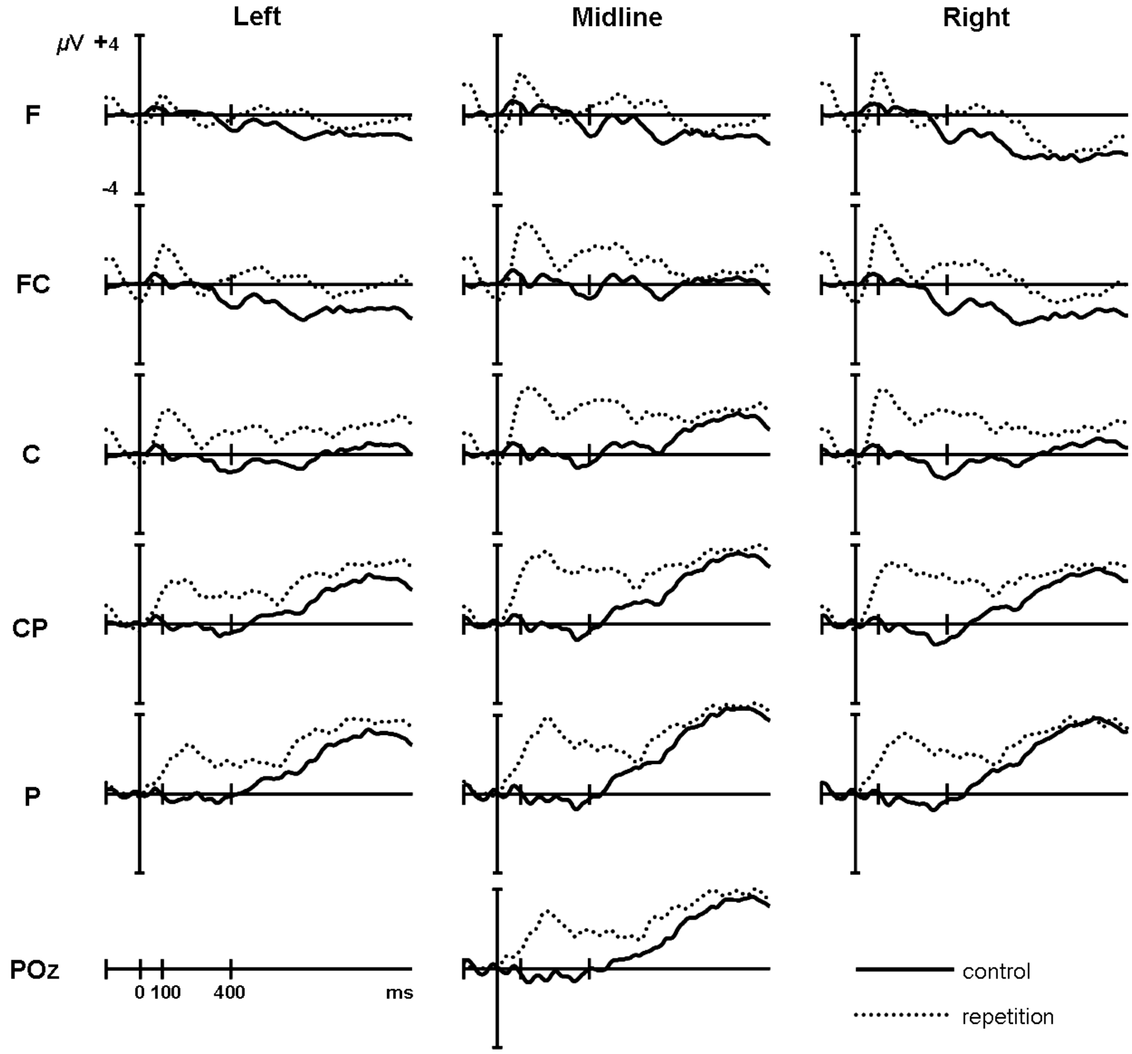

Figure 8.4: Grand average ERPs (n=16) relative to repeated (dotted lines) or fluent control (solid lines) word onsets at frontal (F), fronto-central (FC), central (C), centro-parietal (CP), parietal (P), and occipito-parietal (PO) locations, for electrodes at the midline and grouped over left (electrodes 1, 3, 5) and right (electrodes 2, 4, 6) hemispheres. Repeated words show a relative positivity which is broadly distributed over the scalp, but appears larger over central/centro-parietal sites with a slight right hemisphere bias. The effect emerges at around 50ms, is larger over 100–400ms, and continues until around 600ms.

100–400ms

The hemispheric analysis suggested a laterality difference. There was a main effect of repetition [$F(1, 15) = 15.513$, $\eta_p^2 = .508$, $p = .001$], and interactions between repetition and location [$F(4, 60) = 24.032$, $\eta_p^2 = .616$, $p < .001$], between repetition and hemisphere [$F(1, 15) = 11.109$, $\eta_p^2 = .425$, $p = .005$], and between repetition

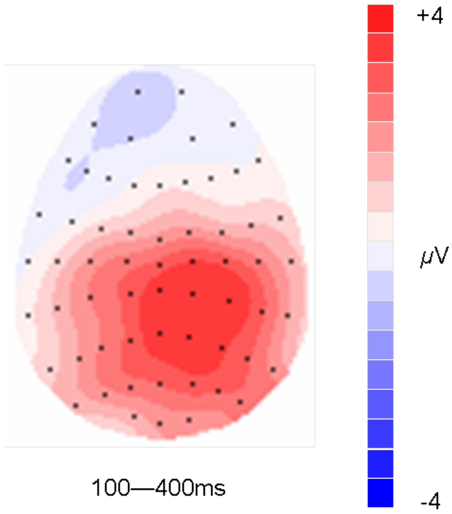

Figure 8.5: Scalp topography (n=16) showing the repetition effect over 100–400ms. There is a positive effect with a broad scalp distribution, which is focused over central/centro-parietal sites with a slight right hemisphere bias.

and site [$F(2, 30) = 20.848$, $\eta_p^2 = .582$, $p < .001$], reflecting a relative positivity which was larger at central/centro-parietal locations, over the right hemisphere, and at superior sites. There were also three-way interactions between repetition, location and site [$F(8, 120) = 4.690$, $\eta_p^2 = .238$, $p = .009$] and between repetition, hemisphere and site [$F(2, 30) = 4.288$, $\eta_p^2 = .222$, $p = .046$], reflecting a greater right hemispheric positivity over superior sites, and at central/centro-parietal locations.

8.2.6 Recognition memory results

Memory performance was quantified as the probability of correctly identifying "old" words, by fluency and predictability. Overall, 61% of the old words were correctly recognised[1] (22% false alarms). Figure 8.6 shows the recognition probability of utterance-final words by fluency and predictability.

Memory analyses were performed following the strategy described in the General Methods Chapter (section 5.7) and used in Experiments 1–3. Words that had been

[1]One word ("party") was never responded to by participants (within the allocated time), resulting in no data for this item. Additionally, presentation of one other word ("garden") was corrupted. These items were excluded from the analyses.

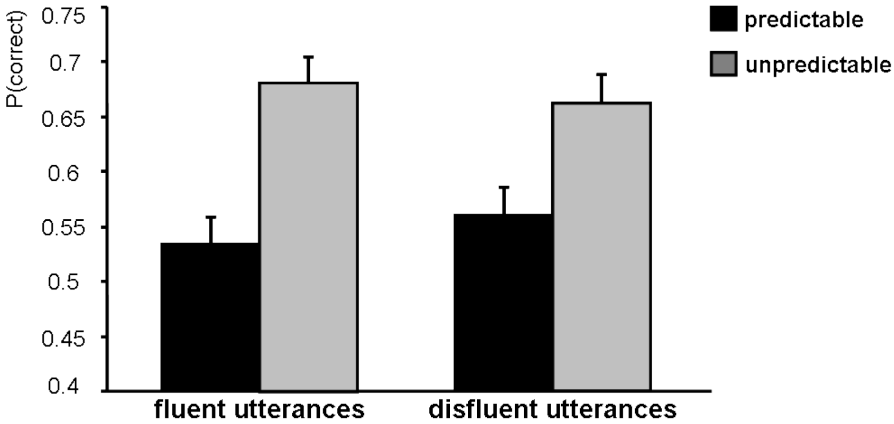

Figure 8.6: Probability of correctly recognising utterance-final words that were originally predictable (black) or unpredictable (grey) in their contexts, for fluent and disfluent utterances. Error bars represent one standard error of the mean. Unpredictable words were more likely to be correctly recognised than predictable words for both fluent and disfluent utterances. There were no effects of disfluency.

unpredictable in their context were more likely to be correctly recognised than words that had been predictable [67% vs. 55%: $F(1, 133) = 27.12$, $\eta_p^2 = .169$, $p < .001$], in both fluent [68% vs. 54%: $t(133) = 4.757$, $p < .001$] and disfluent [66% vs. 56%: $t(133) = 3.399$, $p < .001$] utterances. There was no effect of disfluency [$F < 1$], and no interaction between predictability and disfluency [$F(1, 133) = 1.34$, $\eta_p^2 = .010$, $p < .249$].

8.2.7 Discussion

The aim of the experiment was to investigate whether the effects of the filled pause *er* on language comprehension that were observed in Experiment 2 (see section 6.3) are sensitive to the nature of the disruption to speech. The effects of repetition disfluencies on language comprehension were investigated using a modification of Experiment 2: participants listened for understanding to a series of short utterances which ended in either predictable or unpredictable target words. Utterances were either fluent or disfluent, containing a repetition before the target words.

Summary and interpretation

ERPs showed the presence of an N400 effect for both fluent and disfluent utterances. In contrast to the results for *ers* (Experiment 2), the amplitude of the N400 effect was unaffected by the presence of a repetition. There is therefore no evidence that repetition disfluencies affect the difference between the ease of integrating predictable and unpredictable words. The data are compatible with the possibility that the nature of the disruption to speech is important.

One possibility is that repetitions do not disrupt the integration of either predictable or unpredictable words. Another possibility is that repetitions do disrupt integration, but have an equal effect on predictable and unpredictable words. The current data do not allow these two possibilities to be distinguished because the differing pre-target baselines for fluent and disfluent conditions (disfluent baselines were typically obtained mid-repetition) meant that a direct comparison of the target words between fluent and disfluent utterances could not be made. The waveforms (Figures 8.1 and 8.2) show a greater negativity (relative to the pre-target baseline) for target words from the disfluent utterances which may indicate greater integration difficulties. However, if this were the case, it might be expected that targets from disfluent utterances would have been more likely to be remembered. There was no indication of a longer-term effect of disfluency: in a subsequent memory test, unpredictable words were more likely to be recognised than predictable words for both fluent and disfluent conditions. This finding is in contrast to results showing that both *ers* (Experiments 1 and 2) and silent pauses (Experiment 3) had an effect on recognitition memory.

If as suggested previously (section 6.3.5), the attenuation of the N400 for targets preceded by an *er* reflects a reduction in the extent to which listeners make predictions following such disfluencies, the lack of an attenuation for targets preceded by a repetition suggests that repetition disfluencies do not affect predictive processes. Prediction is discussed further in section 10.3.1. The lack of an effect of repetitions on recognition memory provide no evidence to support the possibility that repetitions heighten listeners' attention to upcoming speech. These findings are compatible with Fox Tree's (1995) results which showed that response times were unaffected by the presence of a preceding repetition.

Analyses of the ERPs after the N400 effect provided evidence for a processing difference between fluent and disfluent utterances. For fluent utterances, the N400 effect was still significant (although weaker) in the 600–900ms time window. For disfluent utterances, unpredictable words elicited a positivity over frontal sites. The effect is similar to the positivity observed in response to unpredictable words in disfluent utterances which included an *er* (Experiment 2; Figure 6.5) and which included a silent pause (Experiment 3; Figure 7.3); the timing and distribution of the effect are compatible with its identification as a Late Positive Complex (LPC), which may reflect aspects of memory retrieval and control (Federmeier et al., 2007; Van Petten et al., 1991, see section 4.3.3). Previously it was suggested (section 6.3.5) that these processes are engaged as participants attempt to resume structural fluency of the message after the interruption and the presence of the effect following a repetition disfluency suggests it may be due to the disruption which disfluency causes.

Comparing repetitions with ers

Differences between the effects of repetition disfluencies and *ers* may be because repetitions are lexicalised and are therefore not immediately distinct from the surround lexical context unlike *ers* which are non-lexicalised interruptions. An account which focuses on the lexical nature of the disfluency would predict that the effect of other lexical interruptions, for example lexical fillers such as *like* or *y'know*, would be similar to those of repetitions.

The disfluent utterances here differed from those which included a pause by more than just the type of disfluency preceding the target. In the *er* and silent pause experiments, the disfluent utterances also included other features of disfluency such as vowel lengthening in the sentence frame leading up to the pause, where deemed natural for the speaker. In the present experiment, the disfluent utterance frames were identical to the fluent utterance frames; the only sign of disfluency was the repetition itself. It is possible, therefore, that the similar increase in memory obtained for words following both *ers* and silent pauses were (at least partly) due to the disfluent features in the sentence frame and that the presence of a pause alone would not have caused the observed effects. This possibility seems unlikely because not all the disfluent utterances included additional disfluent features, but it remains a question for future research.

The repetition effect

The processing of a repetition itself was associated with a relative positivity indicating listeners' sensitivity to the repetition. This effect differs to results from the only other investigation of the effects of processing disfluent repetitions (McAllister et al., 2001), an ERP study, which indicated a relative negativity for repeated words. However, the statistical support for the negativity in that study is weak. It is also possible that differences between the results here and those of McAllister et al. (2001) are due to differences in the types of words that were repeated. In the current study, repetitions were of function words and it is possible that this precluded the possibility of observing an N400 effect, since function words elicit only very small N400s (Van Petten & Kutas, 1991); McAllister et al. (2001) give no details of the repetitions in their stimuli.

One possibility is that the positivity is a P600 effect, associated with syntactic repair or reanalysis following the detection of the syntactic violation created by the repetition (e.g., Friederici, 1995, 2002; Friederici et al., 1996; Gunter et al., 1997, see section 4.4.1). If the processing of repetitions requires the use of resources to recover from the syntactic violation, this could limit the resources available for heightening attention to subsequent words, and account for the limited effects that were observed on the processing and representation of post-repetition words.

The early onset of the positivity (around 50ms), may reflect an overlapping but distinctive earlier positive effect such as the P2, an ERP component related to sensory or perceptual processing of stimuli, although this could not be statistically verified in the current data. The presence of such an effect would be compatible with previous studies that have shown relative positivities to repeated stimuli. For example, an early positivity (30–250ms) with a bilateral fronto-central distribution has been shown in response to repeated tones (Haenschel, Vernon, Dwivedi, Gruzelier, & Baldeweg, 2005) and this has linked to sensory memory formation.

8.3 Conclusions

The chapter reported an experiment that investigated the effects of repetition disfluencies on language comprehension. There was no evidence that repetitions reduced the standard N400 effect and therefore unlike *ers* no evidence that repetitions affected the relative ease of integrating subsequent predictable compared to unpredictable words. There was also no effect of repetitions on the subsequent recognition probability of words, providing no evidence for longer-lasting consequences for representation. However, unpredictable words following a repetition elicited a Late Positive Complex (LPC) which was not present for fluent utterances. In addition to the effects of repetitions on the processing of subsequent words, repetitions themselves elicited a right posterior positivity reflecting reflecting detection of the disfluency and possibly syntactic reanalysis.

The data the data do not provide strong evidence to rule out the possibility that the effects of the filled pause *er* on some aspects of comprehension observed in Experiment 2 can be attributed to the phonological form of the filler and are compatible with the suggestion that the the nature of the disfluent disruption is an important feature in disfluency processing.

The experiments reported so far have focused on the effects of disfluencies on the processing of subsequent words that were straightforward continuations of the pre-interruption speech. The final experiment (Chapter 9) reconsiders the effects of *er* pauses, this time investigating the processing of subsequent words that are repair disfluencies, rather than the fluent continuations used in Experiments 1 and 2 (Chapter 6).

Chapter 9

The effects of repairs on language comprehension

9.1 Introduction

Experiments 1–4 demonstrated that *er* pauses, silent pauses, and repetitions, affect language comprehension. These experiments focused on post-disfluency words that were straightforward continuations of the pre-interruption speech. However, disfluencies sometimes occur in succession. For example, *ers* may occur during the edit interval of a repair disfluency, before the repair. In this chapter I describe an experiment that investigated the effects of *ers* on the processing of subsequent repair words.

9.2 Experiment 5. Investigating the effects of repairs on language comprehension

Experiment 5 investigated the effects of *ers* on the processing of subsequent repair words using ERPs.

9.2.1 Background literature

Repair disfluencies occur when speakers detect and correct mistakes in their speech and are sometimes accompanied by a filled pause during the edit interval (Levelt, 1983; Blackmer & Mitton, 1991). Because repairs are sequences of words which usually constitute a syntactic violation, the processing of repairs might be expected to cause syntactic processing difficulties which require additional resources to be resolved. Fox Tree (1995) investigated the processing of repairs using a word monitoring task (see section 2.4.3). Response times to correctly identify repairs were slower than for control words in fluent utterances where the reparanda had been excised, or replaced by silent pauses of the same duration. These results suggest that the processing of repairs incurs a cost, perhaps reflecting the additional resources or time required for semantic or syntactic integration, although response times do not enable conclusions to be drawn about the specific aspects of language processing which are affected.

Ferreira et al. (2004) proposed that the processing of repairs may employ mechanisms similar to those involved in the recovery from garden path sentences because both appear to require structural reanalysis. This claim is supported by results from a study by McAllister et al. (2001) which used ERPs to investigate the processing of repairs (see section 2.4.3). Participants listened to recordings of speech which were immediately followed by the presentation of a tangram picture, and had to decide whether or not the picture matched the description. Some of the descriptions included repairs. Relative to fluent control words, repairs elicited a positivity over posterior scalp locations in the time period 0–800ms after the word onset, which the authors interpret as a P600 effect associated with syntactic reanalysis and repair (section 4.4). It should be noted that the authors only present analyses from midline electrodes and do not show the scalp distributions of the ERP effect and so the findings require replication.

It has been suggested that filled pauses, during the edit interval of repairs, aid processing by helping listeners to detect the presence of repairs. For example, Levelt (1989, p. 481) suggests that an editing expression like *er* or *uh* may "warn the addressee that the current message is to be replaced". Hindle (1983) suggests that an *er* is an "edit signal". Brennan and Schober (2001) investigated this possibility, using a referential communication task (see section 2.4.3). Listeners followed instructions to press a button to select a referent from a set of geometric objects displayed on a computer screen. Some of the instructions included repairs, of which some included a filled pause during the edit interval. Response times were faster, and accuracy higher, for repairs that were preceded by an *uh*, indicating a processing benefit. There are two main limitations of Brennan and Schober's (2001) study. First, the experiment was confounded by the limited set of referents: because the targets were repair words, the reparanda were particularly informative in reducing the number of potential referents. When the number of referents was increased from two to three and so reduced the informativeness of the reparanda as a cue, the advantage afforded by the filled pause, although still present, was attenuated. Secondly, as for Fox Tree (1995), response times do not enable conclusions to be drawn about the specific aspects of language processing which are affected.

The evidence suggests that the processing of repairs incurs a cost, but there is little evidence to support claims that specifically syntactic aspects of language processing are affected. Claims that *ers* can aid the processing of repairs are only weakly supported. These issues are addressed in Experiment 5.

9.2.2 Experimental rationale

The aim of the experiment was to investigate the effects of processing repairs and to assess whether a pre-repair *er* would alter these effects. The online processing of repairs was assessed using ERPs.

Participants listened to utterances which sometimes included a verb-verb repair mid-utterance, while EEG was recorded. ERPs were formed relative to targets which were either repair or non-repair control words. It was predicted that (relative to control words) repairs would elicit a Left Anterior Negativity (LAN) associated with the detection of a syntactic word category violation (section 4.4.2) and a P600 effect associated with the subsequent reanalysis required (section 4.4.1).

If filled pauses prepare listeners for the possibility of an upcoming repair (Brennan & Schober, 2001; Hindle, 1983; Levelt, 1983), it is hypothesised that listeners may no longer perceive the repair as a syntactic violation, and the LAN effect for repairs compared to control words observed in utterances without an *er*, will be reduced by the presence of a pre-repair *er*. Further, if listeners are prepared for the repair, it is hypothesised that they may be able to commence the reanalysis more promptly and the P600 may onset earlier.

Additionally, an exploratory investigation of the ERP effects associated with the processing of an *er* itself was performed. The ERPs for *ers* were compared with those for control words which appeared in the same location from utterances that did not include an *er*, both for utterances without a repair (control utterances) and for repair utterances. Given the exploratory nature of the investigation and the absence of previous studies investigating the effects of *ers* themselves, no predictions were made.

9.2.3 Methods

Stimuli

The stimuli (see section 5.3 and Appendix C) were utterances that were either control utterances that did not include a repair, or repair utterances that included a repair disfluency comprising two consecutive verbs, the second of which was the repair target. To maximise the chance of observable effects, repairs were two consecutive verbs which usually results in syntactic incongruency full word repairs rather than partial word repairs because these have previously been shown to be more disruptive to processing (Brennan & Schober, 2001). Utterances were either *er*-free ($-er$ utterances) or included an *er* ($+er$ utterances) before the repair or control target word. The mean duration of the *er* was 413ms. Targets were followed by a few words to end the sentence. The $+er$ utterances were identical to the $-er$ utterances up until the point of the *er*, and did not include any other features of disfluency. This design made it possible to compare the effect of processing the *er* itself, with the processing of a control word appearing in the same location. Table 9.1 shows an example stimulus set.

Table 9.1: Example stimulus set comprising control utterances or utterances with a repair. Repair and control word targets are shown in bold. Half of the utterances included an *er* before the target word, indicated in square brackets.

control	It was warm today until the sun	[*er*]	**hid**	behind the clouds
repair	It was warm today until the sun went	[*er*]	**hid**	behind the clouds

Participants

Twenty right-handed native English speakers (12 male, mean age 21 years, range 18-36 years) took part in the experiment.

Procedure

The procedure followed the first part of the procedure that was described in the General Methods Chapter (section 5.5) and which was followed in Experiments 2 and 3. There was just one part to the experiment, designed to investigate the effects of repairs on online processing. These were assessed using ERP data collected during natural listening.

One hundred and sixty experimental utterances (40 each of $-er$ control, $-er$ repair, $+er$ control, $+er$ repair) were presented auditorily. Experimental utterances were interspersed with 160 filler utterances. Utterances were presented in eight blocks lasting approximately 4 minutes each, separated by a break of a few minutes. EEGs were recorded from 62 scalp electrodes using a central scalp reference and re-referenced off-line to the average of left and right mastoid recordings.

