Lexical Influences on Disfluency Production

Michael J. Schnadt, B.A.(Hons), M.Sc.

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

November 2009

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.

Michael J. Schnadt

Abstract

Natural spoken language is full of disfluency. Around 10% of utterances produced in everyday speech contain disfluencies such as repetitions, repairs, filled pauses and other hesitation phenomena. The production of disfluency has generally been attributed to underlying problems in the planning and formulation of upcoming speech. However, it remains an open question to what extent factors known to affect the selection and retrieval of words in isolation influence disfluency production during connected speech, and whether different types of disfluency are associated with difficulties at different stages of production.

Previous attempts to answer these questions have largely relied on corpora of unconstrained, spontaneous speech; to date, there has been little direct experimental research that has attempted to manipulate factors that underlie natural disfluency production. This thesis takes a different approach to the study of disfluency production by constraining the likely content and complexity of speakers utterances while maintaining a context of naturalistic, spontaneous speech.

This thesis presents evidence from five experiments based on the Network Task (Oomen & Postma, 2001), in addition to two related picture-naming studies. In the Network Task, participants described to a listener the route of a marker as it traverses a visually presented network of pictures connected by one or more paths. The disfluencies of interest in their descriptions were associated with the production of the picture name. The experiments varied the ease with which pictures in the networks could be named by manipulating factors known to affect lexical or pre-lexical processing: lexical access and retrieval were impacted by manipulations

of picture-name agreement and the frequency of the dominant picture names, while visual and conceptual processing difficulty was manipulated by blurring pictures and through prior picture familiarisation. The results of these studies indicate that while general production difficulty does reliably increase the likelihood of disfluency, difficulties associated with particular aspects of lexical access and retrieval have dissociable effects on the likelihood of disfluency. Most notably, while the production of function word prolongations demonstrates a close relationship to lexical difficulties relating to the selection and retrieval of picture names, filled pauses tend to occur predominantly at the beginning of utterances, and appear to be primarily associated with message-level planning processes. Picture naming latencies correlated highly with the rates of observed hesitations, establishing that the likelihood of a disfluency could be attributed to the same lexical and pre-lexical processes that result in longer naming times. Moreover, acoustic analyses of a subset of observed disfluencies established that those disfluencies associated with more serious planning difficulties also tended to have longer durations, however they do not reliably relate to longer upcoming delays.

Taken together, the results of these studies demonstrate that the elicitation of disfluency is open to explicit manipulation, and that mid-utterance disfluencies are related to difficulties during specific production processes. Moreover, the type of disfluency produced is not arbitrary, but may be related to both the type and location of the problem encountered at the point that speech is suspended. Through the further exploration of these relationships, it may be possible to use disfluency as an effective tool to study online language production processes.

Acknowledgements

Many people and organisations have been instrumental to the completion of this thesis, and without whose assistance I would not be where I am today.

First I would like to acknowledge a number of funding sources that have supported me through my PhD. My research was funded through an ESRC postgraduate research studentship (no. PTA-030-2002-01229). In addition, the ESRC funded an overseas visiting studentship to work in in the Language Production Lab at UCSD, in San Diego, CA. The School of Philosophy, Psychology and Language Sciences, University of Edinburgh has kindly provided support in the form of several small projects grants for travel to international conferences; a Grindley grant funded my attendance at CogSci2006, in Vancouver Canada. The research presented in this thesis has greatly benefited from the feedback I received and discussions I had with researchers at these conferences.

In terms of people who have had a fundamental impact on the last four years as a PhD student at Edinburgh University, first and foremost, I would like to thank my supervisor, Dr. Martin Corley, who has stimulated, inspired and encouraged me (even reprimanded me!). But most of all he has been an enormous support on both a professional and personal level throughout my studies, and I owe him a huge debt of gratitude (and possibly a pint!).

I also want to thank Dr. Vic Ferreira for inviting me to work with him in San Diego, and to Bob Slevc for taking me under his wing there.

I am grateful to the assistance provided by a number of colleagues who helped with cross-coding and transcription, stooging and general assistance with many different aspects of my experimental work. Most notably, Hannah Clark and Lucy Furness, who assisted with the running of an early experiment; Katelyn Zmich who helped me with coding, Lucy MacGregor and Ollie Stewart, who graciously volunteered to transcribe and cross-code some of my data; and to Esther Mead, Gemma Rundell, Sally Pring and Jess Buck, who performed admirably as innocent stooges, and made many an experimental day more enjoyable.

Life would have been pretty boring if it had not been for my office mates, fellow EDGers, and good friends Dr. Corey McMillan, Dr. Lucy MacGregor, Suzy Moat and Dr. Phil Collard, who have all contributed enormously in terms of enlightening discussion and cups of tea.

I would also like to thank the other members of the Edinburgh Disfluency Group, and in particular Robin Lickley, for continued input and stimulating discussion.

And finally I would like to thank my mother, Liza Hollingshead, and Anabelle Goujon for providing much love and support throughout these last months. I really couldn't have done it without you.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The goal of communication is the transfer of ideas, thoughts, and emotions from individual to individual. Human language achieves this end by encoding the proposition to be expressed in a form that can be transmitted via an appropriate medium (vision, or sound) to its intended audience. However, this system is subject to error. This thesis undoubtedly contains typos, and other unintentional mistakes. Similarly, speech is often affected by errors, from the exchange of phonemes to the unintentional substitution of words (e.g., Levelt, 1989). However, unlike writing, speech also retains evidence of delays that beset the speaker when formulating an utterance, even if the utterance is produced as intended. This evidence is in the form of *disfluencies*, or the hesitations, repetitions, *um*s and *uh*s that are common in spontaneous unplanned speech.

This thesis is concerned with the production of disfluencies in the face of difficulty in formulating an utterance. Specifically, we focus on the choice of words: If a word is difficult to access (for example, because it is a low frequency word), is it more likely to be preceded by disfluency? Do different types of disfluency (such as prolongations or *um*s) accommodate different types of lexical difficulty (e.g., selecting a word vs. retrieving the associated phonological information), or are they determined by the degree of difficulty encountered? The thesis presents six experiments designed to answer these questions. Five of the experiments use a task in which speakers describe a route taken by a marker over a sequence of pictures. By manipulating the properties of the pictures that speakers have to name, we are able to determine

which aspects of the selection and retrieval of the appropriate picture names are likely to result in different types of disfluency. As an alternative measure of difficulty in naming the pictures, one experiment (and an additional sub-experiment) requires participants to name pictures in isolation, in order to examine whether the time taken to select and retrieve the picture names when they are presented in isolation correlates with the likelihood of disfluency when the pictures are named in the context of spontaneous speech.

1.1 Thesis Plan

The thesis is structured as follows. In **Chapter 2**, we survey the literature on disfluency, asking what disfluencies are, why they are of interest, in what circumstances they tend to be produced, and what is known both about influences that are local to a disfluency (i.e., associated with the immediately surrounding speech), and those that relate to the nature of discourse and communication. **Chapter 3** provides an overview of current models of language production, and specifically could account for the production of disfluency, and how disfluencies in continuous speech may be associated with delays associated with different stages of observed **Chapter 4** introduces the the experimental methodology that will be used throughout.

In **Chapter 5** we outline three experiments based on the Network Task, in which participants describe the route taken by a marker through a network of images interconnected by one or more paths. Additionally we introduce a naming study in which we measure the naming latencies for all of the items in the Network Task experiments. We show that prolongations tend to be the disfluency most commonly associated with difficult lexical items. Further, their production is consistently affected by factors thought to influence the speed of lexical access, although in general disfluencies are more likely to occur when the pictures have low name agreement (i.e., correspond to several possible names), rather than low frequency names.

Chapter 6 introduces two further Network Task experiments in which the studies reported in Chapter 5 are refined and extended. The final Network Task experiment includes a condition in which participants first complete a naming task. Using different names in the naming and Network tasks increases the probability of being disfluent, confirming that disfluencies are largely driven by competition, either explicitly or implicitly in the case of pictures which correspond to several names.

In **Chapter 7** we present additional acoustic analyses of the disfluencies recorded in the first experiment in **Chapter 6**, which set out to examine whether, and if so, how different types of disfluency are reliably related to different lengths of upcoming delay. We test two specific hypotheses about disfluency due to Herbert Clark and his colleagues (Clark & Fox Tree, 2002; Fox Tree & Clark, 1997), namely that *um*s are likely to be followed by longer silences than *uh*s, and that prolongations which include full vowels (*the* pronounced "thee") are more likely to be followed by a suspension of speech than reduced prolongations. Neither hypothesis was borne out by our data. **Chapter 8** discusses these and other findings, and presents the conclusions of the thesis.

Chapter 2

Background

2.1 Introduction

Speaking is not easy. When we speak, the words that we utter are the result of a complex process that transforms the concept of an intended message into sets of speech sounds that are formalised by grammatical and phonetic rules. According to Levelt (1989), this process has three main stages: speakers must first prepare a message at the conceptual level, then formulate a syntactic plan and retrieve lexical and phonological representations that correspond to the pre-verbal message, and finally, program articulatory movements that result in connected speech sounds.

This process of speech production is thought to be incremental in nature (Deese, 1984; Kempen & Hoenkamp, 1987; Levelt, 1989; Wheeldon & Lahiri, 1997). While a pre-verbal message may be prepared in advance of the initiation of speaking, the concepts and words that are used to express this message are selected and retrieved incrementally to fit into the evolving structure of the utterance, and then passed on in preparation for articulation. As a result, errors or delays in the conceptual, formulatory or articulatory processing of speech can (and often do) lead to the production of disfluencies.

2.2 Disfluencies

While the act of communication between a speaker and listener often feels effortless and fluent, as psycholinguists, we know that this is not the case. Spontaneous speech is littered with hesitations, pauses, repeated words and restarted or repaired phrases. Such speech phenomena have collectively been called **disfluencies**, a term that was initially used by Johnson (1961) to reflect deviations from fluent speech observed in normal speech and that of people who stutter. Brutten (1963, p. 41) originally defined disfluencies as "interruptions and breaks in the flow of the speech signal," and more recently Fox Tree (1995, p. 709) considered disfluencies to be "phenomena that interrupt the flow of speech and do not add propositional content to an utterance." However, Postma, Kolk, and Povel (1990) have made a distinction between *disfluencies*, which they classed as interruptions to the execution of the speech plan, and *self-repairs*, which they defined as corrections of speech errors, or already articulated deviations in the intended speech plan. While the distinction that they make is an important one, for the purposes of this thesis, use of the term disfluency will include both hesitation and repair phenomena.

The consistency of the general definition of disfluency over several decades of research masks the considerable variation in terminology that has been ascribed to these phenomena, and consequently both what has been considered a disfluency, and the definition of different types of disfluency. Phenomena that fall under the classification of disfluency have been previously referred to as "hesitations" (e.g., Maclay & Osgood, 1959), "pauses" (e.g., Goldman-Eisler, 1958b), "speech disturbances" (e.g., Mahl, 1957), "nonfluency" (e.g., Miller & Hewgill, 1964), "dysfluency" (e.g., Culatta & Leeper, 1988), and "self-repairs" (e.g., Levelt, 1983), among other terms, and this breadth of terminology is at least in part due to the variety of fields of study that disfluency research has come under.

2.3 Classes of disfluency

One of the earliest and most influential disfluency classification structures was developed by Wendell Johnson in his pioneering research into stuttering in the 1950s (Johnson, 1955; Johnson & Associates, 1959). This category structure became the standard for future research into stuttering and disfluency. His eight categories of speech disfluency, observed in both stutterers and "normal" adults were: (1) interjections of sounds, word or phrases, including interjections such as *uh, ah* and *um*; (2) partial-word repetitions; (3) whole word repetitions; (4) phrase repetitions; (5) revisions; (6) incomplete phrases; (7) broken words; (8) prolonged sounds.

Another early structure for categorising different classes of disfluency from the field of psychotherapy was provided by Mahl (1957). Examining the relationship between patient anxiety and "speech disturbances" observed in a clinical setting, Mahl defined eight different categories of what he considered to be "disturbances": (1) "ah"; (2) sentence correction; (3) sentence incompletion; (4) repetition of words; (5) stutter; (6) intruding incoherent sound; (7) slips of the tongue; (8) whole or partial word omission.

In the field of psycholinguistics, Maclay and Osgood (1959), in an early seminal study examining hesitation phenomena in spontaneous speech, further refined these categories to four, which they believed represented the main types of hesitation phenomena in normal spontaneous speech:

1. filled pauses, which equate to Mahl's "ah," but also include other vocalisations, such as "uh," "eh," "um" or "mm".
2. unfilled or silent pauses.
3. repeats of phrases, words or partial words (including part-word stutters).
4. false starts, either retraced or non-retraced, which corresponded to Mahl's sentence corrections and incompletions.

Maclay and Osgood (1959) omitted Mahl's other categories of sound intrusions, tongue slips and word omissions from their categorisation, largely because of a lack

of evidence for them in their corpus. As Postma et al. (1990) suggest, while slips of the tongue and other speech errors such as exchange errors and omissions are deviations from an intended speech plan, if they go uncorrected then they do not constitute a speech interruption. It is only the interruption of speech to repair such an error that would be considered a false start under Maclay and Osgood's category structure.

For the purposes of this thesis, I will use a disfluency classification structure broadly conforming to that of Maclay and Osgood (1959), but containing five disfluency classes, as detailed below. It should be noted that throughout this thesis I will make a distinction between *classes* of disfluency, i.e., the categories to which a particular disfluency conforms, and *types* of disfluency, i.e., differing types of disfluency within a particular disfluency class (e.g., Clark & Fox Tree, 2002, have made a distinction between the fillers *um* and *uh*). The five classes of disfluency examined in this thesis are:

1. **Filled pauses**: vocalised hesitations that interrupt ongoing speech, such as *uh* and *um*.
2. **Silent pauses**: silences of unusually long length occurring within the context of a phrase.
3. **Prolongations**: vowel or consonantal lengthening, such as "the" pronounced *thee*, or "a" pronounced *ay*.
4. **Repetitions**: repeated partial words, whole words or phrases that violate the syntax of the constituent utterance, but do not contain any repaired speech.
5. **Repairs**: Interruptions to an utterance to revise previously produced speech. Repairs may correct speech errors already articulated or replace an erroneous word (*error repair*), or may alter the meaning or further specify part of an intended message (*appropriateness repair*). Different types of repair include word or phrase insertions, substitutions and deletions (as classified by Shriberg, 1994, see section 2.3.5).

Other speech phenomena that have also been included within the umbrella of disfluency include lexical fillers or editing expressions, which are common phrases that have no semantic content within an utterance, such as "I mean" and "you know". The use of such lexical fillers has become increasingly common in both British and American English, and in some ways their use bears many similarities to that of filled pauses. They are often observed at the beginning of utterances and at interruption points within self-repairs, and in this instance are thought to provide additional time for the resolution of planning and re-formulation processes. However, they primarily tend to be used as discourse markers elsewhere within an utterance, and have additional basic meaning, either inviting additional inferences or forewarning of later adjustments (Fox Tree & Schrock, 2002; see also Schiffrin, 1987), suggesting that they are not purely hesitation phenomena. Relatively little research has explicitly examined the production of lexical fillers, and they will not be addressed further within this thesis.

As mentioned previously, these classes of disfluency broadly fall into two categories: hesitation disfluencies and repair disfluencies. Hesitations involve an interruption to speech, but do not result in any back-tracking or repairing of material, while repairs involve interruption to correct previously articulated speech. Hesitations also commonly occur at the interruption point within repairs, providing time for the reformulation of a repair (Levelt, 1983). Levelt (1983, 1989) and Postma and Kolk (1993, Postma et al., 1990) have suggested that hesitations, when observed outside of a self-repair, are actually instances of "covert repair". In these instances, if speakers have identified an error in the internal speech plan, they can interrupt the ongoing speech to correct this error prior to its articulation, resulting in the production of a hesitation (often a filled pause or repetition) which masks this repair process. While this may be one cause of hesitations in spontaneous speech, it is by no means the only one. Filled pauses are most commonly observed at utterance initial positions (Shriberg, 1996), where there is no prior speech to be interrupted, and as this thesis aims to show, the production of mid-utterance hesitation is also closely associated with delays in the formulation of the speech plan, and not just the internal monitoring and repair of this plan.

2.3.1 Filled Pauses

Filled pauses (or fillers, as they are also often termed) are probably the most distinctive form of hesitation in spontaneous speech. They are also thought to be a near-universal form of general hesitation, and have been studied in their different forms in languages such as German, Dutch, Swedish, Norwegian, Spanish, French, Hebrew and Japanese, among others (for a summary of studies, see Clark & Fox Tree, 2002). In each of these languages, fillers take a slightly different form. In English (and the general psycholinguistic literature) they have been characterised as having the form *uh* and *um*, although dialectical variations, such as those found in some British English accents, include *er, eh, erm* and *em*. Some early work examining hesitations (e.g., Mahl, 1957) also included vocalisations such as *ah* and *mm* as filled pauses, however, these common interjections have differing expressive meanings (often used to express surprise or recognition) than that of hesitant fillers. Within these studies, it is not clear whether these were included as variations of the traditional fillers *um* and *uh*, or as different interjections in their own right. As a result, this lack of clarity makes it difficult to compare results from some of the early studies on disfluency with more recent findings.

Clark and Fox Tree (2002) have made a distinction between the use of *uh* and *um*, arguing that they are differentially used by speakers to signal, respectively, either a minor or major anticipated delay in speaking. They contend that *um* is not simply an elongated version of *uh*, but that the speaker's choice of filler needs to be planned prior to the interruption of speaking, and therefore is a signal of the length of the anticipated upcoming delay in speaking. Additionally, they argue for a further variation of each of these fillers that includes a prolonged schwa vowel (which they denote as *u:h* and *u:m*), and that the prolongation of fillers signals the "continuation of an ongoing delay". However, one issue with their argument is that they only address both the length of the fillers and subsequent pauses in terms of "prosodic units" as coded by the transcribers of the corpus they use. There was no objective measurement of either the length of different types of filler nor of the following pauses.

However, in an earlier study, Smith and Clark (1993) did measure length of pauses associated with utterance initial *um*s and *uh*s when speakers answered questions with varying degrees of certainty. They found that, on average, *um*s were followed by a 4.12 second pause, while *uh*s were followed by a much shorter (1 second) pause, lending support to their claims about the selective use of fillers to signal differing lengths of upcoming delay. It should be noted though, that four second pause lengths are uncharacteristically long in spontaneous speech: for example, in a travel agent dialogue, O'Shaughnessy (1992) found "ungrammatical" silent pauses (i.e., pauses that occur within minor syntactic phrases) to last, on average, 490ms, while pauses occurring at major syntactic boundaries lasted, on average, 790ms. In a study of silent pauses in which participants described Rorschach inkblot plates, Kircher, Brammer, Levelt, Bartels, and McGuire (2004), after selecting the 85 longest silent pauses in their corpus, observed a mean pause length of 1261ms. The length of observed pauses in these and other studies suggests that the pause lengths observed by Smith and Clark (1993) may have been due to their location within answers to questions, and may not be consistent with general pause lengths in spontaneous conversational speech.

For the purposes examined here, and to further evaluate whether the filled pauses *um* and *uh* do reflect different lengths of delay, potentially associated with different processing difficulties, I will treat *um* and *uh* as two different *types* of the class of filled pause. As (Clark & Fox Tree, 2002) do not provide any objective method by which to categorise them, no distinction will be made between prolonged and non-prolonged forms of fillers.

2.3.2 Silent pauses

Silent pauses have been defined as "a period of vocal inactivity of a certain duration embedded within the stream of speech" (Heike, Kowal, & O'Connell, 1983). They have long been studied as part of the speech production process, and in early work, Goldman-Eisler (1961) claimed that as much as 50% of a person's time spent speaking is made up of silence. However, more recent re-evaluations of her work have

found her method of measurement to be flawed. O'Connell (1988) argued that she obtained this remarkably high rate because she included all "irrelevant vocal productions", including filled pauses and word repetitions as part of her measurement of silence.

A major issue for the study of silent pauses in spontaneous speech is determining criteria for the consistent identification of silent pauses themselves. There are two facets to this problem. The first has to do with correctly identifying what is a "hesitant" (or performance-based) pause and what is a pause based on a speaker's natural prosody. While some pauses may be associated with delays in planning and production processes, and may thus be considered to be hesitation pauses, pauses are also part of the rhythmic structure of spontaneous speech, separating utterances into intonational phrases, and may not necessarily be directly related to planning processes (Ferreira, 1993). Indeed, most pauses in natural speech tend to occur at clausal boundaries (for example, Hawkins (1971) observed that two thirds of all pauses and three quarters of all pause time are located at such constituent boundaries). Ferreira (2007, 1993) has argued that an important distinction between prosodic and performance-based pauses is that performance-based pauses are associated with the linguistic material that comes after the pause, whereas prosodic pauses are determined by their relationship to prior linguistic material. Prosodic pauses predominantly occur at intonational phrase boundaries and are thought to reflect what is left between the intrinsic length of a word and any word-final lengthening and the timing interval assigned to a phonological phrase within the metrical structure of an utterance. Performance-based pauses, however, can occur at any point at which a speaker encounters difficulty or has to plan upcoming material. As speech planning is thought to occur incrementally (Levelt, 1989; Wheeldon & Lahiri, 1997), an utterance may be passed from conceptualisation to formulation in fragments that reflect the major syntactic constituents of the utterance. Performance-based pauses that occur at grammatical junctions are therefore likely to reflect formulatory processes associated with an upcoming constituent, while pauses within a clause are thought to correspond to delays in lexical retrieval (Levelt, 1989). This distinction between the function of pauses

within clauses and at clause boundaries has also been supported by neuroimaging studies. Kircher et al. (2004) observed an increase in activation of the right inferior frontal gyrus during pauses at clausal boundaries, which they suggested may reflect memory retrieval and search processes related to conceptual organisation, whereas pauses within clauses were associated with activation of the superior and middle temporal gyri, areas previously implicated in lexical retrieval (Indefrey & Levelt, 2004; Kircher, Brammer, Williams, & McGuire, 2000). In fact, it becomes difficult to argue that pauses at clausal junctions are "disfluent" at all as they occur at natural prosodic boundaries and so may be part of normal prosodic segmentation and syntactic planning processes (Gee & Grosjean, 1983). In this case, "hesitant" pauses cannot be reliably identified at clausal constituent boundaries, and so their definition must be restricted to unexpected silence within constituents themselves.

The second part of this issue has to do with the minimum length of silence that is considered to be a silent pause. In Goldman-Eisler's (1968) work, she only included silent pauses with a minimum duration of 250ms. Gee and Grosjean (1983), in their examination of "performance structures" such as pausing, only identified pauses longer than 200ms, and yet Ferreira (2007) has argued that silences as short as 80ms can be associated with processes of planning or prosody. Clearly such short pauses could not be argued to be "hesitant" in nature, but it does raise the question of at what point one would consider a silent pause in speech to be hesitant. This duration would also depend on factors such as an individual's speech rate. Furthermore, a pause may only be considered to be hesitant when it is "noticeable" within its spoken context. Yet, perceptual coding of silent pauses brings its own set of problems, particularly in the consistent and accurate identification of shorter pauses (Kowal & O'Connell, 2000). Clark and Fox Tree (2002) made their analyses of silent pauses in the London-Lund corpus (Svartvik & Quirk, 1980) in terms of "pause units", which were perceptually estimated by the corpus transcribers. However, there is no clear temporal definition of their pause lengths beyond "one light foot" for a brief pause and "one stress unit" for a full (unit) pause (Clark & Fox Tree, 2002, p. 80), rendering any direct comparison with other silent pause research impossible.

As Clark and Fox Tree (2002, Fox Tree & Clark, 1997) have noted, silent pauses are very commonly associated with other types of disfluency, sometimes occurring before, but more often after, other types of hesitation. Because these pauses are clearly associated with the neighbouring disfluency, this raises the question of whether they should be considered to be separate disfluencies in their own right. And yet the fact that many silent pauses are observed in isolation, without associated disfluencies also raises the question of why, if many pauses in spontaneous speech are "filled", others are not.

2.3.3 Prolongations

Prolongations can generally be defined as speech sounds that are stretched out for longer than would be anticipated in normally paced speech. They featured prominently in the early stuttering literature (e.g., Johnson & Associates, 1959) as an early indicator of potential stuttering, and have also been termed *elongations, drawls* or *phonemic lengthening*. Despite their early recognition as a separate hesitation phenomenon, prolongations have often been overlooked or conflated with other disfluency categories in many studies of disfluency (e.g., Beattie & Butterworth, 1979; Blackmer & Mitton, 1991; Maclay & Osgood, 1959; Bortfield, Leon, Bloom, Schober, & Brennan, 2001). One possible reason for this, which is similar to the problem outlined with silent pauses, is that it is often difficult to determine objectively what constitutes a prolongation from purely durational data, as average phone lengths vary among individuals, and also by factors such as speech rate, stress and position within an utterance. Additionally, it can be difficult to reliably distinguish a hesitant prolongation from purely prosodic features such as phrase-final lengthening (Klatt, 1975). As a result of difficulties with automatic or objective prolongation detection, several studies that have explicitly examined prolongation production (e.g., Eklund, 2001, 2004) have coded them perceptually, identifying them as unusually long phones within the context of their surrounding speech, and then examined their temporal and phonetic properties.

While almost any speech sound can be prolonged, the most commonly occurring prolongations are of vowels at word-final positions (Eklund & Shriberg, 1998). In particular, many prolonged words are short function words, such as *the, a* or *to*, which tend to be produced prior to content words within noun or verb phrases. In addition to phone lengthening, recent studies have also focused on non-reduction of vowels as a signal of hesitation. Fox Tree and Clark (1997) examined a specific instance of prolongation, of the word *the*, pronounced with a non-reduced vowel as "thee", as a signal of problems during speech. They found that the non-reduced form, "thee", was associated with both subsequent delay and the production of other disfluencies. Fox Tree and Clark argued that because *the* is usually produced in a reduced form with a schwa vowel, when speakers produce the non-reduced form,they do so to signal an upcoming suspension of speech, reflecting underlying local production difficulty. Bell et al. (2003) extended Fox Tree and Clark's (1997) work, examining the vowel quality and duration of other function words, including *a, in, of, to, and* and *that*. They observed that these function words had longer durations and fuller forms when they were co-located with a disfluency. Moreover, when vowel reduction was controlled for, effects of subsequent disfluencies on word duration were still observed when they contained both reduced and non-reduced vowels. This indicates that prolongation of both reduced and non-reduced forms of function words can be associated with upcoming production difficulty. It should also be noted that Bell et al. (2003) were agnostic as to whether such disfluencies were signals of upcoming planning problems, or part of production mechanisms to gain time to resolve planning problems.

As indicated above, both duration and vowel reduction appear to be separate facets of prolongation phenomena. Previous studies have focused on American English, which may show different reduction patterns to those observed for function words in British English. Therefore, in this thesis I will focus on prolongation of function words that include both reduced and non-reduced forms within the class of prolongations.