Before offline averaging, the continuous EEG files for each participants were segmented into 2152ms epochs starting 152ms before the target onset, and screened for artefacts. Artefact rejection resulted in the exclusion of 24% of the trials. Grand average ERPs were formed time-locked to the onsets of repair words and fluent control words ("word-locked"), from utterances without and with an *er* making four conditions: $-er$ control, $-er$ repair, $+er$ control, $+er$ repair, with mean trial numbers of 30, 31, 30, and 30 respectively.

Grand average ERPs were also formed, time-locked to the onsets of the *ers* ("*er* locked") from the $+er$ control utterances and the $-er$ control utterances, making 2 additional conditions: *ER* control, *ER* repair, with mean trial numbers of 31 and 30 respectively.

9.2.4 ERP results: word-locked

Figure 9.1 shows the grand average ERPs time-locked to repair and control word onsets for utterances without an *er* ($-er$ utterances). Figure 9.2 shows the grand average ERPs time-locked to repair and control word onsets for utterances with an *er* immediately before the target word ($+er$ utterances).

In both $+er$ utterances (Figure 9.1) and $-er$ utterances (Figure 9.2), relative to control words, repair words show a positivity over posterior sites with a right hemisphere bias, which emerges around 100ms and continues until around 800ms. In

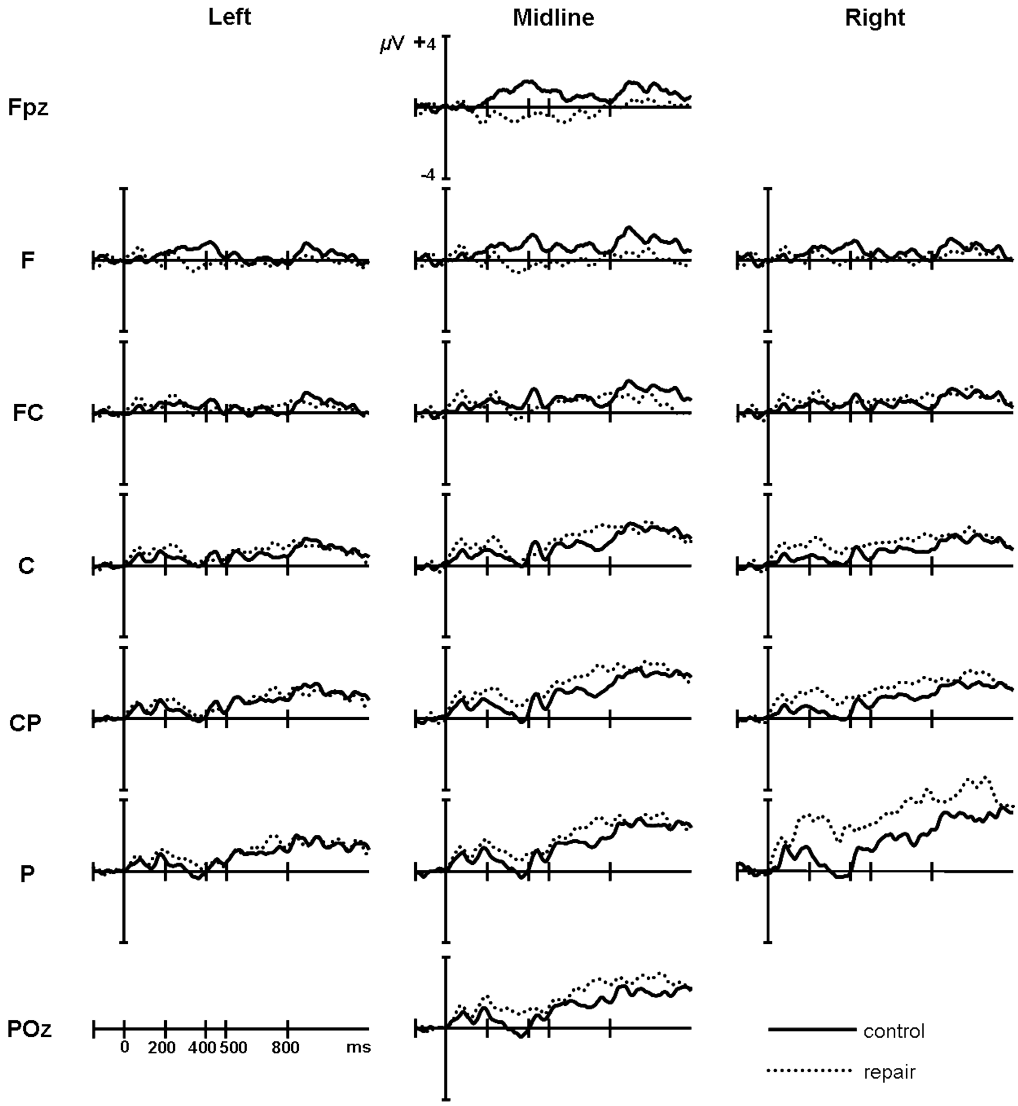

Figure 9.1: Grand average ERPs (n=20) for utterances without an *er* (−*er utterances*) relative to control (solid lines) or repair (dotted lines) word onsets at frontal pole (Fp), frontal (F), fronto-central (FC), central (C), centro-parietal (CP), parietal (P), and occipito-parietal (PO) locations, for electrodes at the midline and grouped over left (electrodes 1, 3, 5) and right (electrodes 1, 3, 5) hemispheres. Repair words show a relative negativity at frontal pole/frontal sites which emerges around 100ms and continues until around 800ms. The negativity is accompanied by a relative positivity over posterior sites with a right hemisphere bias.

−*er* utterances the positivity is accompanied by a relative bilateral negativity over the most frontal sites. In +*er* utterances there is no indication of a negativity.

ERPs were quantified by measuring the mean amplitude of the ERP difference between repair and controls words over three time windows: the standard Left An-

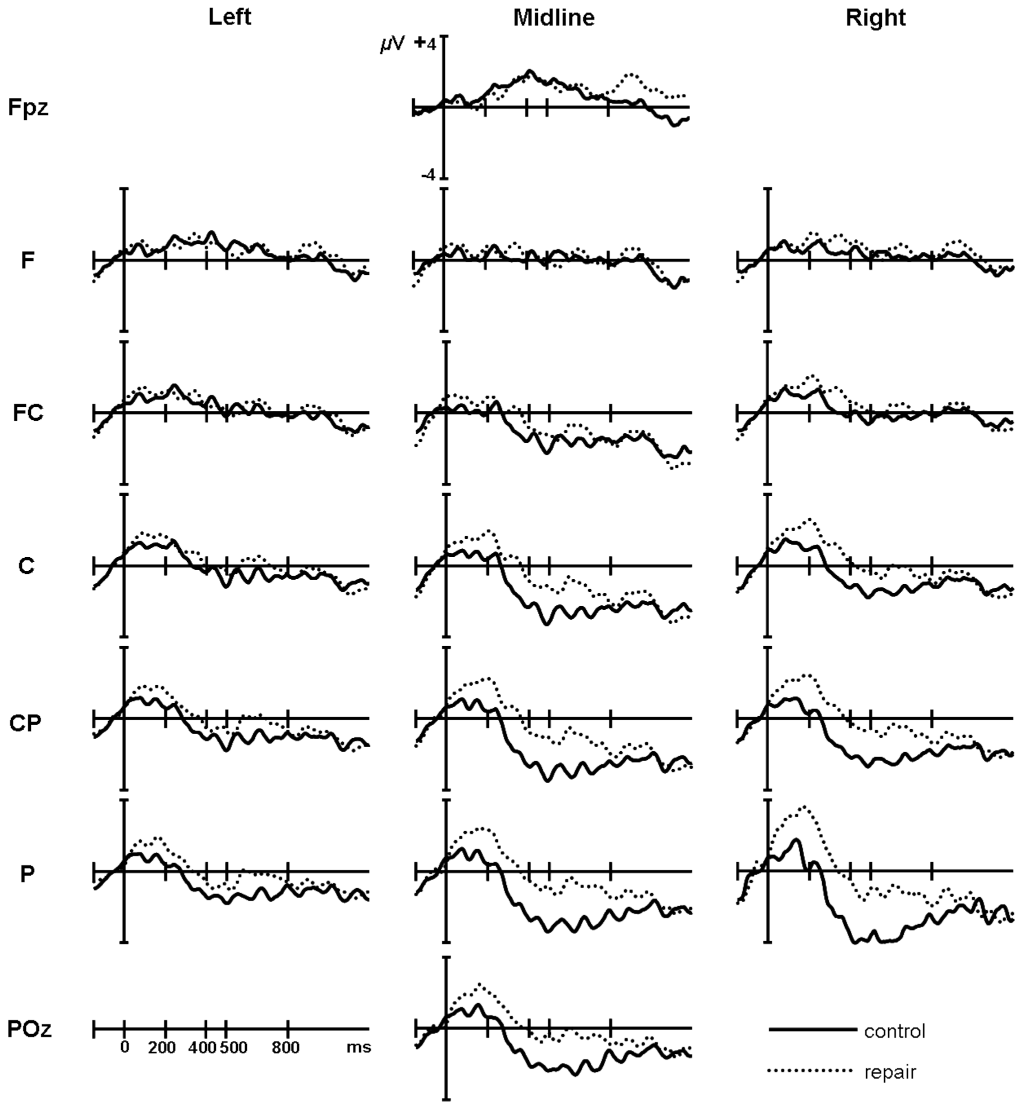

Figure 9.2: Grand average ERPs (n=20) for utterances with an *er* before the target word (+*er utterances*) relative to control (solid lines) or repair word (dotted lines) onsets at frontal pole (Fp), frontal (F), fronto-central (FC), central (C), centro-parietal (CP), parietal (P), and occipito-parietal (PO) locations, for electrodes at the midline and grouped over left (electrodes 1, 3, 5) and right (electrodes 1, 3, 5) hemispheres. Repair words show a relative positivity over posterior sites with a right hemisphere bias which emerges around 100ms and continues until around 800ms. Unlike −*er utterances* there is no indication of a negativity.

terior Negativity (LAN) time window (200–400ms), the standard P600 time window (500–800ms), plus an earlier time window (0–200ms). Topopgraphic distributions of the effects (difference between ERPs for repair and control words) for +*er* and −*er* utterances over these time windows are shown in Figure 9.3.

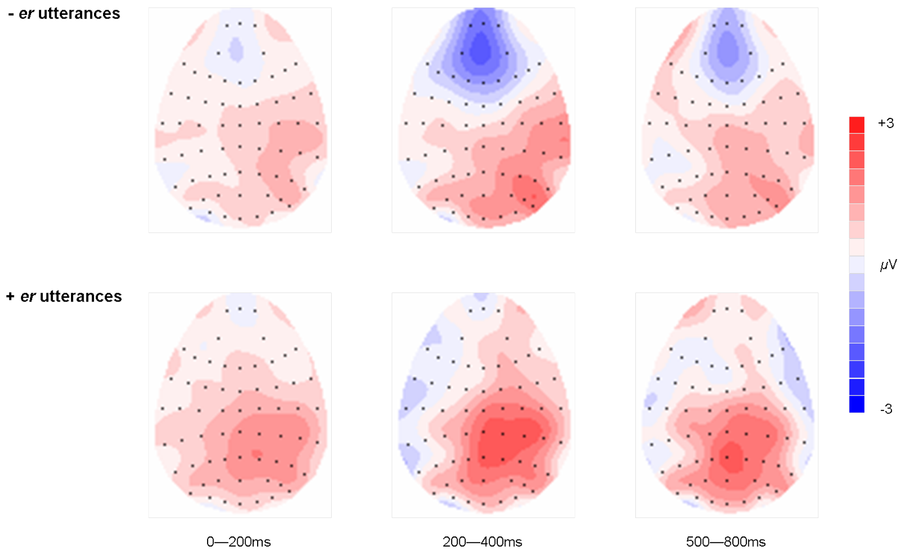

Figure 9.3: Scalp topographies showing the repair effects over three time windows: 0–200ms, 200–400ms and 500–800ms, for utterances with no *er* ($-$*er utterances*) and for utterances with an *er* before the target word (+*er utterances*). $-$*Er utterances* show a negative effect over frontal sites bilaterally and a positive effect over posterior sites with a right hemisphere bias over 200–400ms. Over 0–200ms and 500–800ms, there are indications of similar but weaker negative and positive effects. +*Er utterances* also show a positive effect over posterior sites with a right hemisphere bias over 200–400ms but it also appears over 0–200ms. Over 500–800ms, there is indication of a simliar posterior positivity with a more bilateral distribution. There is no indication of a negative effect.

Magnitude and topographic analyses were performed following the strategy described in the General Methods Chapter (sections 5.6.1 and 5.6.2) and used in Experiments 2–4, with some changes made to the factors. The analyses included a factor of repair [fluent, repair] rather than predictability. The "hemispheric" analysis included factors of repair [fluent, repair], location [F, C, P], hemisphere [left, right], and site [superior: electrode 1/2, medial: electrode 3/4, inferior: electrode 5/6]; the "midline analysis" included factors of predictability [repair, fluent], location [anterior, posterior], and site [superior: FC/CP, mid: F/P, inferior: FP/ PO]). Quantitative comparisons between the magnitude of the effects between +*er* and $-$*er* incorporated a factor of *er* [+*er*, $-$*er*] into the "hemispheric" and "midline" analyses. Figure 9.4 shows the electrodes used in the analyses.

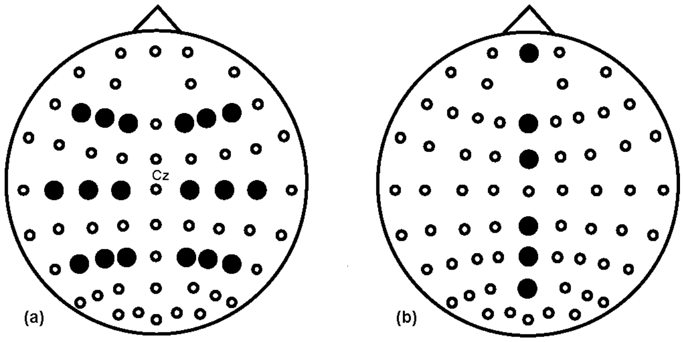

Figure 9.4: Schematic maps of the 62 electrodes sites with highlighted sites from which ERPs were analysed. Panel (a) shows the electrodes used in the "hemispheric analysis" which incorporated factors of predictability [predictable, unpredictable], location [F, C, P], hemisphere [L, R], and site [superior: electrode 1/2, medial: electrode 3/4, inferior: electrode 5/6], and panel (b) shows the electrodes used in the "midline analysis" which incorporated factors of predictability [predictable, unpredictable], location [anterior, posterior], and site [superior: FC/CP, mid: F/P, inferior: Fp/PO]. Electrode Cz is labelled for reference.

Word-locked: 0–200ms

Over 0–200ms, the topographic analysis incorporating all 62 electrodes provided no evidence of distributional differences between the effects for $-er$ and $+er$ utterances (see Figure 9.3). The lack of distributional differences was also supported by midline and hemispheric topographic analyses. Thus there was no reason to suppose that different neural generators were responsible for the recorded effects of a repair and no reason to suppose the engagement of different cognitive processes during this time period. The effects for $-er$ and $+er$ utterances were analysed separately, but because there were no significant effects for $-er$ utterances (see below), the effects for $-er$ and $+er$ utterances were not quantitatively compared.

For $-er$ utterances, the midline and hemispheric analyses showed no significant effects. For the $+er$ utterances, a hemispheric analysis revealed a marginally significant effect of repair [$F(1, 19) = 3.445$, $\eta_p^2 = .153$, $p = .079$] and a marginal interaction between repair and location [$F(2, 38) = 3.251$, $\eta_p^2 = .146$, $p = .065$], reflecting a relative positivity at parietal sites. This was supported by the midline analysis which showed an interaction between repair and location [$F(1, 19) = 5.745$,

$\eta_p^2 = .232$, $p = .027$], and between repair, location and site [$F(2, 38) = 5.418$, $\eta_p^2 = .222$, $p = .019$], reflecting greater relative positivity at the posterior location, particularly at inferior sites.

Word-locked: 200–400ms

Over 200–400ms, the hemispheric topographic analysis suggested a distributional difference between the effects for $+er$ and $-er$ utterances (although distributional differences were not revealed in the topographic analysis incorporating all 62 electrodes, or in the midline topographic analysis; see Figure 9.3). There was an interaction between *er* and site [$F(2, 38) = 7.683$, $\eta_p^2 = .288$, $p = .010$], reflecting greater relative positivity for $+er$ utterances than for $-er$ utterances at inferior sites. Given their distributional differences, the effects for $+er$ and $-er$ utterances were analysed separately and not quantitatively compared.

For $-er$ utterances, the hemispheric analysis provided no evidence of any laterality differences. There was however, an interaction between repair and location [$F(2, 38) = 5.392$, $\eta_p^2 = .221$, $p = .027$], reflecting a relative negativity for repairs at the frontal location but a relative positivity at central and parietal locations. This finding was supported by the midline analysis which also showed an interaction between repair and location [$F(1, 19) = 6.468$, $\eta_p^2 = .254$, $p = .020$], reflecting a negative effect at the anterior location and a positive effect at the posterior location.

Because of the apparent presence of two distinct effects—a frontal negativity and parietal positivity—analyses were performed at more discrete locations to assess for the presence of these two effects independently. At the frontal location, a hemispheric analysis showed an interaction between repair and site [$F(2, 38) = 7.757$, $\eta_p^2 = .290$, $p = .004$], reflecting a relative negativity at sites close to the midline. This was supported by a midline analysis at the anterior location which showed an interaction between repair and site [$F(2, 38) = 6.213$, $\eta_p^2 = .246$, $p = .013$], reflecting a relative negativity at the sites away from the centre. At the parietal location, a hemispheric analysis showed a marginal effect of repair [$F(1,19) = 3.779$, $\eta_p^2 = .166$, $p = .067$] and a marginal interaction between repair and hemisphere [$F(1,19) = 3.609$, $\eta_p^2 = .160$, $p = .073$], reflecting a weak relative positivity, particularly over the right hemisphere. A midline analysis at the posterior location showed no significant effects.

For $+er$ utterances, the hemispheric analysis suggested laterality differences. There was an interaction between repair and hemisphere [$F(1,19) = 9.700$, $\eta_p^2 = .338$, $p = .006$], reflecting greater relative positivity over the right than over the left hemisphere. There was also an interaction between repair and site [$F(2,38) = 6.007$, $\eta_p^2 = .240$, $p = .020$], reflecting greater positivity over sites closer to the midline. The midline analysis revealed an interaction between repair and location [$F(1,19) = 12.622$, $\eta_p^2 = .399$, $p = .002$], reflecting greater relative positivity at the posterior than at the anterior location. There was also a three-way interaction between repair, location and site [$F(2,38) = 4.050$, $\eta_p^2 = .176$, $p = .049$], reflecting a larger effect of location on the relative positivity over superior sites.

Word-locked: 500–800ms

Over 500–800ms, the hemispheric topographic analysis suggested distributional differences between the effects for $+er$ and $-er$ utterances (although distributional differences were not revealed in the topographic analysis incorporating all 62 electrodes, or in the midline topographic analysis; see Figure 9.3). There was an interaction between *er* and site [$F(2,38) = 4.333$, $\eta_p^2 = .286$, $p = .043$], reflecting greater relative positivity for $+er$ utterances than for $-er$ utterances at superior sites. Given their distributional differences, the effects for $+er$ and $-er$ utterances were analysed separately and not quantitatively compared.

For $-er$ utterances, the hemispheric analysis revealed a three-way interaction between repair, location and site [$F(4, 76) = 4.369$, $\eta_p^2 = .187$, $p = .016$], reflecting the relative negativity at frontal superior sites. The midline analysis revealed an interaction between repair and location [$F(1, 19) = 4.743$, $\eta_p^2 = .200$, $p = .042$] reflecting the relative negativity at the anterior location and the relative positivity at the posterior location.

As for the 200–400ms time window, because of the apparent presence of two distinct effects: a frontal negativity and parietal positivity, analyses were performed at more discrete locations to assess for the presence of these two effects independently. At the frontal location, a hemispheric analysis showed an interaction between repair and site [$F(2, 38) = 7.117$, $\eta_p^2 = .272$, $p = .003$], reflecting a relative negativity at sites close to the midline. A midline analysis at the anterior location showed no significant effects. At the parietal location, a hemispheric analysis showed a marginal interaction between repair and hemisphere [$F(1, 19) = 3.657$, $\eta_p^2 = .161$, $p = .071$], reflecting a weak relative positivity, particularly over the right hemisphere. A midline analysis at the posterior location showed no significant effects.

For $+er$ utterances the hemispheric analysis revealed an interaction between repair and site [$F(2, 38) = 4.530$, $\eta_p^2 = .193$, $p = .043$], reflecting the greater relative positivity towards the midline. The midline analysis revealed an interaction between repair and location [$F(1, 19) = 4.861$, $\eta_p^2 = .204$, $p = .040$], reflecting greater relative positivity at the posterior than at the anterior location.

Word-locked: effects over time

For $-er$ utterances, there were significant effects for the 200–400ms and 500–800ms time windows and therefore the effects between these time windows were compared. The topographic analysis incorporating all 62 electrodes provided no evidence for distributional differences between the effects for the two time windows (see Figure 9.3). Thus the effects for the two time windows were quantitatively compared.

Because of the presence of two distinct effects—a frontal negativity and a parietal positivity—analyses were performed at discrete locations. At the frontal location, the hemispheric analysis showed an interaction between repair and site [$F(2, 38) = 10.665$, $\eta_p^2 = .360$, $p < .001$], reflecting a relative negativity at sites close to the midline. A midline analysis at the anterior location showed an interaction between repair and site [$F(2, 38) = 4.730$, $\eta_p^2 = .199$, $p = .032$], reflecting greater negativity at sites away from the centre. There were no interactions with epoch, and therefore no evidence for quantitative changes between two time windows. At the parietal location, a hemispheric analysis showed an interaction between repair and hemisphere [$F(1, 19) = 5.067$, $\eta_p^2 = .211$, $p = .036$], reflecting greater positivity over the right hemisphere. A midline analysis at the posterior location showed no significant effects. There were no interactions with epoch, and therefore no evidence for quantitative changes between two time windows.

For $+er$ utterances, there were significant effects for the 0–200ms, 200–400ms and 500–800ms time windows and therefore the effects between successive time windows were compared. For 0–200ms and 200–400ms, the hemispheric topographic analysis suggested distributional differences between the $+er$ effects, predominantly reflecting the initial right laterality of the relative positivity which became less lateralised over time. There was an interaction between epoch and hemisphere [$F(1, 19) = 5.882$, $\eta_p^2 = .236$, $p = .025$], reflecting the greater right laterality of the relative positivity in the early epoch, and interactions between epoch and site [$F(2, 38) = 4.406$, $\eta_p^2 = .188$, $p = .041$] and between epoch, location and site [$F(4, 76) = 3.933$, $\eta_p^2 = .172$, $p = .031$], reflecting the focus of the repair positivity over superior sites in the later time window, particularly at the parietal location.