2.3.4 Repetitions

Repetitions are another common form of disfluency, that involve the interruption of speech, followed by the repetition of previously produced material, whether that be part of a word, a whole word, or multiple words that have just been produced. A key point about repetitions is that the linguistic output is unaltered from the first to the second repetition, e.g., (1):

(1) "... over the top curve down to the- **to the** left hand side"[1]

In English, function words are repeated much more often than content words and are also often accompanied by other hesitations at the interruption point between the repeated words (Maclay & Osgood, 1959; Clark & Wasow, 1998). Clark and Wasow (1998) examined repetition rates in the Switchboard corpus and found that function words were repeated over ten times as often as content words (25.2 vs 2.4 repetitions per thousand words). They suggested that this higher rate of repetition of function words is largely due to their occurrence at the beginning of constituents, whereas content words tend to occur later in a constituent. They propose that repetitions represent a preliminary commitment to speaking to avoid unnecessary silence. If a speaker commits to producing an utterance before it is fully planned, they may have to suspend speaking to complete the planning of their utterance, and repeat the first words of the utterance to restore continuity. However, not all repetitions occur at the beginning of a constituent, nor do they exclusively involve function words, and so such an account does not cover all types of observed repetitions. Heike (1981) proposed that there were two broad types of repetitions: what he termed *prospective* and *retrospective* repetitions. Prospective repetitions are essentially hesitation devices similar to filled pauses or prolongations that delay production to allow the resolution of upcoming problems (e.g., the resolution of lexical search), while retrospective repetitions provide a "bridging device" to allow the resumption of fluent speech following some kind of interruption.

[1] All examples given throughout this thesis are actual speech excerpts taken from the experiments detailed herein

Alternative accounts of the production of repetitions have proposed that they are a form of *covert repair* (Levelt, 1983; Postma & Kolk, 1993), in which an error in the speech plan is identified prior to its articulation, inducing an interruption of speech which is retraced back to the nearest constituent boundary (see below). Shorter sub-word repetitions have also been thought to be due to an "autonomous restart capacity" of the articulatory buffer which repeats prior material if no new material is passed into the articulatory buffer by the time prepared material has been articulated (Blackmer & Mitton, 1991). Plauché and Shriberg (1999) provide evidence based on differing prosodic features for three types of whole-word repetition, which they suggest correspond to prospective and retrospective repetitions and covert repairs. However, such a distinction is not possible without a detailed prosodic analysis of each repetition, or by making subjective classifications based on their surrounding context. Therefore, repetitions examined within this thesis will not be broken down into different types, but will be treated as a single class of disfluency.

2.3.5 Repairs

Repairs are perhaps the most complex class of disfluency. While the forms of disfluency detailed above result in an interruption to ongoing speech, they do not involve any explicit change to the speech that has already been produced, or to the intended message to be conveyed. Repairs, on the other hand, correspond to what Maclay and Osgood (1959) termed "false starts", but include a much wider range of phenomena. Broadly, they involve the interruption of speech to backtrack and alter the message that has already been produced, for one (or more) of several reasons. When speakers produce a repair, they tend to be very consistent in reformulating the repair so that the intended speech plan maintains syntactic and semantic coherence with the speech already produced.

Levelt (1983) provided a detailed description of the structure of a repair, in which he identified three major parts to any repair (see Figure 2.1). Each repair consists of the *original utterance* (OU), which includes all speech from the beginning of

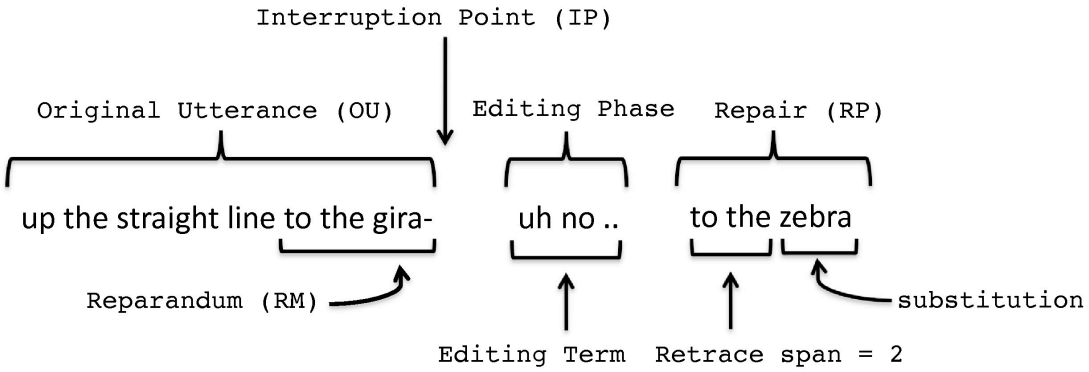


Figure 2.1: Structure of a simple error repair. The reparandum is the partial-word "gir-" of *giraffe*, which is replaced by "zebra". The repair retraces two words prior to the reparandum, and the edit interval between the reparandum and the repair includes the filler "uh", as well as the editing term "no". Adapted from Levelt (1983, p.45).

the current sentence until the *point of interruption* (IP). The OU includes the *reparandum* (RM), the portion of the utterance to be repaired. This is followed by the *edit phase* (EP), which is the period that follows interruption prior to the resumption of speech, and may contain silence, a filled pause or an editing term. Following the edit phase is the *repair* proper (RP), which is the portion of speech that replaces the reparandum to produce a fluent continuation, and may include some retracing of words from the original utterance. While some repairs may follow this general structure, others do not, and it is not uncommon to observe repairs in which the whole OU becomes the reparandum and is abandoned, or where there is no reparandum at all (i.e., what Levelt, 1983, would call a covert repair). Indeed, many repairs may be much more complex than this and include additional nested repairs. When this occurs, speakers are surprisingly capable of managing such repairs, and will normally structure complex repairs so that a fluent utterance can be constructed out of the original utterance and the repaired speech.

Several different schemes have been formulated for the classification of repairs, broadly based on the repair structure detailed above (for an overview, see Lickley, 1994). These fall into two groups: **functional** and **structural** classification

schemes. Levelt (1983) originally proposed a functional classification scheme based on the speaker's perceived motives for different types of repair. He argued that the different types of repairs are produced to correct errors or discrepancies between the intended speech plan and speech that has just been, or is about to be articulated. According to Levelt, repairs fall into 5 categories based on the reason that a speaker might produce them:

1. **D-Repairs** are produced when a speaker aborts what they are currently saying in order to say something *different*. These would correspond to Maclay and Osgood's (1959) false starts:
   e.g., "[goes on a- **from a flag**] on the curved line to the left..."
2. **A-Repairs** are produced when the speaker realises that the speech already produced needs to be qualified in some way to make it more *appropriate*, given the context or a listener's knowledge state. A-repairs may be the result of an overly *ambiguous* reference (*A-A repairs*), the need to adjust the *level of precision* to clearly refer to a concept (*A-L repairs*), or to ensure that what has just been said is *coherent* with the prior discourse or terminology (*A-C repairs*).
   e.g., "uh [down to thee- . . **down and to the right**] to the bone"
3. **E-Repairs** are produced when the speaker identifies an *error* in already produced speech and attempts to correct it. Such repairs may replace *lexical* errors (*E-L repairs*), may correct an incorrect *syntactic* construction or syntactic error (*E-S repairs*), or may be *phonetic* repairs (*E-F repairs*) to correct slips of the tongue and other speech errors.
   e.g., "along on the [top right- uh **top left**] curved line to a brick wall"
4. **C-Repairs** are *covert*, and contain only an interruption plus an editing term, hesitation or repetition, without changing any part of the OU in the repair. This category corresponds to the classes of filled pauses and repetitions described above. Levelt (1983) argued that covert repairs are the result of monitoring some level of 'inner speech', although because such repairs are inherently opaque, it is difficult to make strong claims that this is the case.

e.g., "over the top curved one to [the- um **the**] screw thing"

5. **R-Repairs** are the *rest* of the repairs, including any instances of repair that cannot be systematically categorised into any of the above categories.

While such a classification scheme based on semantic distinctions of the speaker's intent is appealing as it is addresses *why* speakers produce different types of repair, it is also inherently subjective, as it relies on judgements based on the speaker's model of the listener's state of knowledge within the discourse (Shriberg, 1994). This leads to ambiguities in which certain repairs could be classified as both E-repairs and A-repairs, depending on the current discourse state and history, or in the case of part-word repairs, where the reparandum cannot be readily identified, and requires the coder to provide a subjective interpretation of each repair.

Blackmer and Mitton (1991) also used a functional coding scheme based on that of Levelt (1983), but rather than focus on the intentions of the speaker in formulating a repair, made a distinction between what they classed as **conceptual** and **production** based repairs. This distinction was related to the perceived origin of the error within the production system. Conceptually-based repairs correct problems occurring during message conceptualization, and include Levelt's categories of D-repairs and A-Repairs, which were subdivided into *appropriateness replacements*, where some of the prior material is replaced, and *appropriateness inserts*, where additional material is inserted into the repaired message. Production-based repairs correct problems due to formulation or articulatory processes, and correspond to Levelt's category of E-repair. Covert repairs are divided into three groups: those containing a repetition, those containing an editing term or hesitation, and those containing both. This classification scheme, while similar to Levelt's, allows classification based on the perceived locus of the production problem, and therefore repairs (and the time between speech interruption and the onset of the repair) can be more closely tied to particular production processes. It also includes a degree of subclassification based on structural components of a repair. By the authors' own admission, this scheme resulted in a higher proportion of unclassifiable repairs (or R-repairs), largely because it was applied to open domain spontaneous speech which

tended to produce more complex and disfluent utterances than the closed domain set of network descriptions that Levelt (1983) used. Blackmer & Mitton's coding scheme, however, is still open to the same criticisms about the subjectiveness of repair classification as Levelt's scheme.

An alternative to functional coding schemes like those described above are structural coding schemes, which code repairs (and other disfluencies) according to the surface structure of the elicited repair. Such coding schemes avoid any pre-theoretical groupings that are based on semantic features or subjective assessments of the cause of a repair, resulting in a objective coding scheme that can be reliably applied across coders and domains. Shriberg (1994) developed a structural coding scheme which contained 7 main types of disfluency, and can be applied to almost all repairs observed in spontaneous speech:

1. **Repetitions** are repairs in which the RP and RM contain exactly the same words (with the exception of any intervening filler) on either side of the interruption point.
   e.g., "and then [down- **uh down**] to the windmill"
2. **Insertions** are repairs in which an additional word or words are inserted into the RP compared to the RM.
   e.g., "up and [acro- **horizontally across**] to a fish"
3. **Substitutions** are repairs where one or more words in the RM are replaced by other words in the RP. Substituted words must show syntactic and semantic correspondence with the words that they replace.
   e.g., "and down to [a bottle- **a glass**] of wine"
4. **Deletions** are repairs where words that are deleted from the RM have no corresponding words in the RP. Deletions include instances where the reparandum corresponds all or just part of the OU. Whole utterance deletions are also commonly called false starts, as the entire original utterance is deleted and the speaker starts afresh, such as 4a, below. Shriberg's coding scheme does not distinguish between these and partial deletions, where the RP mirrors the RM, but with one or more removed words, as in 4b.

(a) "[**its uh going straight-**] no its taking the left one down to the farm-house"

(b) "to a giraffe [on a straight curved line- sorry **on a straight line**]"

5. **Misarticulations** are repairs that replace a misarticulation or speech error with a correctly articulated replacement.
   e.g., "on an arc downwards to a sheaf of [wheath- **wheat**]"
6. **Complex repairs** are repairs that have multiple IPs, and contain more than one of the above types of repair. They are composed of multiple repairs that can nest within each other, or where a second repair alters the already repaired speech. Complex repairs can be though of in terms of a tree structure, where the internal repair is dealt with first to construct the fluent intended utterance. In the example below, *back* is substituted for *up* before the whole beginning of the utterance (which would correspond to *then back to the*) is deleted, and the utterance is started afresh.
   e.g., "[then [up to- **back to**] the-] from the girl up to the walkman"
7. **Filled pauses** are included within the repair coding scheme as they are also considered to be genuine speech interruptions (although no assumptions are made as to their status as covert repairs). However, filled pauses are only counted as separate disfluencies if they occur on their own within an utterance, and not when they are within the range of another repair.
   e.g., "straight left to thee [**uh**] jug of water"

While Shriberg's (1994) coding scheme does not make any theoretical claims about the relationship between repair structure and function, some parallels between the functional and structural coding schemes are worth noting. Generally, insertions would fall into Levelt's class of A-repairs. As insertions do not include the removal of any erroneous material, the insertion of additional material would only occur to improve the appropriateness of the utterance, facilitating listener comprehension of the intended message. Deletions, when resulting in a complete replacement of the OU, correspond to Levelt's D-repairs. Mid-utterance deletions are strong candidates for E-repairs, as it is unlikely that the removal of previously produced

material would increase the appropriateness of the repaired utterance, and so would generally be used to remove erroneous material. Misarticulations would also clearly be instances of Levelt's E-repairs. Substitutions are more flexible, and could be used to either replace erroneous material, or to further specify a concept with a more appropriate word. Complex repairs could similarly be either E-repairs or A-repairs, and are also most likely to constitute most of the unclassifiable group of R-repairs.

2.3.6 Summary: Classes of disfluency

Above I have outlined the five main classes of disfluency that are observed in natural speech: filled pauses, silent pauses, prolongations, repetitions and repairs. These different disfluency classes reflect two broad categories of disfluency: hesitation disfluencies, in which there is a delay but no retracement of material, and repair disfluencies, in which speakers backtrack and edit speech that has already been overtly produced. While it appears that hesitations and repairs may be used to accommodate different types of underlying production difficulty, other accounts have suggested that hesitations and repairs reflect similar problems in production (i.e., the planning of erroneous speech), however differ only in the stage at which the error is identified by the speech monitor. A related question is why speakers use different types of hesitations in different contexts and at different stages of an utterance? Do different types of hesitation reflect different types of underlying difficulty, or is the choice of disfluency simply related to speaker preference and the surrounding linguistic context. Therefore it is important to examine both the distribution of disfluency in spontaneous speech and to better understand the role of disfluency in speech planning and production in order to evaluate whether different classes of disfluency may reflect the accommodation of different type of production delays.

2.4 The role of disfluency in speech

Disfluency is a ubiquitous part of natural spoken language, and therefore an understanding of the role that disfluency plays in conversation can shed light on the

broader processes that influence communication. Although disfluency is thought to reflect the accommodation of planning difficulty on behalf of the speaker, it has also been shown to affect listeners' comprehension processes and their metalinguistic judgements of a speaker's communicative state, suggesting that disfluency plays a more general role in natural communication (Brennan & Williams, 1995; Clark, 1994; Fox Tree, 2001, 2002). Research on disfluency is therefore of interest for several reasons. First, disfluencies represent instances in which the "normal" processes of speech planning and formulation break down, and therefore can provide useful information about the architecture and constraints within the speech production system (e.g., Blackmer & Mitton, 1991; Cutler, 1982; Dell, 1986; Fromkin, 1973; Garrett, 1975; Levelt, 1989). Second, disfluency has been found to influence listeners' processing of subsequent words (e.g., Bailey & Ferreira, 2003; Brennan & Schober, 2001; Corley, MacGregor, & Donaldson, 2007; Fox Tree, 2001) and predictions about intended referents (e.g., Arnold, Tanenhaus, Altmann, & Fagnano, 2004; Arnold, Hudson Kam, & Tanenhaus, 2007; Barr, 2001), as well as inferences about the metacognitive state of the speaker (e.g., Brennan & Williams, 1995; Christenfeld, 1995; Fox Tree, 2002). Disfluency is not simply treated as noise and filtered out by listeners, but has an influence on a listener's comprehension processes and judgements about the speaker. Finally, disfluency is thought to have a role in structuring conversations, both as a floor-holding device (Maclay & Osgood, 1959; Smith & Clark, 1993), or as markers of discourse structure (Swerts, 1998), and it has been argued that disfluency provides collateral signals that both speakers and listeners use to manage conversational turn-taking in dialogue (Clark, 1994, 2002; Fox Tree, 2002; Sacks, Schegloff, & Jefferson, 1974), as well as to signal problems in speaking (Clark & Wasow, 1998; Clark & Fox Tree, 2002; Fox Tree & Clark, 1997; Smith & Clark, 1993).

This section outlines research examining the distribution of disfluency and findings on inter- and intra-speaker variation in disfluency production, as well as research investigating the effects of disfluency on comprehension processes, and arguments relating to the communicative nature of disfluency to provide an overview of general

patterns of disfluency in spontaneous speech and how disfluencies influence natural communication.

2.4.1 The distribution of disfluency in spontaneous speech

It is evident that disfluencies are relatively common within natural conversation. According to Fox Tree (1995), who collated results from several early disfluency studies, around 6 in every 100 spontaneously produced words are characterised as disfluent. This assessment of the general frequency of disfluency is supported by more recent corpus studies, such as those of Shriberg (1996), who observed a mean disfluency rate of around 6.4% of produced words taken from a corpus of telephone conversations, and Bortfield et al. (2001), who observed an overall disfluency rate of 5.9% within a corpus of face-to-face conversations. This overall rate of disfluency is also consistent in languages other than English (for example, Eklund & Shriberg, 1998, observed an overall disfluency rate of 6.4% within a Swedish human-human dialogue corpus). In contrast, other types of speech errors, such as slips of the tongue, occur much less frequently (Shallice & Butterworth, 1977, estimated that exchange errors occur in around 16 out of every 10,000 spoken words), suggesting that disfluencies accommodate more than just lexical or phonological speech error repair.

To investigate the distribution of disfluency under varying conversational situations, Shriberg (1994, 1996) studied three different corpora of spontaneous speech, consisting of the Switchboard corpus of open domain human-human telephone conversations, the AMEX corpus of human-human task-based travel agent dialogues, and the ATIS corpus of human-computer travel agent dialogues. Across all of these corpora, she observed that the likelihood of a disfluency increased with the length of utterances produced, although the number of disfluencies observed and the strength of this relationship was much greater in human-human free and goal-oriented dialogues than in human-machine dialogues (a finding also supported by Oviatt, 1995, in a study of human-human and human-computer speech). However, the disfluency rate *per word* was constant for sentences longer than about 5 words in both

human-human corpora, suggesting that the cognitive load associated with planning longer sentences was influencing disfluency rates, rather than any increase in the difficulty of planning later parts of an utterance. Furthermore, disfluencies were not evenly distributed within the speech examined. Shriberg (1996) found disfluencies were more likely to occur in a sentence-initial position than in sentence-medial positions[2], and that the likelihood of both sentence-initial and sentence-medial disfluencies increased with overall sentence length.

When Shriberg (1994) examined different classes of disfluency, the overall rates of filled pauses, repetitions and deletions were found to be higher than other types of disfluency. However, rates of repetition and repair were found to correlate with sentence-related measures such as utterance length, while the rate of filled pauses did not. When the position of a disfluency within an utterance was examined, deletions occurred more often in sentence initial positions in the Switchboard corpus, reflecting higher rates of "false starts" of utterances, while repetitions and filled pauses tended to occur more often within an utterance than at the beginning (although all three types of disfluency were more likely to occur initially when comparing the likelihood of a disfluency being associated with an initial word rather than a medial word, see footnote, above). In addition, while the filler *um* tended to occur more often initially, *uh*, which was observed more frequently, occurred more often within an utterance than at the beginning. Shriberg argued that these results provide evidence for regularities in the distribution of different types of disfluency, which could be attributable to the different types of problems speakers encounter during different parts of an utterance. Yet it should also be noted that there is still much flexibility in the production of different types of disfluency, and that the location of a disfluency, rather than its form, can often provide stronger evidence

[2]It should be noted that Shriberg's (1994) measure of disfluency likelihood was based on the number of possible sites that a disfluency could occur in that position. For sentence initial disfluencies, as they could only occur prior to the first word, the disfluency rate corresponded to the number of disfluencies observed divided by the total number of sentences. For medial disfluencies, the disfluency rate was determined by the number of medial disfluencies observed divided by the total number of words in all sentences less the number of sentences, as a disfluency could be associated with any word in a sentence. Therefore, her data on initial-medial disfluency rates do not correspond to an absolute likelihood of a disfluency occurring initially or medially, but to the likelihood of a disfluency being associated with a particular initial or medial word.

about the problem the speaker is attempting to resolve. Hesitation phenomena, such as the filler *uh* and repetitions that tend to occur within an utterance are likely to be associated with local micro-planning processes. When fillers occur at the beginning of an utterance, their production is thought be associated with macro-planning of upcoming speech[3]. However, as no relationship was observed between filled pause production and sentence-related variables, these disfluencies may not be purely associated with production processes, but may also be related to discourse and socio-linguistic factors associated with managing the dialogue.

Swerts (1998) provided further evidence that utterance initial filled pauses are associated with macro-planning processes. He investigated the relationship between disfluency and discourse structure by examining the occurrence of filled pauses at strong and weak discourse boundaries within a corpus of spontaneous Dutch monologues. Strong discourse boundaries are thought to be points at which major conceptualization and message planning occur (Chafe, 1980; Levelt, 1989). Swerts found that 67% of phrases following strong discourse boundaries (i.e., boundaries that at least 75% of coders perceived to be a paragraph transition) contained an utterance initial filled pause, compared to only 17% of phrases produced after weak discourse boundaries (i.e., that less than 75% of coders perceived as a discourse boundary). Furthermore, phrases following weak discourse boundaries were much more fluent overall (60% contained no fillers, compared to 22% of phrases following strong boundaries), and the fillers observed in the disfluent utterances tended to occur in a phrase-internal position. Swerts also found that the filler *um* occurred in an utterance initial position most of the time, while *uh*s were mostly observed within an utterance. He argued that the presence of a discourse boundary makes a hesitation more likely, even though a boundary cannot be predicted by a hesitation alone. This would be in line with the proposal that utterance initial fillers (in this case, predominantly *um*s) reflect macro-planning of an upcoming major discourse segment. However, it is also possible that the presence of an utterance initial filled

[3]Levelt (1989) made a distinction between macro-planning and micro-planning processes in speech production. During macro-planning, the speaker engages in information retrieval and inference as they decide on the basics of what information to express and how to order it. During micro-planning, the speaker shifts attention to lexicalizing the message to be expressed.

pause simply makes this boundary more salient to listeners, particularly as these fillers had longer durations and were more likely to be surrounded by silence than mid-utterance fillers, and therefore may have been more likely to be identified as a major boundary.

Discourse effects on disfluency

The relationship between disfluency and aspects of a discourse, such as the nature of the topic or the role of the speaker have also been investigated to further understand how disfluency relates to conversational structure. For example, the nature of a topic being discussed can impact the processing load associated with macroplanning of upcoming speech. In early work, Goldman-Eisler (1968) observed an increase in hesitations and silent pauses when speakers had to interpret rather than simply describe cartoons, a task that would require greater cognitive effort prior to speaking. Siegman and Pope (1966) also found that speakers produced more hesitations when describing conceptually ambiguous pictures. Schachter, Christenfeld, Ravina, and Bilous (1991) examined the number of filled pauses produced during humanities and sciences lectures and found that humanities lecturers produced more disfluencies than either social scientists or natural scientists, despite the fact that all groups of lecturers produced similar numbers of disfluencies when interviewed about general topics. Schachter et al. concluded that differences in lecture disfluency rates were due to the fact that "hard" sciences were more informationally structured and presented fewer linguistic options. They argued that the greater descriptive choice in the humanities gives rise to increased cognitive planning demands as message level representations are formulated into linguistic expressions, and hence a higher likelihood of hesitation.

Disfluency also appears to be associated with the familiarity of a referent. In a study in which participants described abstract shapes that they either had or had not described previously, Barr (2001) found that speakers were more likely to produce a filled pause when describing a new shape than one previously referred to. Moreover, descriptions of new referents were more than twice as likely to begin with an *um*

than those of old referents, while *uh*s were no more likely to begin descriptions of new referents than of old referents. Referring to something that has not been previously described takes significant additional resources for conceptualization and planning, processes associated with the production of *um*, here. While referring to an old referent may be less cognitively demanding, it would still require retrieval of a previously generated message from short-term memory. These results suggest that such a process is more often accommodated through the production of an *uh*.

The role of the speaker may also affect the production of disfluency, not least because conversational directors tend to produce longer utterances (Bortfield et al., 2001), which, as noted above, result in greater planning load and are associated with higher rates of disfluency. However, as Shriberg (1996) found no relation between the production of filled pauses and sentence length, they may not simply be due to planning load, but could also be the result of coordination between speakers (e.g., Clark, 1994). Indeed, speakers' production of disfluency appears to be related to who (or what) they are speaking to. Oviatt (1995) observed that speakers were more disfluent in dialogues with human partners than when producing monologues, however they were even less disfluent when communicating with machine partners (see also Shriberg, 1996). This may be due to different coordination requirements in different spoken contexts. For example, in a human-human dialogue situation, speakers may produce filled pauses to secure the attention of a listener (e.g., Goodwin, 1981), to hold the conversational floor (Maclay & Osgood, 1959), or to signal that they are having difficulty (Clark & Fox Tree, 2002). Such coordination may be less relevant in a monologue situation, and in human-computer speech, speakers' may be aware that overt disfluency is detrimental to effective communication, and so may shorten utterances, over-enunciate and pre-plan their speech in an attempt to limit disfluency (Wade, Shriberg, & Price, 1992). This is also supported by studies that demonstrate that disfluency is under a degree of intentional control. For example, Siegel, Lenske, and Broen (1969) demonstrated that speakers can suppress the production of disfluencies if instructed or incentivised to do so.

In an investigation of exactly how disfluency relates to situational factors, Bortfield et al. (2001) analysed a corpus of task-oriented conversations in which these factors were explicitly manipulated. They examined the relationship between disfluency rate and the difficulty of the topic domain (discussing pictures of tangrams or of children), task roles (director vs. matcher) and partner familiarity (spouse vs. stranger), as well as demographic factors such as gender and age (see below). They found a strong effect of the roles of partners, with directors exhibiting a higher rate of disfluency (and in particular filled pauses), yet there was no significant effect of partner familiarity. While increases in repetitions and restarts were largely accounted for by differences in utterance length (and therefore attributed to effects of planning), directors produced more fillers even when utterance length was controlled for.

Bortfield et al. suggested that while filler rates may not be independent of planning processes, the higher rate of fillers for directors may be related to processes of interpersonal communication. They also found higher rates of repetitions and repairs when speakers discussed tangrams rather than children, again supporting the idea that these disfluencies are predominantly associated with increased cognitive load due to greater task demands. However, filled pauses occurred more often when discussing children, an effect that was driven entirely by male speakers. This suggests that the production of fillers was not simply due to increased task demands (unless men find discussing children more difficult than discussing abstract shapes), but may have been associated with conversational coordination and possibly interpersonal meta-communication between a male director and a female matcher.

Speaker differences

As detailed above, overall rates of disfluency production and the types of disfluency produced vary widely by a speaker's conversational role and the complexity of what they say. However they also vary significantly among different individuals. For example, speakers display consistent differences in their overall rate of disfluency production, and their relative "preference" for different types disfluency. Maclay

and Osgood (1959) observed that slower speakers tended to be more disfluent overall, and found a negative correlation between an individual's speech rate and the rate of production of filled pauses, even when silent pauses were accounted for.