Given their distributional difference, the effects for the two epochs were not quantitatively compared.

For 200–400ms and 500–800ms, the hemispheric topographic analysis suggested distributional differences between the +*er* effects, again predominantly reflecting the increasing bilaterality of the relative positivity over time. There was an interaction between epoch and hemisphere [$F(1, 19) = 7.400$, $\eta_p^2 = .280$, $p = .014$], and between epoch, hemisphere and site [$F(2, 38) = 7.478$, $\eta_p^2 = .282$, $p = .006$], reflecting the greater right laterality of the positivity in the earlier time window. Given their distributional difference, the effects for the two epochs were not quantitatively compared.

9.2.5 ERP results: er-locked

Figure 9.5 shows the grand average ERPs time-locked to *er* and word onsets for utterances without a repair (control utterances). Figure 9.6 shows the grand average ERPs time-locked to *er* and word onsets for utterances with a repair after the *er* (repair utterances).

For both control utterances and utterances including a repair, relative to control words, *ers* show a negativity which is broadly distributed over the scalp, but appears larger over frontal and fronto-central sites. The effect onsets around 150ms and continues until around 250ms. Following the negativity, there is a relative positivity which is broadly distributed over the scalp, but appears larger over central and centro-parietal sites. The effect onsets around 300ms and continues until around 900ms. There is some indication of a relative positivity preceding the negativity, which is larger over frontal and fronto-central sites, and lasts from around 50–150ms.

ERPs were quantified by measuring the mean amplitude of the ERP difference between *ers* and control words over two time windows: 152–252ms and 300–600ms.

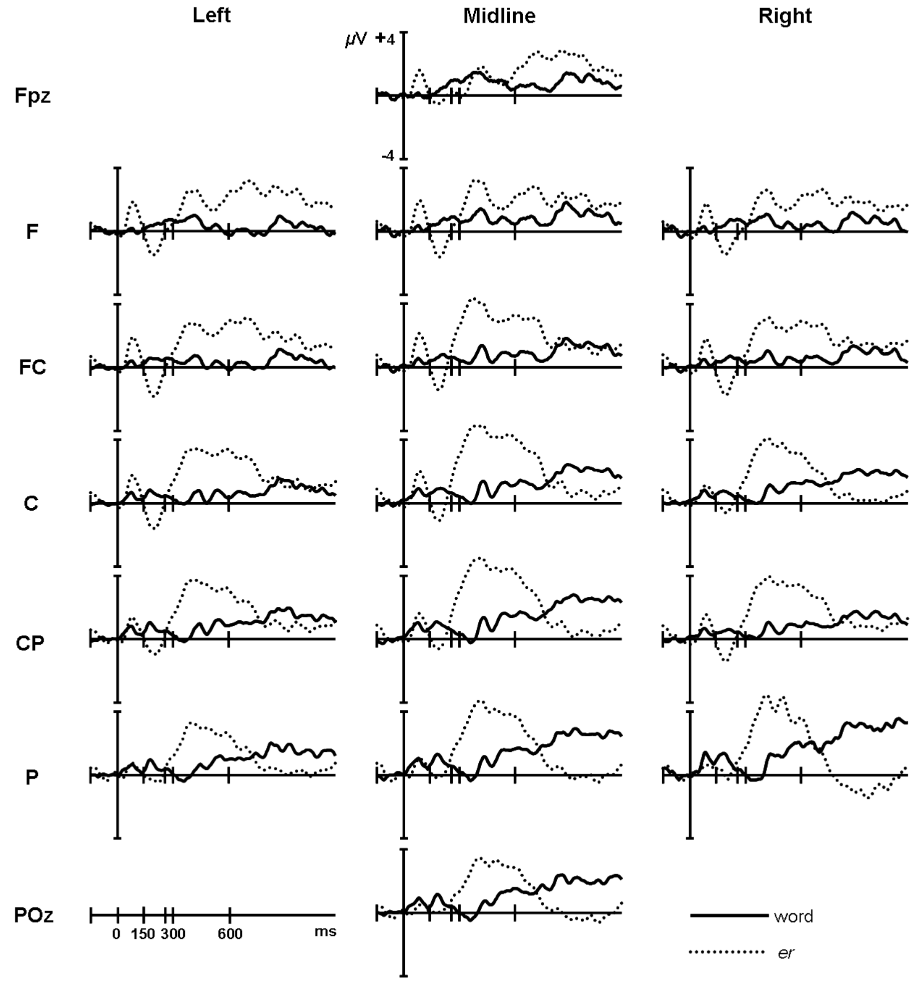

Figure 9.5: Grand average ERPs (n=20) for utterances without a repair (*control utterances*) relative to word (solid lines) or *er* (dotted lines) onsets at frontal pole (Fp), frontal (F), fronto-central (FC), central (C), centro-parietal (CP), parietal (P), and occipito-parietal (PO) locations, for electrodes at the midline and grouped over left (electrodes 1, 3, 5) and right (electrodes 1, 3, 5) hemispheres. *Ers* show a relative negativity which is broadly distributed over the scalp but larger over frontal and fronto-central sites over 150ms–250ms. The negativity is followed by a relative positivity emerging around 300ms which is broadly distributed over the scalp but larger over central and centro-parietal sites.

ERPs were also analysed in the 52–152ms time window, but showed no significant effects are are not discussed further. The effects were not analysed after 600ms because of the overlap with the word-locked effects (section 9.2.4). Topographic distributions of the effects (difference between ERPs for *ers* and control words) for repair and control utterances over the relevant time windows are shown in Figure 9.7.

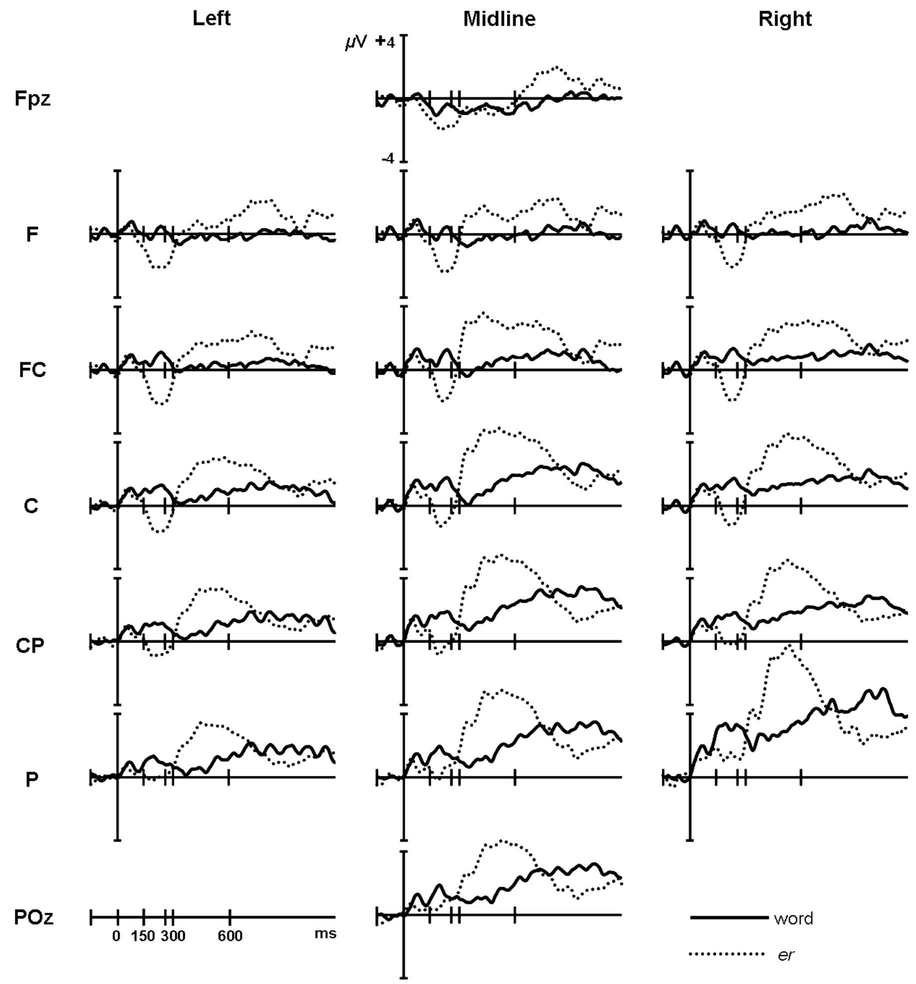

Figure 9.6: Grand average ERPs (n=20) for utterances with a repair (*repair utterances*) relative to word (solid lines) or *er* (dotted lines) onsets at frontal pole (Fp), frontal (F), fronto-central (FC), central (C), centro-parietal (CP), parietal (P), and occipito-parietal (PO) locations, for electrodes at the midline and grouped over left (electrodes 1, 3, 5) and right (electrodes 1, 3, 5) hemispheres. *Ers* show a relative negativity which is broadly distributed over the scalp but larger over frontal and fronto-central sites over 150ms–250ms. The negativity is followed by a relative positivity emerging around 300ms which is broadly distributed over the scalp but larger over central and centro-parietal sites.

Analyses were performed using the same strategy that was used for analysis of the word-locked data (section 9.2.4). Analyses incorporated a factor of *er* [control, *er*] rather than of repair.

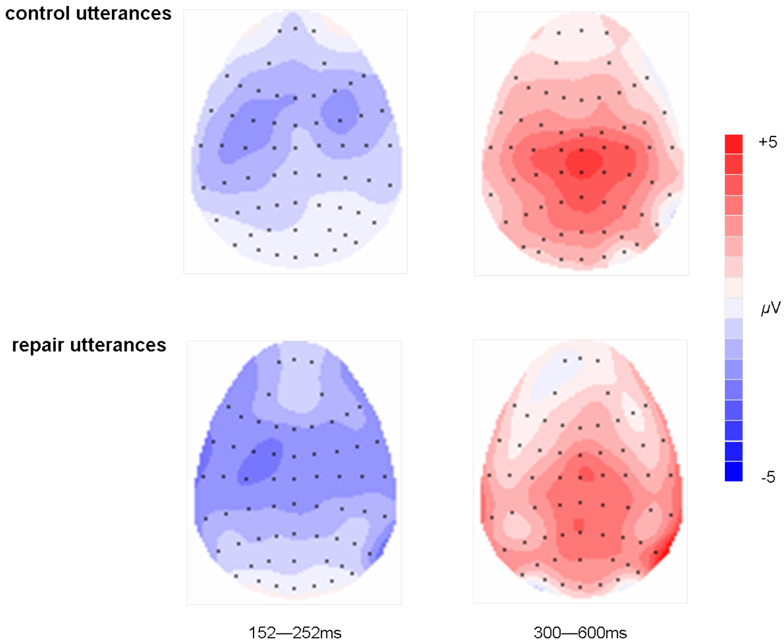

Figure 9.7: Scalp topographies showing the *er* effects over two time windows: 152–252ms and 300–600ms, for utterances without a repair (*control utterances*) and for utterances with a repair after the target *er*/word (*repair utterances*). Both control utterances and repair utterances show two effects: a negative effect over frontal and fronto-central sites bilaterally over the 150–250ms time window, and a positive effect over central and centro-parietal sites over the 300–600ms time window.

Er-locked: 152–252ms

Over 152–252ms, a topographic analysis incorporating all 62 electrodes, and the hemispheric and midline topographic analyses provided no evidence of a distributional difference between the effects for control and repair utterances. The effects for control and repair utterances were therefore analysed separately and then quantitatively compared.

For control utterances, the hemispheric analysis established that the distribution of the *er* effect was not lateralised. There was a main effect of *er* [$F(1, 19) = 6.9937$, $\eta_p^2 = .269$, $p = .016$] and an interaction between *er* and location [$F(2, 38) = 4.282$, $\eta_p^2 = .184$, $p = .038$], reflecting a relative negativity for *ers* which was larger at the frontal and central sites. The midline analysis showed a marginally significant main effect of *er* [$F(1,19) = 3.813$, $\eta_p^2 = .167$, $p = .066$], reflecting a greater negativity for *er*.

For repair utterances, the hemispheric analyses again established that the distribution of the *er* effect was not lateralised. There was a main effect of *er* [$F(1,19) = 16.124$, $\eta_p^2 = .459$, $p = .001$] and an interaction between *er* and location [$F(2,38) = 5.951$, $\eta_p^2 = .239$, $p = .020$], again reflecting a relative negativity for *ers* which was larger at the frontal and central sites. The midline analysis showed a main effect of *er* [$F(1,19) = 16.124$, $\eta_p^2 = .459$, $p = .001$] and an interaction between *er* and site [$F(2,38) = 13.898$, $\eta_p^2 = .422$, $p = .001$], reflecting a relative negativity for *ers* which was larger over sites closer to the centre.

Because the effects were laterally spread over the scalp, the quantitative comparison of the effects between control and repair utterances used the hemispheric analysis. There was a main effect of *er* [$F(1,19) = 26.319$, $\eta_p^2 = .581$, $p < .001$] and an interaction between *er* and location [$F(2,38) = 9.106$, $\eta_p^2 = .324$, $p = .004$], reflecting a relative negativity for *ers* which was larger over frontal and central sites. Importantly, there were no interactions involving repair and *er* and therefore no evidence of a difference between the amplitude of the *er* effect elicited in control utterances [1.20μV] and of the *er* effect in repair utterances [1.90μV].

Er-locked: 300–600ms

Over 300–600ms, a topographic analysis incorporating all 62 electrodes provided no evidence of a distributional difference between the effects for control and repair utterances. The lack of distributional differences was also supported by midline and hemispheric topographic analyses. The effects for control and repair utterances were therefore analysed separately and then quantitatively compared.

For control utterances, the hemispheric analysis showed a main effect of *er* $[F(1,19) = 37.701$, $\eta_p^2 = .665$, $p = .000]$, reflecting a relative positivity for *ers*. There were also interactions between *er* and location $[F(2,38) = 9.986$, $\eta_p^2 = .345$, $p = .001]$ and between *er* and site $[F(2,38) = 19.285$, $\eta_p^2 = .504$, $p < .001]$, reflecting a larger positivity over central and midline sites. There was a three-way interaction between *er*, hemisphere and site $[F(2,38) = 3.714$, $\eta_p^2 = .164$, $p = .041]$, reflecting a larger positivity over the left hemisphere at the inferior sites. The midline analysis showed a main effect of *er* $[F(1,19) = 34.899$, $\eta_p^2 = .647$, $p < .001]$, reflecting a relative positivity for *ers*. There were also interactions between *er* and location $[F(1,19) = 19.211$, $\eta_p^2 = .503$, $p < .001]$, between *er* and site $[F(2,38) = 21.579$, $\eta_p^2 = .532$, $p < .001]$, and between *er*, location and site $[F(2,38) = 7.988$, $\eta_p^2 = .296$, $p = .006]$, reflecting a larger positivity for the posterior location, particularly for sites closer to the centre.

For repair utterances, the hemispheric analysis showed a main effect of *er* $[F(1,19) = 14.719$, $\eta_p^2 = .437$, $p = .001]$, reflecting a relative positivity for *ers*. There were also interactions between *er* and location $[F(2,38) = 5.964$, $\eta_p^2 = .239$, $p = .020]$ and between *er* and site $[F(2,38) = 30.196$, $\eta_p^2 = .614$, $p < .001]$, reflecting a larger positivity at central and midline sites. A midline analysis showed a main effect of *er* $[F(1,19) = 17.302$, $\eta_p^2 = .477$, $p = .001]$, reflecting a relative positivity for *ers*. There were also interactions between *er* and location $[F(1,19) = 7.297$, $\eta_p^2 = .277$, $p = .014]$, between *er* and site $[F(2,38) = 13.714$, $\eta_p^2 = .419$, $p = .001]$ and between *er*, location and site $[F(2,38) = 15.880$, $\eta_p^2 = .455$, $p < .001]$, reflecting a larger positivity for the posterior location, particularly for sites closer to the centre.

Because the effects were laterally spread over the scalp, the quantitative comparison of the effects between control and repair utterances focused on the hemispheric analysis. There was a main effect of *er* $[F(1,19) = 44.148$, $\eta_p^2 = .699$, $p < .001]$ and interactions between *er* and location $[F(2,38) = 26.343$, $\eta_p^2 = .581$, $p < .001]$, between *er* and site [$F(2, 38) = 37.760$, $\eta_p^2 = .665$, $p < .001$], and between *er*, location and site [$F(2, 38) = 3.394$, $\eta_p^2 = .152$, $p < .025$], reflecting a relative positivity for *ers*, which was larger at the central and parietal locations and at sites closer to the midline, and largest at central sites closest to the midline. Importantly there were no interactions involving *er* and repair and therefore no evidence of a difference between the amplitude of the *er* effect elicited in control utterances [2.16μV] and of the *er* effect in repair utterances [1.75μV].

Er-locked: effects over time

Topographic analyses incorporating all 62 electrodes established that the negative and positive effects in the early and later time window respectively had distributional differences suggesting the presence of two distinct effects and the engagement of different populations of neurons.

For control utterances, there was an interaction between between epoch and site [$F(61, 1159) = 12.231$, $\eta_p^2 = .392$, $p < .001$], in addition to a main effect of epoch [$F(1, 19) = 16.352$, $\eta_p^2 = .463$, $p = .001$] and a main effect of site [$F(61, 1159) = 3.518$, $\eta_p^2 = .156$, $p = .009$]. Similarly, for repair utterances, there was an interaction between epoch and site [$F(61, 1159) = 12.989$, $\eta_p^2 = .406$, $p < .001$], in addition to a main effect of epoch [$F(1, 19) = 34.925$, $\eta_p^2 = .648$, $p < .001$] and a main effect of site [$F(61, 1159) = 3.813$, $\eta_p^2 = .167$, $p = .012$].

9.2.6 Discussion

The aim of the experiment was to investigate the effects of repairs on language processing and whether the effects are modified by the presence of an *er* during the edit interval. Participants listened for understanding to a series of short utterances, which included either repair or control target words. Additionally, some of the utterances contained an *er* filled pause before the target.

Summary and interpretation

ERPs for utterances both with and without an *er* showed that relative to control words, repairs elicited a positivity over centro-parietal and parietal sites. For utterances without an *er*, the positivity was present in the 200–400ms and 500–800ms time windows, and was accompanied by a negativity over frontal/frontal polar sites. For utterances with an *er* before the target, the positivity was also weakly present in the 200–400ms and 500–800ms time windows, and also in an earlier 0–200ms time window. There was no indication of a negativity for $+er$ utterances.

The distribution of the positivity for $+er$ utterances is compatible with its identification as a P600 effect, associated with the syntactic repair or reanalysis following the violation of syntactic preferences (e.g., Friederici, 1995, 2002; Friederici et al., 1996; Gunter et al., 1997, section 4.4.1). The positivity for $-er$ utterances is superficially similar and seems likely to be a P600 effect. However, because the positivity was accompanied by a relative negativity, there were distributional differences between the effects for $+er$ and $-er$ utterances. The earlier onset of the positivity in $+er$ utterances suggests that reanalysis processes associated with the repair commenced earlier, suggesting that the *er* provided listeners with a cue to the possiblity of an upcoming repair.

The frontal negativity observed for repairs in $-er$ utterances bears similiarites to two previously identified effects which have been associated with aspects of syntactic processing: the Left Anterior Negativity (LAN) effect and the NRef. The LAN is associated with the detection of a syntactic violation such as word category (e.g., Friederici et al., 1993, 1996; Hahne & Friederici, 1999; Neville et al., 1991, see section 4.4.2). Although LANs are typically left lateralised, this is not always the case (e.g., Hagoort et al., 2003) and therefore the bilaterality of the negativity here does not rule out its identification as a LAN. There are two main interpretations of the LAN. The first is that it reflects difficulties in thematic role assignment (Friederici, 2002). The second is that it reflects general working memory processes (Kluender & Kutas, 1993; King & Kutas, 1995). When repairs are from the same syntactic category as their reparanda, as in the current experiment, this may cause difficulties in thematic role assigment. *Ers* may alleviate this difficulty by preparing listeners for the possibility of an upcoming repair which would replace previous information. Alternatively, repairs may result in a general burden on working memory processes as listeners attempt to reconcile the repair with the prior discourse, by identifying the onset of the repair and locating the reparandum. *Ers* may ease this burden by preparing listeners for the possibility of an upcoming repair, and indicating the onset of the repair. The increase in time which the *er* adds to the signal may offer an additional benefit to processing.

The negativity has a more anterior distribution than a typical LAN, and another possible interpretation of the effect is that it is related to the NRef (Van Berkum, Brown, & Hagoort, 1999b; Van Berkum et al., 2003, see section 4.4.3). The NRef is a frontally maximal sustained negative shift that has been observed in response to ambiguous referents, but there is no reason to suppose that the NRef is specific to referential processes (Van Berkum et al., 2007). One possibility is that it reflects control processes involved in resolving ambiguity, for example inferencing or the search for cues in the episodic memory of the discourse. Another possibility is that it reflects the processing costs associated with the increase in demands on working memory that are imposed when two alternative referential interpretations are stored. This account is similar to the working memory burden account of the LAN (see above). If the frontal negativity reflects an increase in the demands on working memory, more complex repairs, for example those with longer reparanda or which involve a complete deletion of the reparanda, should increase the amplitude of the effect. If the frontal negativity reflects more syntax-specific processes associated with a syntactic violation, such as difficulties with thematic role assignment, the effect might be reduced if the reparandum-repair construction is not necessarily a syntactic violation. Most of the repairs in the current study were verbs and a verb-verb construction is almost always a syntactic violation. By contrast, noun-noun constructions are often syntactically legal and therefore noun-noun repairs might eliminate the negativity.

In the one previous ERP study on the processing of repairs, repairs elicited a relative posterior positivity which onset early (like the positivity in the $+er$ utterances). No frontal negativity (nor any other effects) were reported (McAllister et al., 2001). The study only analysed midline electrodes and only presented data from the Pz electrode, and one possibility is that the authors ignored the presence of a negativity. Another possibility is that differences in the effects are a result of differences between the stimuli used by McAllister et al. (2001) and those of the current study. McAllister et al. (2001) used spontaneous speech which means there may have been cues in the reparanda that a repair was imminent, for example word fragments (Lickley & Bard, 1998) or the absence of co-articulation between successive words (Lickley, 1996). Such cues may have allowed syntactic reanalysis processes to commence promptly, as an *er* cue did in the current experiment. In the current study, the aim was to specifically investigate the role of an *er* as a cue: other cues were avoided by using $-er$ stimuli that were identical until the point of the repair (and $+er$ stimuli that were identical until the point of the *er*). The interaction between *er* and other cues on the processing of repairs remains an area for future research.

The syntactic and semantic relationship between the reparandum and the repair is a potentially important feature in the elicitation of the P600 effect. Lau and Ferreira (2005) showed that the representation (as measured by offline grammaticality judgments) of repair verbs that were ambiguous between a simple past tense or a past participle, were affected by whether the preceding reparanda supported the correct interpretation (see section 2.4.3). These effects on representation suggest that there are effects on processing which could be observed using ERPs. It is predicted that reparanda that do not support the interpretation of ambiguous repairs (e.g., between simple past tense or past participle) would be more costly for syntactic processes associated with reanalysis, leading to a larger P600 effect. It is also worth noting that in the present experiment some repairs were semantically related to the reparanda and hence confirmed or emphasised the original meaning; some repairs contradicted the original meaning. Future work should consider the effects of the relationship between reparanda and repairs on processing.

The er effect

Of secondary importance to the investigation was the effect of the *ers* themselves. Relative to words, *ers* elicited two distinct effects: a negativity over frontal and fronto-central sites (152–252ms), followed by a positivity over central and centro-parietal scalp regions (300–600ms). These effects were present both in control utterances and in repair utterances and showed no qualitative or quantitative differences. Given the timing and distribution of these effects, it seems likely that these are a Mismatch Negativity (MMN) and P300b respectively.

The MMN is associated with automatic neural processes when there is an acoustic mismatch between the incoming stimulus and the sensory memory trace created by previous stimuli known as echoic memory (Näätäanen & Winkler, 1999; Schröger, 1997, section 4.2.1). The P300b, which is often observed following an MMN, is elicited in response to contextually infrequent stimuli and is associated with attention to such deviant stimuli (Donchin, 1981; Donchin & Coles, 1988; Polich, 2004, section 4.2.2). The presence of these effects therefore suggests that the *ers* in the current experiment were perceived to be acoustically deviant with respect to the preceding speech. This is unsurprising because the *ers* were relatively infrequent non-lexical interruptions to an otherwise lexical context. The presence of a P300b is consistent with previous suggestions that *ers* heighten listeners' attention (Brennan & Schober, 2001; Collard, Corley, MacGregor, & Donaldson, in press; Fox Tree, 2001). The results are also compatible with an attentional-based account for the increase in recognition memory performance for words following an *er* reported in Experiments 1 and 2 (Chapter 6). The role of attention in disfluency processing is discussed further in section 10.3.2.

These effects are similar to effects which have been observed when listeners encountered unexpected silent pauses in connected speech (Besson et al., 1997; Mattys, Pleydell-Pearce, Melhorn, & Whitecross, 2005). Besson et al. (1997) presented participants with utterances which were either highly constrained proverbs which ended in predictable words, or unconstrained and hence ended in unpredictable words. Sometimes the utterance-final critical word was delayed unexpectedly by 600ms. ERPs showed that in response to an unexpected delay, there was an emitted potential comprising a negative component which peaked around 100ms after pause onset, followed by positive component which peaked around 350–400ms. The effects, particularly the positivity, were larger when the pause followed a highly constrained utterance (a proverb) than when it followed a weakly constrained utterance. A similar negative-positive complex has been observed using musical phrases as stimuli when the last note of a melody was delayed unexpectedly (Besson & Faita, 1995; Besson, Faita, & Requin, 1994). Mattys et al. (2005) observed a similar negative-positive complex in response to 200ms pauses inserted into utterance-final words in connected speech, when listeners had to explicitly detect the presence of those pauses.

The negative-positive complex in response to unexpected delays has been interpreted as reflecting temporal disruption (Besson et al., 1997; Mattys et al., 2005), although an interpretation based on the acoustic deviance of silent pauses is also possible. If the negative and positive effects observed in reponse to *ers* in the current experiment reflect temporal disruption, the presence of prolongation disfluencies leading up to the *er* should reduce the amplitude of the effects because prolongations also alter the temporal flow of speech.

Whatever the functional interpretation of the *er* effect, importantly, there was no evidence that it differed depending on whether it was in the context of a control utterance or a repair utterance (the two types of utterances necessarily differed). This means that the effect of an *er* on the repair effect (word-locked difference between repairs and controls) was not because of a differential spillover effect of the *er* between the two types of utterances.

Er as a cue

The results demonstrate that *er* affects the processing of repairs and that this may be due to *ers* heightening attention. One possibility, which is supported by the memory results from Experiments 2 and 3, is that the effect of an *er* is related to the increase in time which it adds to the signal. In particular, the time during the edit interval may mark the offset of the reparandum and the onset of the repair, enabling easier processing of the repair. This means that silent pauses at the edit interval of repairs should lead to similar effects on processing as those observed in the current experiment.

Another possibility, not incompatible with a time-based account and supported by the results from Experiment 4, is that the nature of the interruption is also important. The effects of *ers* may be partly driven by the extent to which they heighten attention and this may be related to the extent to which they are distinct from the surrounding lexical speech context. If the distinctiveness of the disfluency from its surrounding context is important, a non-lexical interruption such as a cough should lead to similar effects on processing as those observed in the current experiment.

9.3 Conclusions

The chapter reported an experiment that investigated the effects of processing repairs and whether the effects are modulated by the presence of an *er* during the edit interval. Repairs elicited a relative posterior positivity, identifiable as a P600 effect, suggesting that repairs result in syntactic reanalysis and repair. In the absence of cues for an upcoming repair, the positivity was accompanied by a relative negativity, thought to reflect additional processes engaged because of the syntactic violation. The presence of an *er* eliminated this negativity, demonstrating that *ers* affect the processing of repairs, maybe by preparing listeners for the possibility of an upcoming repair. The distinctiveness of an *er* in a lexical context and the additional time which an *er* introduces before the onset of the repair may reduce the processing cost of initially encountering a syntactic violation. *Ers* may enable easier identification of the onset of repair and of the reparandum, which means syntactic reanalysis can commence more promptly.

Chapter 10

General Discussion

This thesis described a series of experiments that investigated disfluent speech comprehension. The effects of disfluencies on language processing were assessed using response times in a Lexical Decision Task (Experiment 1) and Event-Related Potentials (Experiments 2–5); longer-term consequences on language representation were assessed using a surprise recognition memory test (Experiments 1–4). In this chapter I summarise the main findings from the experiments, suggest possible mechanisms to account for the results, and suggest areas for future research.

10.1 Summary of the main findings

The summary is split into two sections. The first section summarises the results of Experiments 1–4, which investigated the effects of three types of disfluencies—*ers* (Chapter 6), silent pauses (Chapter 7), and repetitions (Chapter 8)—on the comprehension of subsequent words which were straightforward continuations of the pre-interrupted speech. The second section summarises the results of Experiment 5 which investigated the effect of *ers* on the comprehension of subsequent repair words

(Chapter 9). The results from the investigations of the effects of processing repetition (Chapter 8), repair (Chapter 9), and *ers* (Chapter 9) disfluencies themselves are also summarised.

10.1.1 Summary: Experiments 1–4

LDTs were faster for predictable and unpredictable words preceded by an *er*. ERPs showed that the N400 effect elicited in response to contextually unpredictable relative to predictable words was attenuated by the presence of a pre-target *er*, reflecting a reduction in the standard difference where unpredictable words are harder to integrate into their contexts (Kutas et al., 2006). In addition, recognition memory was enhanced for words which had been preceded by an *er*. Silent pauses had no observable effect on the standard N400 but did increase recognition memory for subsequent words. Repetition disfluencies had no effect on the standard N400 and no effect on the recognition memory for subsequent words. For all types of disfluent utterances, relative to predictable words, unpredictable words elicited a Late Positive Complex (LPC), typically associated with aspects of memory control and retrieval (Federmeier et al., 2007; Van Petten et al., 1991, see section 4.3.3). These processes may be activated as listeners try to resume structural fluency of the message, after the interruption.

ERPs also showed an effect of repetition disfluencies themselves. Relative to control words, repetitions elicited an early onsetting sustained posterior positivity reflecting detection of the repetition and possibly syntactic reanalysis (Friederici, 2002, see section 4.4.1).

10.1.2 Summary: Experiment 5

ERPs showed that, relative to control words, repair disfluencies elicited a P600 effect, reflecting syntactic reanalysis (Friederici, 2002), and a bilateral frontal negativity. The functional interpretation of the negativity is not clear, but it may reflect difficulties in thematic role assignment (Friederici, 2002) or more general demands on working memory (Kluender & Kutas, 1993; King & Kutas, 1995, see section 4.4.2), as listeners attempt to reconcile the repair with the prior discourse. A pre-target *er* eliminated the negativity, and led to an earlier onsetting P600 effect for repairs, compatible with *er* providing a cue about the possibility of an upcoming repair.

Relative to words, *ers* elicitated an MMN associated with the identification of deviance in the acoustic environment (Schröger, 1997, see section 4.2.1), probably because they were non-lexical interruptions within a lexical context, and a P300b, associated with the attention to such deviant stimuli (Donchin, 1981; Donchin & Coles, 1988; Polich, 2004, see section 4.2.2).

10.2 Comparing different disfluencies

The series of experiments reported in the thesis was the first attempt at a systematic investigation of the effects of different types of disfluencies. The similar designs of the experiments means that similarities and differences between the effects observed for *ers*, silent pauses, and repetition disfluencies on the comprehension of subsequent words enable conclusions to be drawn about features of disfluencies which may cause the observed effects. The results suggest that both the delay, and the nature of the disruption to speech are important.

10.2.1 The importance of delay

Both *ers* and silent pauses, but not repetitions, increased recognition memory for subsequent words. In addition, the effects of processing *ers* themselves were similar to those elicited in response to unexpected delays in speech which have been observed in previous studies (Besson et al., 1997; Mattys et al., 2005, see section 9.2.6).

The similarities between the effects of *ers* and silent pauses on language comprehension may be driven by the time which pauses add to the signal (see also Bailey & Ferreira, 2003; Brennan & Schober, 2001). *Ers* and silent pauses delay the onset of subsequent new information and introduce a delay between the post-disfluent word and the context into which it must be integrated. Although repetitions delay the onset of subsequent new information, the repeated word is probably part of the context into which the post-disfluent word is integrated. Therefore, conceptualisation of repetitions as a form of delay may not be accurate. Alternatively, the effects of *ers* and silent pauses may reflect temporal disruption, compatible with the suggestion made by Besson et al. (1997); Mattys et al. (2005).

An account that focuses on delay would predict that both non-lexical interruptions (e.g., a cough or a dog bark) and lexical fillers (e.g., *like* or *y'know*) which are not part of the critical context, would lead to similar effects as pauses. Pre-target word prolongation disfluencies should lead to similar effects as repetitions because they do not introduce a delay and are part of the critical context into which the subsequent word would be integrated.

Future work could quantify the critical parameters of the influence of time. It is unknown what the minimal duration of the delay is for effects to be observed, but it may be dependent on the rate of speech. It is also possible that it is longer for silent pauses, which are not necessarily disfluent, than for filled pauses which are unambiguous signs of disfluency. It is also unknown whether the effects of a delay on comprehension are all-or-nothing, or graded depending on the duration of the delay. If the effects of delay are graded, then there may be a critical duration after which additional delay has no additional effects as listeners assume that the speaker is not going to continue. These parameters could be investigated in an experiment which manipulates the duration of the delay.

10.2.2 The importance of the nature of the disruption

The similar effects of *ers* and silent pauses on recognition memory may not be driven by the increase in time which these disfluencies introduce between the post-disfluent word and the preceding context, but be because they are non-lexical interruptions which are immediately distinct from the surrounding lexical context. As a result of their distinctiveness, attention may be heightened. By contrast, repetitions are lexicalised and furthermore are probably part of the context into which subsequent words are integrated.

An account that focuses on the nature of disfluencies would make some similar predictions to an account that focuses on delay: non-lexical interruptions (e.g., a cough or a dog bark) should lead to similar effects as pauses and pre-target word prolongation disfluencies should lead to similar effects as repetitions. Critically the predictions for lexical fillers (e.g., *like* or *y'know*), which are lexicalised but not part of the critical context, would differ. If the important feature of disfluency is whether it is lexicalised, lexical fillers would have similar effects as repetitions.

The effect of *ers* on the N400, but not of silent pause or repetitions also demonstrates the importance of the nature of the disruption on aspects of language comprehension. The findings provide no evidence against the possibility that the reduction of the standard N400 is driven by the phonological form of the filler. One possibility is that the effect is due to the unambiguously disfluent signal conveyed by an *ers*; by contrast, silent pauses and repetitions are not always disfluent.

Importantly, in all types of disfluent utterances, unpredictable words elicited an LPC irrespective of an effect of the N400 or recognition memory and suggest an independence between the cognitive processes which the effects reflect. The presence of the LPC demonstrates a similarity between the effects of *ers*, silent pauses and repetitions on comprehension which may reflect the disruption to the speech signal which they all introduce. The LPC is commonly associated with aspects of memory retrieval and control (section 4.3.3). When an unpredictable word is presented following a disfluency—an *er*, silent pause, or a repetition—such processes may be engaged as listeners are required to maintain the word in memory, activate the pre-interruption context, and resume structural fluency of the message. As the severity of the disruption increases, and greater resources would be required to resume fluency, the amplitude of the LPC should increase. This possibility could be tested by having mid-utterance disruptions and disruptions of longer durations. Repairs which involve deletion of the reparandum might be expected to be particularly disruptive.

It is important to note that in the experiments reported here, the disfluent utterances containing an *er* or a silent pause also included prolongation disfluencies in some utterances. The presence of prolongations was not systematically manipulated, rather the speaker produced them where it felt natural. However, the disfluent utterances containing a repetition did not include any other disfluent features. It is possible that the N400 and recognition memory effects observed for *ers* and silent pauses are, at least in part, due to these disfluent features, and not to the pause itself. The absence of effects for repetition disfluencies may be because a repetition by itself was not a sufficiently disfluent cue. Future work should examine this possibility, for example by introducing disfluent pauses in the absence of other disfluent features and including disfluent markers in the absence of pauses.

10.3 An account of disfluency processing

The results of the current experiments support an account of disfluency processing that incorporates effects on predictive processes and on attention.

10.3.1 Disfluencies affect prediction

One account of the attenuation of the N400 effect following an *er* is that *ers* reduce the extent to which listeners use the context to make predictions about upcoming words, effectively rendering predictable words unpredictable. Thus if a predictable word is encountered, there is a greater cost for integration processes and the standard difference between the ease of processing predictable and unpredictable words is reduced. This processing change can account for the increase in subsequent recognition probability of words, particularly those that were originally contextually predictable.

The prediction-based mechanism assumes that language comprehension involves anticipatory processing, which until recently, was an issue of some contention (DeLong et al., 2005). However, there is increasing evidence, in particular from ERPs, that listeners make predictions online during language comprehension. DeLong et al. (2005) provided evidence for prediction during language processing, by utilising the phonological variants of the indefinite article "a" and "an" in English, whose occurrence depends on whether the following noun begins with a vowel ("an") or a consonant ("a"). Participants read sentences (see 1a and 1b) whose varying levels of constraint led to the expectation of either a vowel-initial or a consonant-initial noun. Sentence ended in either the most expected noun ("kite", 1a), or an unexpected but plausible alternative ("plane", 1b). Sentence-final words were preceded by either "a" or "an", which were semantically identical and both congruent within the context, but either matched or mismatch phonologically with the expected noun.

(1a) The day was breezy so the boy went outside to fly **an**/(a) kite.

(1b) The day was breezy so the boy went outside to fly **an**/(a) airplane.

ERPs showed that indefinite articles whose phonological form mismatched the expected noun ("an") elicited an N400 relative to indefinite articles that matched ("a"). The amplitude of the effect was graded as a function of the expectancy for the noun (using an offline cloze probability test), demonstrating the sensitivity of the N400 to prediction during language processing, and reflecting the contribution of anticipatory processing to the ease of integrating the indefinite article, within the context.

In DeLong et al.'s (2005) study, words were presented every 500ms and it is possible that the results are a function of this slow stimulus presentation rate. However, there is further evidence for prediction from other ERP studies that have used a natural speech rate. For example, Van Berkum et al. (2005) explicitly investigated whether language processing involves predictions, by utilising the arbitrary gender marking of adjectives in Dutch (e.g., "big" can be "groot" or "grote"). Participants listened to constraining mini stories like (2) which ended in the most expected noun ("painting") or a less expected, but plausible alternative ("bookcase"). Discourse-final words were preceded by an adjective ("big") whose gender marking either did or did not agree with the expected noun.

(2) The burglar had no trouble locating the secret family safe. Of course, it was situated behind a . . . **big** painting/(bookcase).

An ERP difference (right lateralised positivity) was observed between adjectives which agreed and adjectives which disagreed with the gender of the expected noun. Although the functional interpretation of this positivity is unknown, the ERP difference itself is evidence that listeners formed specific predictions (for further ERP evidence for prediction in language processing, see Wicha, Bates, et al., 2003; Wicha, Moreno, & Kutas, 2003; Wicha et al., 2004).

These studies demonstrate how ERPs can be used to explicitly test prediction and they could be adapted to test the proposal that *ers* affect prediction during language comprehension. If *ers* reduce the extent to which listeners make specific predictions, then the ERP difference between gender marked adjectives which match and adjectives which mismatch with the expected upcoming noun should be attenuated by a pre-adjective pause.

An alternative account of the attenuation of the N400 effect following *ers* focuses on purely post-lexical factors. *Ers* may cause an increase in the difficulty with which any word can be integrated into its context, and the difficulties might be most apparent for predictable words which are very easy to integrate in the fluent utterances. This processing change can account for the increase in subsequent recognition probability of words, particularly those that were originally contextually predictable. Critically, for a modification of Van Berkum et al.'s (2005) study, a post-lexical account of the N400 attenuation would predict that the relative positivity observed for adjectives which mismatch with the contextually expected upcoming noun, would not be affected by a pre-adjective pause.

The effect of an *er* on the frontal negativity and on the P600 elicited in response to repairs is also compatible with the suggestion that *ers* affect predictive processes. *Ers* may reduce the extent to which listeners predict a fluent continuation, so the syntactic violation created by the repair is no longer perceived as an outright violation, but repair processes can commence more quickly. In other words, *ers* may prepare listeners for the possibility of words which are less contextually predictable, both semantically and syntactically.

10.3.2 Disfluencies affect attention

The increase in recognition memory for words which were preceded by an *er* or a silent pause, and the faster response times to targets in an LDT are compatible with the possibility that disfluent pauses heighten listeners' attention. Stronger evidence that *ers* heighten attention is provided by the ERPs elicited in response to *ers* (Experiment 5), which suggest that *ers* are perceived to be acoustically deviant within a speech context such that attention is drawn to them.

Collard et al. (in press) have recently investigated the role of attention explicitly, using a modification of Experiment 2 (see also Corley et al., 2007). Participants listened to utterances ending in predictable words, which occasionally had been altered to deviate acoustically from this context (the manipulation made the speech sound momentarily compressed and similar to speech over a poor telephone line). Half of the utterances included an *er* before the utterance final word. In fluent utterances, the acoustically deviant words elicited standard attention-related MMN and P300b effects. Critically, although an (albeit attenuated) MMN was present, the P300b was eliminated by the presence of a pre-target *er*, demonstrating that attention was no longer oriented to deviant words in disfluent cases. These results suggest that attention had already been focused by the *er*. A subsequent recognition test showed that the non-manipulated words were more likely to be remembered if they had been preceded by an *er*, replicating results shown previously (Experiments 1 and 2, see also Corley et al., 2007), and demonstrating a longer-term effect on representation.

One question for future research concerns the duration of time for which attention is heightened by *ers* and silent pauses. In the experiments reported in the thesis, memory was only tested for words which immediately followed the disfluency. Future work could investigate whether *ers* and silent pauses would affect the memory for words presented further downstream, by inserting disfluencies mid-utterance.

10.3.3 Linking prediction and attention

In the previous sections I considered prediction and attention separately. The challenge is to integrate prediction and attention into a coherent account of disfluency processing, to explain the effects on language comprehension which can be observed during processing and on the longer-term representation of the message.

The distribution of disfluencies in speech is not arbitrary. Disfluencies are commonly associated with the production of words of lower accessibility for the speaker, for example because they are not the most unpredictable in the context. Listeners are sensitive to this non-arbitrary distribution (either due to probabilistic associations, or because they are sensitive to the speaker's production state). *Ers*, unlike silent pauses and repetitions have an umambiguously disfluent phonological form and hence *ers* can be used as a cue that upcoming information may be less contextually predictable, or may not be a straightforward continuation of the pre-interrupted speech. In the absence of a set of possible words, listeners no longer use the context to form a strong prediction and a subsequently-presented contextually predictable word will be harder to integrate (the standard N400 will be attenuated); if there is a restricted set of possible referents, they may predict the mention of the less accessible item (cf. Arnold et al., 2004, 2007). Because of the possibility that upcoming information will be less contextually predictable and therefore of higher informational content for the meaning of the message, listeners pay more attention. For a similar account, see Fox Tree and Schrock (1999).

Because disfluency heightens listeners' attention during processing, the post-disfluency words are more likely to have a lasting effect on the representation of the message, and are more likely to be correctly recognised in a memory test. Memory is most enhanced for a predictable word following a constrained utterance, which, when presented in fluent utterances would be highly redundant and would not engage attention. Similar effects of heightened attention on memory may be observed by any disruption to speech which is distinct from the surrounding lexical context, for example a silent pause. *Ers*, silent pauses, and repetitions all disrupt fluency, causing difficulties for resuming structural fluency (as indexed by the LPC).

Following a disfluency, if a contextually unpredictable word is then heard, it is, as always, hard to integrate into the context and will elicit an N400 relative to predictable words. The disruption caused by the disfluency means that memory retrieval and control processes are required to maintain the word in memory, activate the pre-interruption context, and resume structural fluency of the message. This leads to an LPC. If a repair word is heard which corrects the pre-interrupted speech, the syntactic reanalysis processes (as indexed by the P600) can commence promptly. If a contextually predictable word is heard and predictions have been reduced, the predictable word is no longer so easy to integrate and the standard N400 difference is reduced. Because of the good semantic fit between the predictable word and the pre-interrupted message, listeners have few difficulties resuming fluency.

10.4 Linking processing and representation

Although the results from Experiment 3 are equivocal, in Experiments 2, and 4, the predictability effect during processing, as indexed by the N400, mirrored the predictability effect on representation suggesting a possible link between processing and representation. Specifically, in fluent utterances, unpredictable words elicited a relative negativity during processing and were more likely to be correctly recognised in a subsequent memory test. When disfluency reduced the amplitude of the N400 and reduced the ERP difference between unpredictable and predictable words (*ers*), there was an associated reduction between the memorability of the predictable and unpredictable words.