Shriberg (1994) found an inverse correlation between the production of filled pauses and the production of deletions and substitutions, indicating that speakers who tend to produce more filled pauses tend to produce fewer repairs, and vice versa. In addition, Shriberg (1994) observed different patterns of repetition and deletion between speakers, suggesting that some speakers tend to be "repeaters" while others tend to be "deleters". She also found that while both groups have similar overall disfluency rates, deleters tend to produce faster speech, on average, than repeaters, and suggested that this could reflect the tendency of faster speakers to "get ahead of themselves", resulting in more backtracking, while slower speakers take more time to plan what they are going to say, and so produce less overt errors, but commensurately more hesitations. In contrast with previous studies which have linked filled pause production to ongoing planning processes (such as that of Maclay & Osgood, 1959), Shriberg (1994) found no relationship between speech rate and the production of filled pauses, although there was significant variation in speakers' preferences for using *um*s and *uh*s in initial and medial utterance positions.

It should be noted that changes to an individual's speech rate also appear to affect the rate of production of different types of disfluency. In a study manipulating speech rate within subjects, Oomen and Postma (2001) observed an increase in repetitions and lexical and phonological repairs during faster speech, but found no difference in the rate of production of filled pauses. They argued that changes in an individual's speech rate can influence the likelihood of repetition and repair, primarily as a result of limitations in monitoring resources that affect the accuracy of pre-articulatory error detection during faster speech. The tendency to produce filled pauses, on the other hand, is relatively consistent regardless of how fast in individual is speaking, as these tend to reflect inter-clausal planning processes that are relatively invariant to the speed of surrounding speech.

Other inter-speaker variables that have been found to correlate with disfluency in spontaneous speech corpora include gender and age. In analyses of the Switchboard corpus, Shriberg (1994) found that men produced significantly more filled pauses than women, and suggested this may reflect different priorities among men and women for holding on to the conversational floor. Bortfield et al. (2001) also observed a significant relationship between gender and disfluency production. In their study, socio-economic status, a possible confound among Shriberg's male and female speakers, was controlled for, and gender was examined in combination with the role of the speaker. Bortfield et al. (2001) found that men produced more disfluencies, on average, than women, and that this was primarily due to increased production of filled pauses and repetitions. In addition, speaking in the role of a conversational director further increased this difference, primarily as a result of a relative increase in the production of filled pauses by men when speaking in this role. They tentatively suggested that men may be more overt in their signalling of production difficulties and collateral requests for assistance than women, although they acknowledge that any such claims require further corroboration.

Bortfield et al. (2001) also observed an effect of the age of a speaker on the rate of disfluency production, with older speakers tending to be more disfluent than younger speakers. However, this difference was only significant between speakers older or younger than 63 years of age, as age groups younger than 63 did not show any differences. In particular, older speakers tended to produce higher rates of within-phrase fillers, which may reflect lexical retrieval difficulties associated with upcoming words, consistent with findings that older speakers have more trouble retrieving words (Obler & Albert, 1984).

2.4.2 Local effects on the production of disfluency

So far, this chapter has discussed the *causes* of disfluency in relatively general terms. Many of the studies discussed in section 2.4.1 examined the distribution of disfluency in the context of entire utterances, measuring disfluency rates as a proportion

of the total number of words produced. Yet the production of disfluency is a local phenomenon: their occurrence varies throughout an utterance, as disfluencies reflect difficulties associated with immediately upcoming (or in the case of repairs, just produced) speech. Beyond associating disfluency with increased planning load, these studies do not attempt to relate the production of disfluency to problems associated with specific production processes. Linking disfluency to particular aspects of speech production requires an assessment of the momentary difficulties encountered by the speaker as they select and utter words.

Choice, uncertainty and disfluency

One facet of speech that has been closely associated with disfluency production is the choice of what to say next. This can be in terms of general uncertainty, or in other words, a lack of confidence about a statement being made: for example, Smith and Clark (1993) found that when speakers were asked to rate their confidence in answers they had made to a series of questions, they produced more disfluencies in answers to questions in which they subsequently exhibited a weaker "feeling of knowing". Yet uncertainty does not have to relate only to confidence about what to say next. It can also relate to the number of alternative options available to a speaker. Take Schachter et al.'s (1991) finding that humanities lecturers are more disfluent than science lecturers, because they have a greater range of descriptive options with which to convey their message. These studies suggest that choice and uncertainty can influence disfluency on a global level, but there is also a large body of evidence that implicates disfluency as a marker of local choice and uncertainty within speech.

Several early studies of disfluency production focused on the relationship between hesitation and the uncertainty of subsequent words. In early work, Goldman-Eisler (1958b, 1958a, 1961) proposed that hesitation pauses appear to be cognitive in origin, and characterise locations of speaker uncertainty. Using a sentence completion task to measure word predictability in which participants heard fragments of spontaneously produced sentences that contained hesitations and attempted to

predict subsequent words in the sentence, she observed that words following silent hesitations tended to be less predictable given their prior context, and took longer to replace than words uttered in fluent contexts. Further, words before silent hesitations were more predictable than normal, leading her to suggest that hesitations tend to occur at encoding choice points, which are marked by transitions between high and low contextual predictability. Tannenbaum, Williams, and Hillier (1965) extended this work, and when they included other types of hesitations, such as filled pauses, repetitions and false starts, also found that hesitations tended to precede unpredictable words.

The predictability of a word, and hence the location of hesitations, is also likely to be related to its grammatical class. Confirming this, Maclay and Osgood (1959) found that filled and unfilled pauses occurred more often before content words than function words. Filled pauses, however, tended to occur more often at phrase boundaries than unfilled pauses. They also found that speakers tended to repeat function words, but that most of these repetitions occurred immediately prior to a lexical item, suggesting that they are distributed, and function, in a similar way to pauses. Maclay and Osgood concurred with Goldman-Eisler (1958b, 1958a), suggesting that hesitations of these types tend to occur at points of highest uncertainty within an utterance, and that their production is related to the dynamics of grammatical and lexical selection. J. G. Martin and Strange (1968) also found that speakers hesitated more often before content words than function words, suggesting that the greater informational content of these words may result in greater speaker difficulty in making appropriate choices in matching words with their intentions.

Beattie and Butterworth (1979) provide further support for the idea that hesitations in speech are associated with the resolution of choice during lexicalisation. Using a judgement procedure to determine measures of contextual probability of content words, they observed a relationship between both the contextual probability of a word and its frequency with subsequent hesitations. Frequency would also be implicated by a distinction between function and content words, as informationally high content words tend to also be much more infrequent than function words

in spontaneous speech. When only low frequency words were examined, speakers produced more hesitations preceding words of low contextual probability. However, when contextual probability was held constant, they found no difference in the frequencies of words in fluent and hesitant contexts. They argued that the hesitations observed in their study reflect the resolution of a choice between semantically related lexical items that could be appropriately used given the preceding context, and that a words contextual probability is an important factor guiding the lexical selection process.

Retrieval difficulty

Beattie and Butterworth (1979) found no systematic frequency effect: According to Levelt (1983,1989, Jescheniak & Levelt, 1994), the frequency of a word is thought to affect its speed of lexical retrieval and phonological encoding. Therefore, the results from Beattie and Butterworth's study would suggest that mid-utterance hesitations are related to issues of choice and uncertainty in word selection during earlier formulatory planning and selection processes, but not necessarily due to difficulty during the retrieval of words. However, Beattie and Butterworth's (1979) frequency classification was atypical: Items deemed low frequency could occur as often as 100 times per million words, which is comparable to the high frequency conditions of many more recent studies examining frequency effects in picture naming (e.g., Griffin & Bock, 1998; Jescheniak & Levelt, 1994; Shatzman & Schiller, 2004).

Levelt (1983), however, did observe a relationship between the frequency of colour names and associated hesitations using a task in which participants described visual patterns comprising coloured circles connected by vertical and horizontal lines to elicit a corpus of naming errors and repairs. In a post-hoc analysis examining the occurrence of pre-lexical hesitations (or what he termed covert repairs) immediately preceding colour names, Levelt observed a correlation between the presence of a hesitation and the frequency of the colour name being referred to. He suggested that *uh* is a "symptom of the actuality or recency of trouble", and that such hesitations reflect trouble during word-form retrieval. However, this is hardly reliable evidence

of a relationship between word frequency and hesitation, as only 11 different colour names were used, and some colours were presented in the networks more often than others, potentially confounding general word frequency effects with local contextual frequency. A much larger corpus of words would be needed to demonstrate a genuine relationship between the frequency of a word and the production of disfluency.

To summarise, the distribution of disfluency in speech is non-random. Disfluencies display systematic general patterns of occurrence that vary by the properties of an utterance, as well as by the speaker and to whom the speech is directed. Evidence indicates that many disfluencies that occur at the beginning of an utterance are associated with macro-planning processes. While there is significant variation amongst the occurrence of different classes of disfluency, repetitions, repairs and mid-utterance fillers appear to be primarily attributed to local processes of micro-planning and error correction. However, not all disfluency can be exclusively attributed to planning, and it appears that utterance-initial fillers such as *um*s may also function as a discourse structuring device and have a role in coordination among interlocutors. Indeed, influences of planning and coordination on utterance-initial disfluency may interact, and previous research does not clearly separate these factors. This raises issues for the interpretation of utterance-initial fillers, in particular. Establishing a distinction between planning and coordination functions would be beneficial for the integration of disfluency in both models of production and dialogue, however this is beyond the scope of this thesis.

2.5 Effects of disfluency on listeners

As detailed above, disfluencies exhibit a relationship to the planning and correction of problems during speech. Moreover, a speaker's production of disfluency also appears to be associated with the effective management of a conversation with a listener. It stands to reason that, given their frequency and relationship to production difficulty, listeners have also developed ways to accommodate disfluencies during comprehension. But are listeners sensitive to the disfluencies that speakers produce?

A number of studies have shown that listeners are poor at detecting disfluencies. For example, Lickley (1995) found that not only are disfluencies inconsistently identified, but some disfluencies may be more salient to listeners than others. Participants listened to monologues that contained disfluencies and had to mark any deviations they identified between the monologue and a fluent transcription. While they correctly identified just over 50% of filled pauses, they performed much worse on other disfluencies, identifying about 40% of false starts and less than 30% of repetitions. Lickley also observed that fillers between sentences were more reliably detected than those within sentences, possibly because they occurred in a more prominent prosodic context. J. G. Martin and Strange (1968) found that listeners who were played spontaneous utterances and asked to repeat exactly what they heard reproduced remarkably few hesitations. Those that they did identify were often displaced towards clausal boundaries, suggesting that hesitations are not processed as a message-level part of speech. Furthermore, Christenfeld (1995) observed that speakers were only sensitive to the presence of *um*s when they were explicitly told to focus on the linguistic style of a speaker, but not when they were told to focus on the content of the speech, or given no instructions at all. Therefore, it appears to be difficult for listeners to consciously identify the location of disfluencies under normal listening conditions, except when the task demands that they attend to them.

However, listeners also tend to forget the surface forms of sentences almost immediately upon hearing them (Jarvella, 1971), so it is unclear whether these observations are directly associated to the memorability of disfluencies themselves, or due to listeners' tendency to discard disfluencies alongside other structural information once the message has been processed. In any case, being unable to accurately remember the location or presence of a disfluency does not mean that the listener was unaware of it at the moment it was encountered, or that the disfluency had no impact on the listener's processing or representation of speech at that time.

Indeed, there is evidence to suggest that disfluencies influence listeners' perceptions of speech, whether they are consciously identified or not. For example, in

Christenfeld's (1995) study, a disfluent passage of speech was rated as being less eloquent than one in which the disfluencies were excised, but more relaxed than one where they were replaced by silent pauses, despite the fact that listeners' estimates of the number of disfluencies were similar for all three passages. Other studies have demonstrated that disfluencies can also influence listeners' perceptions about a speaker. Fox Tree (2002) found that listeners judge disfluent speakers as having more production difficulty, being less comfortable about their topic, and exhibiting less honesty about what they say. Additionally, Brennan and Williams (1995) found that listeners rated speakers as being less confident about their answers when the answers were preceded by silence or a filled pause, and suggested that listeners may use disfluencies as a source of collateral evidence about the mental state of the speaker. However, such studies rely on post-hoc interpretations of listener judgements, and importantly, do not address whether disfluencies directly affect a listener's processes of comprehension and message representation at the time that they are encountered.

2.5.1 Effects of disfluency on the comprehension of subsequent words

One commonly held view of the role of disfluency in comprehension essentially treats disfluencies as noise (Brennan & Schober, 2001; Bailey & Ferreira, 2003). This view assumes that disfluencies would have no effect on a listener's processing of speech as they are filtered out prior to comprehension, possibly because they are not recognised as linguistically valid input (Lickley, 1996). Yet, disfluencies interrupt the flow of speech, introduce delay and prosodic discontinuities (Plauché & Shriberg, 1999; Shriberg, 1994), and may result in local ungrammaticality which can hamper syntactic and semantic processing, particularly in the case of repairs (Levelt, 1989). Therefore, it is possible that disfluencies impede comprehension, as the resolution of disfluent speech places additional processing demands on the comprehension system.

To investigate the disruptive effects of repetitions and repairs, Fox Tree (1995) performed a word monitoring study in which she examined participants' response

times to identify target words when they occurred following a repetition or after the interruption point of a repair, compared to when the reparandum was removed or replaced by silence. All of the repairs were false starts (i.e., deletions): the initial part of the utterance was interrupted and the target word started a completely new utterance. While no differences in recognition time were observed between items that followed repetitions and those presented in fluent contexts, she found that recognition times were longer when the target word followed an interruption as part of a repair. Fox Tree's results suggest that the processing of repairs incurs an integration cost in terms of syntactic and semantic processing, and that more cognitive resources may be required to hold the reparandum in memory while processing the repair. The lack of an effect on repetitions indicates that this effect is not purely due to interruption or to a syntactic violation, unless syntactic re-evaluation and integration is more taxing for repairs than it is for repetitions.

Although repairs can impede the integration of subsequent words into an utterance, disfluency is also thought to have a facilitatory effect on language comprehension. Listeners appear to be sensitive to, and to take advantage of, prosodic cues to detect both repairs and hesitations at the point of interruption (Lickley, 1994; Lickley & Bard, 1998). By alerting listeners to an upcoming hesitation or speech interruption, such cues may alleviate integration costs by helping listeners avoid prematurely treating a repair as a continuation of fluent speech.

Brennan and Schober (2001) demonstrated that disuency may facilitate comprehension in a study that investigated how repairs and the filler *uh* influence the speed of processing of a subsequent target word. They used an on-line referential communication task in which participants followed instructions to select an item from a set of geometric shapes that differed by colour. Response times and accuracy were used as measures of the ease of processing of the disambiguating target word. Brennan and Schober found that listeners were slower to respond to the target word in a fluent utterance (e.g., move to the *purple* square) than when it occurred immediately following the interruption point of a repair (e.g., move to the *yellow- purple* square). Further, response times were faster following mid-word interruptions that

contained the filler *uh* at the interruption point (e.g., *yell- uh purple*) than either between-word interruptions (e.g., *yellow- purple*), or mid-word interruptions without a filler (e.g., *yell- purple*). They also observed that error rates for instructions containing a filler were no different than for fluent controls, and were significantly lower than those observed for repairs containing both whole and mid-word interruptions (higher error rates would be expected following a repair, as listeners must resolve the processing of referents in both the reparandum and repair).

These results could be accounted for if the presence of an *uh* provides additional information to the listener that facilitates the rejection of the previous word and processing of the repair. However, no difference in either response times or error rates was found when the filler *uh* was replaced with a silence of equal length. This indicates that the additional time that the pause affords the listener, rather than the phonological form of the filler, facilitates processing of the repair. In utterances without a filled or silent pause, Brennan and Schober suggested that listeners may be sensitive to the contrastive stress of the repair word as a signal that what came before the interruption was erroneous, yielding faster response times to the stressed word. However, it should also be noted that because there were only two referents, the interruption of naming of the first item immediately signalled that the second shape was the intended target, allowing earlier preparation of a response. When three shapes were used, the benefits associated with different forms of repair decreased but were not eliminated, presumably as they still discounted one of the potential options. Therefore, it appears that listeners were using information in the repair to eliminate alternative referents, which is helpful in this particular task, but may be less useful when repairs are encountered in spontaneous speech where there is potentially an unlimited number of alternatives.

Other studies have suggested that some types of filler provide a benefit to comprehension that is not purely attributable to additional processing time. Fox Tree (2001) evaluated whether filled pauses facilitate the incorporation of subsequent words into the ongoing message-level representation using a word monitoring task,

in which listeners pressed a button upon hearing a target word. The speed of target response was considered to be an index of the ease of syntactic and semantic integration with prior speech (Fox Tree, 1995; Marslen-Wilson & Tyler, 1980). Listeners were faster to identify a target word when it was preceded by an *uh*, but not an *um*, compared to utterances in which the disfluency was digitally removed. If this benefit was purely due to additional processing time, it should be observed for both types of filled pause. Fox Tree (2001) proposed that the different results for *um*s and *uh*s arise because they signal differing lengths of upcoming delay (Clark & Fox Tree, 2002; Smith & Clark, 1993). She suggested that *uh*s afford a processing benefit by focusing attention on the subsequent word, which would be beneficial to a listener when the delay prior to the upcoming word of interest is anticipated to be short. When the length of subsequent delay is longer, or potentially indeterminate, as may occur after an *um*, immediate focusing of attention provides less of a benefit.

Fox Tree's (2001) study compared disfluent speech to control utterances in which the filler had been excised, but a significant silent pause (on average, 704ms) remained. This does not represent a fluent utterance as the silence is long enough to be interpreted as a disfluent signal. However, it may also be interpreted as an extended discontinuity that hampers subsequent processing. In this case, an extended silence or an *um* might slow down processing of the following word, as they reflect a higher likelihood of discontinuity and therefore less need to attend to further speech, while *uh* signals likely continuation, similar to in a fluent utterance. The inclusion of an additional fluent control condition in which all silence was excised, and therefore continuity was assumed, could shed light on whether the effect observed is the result of facilitation or delay of processing of the following word.

These studies demonstrate observable effects of hesitations on comprehension, and suggest that filled pauses can facilitate the processing of surrounding speech, either (in the case of *uh*s) by focusing attention on a subsequent word, or by providing additional time for the processing of a repair. While prosodic features of a repair may also facilitate the identification and abandonment of the reparandum, it

appears that repairs in general have a negative impact on the processing of subsequent speech, as they increase the cognitive load associated with correctly parsing a speaker's utterance. However, both studies have limitations in their scope of interpretation, not least because neither task (word monitoring or button pressing) provides a close analogue to natural comprehension processes.

2.5.2 Disfluency and reference resolution

While these studies provide some evidence that filled and silent pauses may facilitate the processing of upcoming words by focusing attentional processes, there is stronger evidence that disfluencies, and in particular fillers, can influence listeners' attentional responses to different types of referents. In one paradigm, Barr (2001, see 2.4.1, above) presented listeners with two abstract shapes on a screen, one of which had been previously described. They were played a description of a shape which either did or did not include an utterance-initial *um*, and they had to click on the described object as quickly as possible. Participants were reliably faster to click on a new referent when its description was preceded by an *um*. Additionally, mouse movements toward the new picture were often initiated during the *um*, before the actually description began. Barr argued that listeners readily exploit disfluent signals to enhance linguistic and conceptual coordination, and suggested that speakers may be sensitive to different interpretations depending on the kind of trouble the speaker is perceived to be dealing with.

Arnold et al. (2004) used a visual world paradigm to perform a related experiment investigating the effects of disfluency on listeners' judgements about the discourse status of referents. In this study, participants' eye movements were monitored as they followed verbal instructions to manipulate four objects on a screen. Eye fixations on objects are thought to reflect lexical access and can therefore be used to track the time course of continuous speech processing (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Within the matrix of objects, two picture names shared an onset phoneme (e.g., candle and camel). One of these items was a new referent, while the other had been mentioned in the previous trial. When an

instruction contained a disfluency (e.g., *thee uh...*), participants were more likely to glance at the discourse-new referent prior to the point of disambiguation between names. Conversely, when the instruction was fluent, they were more likely to glance at the discourse-given object, suggesting that listeners were exploiting disfluent signals to make predictive judgements about upcoming referents. However, it is possible that listeners are simply very sensitive to the general distributions of disfluency in spontaneous speech, and that this information only affects comprehension processes in a reflexive way.

In a second study, Arnold et al. (2007) used a similar visual world paradigm, but instead manipulated the familiarity of presented objects. Some objects were familiar, concrete pictures (e.g., an ice cream cone), while others were unfamiliar, abstract shapes (e.g., a squiggly shape). Here, they found that a disfluency (e.g., *thee uh...* prior to the object description resulted in more looks towards the unfamiliar object, biasing listeners' interpretations about an initially ambiguous referential expression towards the unfamiliar object. However, in this instance fluent expressions did not bias towards either familiar or unfamiliar referents. Arnold et al. (2007) proposed that this bias toward the unfamiliar item is due to listeners' ability to rapidly interpret the information that disfluencies imply about the nature of the speaker's underlying production difficulty: If the speaker is being disfluent, it is likely that they are trying to refer to something hard to access or describe. Indeed, when listeners' expectations about the source of the difficulty were modified (by informing them that the speaker had object agnosia), they made no reflexive predictions about the object being referred to (as measured by eye fixations). This implies that listeners can modulate the assumptions by which they determine their expectations about the causes of disfluency, demonstrating more than just a simple predictive association between disfluency and hard-to-name objects.

These results provide further evidence that disfluencies can cause listeners to update their expectations about upcoming referents in real time, often before they acquire explicit information enabling identification. Considering that there is increasing evidence that listeners make online predictions about a wide variety of syntactic

and thematic information (e.g., Altmann & Kamide, 1999; Kamide, Scheepers, & Altmann, 2003; Pickering & Garrod, 2007), Arnold and Barr's studies provide strong evidence that listeners' inferences about the causes of disfluency allow them to rapidly and dynamically modify their predictions about upcoming speech and the state of the speaker's production system. However, in all of these studies participants must make a choice from a restricted set of images, and influences on predictive processes may be due to the availability of clearly defined alternative referents. Natural language does not impose such constraints, as at any time a speaker could use one of any number of syntactically coherent words that are more or less expected given the context. This raises the question of whether disfluency affects prediction only when under such representational constraints, and whether evidence for these effects can be observed in wider domains of natural comprehension.

Recently, Corley and colleagues (Corley et al., 2007; Collard, Corley, MacGregor, & Donaldson, 2008) used ERP (event related potential) methodologies to investigate whether integration costs or surprise effects could be affected by the inclusion of a filled pause prior to the target word. ERP studies provide an advantage over experimental methodologies such as those described above, as participants are only required to listen to speech and do not have to perform any secondary tasks. As a result, comprehension can be directly evaluated in a natural context, and effects can be determined without limiting the set of potential referents. Performance is measured in terms of a relative negativity or positivity across the scalp for components of the ERP waveform that are associated with different language comprehension processes (Kutas & Hillyard, 1980). Corley et al. (2007) had participants listen to sentences that contained either a predictable or unpredictable word, given their context. They found that the N400, a negativity associated with the integration of unexpected or incongruent words (Kutas & Hillyard, 1984), was significantly attenuated when an unpredictable word was preceded by the filler *er* (equivalent in British English to *uh*), indicating that the presence of *er* reduced the processing cost associated with integrating an unpredictable word into its context. In Collard et al.'s (2008) study, predictable words were altered so that their acoustic characteristics were incongruous with the preceding speech. This manipulation caused

an increase in the P300, a positivity associated with the orienting of attention and updating of memory (Polich, 2004), as well as a change in the MMN (mismatch negativity) associated with the detection of change (Alho, 1995). When incongruous words were preceded by an *er*, they found no effect on the MMN, but observed a reduction in the P300. They argued that this was due to attention orienting effects of the *er*, which had already directed attention to the upcoming word. Importantly, in both studies, words that had previously been preceded by disfluency were remembered better by participants in post-experiment recognition tests, suggesting that not only did the presence of a filler affect processing of the subsequent word, it also affected representational encoding. These experiments establish not only that disfluencies have effects in more "naturalistic" settings where participants do not have to respond to instructions, but also that these effects have longer-lasting consequences for the representation of what the speaker has just said.

2.6 Is disfluency a signal of production difficulty?

The studies detailed above demonstrate that disfluencies affect listeners' perceptions of a speaker, as well as their processing and representation of speech. However, an important question with regard to the role of disfluency in natural communication is whether listeners are simply efficient at extracting information from the speech stream and integrating it into their perceptions and predictions about upcoming events, or whether they are responding to explicit (or implicit) signals produced by the speaker to convey meta-linguistic information about the state of their production system. In other words, are disfluencies intentional signals from the speaker of upcoming delay, or are they just symptoms of underlying production difficulty that listeners are able to exploit?

So far, we have primarily been discussing disfluencies as epiphenomena, essentially as by-products of problems in language production. However, there is also a long held view that disfluencies are intentional signals that are planned by speakers to accommodate production problems in speaking. Maclay and Osgood (1959) originally suggested that hesitations, such as filled pauses, are used by speakers to "hold

the conversational floor" while they resolve syntactic and lexical formulation difficulties, implying that their production is at least somewhat intentional. Clark and colleagues (Clark, 1994, 1996; Clark & Wilkes-Gibbs, 1986; Clark & Wasow, 1998; Clark & Fox Tree, 2002; Fox Tree & Clark, 1997; Smith & Clark, 1993) developed a more complete theory, and argued that disfluencies are intentional acts within their "strategic modelling view" of collaborative dialogue. This theory proposes that speakers employ intentional tactics to manage a conversation, constantly checking their speech against a model of what the listener might know, while formulating an utterance for the listener (Clark, 1994; Clark & Wilkes-Gibbs, 1986). In other words, speakers actively coordinate with listeners and provide feedback about their own production processes in order to maintain a successful dialogue.

Clark (1994, 1996) argued that disfluencies are an integral part of this process, and that speakers manage problems in conversation by using them as a "collateral signal" of current or upcoming production difficulty that may involve some delay before the resumption of fluent speech. Filled pauses may be used, for example, as a device to hold the conversational floor, informing listeners of the speaker's trouble, but indicating that they will resume speaking shortly and should not be interrupted. In support of this interpretation, Smith and Clark (1993, see also Clark and Fox Tree, 2002) observed that the filler *um* was more commonly associated with a subsequent pause than the filler *uh*, and that pauses following *um*s were significantly longer than those following *uh*s. They argued that speakers strategically use *um* and *uh* to signal different things to a listener: *uh* to signal a minor problem, and hence a short upcoming delay in speaking, and *um* to signal a longer delay. Clark and Fox Tree (2002) argued that the fillers *um* and *uh* are used by speakers as conventional words, and while they do not add propositional content to an utterance, they convey meta-linguistic information in a similar way to other interjections, such as "oh" (which is used as an expression of surprise). Their arguments have also been applied to the production of other forms of hesitation, such as repetitions (Clark & Wasow, 1998) and prolongations (Fox Tree & Clark, 1997), arguing that these disfluencies are also planned by speakers to convey information about, and to comment on, the state of their production processes.