When disfluency had no effect on the N400 effect (repetitions) and the ERP difference between unpredictable and predictable words was similar for fluent and disfluent utterances, there was also no effect on the memorability of predictable and unpredictable words: unpredictable words were more likely to be correctly recognised than predictable words in both fluent and disfluent utterances.

The reduction of the predictablity effect for recognition memory was driven by an increase in the memorability of the predictable words. If the memory effects reflect the N400 effect during processing, this suggests that the reduction of the N400 following pauses was driven by an increase in the relative negativity of the predictable words, reflecting an increase in the difficulty of integrating predictable words. This is compatible with both the accounts of disfluency which focus on the effects on linguistic processes which were proposed: those which focus on prediction, and those which focus on post-lexical integration difficulties (section 10.3.1).

Future work could use ERPs to investigate the relationship between the processing and longer-term representation of disfluent speech using the subsequent memory or "Difference in memory" (Dm) paradigm. The subsequent memory paradigm involves analysing the data acquired during processing, using the data from the subsequent memory test, so that ERP activity for processing can be contrasted between subsequently remembered and subsequently forgotten stimuli.

10.5 Conclusions

The series of experiments reported in the thesis was the first attempt at a systematic investigation of the effects of different types of disfluencies on language comprehension. The novel approach used ERPs to assess the online effects of disfluencies on language processing combined with a subsequent recognition memory test to assess the longer-term effects of disfluencies on language representation. The experiments support an account of disfluency processing which incorporates both prediction and attention: disfluencies can reduce the extent to which listeners engage in normal predictive processes and heighten listeners' attention to upcoming words which may be particularly informative for the meaning of the message. As a consequence, listeners may be more likely to remember words preceded by a disfluency.

Appendix A

Cloze probability pre-test instructions

During the experiment you will see a number of sentences, each with the final word left blank. Your task is simply to read each sentence at your normal rate, and type the first word that occurs to you as a likely end of that sentence.

For example, if you were presented with:

`The party did not end until ...`

possible responses might include `dawn, three, late, midnight,` etc.

Don't try to be unique or average; just be natural. You should keep within the following bounds, however:

1. Only one response word per sentence;
2. The word should "make sense" of the sentence;
3. English words only.

For some sentences, the response will seem obvious; for others, any number of words will seem possible. In each case, the first word that comes to mind and makes the sentence whole and sensible is the appropriate response.

Appendix B

Stimuli for Experiments 1–4

Stimuli for Experiments 1–4 were edited from the following 160 utterances. Targets were the utterance-final words which were either predictable or unpredictable (indicated in bold). Utterances were either fluent or contained a disfluency (an *er*, a silent pause, or a repetition of the pre-target word) before the target word and is indicated by *.

1. My sister had a skiing accident and she broke her * leg/**promise**.
2. She said she wouldn't cheat on him but she broke her * promise/**leg**.
3. The sitting room is really cold so I think you should light a * fire/**cigarette**.
4. We're in a smoking area so now you can light a * cigarette/**fire**.
5. I've got a deadline to meet for tomorrow so I can't afford to waste * time/**paper**.
6. I reuse old envelopes because I hate to waste * paper/**time**.
7. On the front of the card was a lucky black * cat/**tie**.
8. The invite says that the dress code is black * tie/**cat**.
9. She's a bit dippy and thinks there's fairies at the bottom of the * garden/**class**.
10. She messed up in her exams and went to the bottom of the * class/**garden**.
11. I love it when I wake up and the birds are * singing/**talking**.
12. It's distracting in lectures when people are * talking/**singing**.

13. Ben's so stubborn he can never admit when he's * wrong/**drunk**.
14. Ben's only had a few beers, but I think that he's * drunk/**wrong**.
15. I can't post the letter because it hasn't got a * stamp/**name**.
16. It's only got an address and telephone number but it hasn't got a * name/**stamp**.
17. I stood up and started to speak but then my mind went totally * blank/**random**.
18. The list of names had no order to it and appeared to be totally * random/**blank**.
19. She hated the CD but then she's never liked my taste in * music/**clothes**.
20. She hated the jumper but then she's never liked my taste in * clothes/**music**.
21. I'm really thirsty. Lets go to the pub for a * drink/**burger**.
22. I'm really hungry. Lets go to McDonald's for a * burger/**drink**.
23. We had a great day playing at the seaside. It ended perfectly when we sat outside eating fish and * chips/**ice cream**.
24. We had a great time playing games at the party. It ended perfectly when we sat outside eating jelly and * ice cream/**chips**.
25. I think it's true that a lot of people stay at University because they appreciate the student * lifestyle/**discount**.
26. You know, if you take your University card you'll get a ten percent student * discount/**lifestyle**.
27. We've not booked and so the first thing to do when we get there is to find somewhere to * stay/**play**.
28. I've brought the football and so I think we should find somewhere to * play/ **stay**.
29. I want to visit South America but to get the most out of it I should really learn to speak * Spanish/ **properly**.
30. My little sister puts on a silly voice and so my mum is always telling her to speak * properly/ **Spanish**.
31. It's been really busy in the office today, and the phones haven't stopped * ringing/ **working**.
32. I thought the computer had been fixed, but now it's stopped * working/**ringing**.
33. Everyone's got bad habits, and mine is biting my * nails/**tongue**.
34. That drink was too hot; I've just burnt my * tongue/**nails**.

35. She's been taking driving lessons all year, but she's still nervous about taking the * test/**class**.
36. It's hard for the teachers to keep control when there are so many children in the * class/**test**.
37. I want to travel the world, see new places, and meet new * people/**animals**.
38. They're going to be extending the zoo and are planning to get in some new * animals/**people**.
39. I'm going to fill up the kettle for a hot water bottle. Then I'm going to * bed/**work** .
40. I think I've just got time to grab a coffee. Then I'm going to * work/**bed**.
41. I think it's brewed enough by now so pass over a mug and I'll pour you some * tea/**water**.
42. I filled up my bottle from the stream we passed earlier so pass over yours and I'll pour you some * water/**tea**.
43. I can hear someone knocking: please can you answer the * door/**phone**.
44. I can hear something ringing: please can you answer the * phone/**door**.
45. I've got to remember to take that book I borrowed when I go to the * library/**supermarket**.
46. I've got to remember to take those money-off coupons I collected when I go to the * supermarket/**library**.
47. Budweiser is a famous name associated with the King of * beers/**France**.
48. Louis is a famous name associated with the King of * France/**beers**.
49. Jimmy's parents have been told that he can't hear and they're going to send him to a school for the * deaf/**blind**.
50. Jimmy's parents have been told that he can't see and they're going to send him to a school for the * blind/**deaf**.
51. I'm really tired and so I'm going to have an early * night/**start**.
52. I'm going to bed at 9 because tomorrow I've got an early * start/**night**.
53. Daniel would have finished his essay but he ran out of * time/**money**.
54. Daniel would have bought the whole set but he ran out of * money/**time**.
55. I got to my front door and went to open it, but I couldn't find my * keys/**phone**.

56. I went to text her, but I couldn't find my * phone/**keys**.
57. If you hear the firealarm, move to the nearest * exit/**pub**.
58. After work on Fridays, we always meet in the nearest * pub/**exit**.
59. Because the plane was so delayed, we had to spend the whole night at the * airport/**beach**.
60. We're right by the coast so when the weather's good we tend to spend the whole day at the * beach/**airport**.
61. Marion started crying about her haircut as soon as she left the * hairdresser/**vet**.
62. Marion started crying about her dog's health as soon as she left the * vet/**hairdresser**.
63. Carrots are great for helping to see in the * dark/**water**.
64. Those goggles are the best for helping to see under the * water/**dark**.
65. The meeting was boring: Rob kept looking at his * watch/**mirror**.
66. The new haircut was great; Rob kept looking in his * mirror/**watch**.
67. Terry's bored because there's nothing to watch on the * TV/**wall**.
68. Terry's bored: he's been staring at the picture on the * wall/**TV**.
69. If we're going to the cinema we should book in advance because it's a popular * film/**book**.
70. I'll have to recall it from the library because it's a popular * book/**film**.
71. I was caught by the police for speeding in my * car/**boat**.
72. The lake's big but we can cross it in my * boat/**car**.
73. It's raining so I'm planning to stay indoors. I really don't want to go * outside/**tomorrow**.
74. John suggested going today, but it would be better for me to go * tomorrow/**outside**.
75. My grandma died and I still need to get a black dress for the * funeral/**party**.
76. I'm looking forward to my sister's birthday but I still need to get a dress for the * party/**funeral**.
77. It was really hot yesterday and so I wore a hat so I wouldn't get * burnt/**wet**.
78. It was really rainy yesterday and I wore a hat so I wouldn't get * wet/**burnt**.
79. I'm really looking forward to my holiday to Crete and I keep thinking about relaxing and sitting on the * beach/**floor**.

80. When there were no more chairs in the lecture room I thought about sitting on the * floor/**beach**.
81. I don't have time to paint the walls so we should call in the * decorator/**police**.
82. I think next door have been burgled: we should call in the * police/**decorator**.
83. Traditionally, Christians use Sundays for going to * church/**jail**.
84. He's a murderer: he'll be going to * jail/**church**.
85. I'm so sweaty because I've just come from the gym. I'm just going to have a * shower/ **kit-kat**.
86. You know what they say: have a break, have a * kit-kat/**shower**.
87. Watching the bit in the film when Bambi's mother died always makes Caroline * cry/**laugh**.
88. Watching the antics of the cat and mouse in Tom and Jerry cartoons always makes Caroline * laugh/**cry**.
89. My grandmother always drinks her tea from a fine porcelain * cup/**bucket**.
90. My grandfather always mixes wallpaper paste in a big metal * bucket/**cup**.
91. If everyone has their coffee white, we'll have to buy more * milk/**bread**.
92. If everyone wants sandwiches, we'll have to buy more * bread/**milk**.
93. After the car crash, the three of us spent the night in * hospital/**hotel**.
94. After winning the holiday competition, three of us spent the night in a five-star * hotel/**hospital**.
95. We could tell it was a formal occasion because Andrew was wearing a smart suit and a * tie/**scarf**.
96. We could tell it was going to be cold because Andrew was wearing a woolly hat and a * scarf/**tie**.
97. If you haven't got much time it takes only a few minutes to heat the soup in a * microwave/**fridge**.
98. It will take at least an hour for the wine to chill in the * fridge/**microwave**.
99. If you want to know the time, why don't you look at the * clock/**calendar**.
100. If you want to know the date, why don't you look at the * calendar/**clock**.

101. Jack's very short-sighted, and he spent all his birthday money on a new pair of * glasses/**trainers**.
102. Jack's a very keen runner and he spent all his birthday money on a new pair of * trainers/**glasses**.
103. When my teacher tells us off he shouts so loudly and always wags his * finger/**tail**.
104. When my dog greets me, he barks so happily and always wags his * tail/**finger**.
105. One of the things I like about this restaurant is that there are always fresh flowers on every * table/**grave**.
106. One of the things I like about this churchyard is that there are always fresh flowers on every * grave/**table**.
107. If we put more cheese in the trap we'll catch another * mouse/**fish**.
108. If we put more bait on the hook we'll catch another * fish/**mouse**.
109. In the summer, he just sits under the trees in the local * park/**pub**.
110. In the summer, he just sits with a beer in the local * pub/**park**.
111. Watching the cooking programme has made me feel * hungry/**sad**.
112. Watching her leave has made me feel * sad/**hungry**.
113. She was easily the quickest for her age and broke loads of * records/**plates**.
114. She was the clumsiest waitress at the restaurant and broke loads of * plates/**records**.
115. At the end of breaktime we should ring the * bell/**firebrigade**.
116. That smoke is not from a bonfire: we should ring the * firebrigade/**bell**.
117. I pulled a muscle playing hockey and so I'm going home for a long hot * bath/**summer**.
118. It's been a beautifully warm spring and so we're hoping for a long hot * summer/**bath**.
119. I've finished cooking for Christmas. I've just got to put the icing on top of the * cake/**tree**.
120. I've finished decorating for Christmas. I've just got to put the fairy on top of the * tree/**cake**.
121. Management decided to close the pit and the miners have voted to go on * strike/**holiday**.
122. Term finishes next week and so I'm planning to go on * holiday/**strike**.

123. There's a housing shortage in rural areas because so many people have second homes/**degrees**.
124. A BA or a BSc doesn't count for much these days because so many people have second * degrees/**homes**.
125. They've been living together for ages so I wasn't surprised to hear they're going to get * married/**divorced**.
126. They've been separated for ages so I wasn't surprised to hear they're going to get * divorced/**married**.
127. I want my birthday present to be a surprise. I wish my sisters would stop dropping * hints/**litter**.
128. The street where I live is really scruffy: I wish people would stop dropping * litter/**hints**.
129. I'm meeting the estate agent after work. He's going to show me round the * house/**school**.
130. I'm meeting the headteacher after work: he's going to show me round the * school/**house**.
131. He's going bald and he's going to go on a special diet because he doesn't want to lose any more * hair/**weight**.
132. She's too thin and she's going to go a special diet because she doesn't want to lose any more * weight/**hair**.
133. I wish I was better at foreign languages. It was difficult in Paris because I couldn't speak * French/**German**.
134. I wish I was better at foreign languages. It was difficult in Berlin because I couldn't speak * German/**French**.
135. The holiday's all arranged. I've just got to go the the travel agents to collect the * tickets/**rent**.
136. We've signed the tenancy agreement on the flat. The landlord will come on Fridays to collect the * rent/**tickets**.
137. I don't like pastry and I loathe sausage * rolls/**dogs**.
138. I don't like poodles and I loathe sausage * dogs/**rolls**.

139. You can retire when you want, but you have to be 60 to get a * pension/**job**.
140. Today there are so many graduates looking for work that it's hard to get a * job/**pension**.
141. We need to pay; please can you ask the waiter for the * bill/**menu**.
142. We need to order; please can you ask the waiter for the * menu/**bill**.
143. Susie's mum makes the best chocolate * cake/**pie**.
144. Susie's mum makes the best apple * pie/**cake**.
145. If you want to be safe when you're walking home late then stay in a group and keep * together/**fit**.
146. If you want to keep in shape and feel good then go to the gym and keep * fit/**together**.
147. I don't think he remembers who I am because he keeps calling me by the wrong * name/**number**.
148. I tried to call him but I kept getting the wrong * number/**name**.
149. It was too noisy to have a conversation on the dance floor; we had to shout to be * heard/**seen**.
150. It was too crowded to catch his attention, we had to jump and wave to be * seen/**heard**.
151. She wasn't very involved in the business: she was a sleeping * partner/**bag**.
152. I was cold when we went camping: I forgot to take my sleeping * bag/**partner**.
153. Urban land is expensive so the council have started building sky * scrapers/**lights**.
154. We get pretty good views from the attic because of the sky * lights/**scrapers**.
155. It was a formal dinner and Ellie was surprised to find herself at the top * table/**soil**.
156. The earth was barren because the rain had washed away the top * soil/**table**.
157. I need to buy a new coat but I want to get a discount so I'm going to wait until the January * sales/**season**.
158. I need to buy a new coat but I'm going to wait and see the fashions for the new * season/**sales**.
159. I can't pay for the holiday because I don't have any * money/**work**.

160. He's unemployed at the moment. In fact I can't remember the last time he did any
* work/**money**.

Appendix C

Stimuli for Experiment 5

Stimuli for Experiment 5 were edited from the following 160 utterances. Target words (indicated in bold), were either repair free or included a repair (reparanda indicated in parentheses). Utterances were either fluent or contained an *er* before the target, indicated by *.

1. When we were walking, we got to a fence and Ben just (jumped) * **bounded** straight over.
2. Because Lisa was given a lot of detentions at school, she really (disliked) * **hated** the teacher.
3. After Rachael was winning for 2 miles but then she was (overtaken) * **beaten** in the last 200 metres.
4. Because Angie arrived late into town, Phil went to (meet) * **surprise** her at the station.
5. When Craig spoke in class the teacher just (ignored) * **dismissed** his ideas.
6. When Gary lost a lot of weight he didn't (see) * **realise** that his clothes were baggier.
7. After Lesley got her results I told her I always (thought) * **knew** she'd do well.
8. When I was in town I saw the new building being (made) * **erected** next to the market.

9. When I was little I was really worried about being (licked) * **attacked** by the big dog next door.
10. When we flew, the tubulence on the plane really (jolted) * **jiggled** us about in our seats.
11. After the group leader told us to watch our heads, he (led) * **guided** us through the tunnel.
12. When Adam was in town on Friday night, someone (hit) * **punched** him in the face.
13. When Paula's kids are a worry, she (frets) * **fusses** so much it can't be good.
14. When the nurse was putting on the bandage, she really (hurt) * **damaged** the skin on my arm.
15. Because it's Thursday, we're going to the karaoke to (sing) * **shriek** along to some cheesy tunes.
16. After I thought again, I don't think it's (possible) * **practical** to finish it all before June.
17. Because I'm a bit hungry please can you (chuck) * **hand** me the bowl of crisps.
18. After a hard winter of training Jonny was (chosen) * **selected** for the team.
19. If Anne's not driving then we should (use) * **catch** the train.
20. During the 100m final it was amazing to see athletes (race) * **sprint** so fast.
21. Because the vegetables will taste better, we should (roast) * **fry** them with some garlic.
22. It was warm today until the sun (hid) * **went** behind the clouds.
23. When Betty's feeling nosy she (looks) * **peers** out from behind her curtains.
24. After the wheel came off my bike I had to (fix) * **mend** it before the holiday.
25. Because Rufus was rolling in the mud the car now (smells) * **stinks** of wet dog.
26. Because of the bushes, that field is a great place for kids to (run) * **hide** in and have fun.
27. When the neighbours had parties the noise used to (make) * **send** Josie crazy.
28. When people talk in lectures it really (annoys) * **distracts** me and stops me concentrating.

29. When my sister went skiing at Christmas she (twisted) * **fractured** her ankle on the first day.
30. Even though I hadn't seen him since school I (recognised) * **remembered** him straight away.
31. When Fred's uncle went to prison it was because he (aided) * **devised** a tax crime.
32. When Graham told the secretary she had to (arrange) * **organise** the Christmas lunch it wasn't a surprise.
33. If Clare's going to get a job she'll have to (post) * **submit** a few more applications and then hope for an interview.
34. When it's summer I (adore) * **love** the sunshine.
35. Because I'd not been swimming for 3 days I was (keen) * **desperate** to get into the water.
36. After the car crash Lisa was (fortunate) * **lucky** to only suffer whiplash.
37. Because Mary's so narrow minded she never (contemplates) * **compares** the alternatives.
38. When you left last night, Tony and Kelly (argued) * **bickered** about it for over two hours.
39. Because my desk's a mess I need to (sort) * **file** all my papers quite urgently.
40. Because I'd worked quite hard I'd (hoped) * **expected** to do a lot better.
41. When Alice returned from her holiday she had to try not to (dream) * **fantasise** about sitting on the beach.
42. Because Anna takes so long to choose her clothes we always have to (wait) * **hang** around for her.
43. Because Gran wanted attention she used her walking stick to (prod) * **poke** us in the leg.
44. When skaters are on ice they (spin) * **twirl** about so gracefully.
45. Because Katie had such a great time in Australia she (yearns) * **longs** to go travelling again.
46. When Josie talks so much she doesn't (listen) * **concentrate** on what's going on around her.

47. When the waves frightened the boy he just (held) * **grabbed** onto the side of the boat.
48. After we've worked hard all week we deserve to (enjoy) * **relish** the weekend.
49. When the kids have too much sugar it's hard to (settle) * **calm** them down.
50. Because of the brain we're going to be (drenched) * **soaked** to the bone.
51. After the builders looked at the roof they (assured) * **promised** us that it would be done by Christmas.
52. When some children try they can (influence) * **persuade** their parents to do almost anything.
53. When I was dusting the mantlepiece I accidently (dropped) * **smashed** the pot we were given for Christmas.
54. After you weigh the ingredients you have to (grind) * **blend** them together in a big pot.
55. After Brian's success with the band he's planning to (commit) * **dedicate** himself to music.
56. When Paul takes a simple idea he always (complicates) * **elaborates** it extensively.
57. After I bought the flatpacked furniture Thomas (constructed) * **assembled** it in less than an hour.
58. If you're going to stay together you'll have to (confide) * **trust** in her more.
59. After the school expanded, they (connected) * **joined** the outer buildings to the main one.
60. Because everyone's happy we shouldn't (change) * **alter** anything until next year.
61. When our team scores a goal you have to (shout) * **cheer** as loudly as you can.
62. When Sally is upset James tries to (comfort) * **console** her and she always feels better.
63. When Henry's uncle was younger he was (taught) * **educated** at a famous school in England.
64. Because I've no idea about prices please can you (approximate) * **estimate** how much you think it might be.
65. After Hannah got over her fear of snakes she (felt) * **touched** one on holiday.

66. When Ian gets the photos he's going to (add) * **insert** them into the text.
67. Because Ken didn't tell Tom all the details he was (mislead) * **tricked** into doing it.
68. When you're too hot you should take off your jacket and (loosen) * **undo** your tie.
69. When Gran visits on Thursday you'll need to (help) * **assist** her when she gets off the train.
70. When I saw the circus performers they really (excited) * **amazed** the young children.
71. When those policemen saw my friend they really (threatened) * **scared** her with their batons.
72. When Josie stood for the council she needed people to (back) * **support** her nomination.
73. Because there are some problems we have to (tackle) * **address** the causes.
74. After we resolved the issues, we were able to (communicate) * **collaborate** with each other.
75. Because we were happy we (celebrated) * **partied** all night long.
76. After one distracted me the other (slid) * **nicked** my wallet out of my pocket.
77. When the car stopped working, my uncle (inspected) * **tested** the engine.
78. Before we held the event we (advertised) * **promoted** it on the internet.
79. We're holding a comedy show to (collect) * **raise** money for the children's hospital.
80. A critical aspect of a democratic society is the right to (march) * **protest** in a peaceful way.
81. Because we live in a democracy, everyone has the right to (express) * **declare** their opinions but not in an aggressive way.
82. When the compost bins arrive they will (affect) * **reduce** the volume of waste which people throw away.
83. When Paul asked me I was happy to (justify) * **validate** my opinion with a number of examples.
84. When the new traffic system is implemented it should (limit) * **ease** the congestion in the centre of town.

85. I'm not sure what time I'll be home because I have to (revise) * **study** for the exam next week.
86. Because you think he's good you should (propose) * **recommend** him for the position.
87. Because I dislike getting wet I don't like it when it (rains) * **pours** for day after day.
88. Because I love modern fiction I've volunteered to (read) * **review** books for the student newspaper.
89. When it's Sarah's birthday Jill's going to (buy) * **bake** a fancy cake.
90. When Jo forgot to water the plants the flowers (drooped) * **wilted** within a day.
91. When Ken got indigestion it was because he didn't (cut) * **chew** his food properly.
92. When you've got a lot of work just try not to (worry) * **stress** and then it won't seem so bad.
93. When you've decided please tell us what you're going to (bring) * **take** to the picnic.
94. When we tried that new club we (drank) * **danced** for about 3 hours nonstop.
95. When we do a stocktake we have to (close) * **shut** the shop early.
96. Because no-one knows what happened the police are working hard to (establish) * **discover** the cause of the incident.
97. When we have weekly meetings we have to (reflect) * **comment** on the progress that we've made.
98. Because George was disturbed he used to (kick) * **terrorise** the other children.
99. After the group analysed the data they (issued) * **presented** the findings at the meeting.
100. Before the AGM can take place we need somewhere to (host) * **hold** it and someone to take charge.
101. When the athletics club asked, there were many volunteers to help (liase) * **coordinate** the summer championships.
102. When Jen was in the police she was trained to (shock) * **coerce** people to give answers.

103. After I did the experiment, Karl (emailed) * **contacted** me with the final results.
104. When John didn't have the information about Saturday he (speculated) * **presupposed** a few things which ended up being wrong.
105. When the competition took place there was transport to (bus) * **ship** the athletes between venues.
106. When we made our own rafts we (rode) * **floated** down the river to the next village.
107. When we made it on time it was because we (forgot) * **abandoned** the car and walked.
108. Because the weather on holiday was hotter we had to think hard what to (wear) * **carry** when we went away.
109. When Emily was happy she (smiled) * **smirked** all afternoon.
110. Because we're lost we should (halt) * **congregate** together.
111. Because she's always hungry she carries food to (eat) * **snack** on the way.
112. Because Lisa is annoying you'll have to (ask) * **demand** that she stops calling.
113. If the boiler's still broken I'll (phone) * **call** Pete first thing on Monday.
114. When I was preparing for the competition I (swam) * **jogged** every morning.
115. When Cathy went to India she (developed) * **contracted** malaria in the first month.
116. When I last tried that bell it wouldn't (ring) * **work** so you should give it a go.
117. When the obstacle course included ropes you had to (balance) * **swing** on them without touching the ground.
118. When the weather was good we (walked) * **cycled** to the park.
119. When we made a Christmas wreath we (gathered) * **picked** some ivy and berries from the garden.
120. When Roger found out it was healthier he (steamed) * **grilled** all his vegetables.
121. When I speak to Christine she (criticises) * **analyses** everything that I say.
122. After I tried I managed to (avoid) * **prevent** the same mistakes happening again.
123. When the headteacher found out he (regretted) * **condemned** the behaviour.
124. Ever since he was five, Tom has (wanted) * **aspired** to be an astronaut.
125. After Rob applied he found someone to (employ) * **mentor** him from January.
126. When everyone looked tired we (interrupted) * **suspended** the practice for coffee.

127. When Laura's inlaws found out they (interfered) * **intruded** when they weren't wanted.
128. When I got the mosquito bite it was hard not to (itch) * **scratch** it all the time.
129. When the poster came off the wall I just (glued) * **stuck** it back on.
130. When you're in the car with Ellie make sure she (clips) * **fastens** her seatbelt properly.
131. After the hinge was so squeaky I had to (oil) * **grease** it so I wouldn't go mad.
132. When we left so early we (loaded) * **filled** the boot of the car the night before.
133. When the ballet company began, the audience were (mesmerised) * **entranced** within 5 minutes.
134. When you've got so much stuff don't (jam) * **pack** it all in one bag.
135. When I gave Rob the names he (labelled) * **printed** them all on the envelopes.
136. After the last song, the musical (concluded) * **culminated** with all the performers coming on stage.
137. When you have questions just come and (knock) * **bang** on our door.
138. When my friend taught in a school she was (observed) * **assessed** by the OFSTED inspectors.
139. When the company employed the graduate she (launched) * **instigated** a great money-saving scheme.
140. When the kids began school they didn't know how to (act) * **behave** in front of adults.
141. If we're going on holiday we'll have to (save) * **economise** on a few things.
142. After I moved house I didn't have to (travel) * **drive** to work.
143. After Kelly retired from athletics she decided to (coach) * **train** youngsters at her local club.
144. When Bradley held the knife he nearly (chopped) * **sliced** his finger off.
145. Before the guests arrived we had to (hoover) * **sweep** the dining room.
146. After the high jumper was successful the bar was (increased) * **heightened** by 2 cm.
147. When they use wood it needs to be (strengthened) * **reinforced** by steel.

148. When Josh is with the other kids he (pushes) * **shoves** them around.
149. When Carly talks she often (offends) * **insults** whoever's around.
150. While we were in the bar they (reported) * **announced** the winners of the raffle.
151. When the kids made a mess we had to (tidy) * **clean** the kitchen.
152. Although everyone managed to fit in, we all had to (squash) * **squeeze** up close.
153. When the music came on, the guests (tapped) * **clicked** their fingers.
154. While I was in Woolworths, a child (snatched) * **pinched** some sweets.
155. Since Sandra's illness got worse, she's had to (rely) * **depend** on home help.
156. When Liz was round she told me about the play she's going to (produce) * **direct** in the winter.
157. Because Amanda has such a talent, Angie really (admires) * **envies** her success.
158. When the committee met they decided to (allow) * **permit** non members entry for a fee.
159. When he's made a mistake, James will never (admit) * **confess** that he's wrong.
160. When I asked Hannah she didn't mind that I (took) * **borrowed** some of her shampoo.

Appendix D

Corley et al. (2007)

Available online at www.sciencedirect.com
ScienceDirect
Cognition 105 (2007) 658–668
COGNITION
www.elsevier.com/locate/COGNIT

Brief article

It's the way that you, er, say it: Hesitations in speech affect language comprehension

Martin Corley [a], Lucy J. MacGregor [a,*], David I. Donaldson [b]

[a] *School of Philosophy, Psychology, and Language Sciences, University of Edinburgh, Edinburgh EH8 9JZ, UK*
[b] *Department of Psychology, University of Stirling, FK9 4LA, UK*

Received 12 July 2006; revised 29 September 2006; accepted 29 October 2006

Abstract

Everyday speech is littered with disfluency, often correlated with the production of less predictable words (e.g., Beattie & Butterworth [Beattie, G., & Butterworth, B. (1979). Contextual probability and word frequency as determinants of pauses in spontaneous speech. *Language and Speech, 22*, 201–211.]). But what are the effects of disfluency on listeners? In an ERP experiment which compared fluent to disfluent utterances, we established an N400 effect for unpredictable compared to predictable words. This effect, reflecting the difference in ease of integrating words into their contexts, was reduced in cases where the target words were preceded by a hesitation marked by the word *er*. Moreover, a subsequent recognition memory test showed that words preceded by disfluency were more likely to be remembered. The study demonstrates that hesitation affects the way in which listeners process spoken language, and that these changes are associated with longer-term consequences for the representation of the message.

Keywords: Language comprehension; Disfluency; Speech; ERPs

* Corresponding author. Tel.: +44 131 651 1304; fax: +44 131 650 3461.
E-mail address: Lucy.MacGregor@ed.ac.uk (L.J. MacGregor).

0010-0277/$ - see front matter
doi:10.1016/j.cognition.2006.10.010

1. Introduction

Approximately 6 in every 100 words are affected by disfluency, including repetitions, corrections, and hesitations such as the fillers *um* and *er* (Fox Tree, 1995). Moreover, the distribution of disfluency is not arbitrary. For example, fillers tend to occur before low frequency and unpredictable words (Beattie & Butterworth, 1979; Levelt, 1983; Schnadt & Corley, 2006), in circumstances where the speaker is faced with multiple semantic or syntactic possibilities (Schachter, Christenfeld, Ravina, & Bilous, 1991), as well as in cases where other types of uncertainty occur (Brennan & Williams, 1995). But what are the effects of hesitations on listeners and on language comprehension?

Although the majority of psycholinguistic research on speech comprehension has been conducted using idealised, fluent utterances, a number of corpus analyses and behavioural studies suggest that disfluency can affect listeners. Longer-term consequences of disfluency include speakers being rated as less likely to know answers to general knowledge questions when their answers are preceded by hesitations (Brennan & Williams, 1995), suggesting that listeners are sensitive to the uncertainty conveyed by hesitations at a metacognitive level. Offline questionnaire studies additionally reveal that hesitations can influence grammaticality ratings for garden path sentences, reflecting probable differences in the ways in which they have been comprehended (Bailey & Ferreira, 2003).

Investigations of the shorter-term effects of disfluency show that listeners are faster at a word monitoring task when words are preceded by a hesitation (Fox Tree, 2001) and from this it has been argued that hesitations heighten listeners' immediate attention to upcoming speech. Work by Arnold and colleagues (Arnold, Tanenhaus, Altmann, & Fagnano, 2004) attempts to refine an account of how listeners respond to disfluency in real time. Using a visual world paradigm (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), participants' eye movements to depictions of objects on a computer screen were monitored as they responded to auditory instructions to move the objects with a mouse. The presence of a disfluency (*thee uh*) before the target object increased the probability of an initial eye movement to an object that had not been previously mentioned; in contrast, when the instructions were fluent, participants were more likely to look first at a previously mentioned object. Arnold et al.'s (2004) findings suggest that listeners are sensitive to the fact that speakers find it more difficult to retrieve the names of items they have not mentioned before (Arnold, Wasow, Ginstrom, & Losongco, 2000) and can predict that these items are more likely to be mentioned following disfluency. However, there are at least two limitations of the Arnold et al. (2004) study. First, the effects of disfluency may be driven by the nature of the task. In natural dialogue, it is rare for listeners to be presented *a priori* with a limited set of images which provide potential sentence completions (although see Dahan, Magnuson, & Tanenhaus, 2001, for evidence that non-presented items can affect eye-movements). Second, the study does not address possible longer-term consequences of disfluency. Therefore, although the results suggest that listeners can strategically profit from disfluency in a constrained task-driven situation, the question of whether and how disfluency affects listeners on-line and in the longer term, under more natural circumstances, remains unanswered.

Our study addresses both of these issues, using Event-Related Potentials (ERPs) to provide a real-time measure of the processing of disfluent speech, and a surprise recognition memory test to assess the longer-term consequences of disfluency on language representation.

ERPs – neural activity recorded at the scalp, time locked to the onset of a cognitive event of interest and averaged over multiple events – are ideal for investigating the functional and neural basis of spoken language comprehension. They have two particular benefits over eye movement methods. First, there is no need for a contextually relevant visual presentation (with its attendant constraints), and second, participants need not perform any other task other than listen to the experimental stimuli. This means that ERPs provide an ideal means to investigate how listeners process disfluent speech in a situation which is a close analogue to everyday language comprehension.

We focused on the N400, an ERP component associated with the meaningful processing of language (Kutas & Hillyard, 1980, 1984). During comprehension, each word must be integrated with its linguistic context, from which it can often be predicted. Where integration is difficult (for example because a word is not predictable), a negative change in voltages recorded at the scalp relative to more easily integrated words is observed. This difference, the N400 effect, peaks at around 400 ms after word onset, maximally over central and centro-parietal regions.

Because disfluency tends to precede less predictable items in speech (Beattie & Butterworth, 1979; Levelt, 1983; Schnadt & Corley, 2006), we focused on listeners' ability to integrate predictable and unpredictable target words into their preceding contexts. If listeners interpret hesitation as a signal that the following words may not follow from the preceding context, the presence of hesitations before target words should reduce the N400 difference between predictable and unpredictable words. Changes in the N400, indicating differences in the processing of the input, may result in changes to the representation of the message, particularly of the words immediately following the disfluency. An effect in memory for these words would provide evidence for this, as well as a longer-term correlate of any effects observed in the ERP record at the time the utterances were heard.

2. Method

2.1. Materials

Auditory materials were created from 80 pairs of sentence frames, together with corresponding pairs of utterance-final target words, which were the most predictable ending for one sentence frame (mean cloze probability: .84) and an unpredictable ending for the other (0). Predictability was determined using a cloze probability pre-test. Table 1 shows an example material set. Double-counterbalancing ensured that each target word and each sentence frame contributed equally to each of the

Table 1
Example stimulus set comprising two highly constraining sentence frames, crossed with two target words which were predictable or unpredictable in context

Predictable	Everyone's got bad habits and mine is biting my	[*er*]	nails
	That drink's too hot; I have just burnt my	[*er*]	tongue
Unpredictable	Everyone's got bad habits and mine is biting my	[*er*]	tongue
	That drink's too hot; I have just burnt my	[*er*]	nails

Recorded utterances were either fluent or disfluent (containing the filler *er*, indicated in square brackets).

conditions obtained from crossing disfluency with predictability, and that no participant heard any of the sentence frames or target words twice.

Fluent and disfluent versions of the sentence frames were recorded at a natural speaking rate. In each case, the target word was replaced with a 'pseudo-target' (e.g., *pen*) so that actual targets were not predictable from phonotactic cues in the frames. In disfluent versions, the pseudo-target was preceded by an *er* (pronounced [ɜ˞ː]) with prolongations of the previous word (e.g., *thee* [ðiː]), and included prosodic changes where natural for the speaker. Finally, identical recordings of the target words were spliced onto the recorded frames in place of the pseudo-targets. This ensured that any observed ERP differences between conditions would be directly attributable to the contexts, rather than to differences between the recordings of the targets themselves. In each of four versions of the experiment, 80 of the resulting recordings were presented in disfluent form, and 80 were fluent. Recordings of 80 unrelated filler sentences, including some with less predictable words either mid-utterance or at the end of the utterance, were also added to each version. Half of the fillers included disfluencies of various types.

2.2. Participants

Twelve native British English speakers (6 male; mean age 23; range 16–35; all right-handed) with no known hearing or neurological impairment participated for financial compensation. Informed consent was obtained in accordance with the University of Stirling Psychology Ethics Committee guidelines.

2.3. Procedure

There were two parts to the experiment. In the first part, participants were told that they would hear a series of utterances which were re-recorded extracts from previously recorded conversations, and that they should listen for understanding, just as they would in a natural situation. No other task was imposed. One hundred and sixty experimental utterances were presented auditorily, interspersed with fillers. Recordings were presented in two blocks lasting approximately 15 min each, separated by a break of a few minutes. The start of each recording was signalled visually by a fixation cross, used to discourage eye movements.

EEG was recorded from 61 scalp sites using a left mastoid reference, and re-referenced to average mastoid recordings off-line. Electro-oculograms were recorded to monitor for vertical and horizontal eye movements. Electrode impedances were kept below 5 kΩ. The analogue recordings were amplified (band pass filter 0.01–40 Hz), and continuously digitised (16 bit) at a sampling frequency of 200 Hz.

Before off-line averaging, the continuous EEG files for each participant were screened, resulting in a loss of 24.8% of ERP trials due to artefacts, with little variability across conditions. The effect of blink artefacts was minimised by estimating and correcting their contribution to the ERP waveforms (Rugg, Mark, Gilchrist, & Roberts, 1997). Average ERPs (epoch length 1350 ms, pre target baseline 150 ms) time locked to the onsets of target words were formed for each participant (average 26 artefact-free trials by condition, minimum 16), and the waveforms were smoothed over 5 points.

In the second part of the experiment, the 160 utterance-final 'old' (previously heard) words were presented visually, interspersed with 160 frequency-matched 'new' foils, which had not been heard at any point in the first part of the experiment. Participants discriminated between old and new words as accurately as possible by pressing one of two response keys. The start of each presentation of a target word was signalled by the appearance of a fixation cross, which was replaced by the stimulus. After a 750 ms presentation, the screen was blanked for 1750 ms. Responses made later than this were not recorded.

3. Results

ERPs in response to predictable and unpredictable target words in fluent and disfluent utterances were quantified by measuring the mean amplitude within the standard N400 time window of 300–500 ms after word onset. All analyses made use of Greenhouse–Geisser corrections where appropriate, and are reported using corrected F values.

Figs. 1 and 2 show ERPs time locked to the utterance-final word onsets for fluent and disfluent utterances respectively, for midline (Fz, FCz, Cz, CPz, Pz) and grouped left- and right-hemisphere electrodes. Unpredictable words lead to greater negativity over the conventional N400 epoch of 300–500 ms. This negativity is broadly distributed over the scalp, but appears larger over central and midline locations, closely resembling effects shown in previous studies (Kutas & Hillyard, 1980, 1984; Van Berkum, Brown, & Hagoort, 1999; Van Petten, Coulson, Rubin, Plante, & Parks, 1999).

Because the pre-target baselines for fluent and disfluent materials were recorded from different points in the utterances (disfluent baselines are typically obtained mid-*er*), direct comparisons for targets in fluent vs. disfluent conditions could not be made: instead we used an interaction analysis to compare the size of the N400 predictability effect across conditions. In order to establish that this comparison was meaningful, we first ensured that there was no distributional difference between the N400s obtained in fluent and disfluent conditions. To do this, we calculated the mean voltage difference between ERPs for unpredictable and predictable targets over the 300–500 ms time window for each of the 61 electrodes, separately for fluent and disfluent utterances. ANOVA (factors of fluency and location) performed on

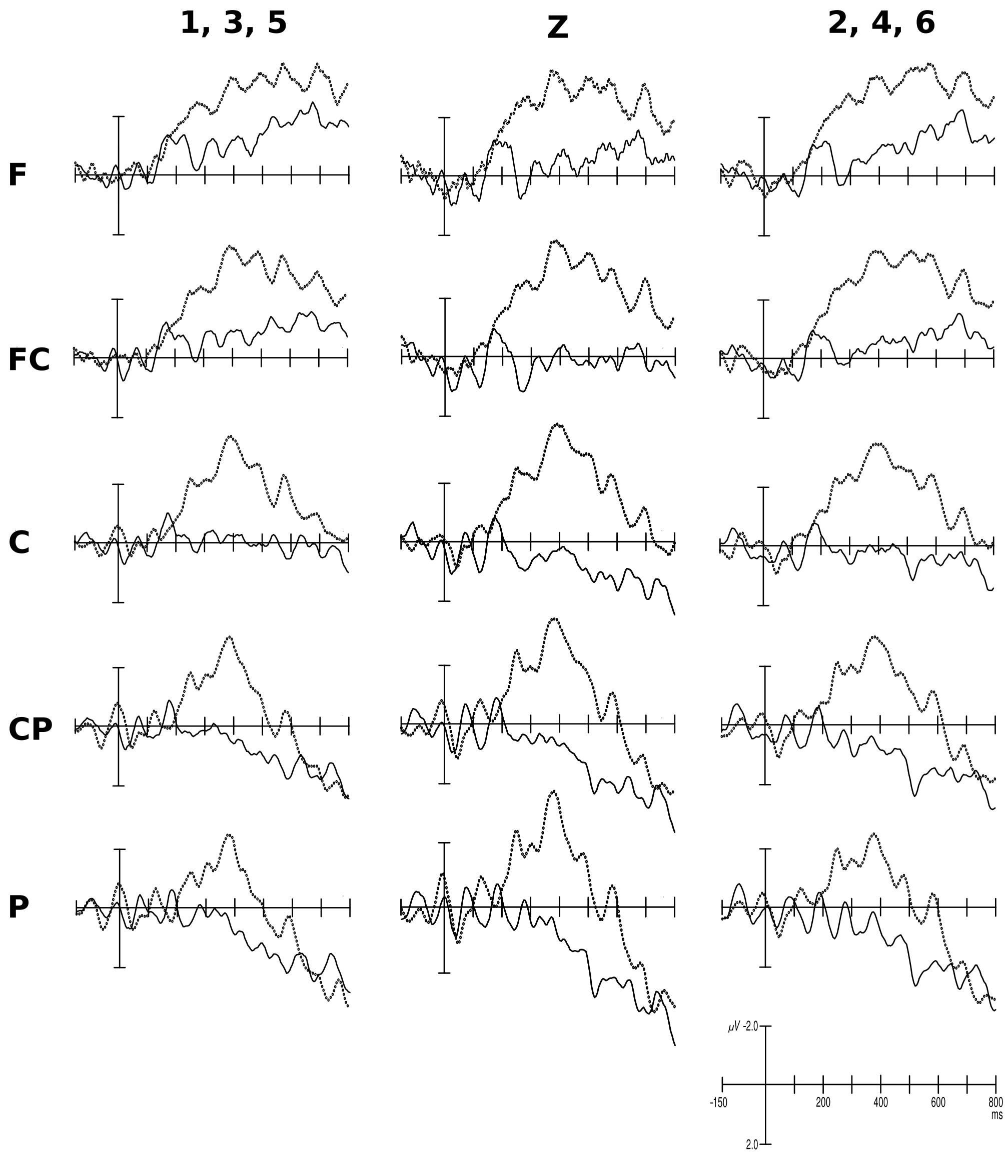

Fig. 1. ERPs for *fluent* utterances relative to predictable (solid lines) or unpredictable (dotted lines) target word onsets. The central column represents the midline sites (from top: frontal, fronto-central, central, centro-parietal, parietal); the left-hand and right-hand columns represent averages of three electrodes to the left or right of the midline, respectively.

these differences, after normalisation for amplitude differences (using the max/min method: McCarthy & Wood, 1985), reveals no effect of location [$F(60, 660) = 1.70$, $\epsilon = .046$, $\eta_p^2 = .134$, $p = .191$], nor of fluency [$F < 1$], nor any interaction between fluency and location [$F < 1$]. The lack of difference in scalp topographies between the fluent and disfluent conditions gives us no reason to suppose that different neural generators are responsible for the recorded effects of predictability.