Alternatively, disfluencies may not be intentional signals, but merely symptoms of underlying production difficulty. According to this view, disfluencies are not under intentional control, but are an involuntary consequence of encountering problems during speaking (Levelt, 1989). For example, in the case of repairs (which Levelt, 1983, 1989, considers to include fillers and repetitions as forms of covert repair), speech is interrupted to resolve an error that has been identified through either internal or external monitoring of speech, not to signal to a listener some kind of production difficulty. Any information that a listener extracts from a hesitation or repair is a result of their own interpretive inferences, rather than explicit signals on behalf of the speaker. This perspective on the role of disfluency is similar to the "cognitive burden" view of collaborative dialogue, which argues that a disfluency is considered to be an unintentional sign of cognitive difficulty, and that modelling a listener's perspective represents an unnecessary additional cognitive cost, given the already high processing demands on the speaker during dialogue (Keysar, Barr, Balin, & Brauner, 2000; Horton & Keysar, 1996).

Clark's (1994, 1996) arguments for disfluency as a signal are appealing, as they provide an attributional explanation that corroborates with our own conscious experience of disfluency in everyday speech. Most listeners, when asked, tend to attribute disfluency to problems in speaking (Fox Tree, 2002), and while many hesitations are related to planning load, the occurrence of others, most notably utterance initial fillers, remains invariant relative to changes that affect the cognitive burden of planning, such as utterance length (Shriberg, 1996), suggesting they may have an alternative role. Additionally, disfluency rates have also been linked with a number of factors related to managing the conversation rather than to purely cognitive processes (e.g., Bortfield et al., 2001). However, if the intentional production of disfluency is an act of communication between speaker and listener, this cannot fully explain disfluencies that occur in monologues, or when speakers are interacting with a computer (Oviatt, 1995). Despite reduced rates in these situations, disfluency is not completely eliminated, even when speakers are under no pressure and are in a situation in which disfluencies will not aid the perceived listener. It could be argued that in these situations speakers still have an audience in mind,

or that speakers' use of disfluency in these settings is a continuation of habits developed during natural conversation, however it is unlikely that all hesitations are intentionally produced solely to signal upcoming delay.

While Clark and Fox Tree (2002) observed a relationship between the type of filled pause and the presence and length of upcoming delay, they acknowledge that they were actually measuring the *perception* of pause length by trained coders. When O'Connell and Kowal (2005) analysed a corpus of spontaneous speech consisting of a series of interviews with Hillary Clinton, they found that in most cases the fillers she produced were not followed by silent pauses. On average, only 40% of *um*s and 20% of *uh*s were followed by a silent pause. They argued that if filled pauses do not reliably predict a pause in speaking, then it makes no sense to produce them (or as a listener, to treat them) as a signal of upcoming delay. Additionally, O'Connell and Kowal found that when silent pauses occurred following fillers, their average duration was only 110ms longer following *um*s than *uh*s (440ms vs. 330ms), which they considered to be insufficient to distinguish between a "minor" and "major" delay in speaking.

What is clear is that disfluencies do have a communicative role in conveying information about the state of the speaker's production system to a listener that helps interlocutors manage an ongoing dialogue. Filled pauses, for example, may well be an attempt to maintain fluency, by maintaining vocalisation that marks a speaker's intent to continue with their speech act. However, evidence to support Clark and Fox Tree's (2002) argument that disfluencies possess metalinguistic and semantic content as signals of production difficulty, and speakers *use* them in a similar way as they use other words, remains circumstantial at best. The fact that listeners are adept at making inferences based on the presence of disfluency does not imbue them with communicative intent on behalf of the speaker. Therefore it is prudent to maintain scepticism about the intentionality of disfluency production, and focus on what is observable, namely how disfluency relates to the nature of the processes that underlie their production.

2.7 Conclusion

Disfluencies are an integral part of the way we speak, which affect listeners in predictable ways. There is considerable evidence concerning the circumstances in which speakers are likely to be disfluent. However, many of the causes attributed to disfluency are at the level of discourse, or derived from post-hoc reasoning about recorded speech. Questions remain as to the local causes of disfluency and the nature of the processes that underlie their production. In the following chapter, we introduce a process-based investigation into the production of disfluency. First, we outline a model of speech production, and then consider how such a model might give rise to disfluency. We then introduce the experimental methodology, based on the Network Task, that will be used throughout the thesis.

Chapter 3

Disfluency and Models of Speech Production

Most models of speech production (e.g., Dell, 1986; Caramazza, 1997; Garrett, 1980; Levelt, 1989) agree that the process of speaking involves three broad mechanisms. Speaking starts with conceptualisation of a pre-verbal message (i.e., planning an utterance's message-level meaning), which is followed by the formulation of this message into a linguistic structure. This linguistic plan is then passed onto the articulatory system, resulting in spoken output. This process of speech production is also thought to be incremental in nature: While an initial message plan may be formulated in advance of the initiation of speaking, the lexical representations that are activated by this plan can be accessed as they are required during the course of production, resulting in the incremental retrieval and encoding of the phonological form of an utterance, which is then buffered for articulation (Deese, 1984; Griffin & Bock, 2000; Griffin, 2001; Kempen & Hoenkamp, 1987; Levelt, 1989). There are many potential sources of difficulty in this process. For example, the speaker may struggle to successfully plan the message they intend to convey, or may have to formulate a complex syntactic structure. They may produce a speech plan which contains errors identified by the internal or external monitors, which require effort to repair (e.g., Levelt, 1983). Or speakers may encounter unexpected delays during the selection and retrieval of lexical representations that correspond to their intended message. In each case, the difficulties may result in a disfluency—filled pauses, silent pauses, prolongations, repetitions, or, in some cases, repairs.

Although disfluencies can reflect a speaker's accommodation of a particular production problem, by their nature they are opaque, in that the production of a particular disfluency does not by itself tell us much about what kind of problem the speaker was attempting to resolve (Clark & Fox Tree, 2002). And yet, the idea that different types of hesitations index different kinds of encoding decisions has long been of interest to disfluency researchers (e.g., Tannenbaum et al., 1965). While studies assessing the distribution of different types of disfluency provide some insight into whether some hesitations may be associated with particular types of production difficulty, there is little research to suggest why this may be the case. The focus of this thesis is therefore on how disfluency varies with factors that are known to affect the ease of selection and retrieval of words during speech, in order to develop a more complete picture of how different types of disfluency relate to local underlying production difficulty.

3.1 Models of Lexical Access

Any theory of the production of words must be able to provide an account of how the production system proceeds from the activation of a concept to the retrieval of the appropriate lexical lexical representations that specify the syntactic and phonological information required to express that concept and integrate into the syntactic frame of an utterance. This process can broadly be broken down into two stages: those of lexical selection and word-form encoding. During lexical selection, multiple lexical concepts that specify the semantic properties of different words will be activated by the message-level plan, and the most appropriate lexical concept will receive the highest level of activation, given a variety of contextual, dialectical and conceptual constraints. This activated lexical concept passes activation on to its corresponding lemma representation, which contains the syntactic and semantic information required to encode and integrate that word into the ongoing utterance. Following the successful selection of the lemma and integration into the ongoing syntactic frame, the corresponding phonological form is retrieved in order to be buffered for articulation. Current models of language production (such as

those of Caramazza, 1997; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Levelt, Roelofs, & Meyer, 1999) make differing claims about the structure of lexical processing, and there is ongoing debate about both the number of processing levels and the directionality and discreteness of information flow during lexical access. The experiments we present in this thesis are based on the model of Levelt et al. (1999), which makes clear claims about both the structure of lexical representations and the locus of effect and temporal consequences of factors thought to influence lexicalisation.

The model of Levelt et al. (1999; see also Roelofs, 1992) proposes that lexical selection occurs through a process of competition between related lemmas, as activation of the to-be expressed lexical concept will also spread to semantically related concepts, which in turn provide activation to their corresponding lemma representations. These lemmas then compete for selection through a non-inhibitory process until one candidate is selected for retrieval (but see Peterson & Savoy, 1998, for evidence suggesting that activation cascades from semantic to phonological levels for non-selected as well as selected lemmas). This model of lexical selection has been motivated by the results of picture-word interference (PWI) studies, which have shown an increase in picture naming latencies when they are presented with an embedded semantically related distractor word, compared to an unrelated control (e.g., Schriefers, Meyer, & Levelt, 1990). Differential interference occurs because semantically related distractor words receive activation from the presented word as well as spreading activation from the target concept, while an unrelated distractor will only receive activation from the presented word.

According to Levelt et al. (1999), lexical retrieval comprises two distinct processes, lemma retrieval and word-from encoding. Once a lemma is selected, it is retrieved from memory, making available the syntactic properties of that word required to integrate it into the syntactic frame of an utterance. Subsequent to this, the lemma activates its corresponding lexeme, which contains the morpho-phonological properties of a word required to construct a phonetic and articulatory program.

3.2 Factors influencing lexical access in language production

In line with such a model, it is likely that a concept's name agreement and the frequency in the language of the word used to represent a concept influence different stages of lexical access.

Name agreement, or codability, as it has also been termed by some authors, reflects the extent to which a concept (or a picture in the picture naming literature) can be ascribed one or more different valid alternative names. While codability has generally been used to refer to the consistency of the names that different individuals provide for visual stimuli, (i.e., pictures of objects; see Lachman, 1973), the term name agreement has been used to refer to the variability in names associated with both *pictures* and their underlying *concepts*. Low name agreement has been shown to result in longer picture naming times (Lachman, 1973; Lachman, Shaffer, & Hennrikus, 1974; Snodgrass & Vanderwart, 1980; Vitkovitch & Tyrrell, 1995), and the locus of this effect is thought to occur during the stage of lexical selection. Naming an object that has multiple possible names is not unlike naming objects in the presence of a semantically related distractor word. According to Levelt et al.'s (1999) model, expression of a low name agreement concept such as *couch* will not only activate the lexical concept *couch*, but also the synonymous lexical concept *sofa*, as well as activating semantically related lexical concepts such as *chair*, and *table*. These activated lexical concepts compete for selection. As the lexical representations for *couch* and *sofa* receive direct activation from the activated concept node, they will have higher activation levels relative to other semantic neighbours, but similar levels of activation relative to each other, and as a result, competition between *couch* and *sofa* may take longer to resolve before a winning lexical item is selected. Alternatively, an underspecified concept may activate several related lexical concepts that, while not synonymous, are sufficiently similar as to be able to adequately represent the intended concept, such as *trophy, cup* or *chalice*. Associatively related items that are often the subject of exchange errors, such as *nut* and *bolt*, may also compete for selection to a sufficient degree to affect naming latencies (Rahman & Melinger, 2007). Even if an individual speaker is heavily biased towards

using one word over another for a particular concept, the fact that an alternative word in their lexicon is directly activated by the same concept suggests that this lexical item will provide stronger competition than activated semantic neighbours, resulting in longer times required for lexical selection and hence longer naming latencies. Lachman's (1973) study supports this interpretation. He found strong effects of name agreement on picture naming latencies, with low name agreement pictures taking as much as 600ms longer to name than pictures of high codability. While such low name agreement items often have low frequency and late-acquired names, in a subsequent multiple regression analysis he found that name agreement effects persisted over and above any effects of frequency or age-of-acquisition.

This also raises the question of whether a low name agreement lexical concept with one strong competitor, such as *couch* and the competitor *sofa* would be expected to produce more competition, and hence result in a longer delay before the resolution of lexical selection, than a lexical concept with multiple weak competitors (such as *boat*, with the competitors *ship*, *sailboat*, *schooner*, *yacht* etc.). On first examination it would make sense that a single strong competitor would provide greater competition. According to Roelofs's (1992) model of lexical selection, activation of lexical concepts occurs through a process of spreading activation: as all activated lexical concepts receive activation from the concept node, this activation is spread by each lexical concept to all of its neighbours. Furthermore, this model specifies that the probability of selection of a target word at a given time step is determined by the activation level of the target word divided by the sum of activation of all words in the system. In the case of a single strong competitor, this competitor will spread activation only to the target lexical concept further increasing its activation state. In the case of multiple weak competitors, all competing lexical concepts would spread activation to each other, as well as the target concept, potentially resulting in a greater amount of total activation within the system, and hence a lower probability of selection of the target word at any given time. This process is reflected in measures of naming uncertainty, such as the H Statistic, which provide a measure of name agreement that is based the use of dominant and subordinate competitors within a population (for further details, see 6.1).

While the effects of name agreement are closely tied to the process of lexical selection, the influence of a word's frequency has been argued to be closely tied to the process of word-form retrieval. In a series of experiments, Jescheniak and Levelt (1994) examined the locus of the word frequency effect in speech production, which has come to be seen as a signature effect of lexical access (Almeida, Knobel, Finkbeiner, & Caramazza, 2007; Griffin & Bock, 1998; Jescheniak & Levelt, 1994; Levelt et al., 1999; Wingfield, 1968). In addition to replicating a consistent frequency effect over multiple repetitions in picture naming, they established that this effect was lexical in nature, ruling out object identification or initiation of articulation as possible loci of effect. In a gender decision task, an initial frequency effect was observed which disappeared after multiple repetitions, in contrast to its persistence during picture naming. Jescheniak and Levelt (1994) argued that upon initial presentation for gender decision both the word's lemma and lexeme were accessed, but on subsequent presentations only the lemma needed to be accessed to determine the gender, and the cessation of the frequency effects over multiple gender decision trials suggests that the locus of the frequency effect is not associated with lemma retrieval. Crucially, in a further experiment, similar naming latencies were observed for low frequency homophones (e.g., *nun*) and for words matched to the cumulative frequency of both the high and low frequency homophones (i.e., the cumulative frequency of *nun* and *none*). If, as they argue, homophones with separate lemmas share a common lexeme, this result suggests that it is the frequency of this shared lexeme that impacts naming latencies.