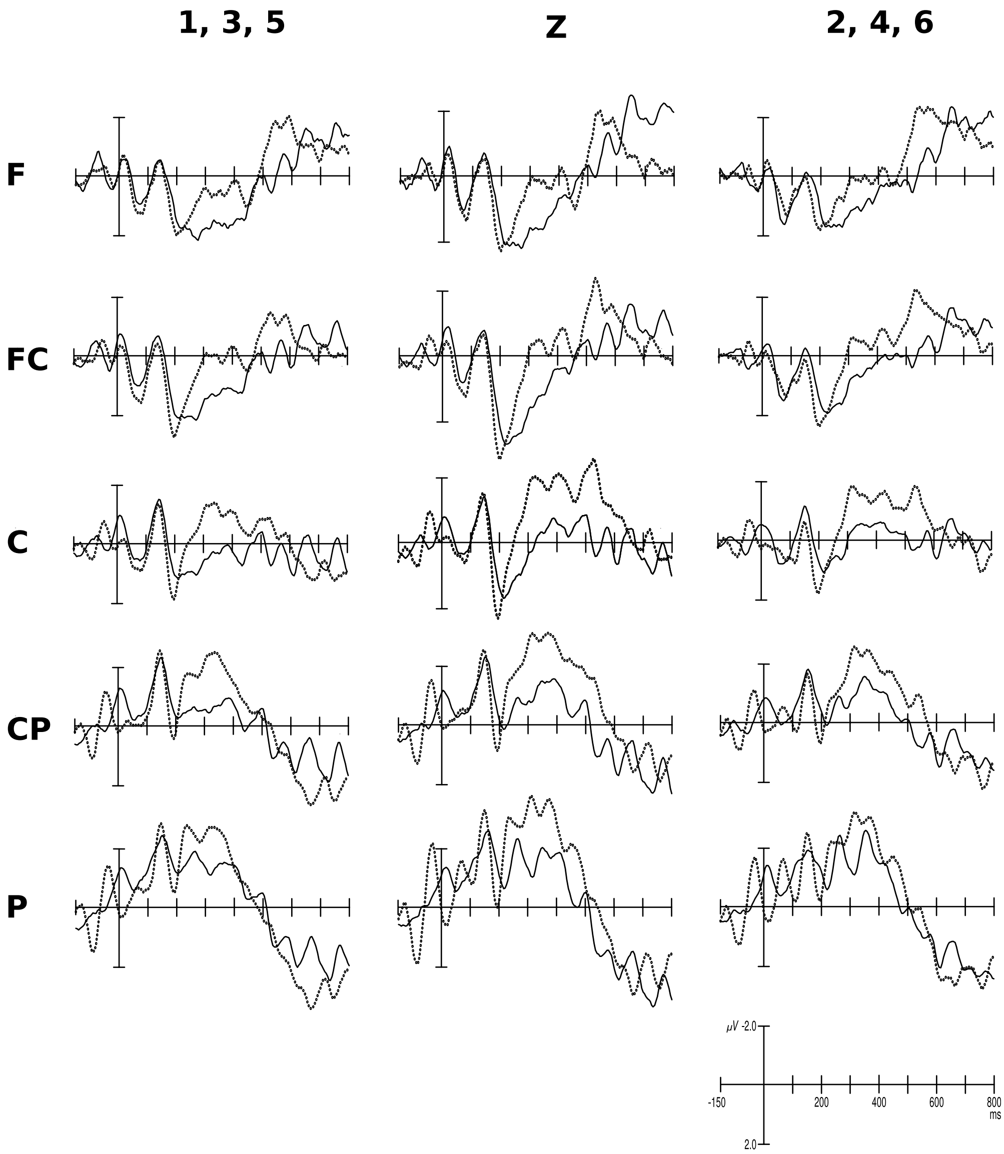

Fig. 2. ERPs for *disfluent* utterances relative to predictable (solid lines) or unpredictable (dotted lines) target word onsets. The central column represents the midline sites (from top: frontal, fronto-central, central, centro-parietal, parietal); the left-hand and right-hand columns represent averages of three electrodes to the left or right of the midline, respectively.

Two further analyses established that the distributions of the fluent and disfluent N400s were not lateralised (factors of predictability, location [F, FC, C, CP, P], hemisphere [L, R], and laterality [1, 2 vs. 3, 4 vs. 5, 6]). For fluent utterances, the analysis revealed a main effect of predictability [$F(1,11) = 43.93$, $\eta_p^2 = .800$, $p < .001$] and an interaction of predictability with laterality [$F(2,22) = 8.95$, $\epsilon = .550$, $\eta_p^2 = .448$, $p = .010$]. No other effect involving predictability was significant.

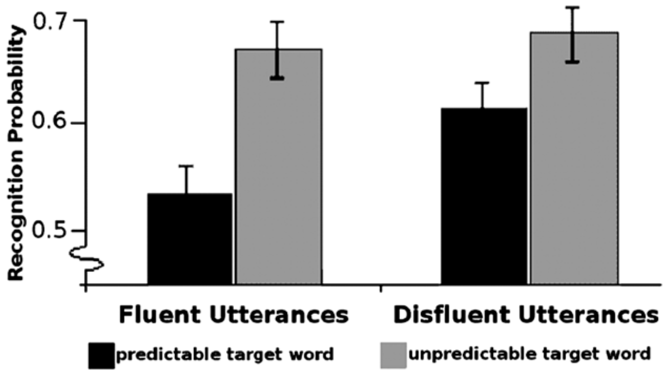

Fig. 3. Memory performance for utterance-final words which were originally predictable or unpredictable in their contexts, by utterance fluency (error bars represent one standard error of the mean).

For disfluent utterances, there was no effect of predictability [$F(1,11) = 2.96$, $\eta_p^2 = .212$, $p = .113$] and no other effect involving predictability reached significance. Since no effects involving hemisphere were found in either analysis, we concentrated on the midline electrodes (Fz, FCz, Cz, CPz, Pz) for the comparison of fluent with disfluent utterances.

An analysis of the midline electrodes (factors of fluency, predictability, location) demonstrated a main effect of predictability [$F(1,11) = 19.39$, $\eta_p^2 = .638$, $p = .001$] and an interaction of fluency with location [$F(4,44) = 13.79$, $\epsilon = .307$, $\eta_p^2 = .556$, $p = .002$], reflecting general frontal positivity relative to the baseline in the disfluent case. Importantly, fluency interacted with predictability [$F(1,11) = 7.93$, $\eta_p^2 = .419$, $p = .017$], establishing that the N400 effect for fluent items (3.14 μV) is reduced for disfluent items (1.19 μV).

As a final check, we performed an ANOVA for the midline N400 effects after normalisation (using the max/min method) to examine whether there were any distributional differences between fluent and disfluent conditions for this crucial interaction. There were no observable differences between fluent and disfluent items [for location: $F(4,44) = 1.34$, $\epsilon = .274$, $\eta_p^2 = .109$, $p = .274$; other Fs < 1].

Memory performance was quantified as the probability of correctly identifying old (previously heard) words. To control for differences in individual memory performance, we treated stimulus identity as a random factor.[1] Overall, 62% of the old words were correctly recognised (false alarm rate 24%). Fig. 3 shows the recognition probability of utterance-final words by fluency and predictability.

ANOVA (factors of fluency and predictability) reveals that words which were unpredictable utterance endings are more likely to be recognised than predictable

[1] Traditional adjustments for individual error-rates, such as d', are inappropriate here, since the properties of 'old' stimuli are determined by their context of occurrence and hence there are no comparable categories of 'new' stimuli. Using stimulus identity as a random factor ensures that per-participant biases to respond 'old' or 'new' are controlled for across the experiment.Twelve target words were inadvertently repeated in the experiment, resulting in 148 distinct targets. Analysis with data from the repeated targets removed did not affect the outcome.

words [69% vs. 58%: $F(1,147) = 23.48$, $\eta_p^2 = .138$, $p < .001$]. Importantly, disfluency also has a long-term effect: words which were preceded by hesitation are better recognised [66% vs. 62%: $F(1,147) = 4.31$, $\eta_p^2 = .029$, $p < .05$], primarily predictable words [62% vs. 55%: $F(1,147) = 4.73$, $\eta_p^2 = .031$, $p = .031$].

4. Discussion

In the presence of disfluency, the N400 effect, traditionally associated with the processing of less compared to more predictable words, was substantially reduced. Hesitation also had a longer-term effect: words following *er* were more likely to be recognised in a subsequent memory test. This suggests that these words have been processed differently as a consequence of hesitation. Since the N400 differences correspond to differences in memory performance, we can additionally conclude that the ERP differences are not due to contamination of the N400 waveform by spillover effects from the processing of the *er*.

Because predictability and ease of integration are often confounded, we are left with two possible accounts of the locus of the N400 attenuation. First, it may be because the *er* affects post-lexical factors, which operate once the target has been heard. Previous research has shown that the N400 is sensitive to differences in the semantic fit of words that do not differ in terms of predictability (Van Berkum, Zwitserlood, Hagoort, & Brown, 2003). We know from the speech production literature (e.g., Levelt, 1989) that fillers such as *er* often co-occur with other disfluent phenomena such as corrections. These are hypothesised to be more difficult to integrate syntactically and semantically, because some kind of revision must take place. A similar process could be responsible for post-*er* integration in the current experiment: hesitation could add to the difficulty with which both predictable and unpredictable words are integrated. Alternatively, the *er* may affect the comprehension system before the target is heard, effectively reducing the extent to which specific predictions are made, and therefore increasing the integration difficulty. In both cases, we might assume that predictable words would give rise to more negativity in disfluent compared to fluent contexts, as suggested by visual inspection of Figs. 1 and 2. Since limitations of the present design prevent a direct comparison from being made, it is particularly relevant that these views also predict that words following disfluency will be better remembered. Such a memory effect is demonstrated in this study, albeit with a small size because of the large number of other factors that are likely to influence the likelihood of later remembering a particular word heard among 240 recorded utterances.

Whatever the detailed mechanism, disfluency clearly affects the processing of language. But what is it about *er* that causes a processing change? One view is that there is nothing intrinsic to *er* that allows it to be understood as a disfluent signal. Instead, the N400 attenuation and subsequent effects on memory might be attributed to timing differences in the fluent and disfluent utterances: in the disfluent utterances, the *er* necessarily introduces more time between the context and the (predictable or unpredictable) target word. This might be particularly salient in the experimental situation, where many utterances end unpredictably. Among competing possibilities, listeners could be sensitive to disfluent 'words' such as *er*, as suggested by Clark and Fox Tree (2002). Although the nature of the signal remains a question for future research (and some hints as to its resolution can be found in, e.g., Bailey & Ferreira's (2003) demonstration of the 'disfluency-like' effects of unnatural interruptions to speech), it is secondary to the primary motivation for the current study, which is to demonstrate that disfluent signals in speech affect listeners.

The effect of disfluency demonstrated in this paper is profound: differences in the processing of words in an utterance are visible immediately after the disfluency is encountered, and after a substantial delay (of up to 55 min after the first few utterances are heard) participants are more likely to recognise words which have been preceded by disfluency. Using a combined ERP and memory approach, we have established an effect of disfluency using a different type of predictability, and a different methodology, to those used by Arnold et al. (2004). Moreover, we have shown that the electrophysiological differences observed following hesitations are not merely epiphenomena, but reflect differences in immediate processing which have lasting effects. In other words, disfluency in speech has both short- and longer-term consequences for listeners.

Acknowledgements

We thank Jos Van Berkum for his helpful comments on an earlier version of this paper. Phil Collard, Alice Foucart, Graham MacKenzie, and Sinéad Rhodes each made useful suggestions.

This research was partially supported by a Staff Research Development Grant (M.C.), the Economic and Social Research Council (L.J.M.), and the Biotechnology and Biological Sciences Research Council (D.I.D.).

References

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new: disfluency and reference resolution. *Psychological Science, 15*, 578–582.

Arnold, J. E., Wasow, T., Ginstrom, R., & Losongco, T. (2000). The effects of structural complexity and discourse status on constituent ordering. *Language, 76*, 28–55.

Bailey, K. G. D., & Ferreira, F. (2003). Disfluencies affect the parsing of garden-path sentences. *Journal of Memory and Language, 49*, 183–200.

Beattie, G., & Butterworth, B. (1979). Contextual probability and word frequency as determinants of pauses in spontaneous speech. *Language and Speech, 22*, 201–211.

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language, 34*, 383–398.

Clark, H. H., & Fox Tree, J. E. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition, 84*, 73–111.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology, 42*, 317–367.

Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language, 34*, 709–738.

Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory and Cognition, 29*, 320–326.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science, 207*, 203–205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161–163.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14*, 41–104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology, 62*, 203–208.

Rugg, M. D., Mark, R. E., Gilchrist, J., & Roberts, R. C. (1997). ERP repetition effects in indirect and direct tasks: effects of age and interim lag. *Psychophysiology, 34*, 572–586.

Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology, 60*, 362–367.

Schnadt, M. J., & Corley, M. (2006). The influence of lexical, conceptual and planning based factors on disfluency production. In *Proceedings of the twenty-eighth meeting of the cognitive science society*. Canada: Vancouver.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.

Van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language, 41*, 147–182.

Van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research, 17*, 701–718.

Van Petten, C. M., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 394–417.

References

Aguirre, G. K. (2003). Functional imaging in behavioral neurology and cognitive neuropsychology. In T. E. Feinberg & M. J. Farah (Eds.), *Behavioral neurology and cognitive neuropsychology.* New York: McGraw Hill.

Allison, T., Wood, C. C., & McCarthy, G. (1986). The central nervous system. In M. G. H. Coles, E. Donchin, & S. W. Porges (Eds.), *Psychophysiology: Systems, processes and applications* (p. 5-26). London: Guildford Press.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247-264.

American Electroencephalographic Society. (1994). Guideline thirteen: Guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, *11*, 111-113.

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., et al. (1991). The HCRC map task corpus. *Language and Speech*, *34*, 351-366.

Arnold, J. E., Fagnano, M., & Tanenhaus, M. K. (2003). Disfluencies signal thee, um, new information. *Journal of Psycholinguistic Research*, *32*, 25-36.

Arnold, J. E., Hudson Kam, C. L., & Tanenhaus, M. K. (2007). If you say thee uh you are describing something hard: The on-line attribution of disfluency during reference resolution. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *33*, 914-930.

Arnold, J. E., & Tanenhaus, M. K. (2007). Disfluency effects in comprehension: How new information can become accessible. In E. Gibson & N. Perlmutter (Eds.), *The processing and acquisition of reference.* MIT Press.

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new. disfluency and reference resolution. *Psychological Science*, *15*, 578-582.

Arnold, J. E., Wasow, T., Losongco, A., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of complexity and information structure on constituent ordering. *Language*, *76*, 28-55.

Bailey, K. G. D. (2004). *Disfluent speech and the visual world: An application of the visual world paradigm to the study of spoken language comprehension.* Unpublished doctoral dissertation, Michigan State University.

Bailey, K. G. D., & Ferreira, F. (2003). Disfluencies affect the parsing of garden-path sentences. *Journal of Memory and Language*, *49*, 183-200.

Barr, D. J. (2001). Trouble in mind: Paralinguistic indices of effort and uncertainty in communication. In C. Cavé, I. Guaïtella, & S. Santi (Eds.), *Oralité and gestualité: interactions et comportaments multimodaux dans la communication* (p. 597-600). Paris: L'Harmattan.

Beattie, G. W., & Bradbury, R. J. (1979). An experimental investigation of the modifiability of the temporal structure of spontaneous speech. *Journal of Psycholinguistic Research*, *8*, 225-248.

Beattie, G. W., & Butterworth, B. L. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech*, *22*, 201-211.

Bentin, S., Kutas, M., & Hillyard, S. A. (1993). Electrophysiological evidence for task effects on semantic priming in auditory word processing. *Psychophysiology*, *30*, 161-169.

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials associated with semantic priming. *Electroencephalography and Clinical Neurophysiology*, *60*, 343-355.

Besson, M., & Faita, F. (1995). An event-related potential (ERP) study of musical expectancy: Comparison of musicians with nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1278-1296.

Besson, M., Faita, F., Czternasty, C., & Kutas, M. (1997). What's in a pause: Event-related potential analysis of temporal disruptions in written and spoken sentences. *Biological Psychology*, *46*, 3-23.

Besson, M., Faita, F., & Requin, J. (1994). Brain waves associated with musical incongruities differ for musicians and non-musicians. *Neuroscience Letters*, *168*, 101-105.

Besson, M., & Kutas, M. (1993). The many facets of repetition: A cued-recall and event-related potential analysis of repeating words in the same versus different sentence contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1115-1133.

Besson, M., Kutas, M., & Van Petten, C. (1992). An event-related potential (ERP) analysis of semantic congruity and repetition effects in sentences. *Journal of Cognitive Neuroscience*, *4*, 132-149.

Blackmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, *39*, 173-94.

Blank, M. A., & Foss, D. J. (1978). Semantic facilitation and lexical access during sentence processing. *Memory and Cognition*, *6*, 644-652.

Bobes, M. A., Valdes-Sosa, M., & Olivares, E. (1994). An ERP study of expectancy violation in face perception. *Brain and Cognition*, *26*, 1-22.

Boomer, D. S. (1965). Hesitation and grammatical encoding. *Language and Speech*, *8*, 148-158.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech*, *44*, 123-147.

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, *44*, 274-296.

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive state of speakers. *Journal of Memory and Language*, *34*, 383-398.

Buckner, R. L. (2003). Functional-anatomic correlates of control processes in memory. *Journal of Cognitive Neuroscience*, *23*, 3999-4004.

Butterworth, B. L. (1975). Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research*, *1*, 75-78.

Cajal, S. R. (1909-1911). *Histologie du systeme nerveux de l'homme et des vertebres.* Paris: Maloine.

Carreiras, M., Salillas, E., & Barber, H. (2004). Event-related potentials elicited during parsing of ambiguous relative clauses in Spanish. *Cognitive Brain Research*, *20*, 98-105.

Christenfeld, N. (1994). Options and ums. *Journal of Language and Social Psychology*, *13*, 192-199.

Christenfeld, N. (1995). Does it hurt to say um? *Journal of nonverbal behaviour*, *19*, 171-186.

Christenfeld, N. (1996). Effects of a metronome on the filled pauses of fluent speakers. *Journal of Speech and Hearing Research*, *39*, 1232-1238.

Christenfeld, N., & Creager, B. (1996). Anxiety, alchohol, aphasia, and ums. *Journal of Personality and Social Psychology*, *70*, 451-460.

Chwilla, D. J., Brown, C. M., & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology*, *32*, 274-285.

Clark, H. H. (1994). Managing problems in speaking. *Speech communication*, *15*, 243-250.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*, 73-111.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, *37*, 368-407.

Coles, M. G. H., & Rugg, M. D. (1995). ERPs: An introduction. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind: Event-related potentials and cognition* (p. 1-26). Oxford: Oxford University Press.

Collard, P., Corley, M., MacGregor, L. J., & Donaldson, D. I. (in press). Attention orienting effects of hesitations in speech: Evidence from ERPs. *Journal of Experimental Psychology: Learning, Memory and Cognition*.

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, *6*, 256-266.

Connolly, J. F., Phillips, N. A., Stewart, S. H., & Brake, W. G. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and Language*, *43*, 1-18.

Connolly, J. F., Stewart, S. H., & Phillips, N. A. (1990). The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain and Language*, *39*, 302-318.

Cook, M. (1969). Transitional probabilities and the incidence of filled pauses. *Psychonomic Science*, *16*, 191-192.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*, 84-107.

Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition*, *105*, 658-668.

Coulson, S., Federmeier, K. D., Van Petten, C. M., & Kutas, M. (2005). Right hemisphere sensitivity to word and sentence level context: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 129-147.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, *13*, 21-58.

Coulson, S., & Wu, Y. C. (2005). Right hemisphere activation of joke-related information: An event-related brain potential study. *Journal of Cognitive Neuroscience*, *17*, 494-506.

Cummings, A., Ceponiene, R., Koyama, A., & Saygin, A. P. (2006). Auditory semantic networks for words and natural sounds. *Brain Research*, *1115*, 92-107.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabalistic word preactivation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*, 1117-1122.

Diaz, M., & Swaab, T. Y. (2007). Electrophysiological differentiation of phonological and semantic integration in word and sentence contexts. *Brain Research*, *1146*, 85-100.

Dien, J. (1998). Issues in the application of the average reference: Review, critiques, and recommendations. *Behavior Research Methods, Instruments, and Computers*, *30*, 34-43.

Dien, J., & Santuzzi, A. M. (2005). Application of repeated measures ANOVA to high-density ERP data setset: A review and tutorial. In T. C. Handy (Ed.), *Event-related potentials: A methods handbook* (p. 57-82). Cambridge, MA: MIT Press.

Donchin, E. (1981). Surprise!…Surprise? *Psychophysiology*, *18*, 493-513.

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Science*, *11*, 355-372.

Duez, D. (1985). Perception of silent pauses in continuous speech. *Language and Speech*, *28*, 377-389.

Eriksen, C. W., Pollack, M. D., & Montague, W. E. (1970). Implicit speech: Mechanisms in perceptual encoding? *Journal of Experimental Psychology*, *84*, 502-507.

Fabiani, M., Gratton, G., & Coles, M. G. H. (2000). Event-related brain potentials. Methods, theory, and applications. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (p. 53-84). Cambridge: Cambridge University Press.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*, 469-495.

Federmeier, K. D., & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 202-224.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75-84.

Ferreira, F. (2007). Prosody and performance in language production. *Language and Cognitive Processes*, *22*, 1151-1177.

Ferreira, F., Lau, E. F., & Bailey, K. G. D. (2004). Disfluencies, language comprehension, and tree adjoining grammars. *Cognitive Science*, *28*, 721-749.

Fischler, I., Bloom, D. G., Childers, D. G., Roucos, S. E., & Perry Jr., N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, *20*, 400-469.

Fischler, I., & Bloom, P. A. (1979). Automatic attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning and Verbal Behavior*, *18*, 1-20.

Ford, M. (1982). Sentence planning units: Implications for the speaker's representation of meaningful relations underlying sentences. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (p. 798-827). Cambridge, MA: MIT Press.

Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, *34*, 709-738.

Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory and Cognition*, *29*, 320-326.

Fox Tree, J. E. (2002). Interpreting pauses and ums at turn exchanges. *Discourse Processes*, *34*, 37-55.

Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, *62*, 151-167.

Fox Tree, J. E., & Schrock, J. C. (1999). Discourse markers in spontaneous speech: Oh what a difference an oh makes. *Journal of Memory and Language*, *40*, 280-295.

Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and Language*, *50*, 259-281.

Friederici, A. D. (2002). Towards a neural basis for auditory sentence processing. *Trends in Cognitive Science*, *6*, 78-84.

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory and Cognition.*, *22*, 1219-1248.

Friederici, A. D., Hahne, A., & Saddy, D. (2002). Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. *Journal of Psycholinguistic Research*, *31*, 45-63.

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

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic morphological syntactic violations. *Cognitive Brain Research*, *1*, 183-192.

Friederici, A. D., Steinhauer, K., & Frisch, S. (1999). Lexical integration: Sequential effects of syntactic and semantic information. *Memory and Cognition*, *27*, 438-453.

Friston, K. J., Price, C. J., Fletcher, P., Moore, C., & Frackowiak, R. (1996). The trouble with cognitive subtraction. *Neuroimage*, *4*, 97-104.

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, *8*, 89-106.

Glaser, E. M., & Ruchkin, D. S. (1976). *Principles of neurobiological signal analysis.* New York: Academic Press.

Goldman-Eisler, F. (1958a). The predictability of words in context and the length of pauses in speech. *Language and Speech*, *1*, 226-231.

Goldman-Eisler, F. (1958b). Speech production and the predictability of words in context. *Quarterly Journal of Experimental Psychology*, *10*, 96-106.

Grainger, J. (1990). Word frequency effects in lexical decision and naming. *Journal of Memory and Language*, *29*, 228-244.

Gratton, G. (1998). Dealing with artifacts: The EOG contamination of the event-related brain potential. *Behavior Research Methods, Instruments, & Computers*, *30*, 44-53.

Greenhouse, S. W., & Geisser, S. (1959). On the methods in the analysis of profile data. *Psychometrika*, *24*, 95-112.

Gunter, T. C., Friederici, A. D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience*, *12*, 556-568.

Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology*, *34*, 660-676.

Haenschel, C., Vernon, D. J., Dwivedi, P., Gruzelier, J. H., & Baldeweg, T. (2005). Event-related brain potential correlates of human auditory sensory memory-trace formation. *The Journal of Neuroscience*, *25*, 10494-10501.

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

Hagoort, P., & Brown, C. M. (1999). Gender electrified: ERP evidence on the syntactic nature of gender processing. *Journal of Psycholingusitic Research*, *28*(6), 715-728.

Hagoort, P., & Brown, C. M. (2000a). ERP effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia*, *38*, 1531-1549.

Hagoort, P., & Brown, C. M. (2000b). ERP effects of listening to speech: Semantic ERP effects. *Neuropsychologia*, *38*, 1518-1530.

Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP-measure of syntactic processing. *Language and Cognitive*

Processes, *8*, 439-483.

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

Hagoort, P., Wassenaar, M., & Brown, C. M. (2003). Syntax-related ERP effects in Dutch. *Cognitive Brain Research*, *16*, 38-50.

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, *11*, 193-204.

Hahne, A., & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research*, *13*, 339-356.

Hahne, A., & Jescheniak, J. D. (2001). What's left if the jabberwock gets the semantics? An ERP investigation into semantic and syntactic processes during auditory sentence comprehension. *Cognitive Brain Research*, *11*, 199-212.

Hald, L., Steenbeek-Planting, E. G., & Hagoort, P. (2007). The interaction of discourse context and world knowledge in online sentence comprehension. Evidence from the N400. *Brain Research*, *1146*, 210-219.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, *56*, 208-216.

Hawkins, R. R. (1971). The syntactic location of hesitation pauses. *Language and Speech*, *14*, 277-288.

Heike, A. E. (1981). A content-processing view of hesitation phenomena. *Language and Speech*, *24*, 147-160.

Heinze, H., Muente, T., & Kutas, M. (1998). Context effects in a category verification task as assessed by event-related brain potential (ERP) measures. *Biological Psychology*, *47*, 121-135.

Hindle, D. (1983). Deterministic parsing of syntactic non-fluencies. In *Proceedings of the 21st annual meeting of the association for computational linguistics* (p. 123-128). Cambridge, MA.

Hoeks, J. C. J., Stowe, L., & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, *19*, 59-73.

Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 721-742.

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in Lexical Decision: A comparison using event-related brain potentials. *Language and Cognitive Processes*, *5*, 281-312.

Holcomb, P. J., & Neville, H. J. (1991). Natural speech processing: An analysis using event-related potentials. *Psychobiology*, *19*, 286-300.

Holmes, V. M. (1988). Hesitations and sentence planning. *Language and Cognitive Processes*, *3*, 323-361.

Howell, P., & Sackin, S. (2001). Function word repetitions emerge when speakers are operantly conditioned to reduce frequency of silent pauses. *Journal of Psycholinguistic Research*, *30*, 457-474.

Howell, P., & Young, K. (1991). The use of prosody in highlighting alterations in repairs from unrestricted speech. *Quarterly Journal of Experimental Psychology*, *43*, 733-758.

Jackendoff, R. (2002). *Foundations of language*. Oxford: Oxford University Press.

Jasper, H. H. (1958). Report to the committee on methods of clinical examination in electroencephalography. Appendix: The ten-twenty system of the international federation. *Electroencephalography and Clinical Neurophysiology*, *10*, 370-375.

Kaan, E., Harris, T., Gibson, E., & Holcomb, P. J. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*, 159-201.

Kaan, E., & Swaab, T. Y. (2003a). Electrophysiological evidence for serial sentence processing: A comparison between non-preferred and ungrammatical continuations. *Cognitive Brain Research*, *17*, 621-635.

Kaan, E., & Swaab, T. Y. (2003b). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal Of Cognitive Neuroscience*, *15*, 98-110.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). Prediction and thematic information in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*, 133-156.

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguisic Research*, *32*, 37-55.

Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (Eds.). (2000). *Principles of neural science*. New York: McGraw-Hill.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*, 205-225.

King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and clause-related ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, *7*, 378-397.

Kircher, T. J., Brammer, M. J., Levelt, W. J. M., Bartels, M., & McGuire, P. K. (2004). Pausing for thought: Engagement of left temporal cortext during pauses in speech. *NeuroImage*, *21*, 84-90.

Kleinman, G. M. (1980). Sentence frame contexts and lexical decisions: sentence-acceptability and word-relatedness effects. *Memory and Cognition*, *8*, 336-344.

Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, *5*, 196-214.

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, *95*, 95-127.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience*, *7*, 302-307.

Koles, Z. J. (1998). Trends in EEG source localization. *Electroencephalography and Clinical Neurophysiology*, *106*, 127-137.

Kolk, H., Chwilla, D. J., van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, *85*, 1-36.

Koster, W. G. (1969). On the speed of mental processes. *Acta Psychologica*, *30*, 412-431.

Kounios, J., & Holcomb, P. J. (1992). Structure and process in semantic memory: Evidence from event-related potentials and reaction times. *Journal of Experimental Psychology*, *121*, 460-480.

Kounios, J., & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 804-823.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23-49.

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, *17*, 117-129.

Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*, *8*, 533-572.

Kutas, M., & Dale, A. (1997). Electrical and magnetic readings of mental function. In M. D. Rugg (Ed.), *Cognitive neuroscience studies in cognition* (p. 197-241). Cambridge, MA: MIT Press.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203-205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161-163.

Kutas, M., Neville, H. J., & Holcomb, P. J. (1987). A preliminary comparison of the N400 response to semantic anomalies during reading, listening, and signing. *Electroencephalography and Clinical Neurophysiology*, *Supplement 39*, 325-330.

Kutas, M., Van Petten, C. K., & Kluender, R. (2006). Psycholinguistics electrified II (1994-2005). In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., p. 83-143). San Diego, CA: Elsevier Press.

Lau, E. F., & Ferreira, F. (2005). Lingering effects of disfluent material on comprehension of garden path sentences. *Language and Cognitive Processes*, *20*, 633-666.

Lau, E. F., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain and Language*, *98*, 74-88.

Ledoux, K., Traxler, M. J., & Swaab, T. Y. (2007). Syntactic priming in comprehension: Evidence from event-related potentials. *Psychological Science*, *18*, 135-143.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*, 41-104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Levelt, W. J. M., & Cutler, A. (1983). Prosodic marking in speech repair. *Journal of Semantics*, *2*, 205-218.

Lickley, R. J. (1994). *Detecting disfluency in spontaneous speech.* Unpublished doctoral dissertation, University of Edinburgh.

Lickley, R. J. (1995). Missing disfluencies. In *Proceedings of the international congress of phonetic sciences* (Vol. 4, p. 192-195). Stockholm, Sweden.

Lickley, R. J. (1996). Juncture cues to disfluency. In *Proceedings of the international conference on spoken language processing* (Vol. 4, p. 2478-2481). Philadelphia, USA.

Lickley, R. J., & Bard, E. G. (1998). When can listeners detect disfluency in spontaneous speech? *Language and Speech*, *41*, 203-226.

Lindsay, J., & O'Connell, D. C. (1995). How do transcribers deal with audio recordings of spoken discourse? *Journal of Psycholinguistic Research*, *24*, 101-115.

Lins, O. G., Picton, T. W., Berg, P., & Scherg, M. (1993). Ocular artifacts in EEG and event-related potentials I: Scalp topography. *Brain topography*, *6*, 51-63.

Luck, S. J. (2005). *An introduction to the event-related potential technique.* Cambridge, Massachusetts: The MIT Press.

MacDonald, M., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676-703.

Maclay, H., & Osgood, C. E. (1959). Heistation phenomena in English speech. *Word*, *15*, 19-44.

Marslen-Wilson, W., & Tyler, L. G. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*, 1-71.

Marslen-Wilson, W., & Welsh, A. (1978). Processing interactions during word recognition in continuous speech. *Cognitive Psychology*, *10*, 29-63.

Martin, J. G. (1967). Hesitations in the speaker's production and listener's reproduction of utterances. *Journal of Verbal Learning and Verbal Behavior*, *6*, 903-909.

Martin, J. G., & Strange, W. (1968). The perception of hesitation in spontaneous speech. *Perception and Psychophysics*, *3*, 427-438.

Mattys, S. L., Pleydell-Pearce, C. W., Melhorn, J. F., & Whitecross, S. E. (2005). Detecting silent pauses in speech. A new tool for measuring on-line lexical and semantic processing. *Psychological Science*, *16*, 958-964.

McAllister, J., Cato-Symonds, S., & Johnson, B. (2001, August). Listeners' ERP responses to false starts and repetitions in spontaneous speech. In *Disfluency in spontaneous speech* (p. 65-68). Edinburgh, Scotland.

McCallum, W. C., Farmer, S. F., & Pocock, P. V. (1984). The effects of physical and semantic incongruities on auditory event-related potentials. *Electroencephalography and Clinical Neurophysiology*, *59*, 477-488.

McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology*, *62*, 203-208.

McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology*, *36*, 53-65.

Mecklinger, A., Schriefers, H., Steinhauer, K., & Friederici, A. D. (1995). Processing relative clauses varying on syntactic and semantic dimensions: An analysis with event-related potentials. *Memory and Cognition*, *23*, 477-494.

Meyer, D., & Schvaneveldt, R. W. (1971). Facilitation in recognising pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*, 227-234.

Miller, G. A., Lutzenberger, W., & Elbert, T. (1991). The linked-reference issue in EEG and ERP recording. *Journal of Psychophysiology*, *5*, 273-276.

Moreno, E. M., Federmeier, K. D., & Kutas, M. (2002). Switching languages, switching palabras (words): An electrophysiological study of code switching. *Brain and Language*, *80*, 188-207.

Münte, T. F., Matske, M., & Johannes, S. (1997). Brain activity associated with syntactic incongruencies in words and pseudo-words. *Journal of Cognitive Neuroscience*, *9*, 318-329.

Näätäanen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, *6*, 826-859.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its equivalent (MMNm). *Psychophysiology*, *38*, 1-21.

Näätänen, R., Gaillard, A. W. K., & Mantysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologia*, *42*, 313-329.

Nagy, M. E., & Rugg, M. D. (1989). Modulation of event-related potentials by word repetition: The effects of inter-item lag. *Psychophysiology*, *26*, 431-436.

Nakatani, C. H., & Hirschberg, J. (1994). A corpus-based study of repair cues in spontaneous speech. *Journal of the Acoustic Society of America*, *95*, 1603-1616.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (p. 264-336). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Neely, J. H., Keefe, D. E., & Ross, K. (1989). Semantic priming in the lexical decision task: Roles of prospective prime-generated expectancies and retrospective relation-checking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1003-1019.

Neville, H., Nicol, J., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Memory and Language*, *3*, 151-165.

Nieuwland, M. S., & Van Berkum, J. J. A. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary change deafness in discourse comprehension. *Cognitive Brain Research*, *24*, 691-701.

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

Okita, T., & Jibu, T. (1998). Selective attention and N400 attenuation with spoken word repetition. *Psychophysiology*, *35*, 260-271.

Oomen, C. C. E., & Postma, A. (2001a). Effects of divided attention on the production of filled pauses and repetitions. *Journal of Speech, Language, and Hearing Research*, *44*, 997-1004.

Oomen, C. C. E., & Postma, A. (2001b). Effects of time-pressure on mechanisms of speech production and self-monitoring. *Journal of Psycholinguistic Research*, *30*, 163-184.

Osterhout, L., & Hagoort, P. (1999). A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King and Kutas (1998) in the P600/SPS-P300 debate. *Language and Cognitive Processes*, *14*, 1-14.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 785-806.

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 786-803.

Osterhout, L., McKinnon, R., Bersick, M., & Corey, V. (1996). On the language specificity of the brain response to syntactic anomalies: Is the syntactic positive shift a member of the P300 family? *Journal of Cognitive Neuroscience*, *8*, 507-526.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, *34*, 739-773.

Otten, L. J., & Rugg, M. D. (2005). Interpreting event-related potentials. In T. C. Handy (Ed.), *Event-related potentials: A methods handbook* (p. 3-16). Cambridge, MA: MIT Press.

Otten, M., & Van Berkum, J. J. A. (2007). What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse-dependent N400 effect. *Brain Research*, *1153*, 166-177.

Paller, K. A., Kutas, M., & Mayes, A. R. (1987). Neural correlates of encoding in an incidental learning paradigm. *Electroencephalography and Clinical Neurophsysiology*, *67*, 360-371.

Patel, A., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, *10*, 717-733.

Pickering, M. J., & Garrod, M. J. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Science*, *11*, 105-110.

Plauche, M. C., & Shriberg, E. E. (1999, July). Data-driven subclassification of disfluent repetitions based on prosodic features. In *Proceedings of the international congress of phonetic sciences* (Vol. 3, p. 1513-1516). Berkley, USA.

Polich, J. (2004). Neuropsychology of P3a and P3b: A theoretical overview. In N. C. Moore & K. Arikan (Eds.), *Brainwaves and mind: Recent developments* (p. 15-29). Wheaton, IL: Kjellberg Inc.

Postma, A., & Kolk, H. H. J. (1992). The effects of noise masking and required accuracy on speech errors, disfluencies, and self-repairs. *Journal of Speech and Hearing Research*, *35*, 537-544.

Postma, A., Kolk, H. H. J., & Povel, D. J. (1990). On the relation among speech errors, disfluencies, and self-repairs. *Language and Speech*, *33*, 19-29.

Ranganath, C., & Paller, K. A. (1999). Frontal brain activity during episodic and semantic retrieval: Insights from event-related potentials. *Journal of Cognitive*

Neuroscience, *11*, 598-609.

Rhodes, S. M., & Donaldson, D. I. (2008). Association and not semantic relationships elicit the N400 effect: Electrophysiological evidence from an explicit language comprehension task. *Psychophysiology*, *44*, 50-59.

Ruchkin, D. S. (1988). Measurement of event-related potentials: Signal extraction. In T. W. Picton (Ed.), *Handbook of electroencephalography and clinical neurophysiology: Vol. 3. Human event-related potentials* (p. 7-43). Amsterdam: Elsevier.

Rugg, M. D. (1985). The effects of semantic priming and word repetition on event-related potentials. *Psychophysiology*, *22*, 642-647.

Rugg, M. D., Allan, K., & Birch, C. S. (2000). Electrophysiological evidence for the modulation of retrieval orientation by depth of study processing. *Journal of Cognitive Neuroscience*, *12*, 664-678.

Rugg, M. D., & Coles, M. G. H. (Eds.). (1995). *Electrophysiology of mind: Event-related potentials and cognition.* London: Oxford University Press.

Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology*, *60*, 362-367.

Schachter, S., Rauscher, F., Christenfeld, N. J. W., & Crone, K. T. (1994). The vocabularies of academia. *Psychological Science*, *5*, 37-41.

Schröger, R. (1997). On the detection of auditory deviants: A pre-attentive activation model. *Psychophysiology*, *34*, 245-257.

Schuberth, R. E., & Eimas, P. D. (1977). Effects of context on the classification of words and nonwords. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 27-36.

Schwanenflugel, P. J., & LaCount, K. L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 344-354.

Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of faciliation for upcoming words. *Journal of Memory and Language*, *24*, 232-252.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving semantic incremental interpretation through contextual representation. *Cognition*, *71*, 109-147.

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

Senkfor, A. J., & Van Petten, C. M. (1998). Who said what? An event-related potential investigation of source and item memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1005-1025.

Shelton, J. R., & Martin, R. (1992). How semantic is automatic semantic priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1191-1209.

Shriberg, E. S. (1994). *Preliminaries to a theory of speech disfluencies.* Unpublished doctoral dissertation, University of California at Berkley.

Shriberg, E. S. (1995, August). Acoustic properties of disfluent repetitions. In *Proceedings of the international congress of phonetic sciences* (Vol. 4, p. 384-387). Stockholm, Sweden.

Shriberg, E. S. (2001). To "errrr" is human: ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association*, *31*, 153-169.

Siegman, A. W., & Pope, B. (1966). Ambiguity and verbal fluency in the TAT. *Journal of Consulting Psychology*, *30*, 239-245.

Smith, V., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, *32*, 25-38.

Spencer, K. M. (2005). Averaging, detection, and classification of single-trial ERPs. In T. C. Handy (Ed.), *Event-related potentials: A methods handbook* (p. 209-227). Cambridge, MA: MIT Press.

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli. *Electroencephalography*, *38*, 387-401.

St. George, M., Mannes, S., & Hoffman, J. E. (1994). Global semantic expectancy and language comprehension. *Journal of Cognitive Neuroscience*, *6*, 70-83.

Swaab, T. Y., Camblin, C. C., & Gordon, P. C. (2004). Electrophysiological evidence for reversed lexical repetition effects in language processing. *Journal Of Cognitive Neuroscience*, *16*(5), 715-726.

Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics*, *30*, 485-496.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*, 545-569.

Tabossi, P. (1988). Effects of context on the immediate interpretation of unambiguous nouns. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 153-162.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. E. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 632-634.

Tannenbaum, P. H., Williams, F., & Hillier, C. S. (1965). Word predictability in the environments of hesitations. *Journal of Verbal Learning and Verbal Behavior*, *4*, 134-140.

Traxler, M. J., Bybee, M. D., & Pickering, M. J. (1997). Influence of connectives on language comprehension: Eye-tracking evidence for incremental interpretation. *Quarterly Journal of Expeirmental Psychology: Human Experimental*

Psychology, *50A*, 481-497.

Urbach, T. P., & Kutas, M. (2002). The intractability of scaling scalp distributions to infer neuroelectric sources. *Psychophysiology*, *39*, 791–808.

Urbach, T. P., & Kutas, M. (2006). Interpreting event-related brain potential (ERP) distributions: Implications of baseline potentials and variability with application to amplitude normalization by vector scaling. *Biological Psychology*, *72*, 333-343.

Van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, *13*, 967-985.

Van den Brink, D., & Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPs. *Journal of Cognitive Neuroscience*, *16*, 1068-1084.

Van Petten, C. M. (1993). A comparison of lexical and sentence-level context effects and their temporal parameters. *Language and Cognitive Processes*, *8*, 485-532.

Van Petten, C. M., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 394-417.

Van Petten, C. M., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition*, *18*, 380-393.

Van Petten, C. M., & Kutas, M. (1991). Influences of semantic and syntactic context on open and closed class words. *Memory and Cognition*, *19*, 95-112.

Van Petten, C. M., Kutas, M., Kluender, R., Mitchiner, M., & McIsaac, H. (1991). Fractionating the word repetition effect with event-related potentials. *Journal* *of Cognitive Neuroscience*, *3*, 131-150.

Van Petten, C. M., & Rheinfelder, H. (1995). Conceptual relationships between spoken words and environmental sounds: Event-related brain potential measures. *Neuropsychologia*, *33*, 485-508.

Van Petten, C. M., Weckerly, J., McIsaac, H. K., & Kutas, M. (1997). Working memory capacity dissociates lexical and sentential context effects. *Psychological Science*, *8*, 238-242.

Van Wijk, C., & Kempen, G. (1987). A dual system for producing self-repairs in spontaneous speech: Evidence from experimentally elicited corrections. *Cognitive Psychology*, *19*, 403-440.

Van Berkum, J. J. A. (2005). Sentence comprehension in a wider discourse: Can we use ERPs to keep track of things? In M. Carreiras & C. J. Clifton (Eds.), *The on-line study of sentence comprehension: Eye tracking, ERP and beyond* (p. 229-270). Psychology Press.

Van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999a). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, *41*, 147-182.

Van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999b). When does gender constrain parsing? evidence from ERPs. *Journal of Psycholingusitic Research*, *28*(5), 555-566.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory and Cognition.*, *31*, 443-467.

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

Van Berkum, J. J. A., Koornneef, A. W., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research*, *1146*, 158-171.

Van Berkum, J. J. A., Zwisterlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research*, *17*, 701-718.

Van Herten, M., Kolk, H. H. J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic reversal anomalies. *Cognitive Brain Research*, *22*, 241-255.

Verleger, R. (1991). The instruction to refrain from blinking affects auditory P3 and N1 amplitudes. *Electroencephalography and Clinical Neurophysiology*, *78*, 240-251.

Verleger, R., Jasowski, P., & Wauschkuhn, B. (1994). Suspense and surprise: On the relationship between expectancies and P3. *Psychophysiology*, *31*, 359-369.

Vissers, C. T. W. M., Chwilla, D. J., & Kolk, H. H. J. (2006). Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. *Brain Research*, *1*, 150-163.

Vos, S. H., Gunter, T. C., Kolk, H. H. J., & Mulder, G. (2001). Working memory constraints on syntactic processing: An electrophysiological investigation. *Psychophysiology*, *38*, 41-63.

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

West, W., & Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: An electrophysiological investigation. *Journal of Cognitive Neuroscience*, *12*, 1024-1037.

Whaley, C. P. (1978). Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, *17*, 143-154.

Wicha, N. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not pope: Human brain potentials to gender expectation and agreement in spanish spoken sentences. *Neuroscience Letters*, *346*, 165-168.

Wicha, N. Y., Moreno, E. M., & Kutas, M. (2003). Anticipating words and their gender: An event-related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in spanish. *Cortex*, *39*, 483-508.

Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy and gender agreement in spanish sentence reading. *Journal of Cognitive Neuroscience*, *16*, 1272-1288.

Wilding, E. L. (2006). The practice of rescaling scalp-recorded event-related potentials. *Biological Psychology*, *72*, 325–332.

Wingate, M. E. (1984). Pause loci in stuttered and normal speech. *Journal of Fluency Disorders*, *9*, 227-235.

Wood, C. C. (1987). Generators of event-related potentials. In A. M. Halliday, S. R. Butler, & R. Paul (Eds.), *A textbook of clinical neurophysiology* (p. 535-567). John Wiley and Sons Ltd.

Wood, C. C., & Allison, T. (1981). Interpretation of evoked potentials: A neurophysiological perspective. *Canadian Journal of Psychology*, *35*, 113-135.

Yamada, Y., & Neville, H. J. (2007). An ERP study of syntactic processing in English and nonsense sentences. *Brain Research*, *1130*, 167-180.

Zellner, B. (1994). Pauses and the temporal structure of speech. In E. Keller (Ed.), *Fundamentals of speech synthesis and speech recognition* (p. 31-62). Chichester: John Wiley.

Zwiterslood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, *32*, 25-64.