3.3 Incrementality and sentence production

Models of lexical access such as that of Levelt et al. (1999) provide detailed predictions about the time course and underlying processes of selection and retrieval of words in isolation. But what happens when words are produced as part of an ongoing utterance? The process of sentence production can be broadly conceived as starting with conceptualisation of a pre-verbal message (i.e., planning an utterance's overall meaning and purpose), which is then passed on to the formulator,

which translates this message into a linguistic structure. This process of grammatical encoding involves the selection of appropriate lemmas, which contain syntactic information that enables the creation of a grammatical surface structure. Phonological representations for each word within this structure are then retrieved in order to create a phonetic plan that can be buffered for articulation.

According to some accounts of sentence production, speakers plan and select all the words in an utterance before they start speaking, and only retrieve the phonological forms of words after speech has been initiated (e.g., Garrett, 1975; Goldman-Eisler, 1968). Such an account would suggest that most linguistic processing would be complete before the initiation of articulation, and hence disfluencies would either arise when preparing the content and structure of an utterance, or would be due to errors retrieving individual words' phonological forms. However, more recent accounts of the process of sentence production have suggested that both grammatical encoding and the selection of words occur incrementally, i.e., during the course of speaking words, are not specified until immediately before they are produced (Kempen & Hoenkamp, 1987; Levelt, 1989). Such accounts would also imply that as grammatical and lexical structures are generated and selected "on the fly", hesitations and disfluencies could result from any processing delays related to the generation of the linguistic output that may be encountered while speaking.

One important source of evidence for this incremental account of sentence production comes from studies that have used eye movements when describing visual scenes as a measure of linguistic planning processes. Meyer, Sleiderink, and Levelt (1998) first used this paradigm to determine whether the length of gaze durations on objects were related to lexical processing of those objects' names. In their study, speakers were visually presented object pairs that varied in terms of the frequency of the object name and the completeness of the picture, while monitoring their gaze durations on each object. When speakers were instructed to name the objects, the completeness of objects and their name frequencies were found to affect both naming latencies and gaze durations to objects. However, when speakers only had to categorize objects, frequency effects on gaze durations disappeared. They argued

that gaze durations of speakers when naming visually presented objects reflect linguistic processing, and that speakers shift their gaze from one object to another only when they have retrieved the phonological form of the object's name. This methodology and conclusion was extended to examine processes of incrementality of production by Griffin (2001), who suggested that there is a high level of synchronisation between speakers' gazes and production processes when describing visual scenes. Griffin presented speakers with displays of three different objects and had them describe their arrangement using sentences of the form "The *A* and the *B* are above the *C*", while monitoring their eye movements. She then manipulated both the name frequency and codability of the objects *B* and *C*, with the premise that if speakers select the names of all objects before the initiation of speaking, the frequency and codability of B and C should affect when speakers begin the utterance, whereas if speakers select and retrieve each name incrementally, gaze durations would reflect the time taken to process each name, but there should be no effects of variations in naming difficulty on the initiation of speaking. She found no difference in speech onset times between different conditions of items B and C, indicating that speakers only prepared the name of object A before beginning speaking. Furthermore, gaze durations during the course of speaking indicated that lexical processing of each object's name was restricted to immediately prior to producing that name. Speakers gazed longer at objects with low frequency and low codability objects immediately prior to naming them in their utterance, but changes in frequency and codability of objects B and C did not affect the length of gazes to object A, suggesting that the names of object B and C were not selected before speakers initiated articulation, but rather were selected incrementally as required in the production of the utterance.

While this provides evidence of incremental lexical processing during the course of speech, it is not to say that all sentential processing is strictly incremental. Indeed, speakers may prepare words further in advance when they are asked to or when the task demands it (Griffin & Bock, 2000; Ferreira & Swets, 2002), and it is likely that the degree of forward planning and preparation is under strategic control. Yet speakers are more likely to adopt incremental production strategies in situations

where they are under pressure to plan and formulate speech. In such instances hesitations are likely to result when speakers run out of preparation time, and as such are likely to reflect delays encountered at the moment of hesitation.

In chapters 5 and 6, we present six experiments designed to investigate the relationship between difficulty in naming a picture and the production of different types of local disfluency. In the following chapter, we introduce the methodological considerations, before presenting the experimental method to be used throughout the thesis.

Chapter 4

Design and Methodology

This chapter presents the methodology chosen to address those questions. Because our emphasis is on the lexical factors that may underlie the production of disfluency, instead of focusing on corpus methods which have often been used in previous research, we used the Network Task, an experimental approach which allows close control over the words that are likely to be uttered in spontaneous, unplanned speech, and therefore experimental control over the lexical factors that may result in disfluency.

4.1 Selection of Method

Linking disfluency to particular aspects of speech production requires us to be able to assess the momentary difficulties encountered by speakers as they select and utter words. Many previous studies of disfluency have utilised corpora of spontaneous speech to examine how different cognitive, linguistic and discourse factors influence the production of disfluency. Although corpora allow us to investigate naturally-occurring speech, they pose difficulties for interpretation, because the antecedent conditions of what is said can be hard to determine. In this chapter, I briefly review some of these difficulties before introducing the main experimental paradigm used in the present thesis, the Network Task. In this task the naming of pictures is performed within the context of an elicited speech paradigm, allowing the experimenter

to exert more control over the conditions in which participants produce fluent or disfluent speech.

4.1.1 Corpus studies of disfluency production

Previous studies of disfluency production have primarily relied upon the analysis of large corpora of spontaneous speech, and broadly fall into two groups. The first of these are based on corpora that consist of large sets of open domain conversations recorded between two or more participants, such as the Switchboard corpus (Godfrey, Holliman, & McDaniel, 1992) or the London Lund corpus (Svartvik & Quirk, 1980). These corpora include conversations about a wide variety of topics, and place little restriction on the content of participants' discussions. Additionally, these corpora have not generally been coded with disfluency research in mind, and so some disfluencies (such as prolongations, silent pauses or word fragments) are inconsistently coded. Studies utilising speech corpora have focused on factors such as utterance length, clausal complexity, position within a sentence frame or clause, their co-location with pauses and other disfluencies, and contextual probability (e.g., Bell et al., 2003; Clark & Wasow, 1998; Clark & Fox Tree, 2002; Fox Tree & Clark, 1997; Shriberg, 1994, 1996). Other studies are based on smaller and more restricted corpora generated specifically for a particular study, and often consist of recordings in which participants (usually in pairs) perform different tasks, such as discussing familiar or abstract objects or pictures (e.g., Barr, 2001; Bortfield et al., 2001), arguing or defending a point of view (Beattie & Butterworth, 1979; Butterworth, 1975), or performing task-based conversations such as arranging a car rental (e.g., Oviatt, 1995). These task-based corpora have been used to investigate how disfluency varies with a variety of factors including partner familiarity, conversational role, topic familiarity or constraint, concreteness, and conceptual ambiguity.

The analysis of spontaneous speech corpora can provide insight into the relationship between discourse factors and disfluency, and how disfluency relates to the surrounding speech, yet they provide very little restriction over exactly what speakers have to say. The findings of these studies often do not result from explicit experimental

manipulations, but tend to be correlational in nature, and require a degree of *a posteriori* interpretation. As a result, corpus-based studies struggle to provide direct, causative evidence about how disfluency is affected by underlying speech planning and production processes.

However, corpus-based studies do highlight issues concerning the natural elicitation of disfluency that are important for any experimental paradigm to consider: natural disfluency production requires speakers to produce spontaneous, continuous, unprepared speech, which is part of a communicative act to another person. Such factors are relevant as disfluency has been shown to have an important conversational role in marking discourse structure and turn-taking in dialogue (Branigan, Lickley, & McKelvie, 1999; Clark & Fox Tree, 2002; Fox Tree, 2002; Swerts, 1998), and patterns of disfluency have also been shown to vary substantially depending on whether the speaker is interacting with a human or computer interlocutor (Oviatt, 1995; Eklund & Shriberg, 1998). As a result, it is important that even if the task used is purely a production task, in order to ensure the elicitation of naturalistic disfluency, it should be presented as an interactive task between a speaker and listener, where the speaker maintains communicative intent.

4.1.2 The Network Task

One experimental paradigm that has the potential to produce constrained, yet spontaneous speech is the network description task originally developed by Levelt (1983) to study patterns of speech error and self-repair. In this study, speakers were shown a series of visual patterns of interconnected coloured circles which were assigned positions on a 3x3 grid and connected by either horizontal, vertical or diagonal lines. From an indicated starting position, subjects had to describe the network, identifying each circle and providing the necessary directions that connected it to the next in sufficient detail that their description could be used to reproduce the pattern on a blank grid (for an example, see Figure 4.1).

This task placed the choice of individual colour names within the context of spontaneously produced utterances, while limiting the complexity of the semantic and

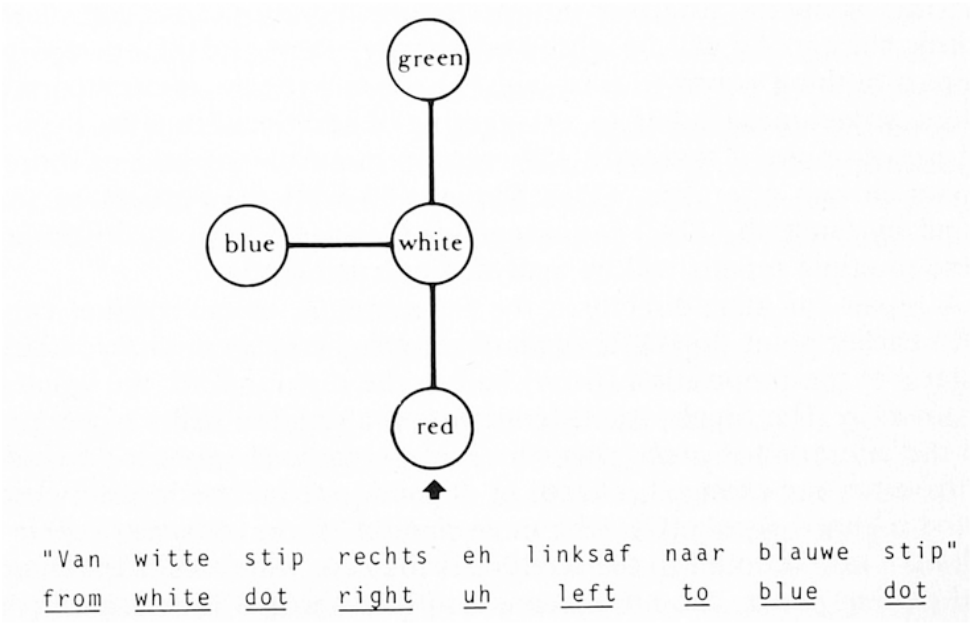


Figure 4.1: An example network and repaired utterance from Levelt's (1983) study. Each network was comprised of coloured circles connected by either horizontal or vertical lines, which were used to generate a corpus of spontaneous repairs (taken from Levelt, 1983, pp. 43).

syntactic context. As a result, more specific claims could be made about the relationship between disfluency and underlying production processes. Levelt observed that in word substitution errors, naming errors tended to be semantically, but not phonologically related to their subsequent repair, and suggested that these substitution errors occur independently of the retrieval of the word's phonological form. In contrast, when examining the occurrence of pre-lexical hesitations immediately preceding colour names, Levelt observed a significant negative correlation between the presence of a hesitation and the lexical frequency of the colour name being referred to. He argued that as word frequency is thought to affect the process of phonological encoding, such hesitations mark "the actuality or recency of trouble" during word-form retrieval.

N. Martin, Weisburg, and Saffran (1989) developed this methodology further, using networks of pictures to examine in detail the relationship between the semantic and phonological similarity of the picture names and the resulting substitution errors

observed. N. Martin et al. (1989) used the same task as Levelt (1983), but increased the sizes of the networks and used a larger number of coloured pictures instead of circles as items within the network, allowing a more precise manipulation of the semantic and phonological relationships between the items used. They examined content-word substitutions produced by speakers describing these networks, and found an interactive influence of both semantic and phonological variables, which they took as support for an interactive (as opposed to a discrete) model of language production. However, they did not examine the use of disfluencies such as filled pauses or repetitions, and so did not address whether the correlation between lexical frequency and filled pause rate Levelt (1983) observed for colour names also held for picture names.

More recently, Oomen and Postma (2001) adapted Levelt's (1983) network description paradigm to examine the effects of time pressure on the mechanisms of speech production and self-monitoring. In order to explicitly manipulate time pressure, Oomen and Postma (2001) used a variation of the task, in which speakers describe the path a marker takes as it traverses a network of pictures connected by multiple paths, while ensuring that their speech keeps pace with the position of the marker in the network. Levelt's (1983) task provided no way to manipulate speech rate, which can be an important factor affecting error elicitation in a production task. Under conditions of increased time pressure, the internal monitor would have less time to monitor the phonetic plan and accurately detect upcoming errors prior to articulation, which Oomen & Postma argued would lead to more speech errors and result in shorter pauses, as well as delaying the interruption point relative to normally paced speech. They found that at a higher speech rate speakers produced more overt lexical and phonological errors and syntactic omissions compared to normally paced speech. However, they did not find any effect on the number of filled pauses produced, leading them to argue that such hesitations are not simply instances of covert repair, and may be regulated by different processes. While speech rate can have a substantial impact on planning, formulation and monitoring processes, Oomen and Postma (2001) did not manipulate the lexical properties of the pictures

used in the networks, which could be used to examine how difficulties during lexical access impact resultant disfluency. Under time pressure speakers would also be expected to produce more disfluencies associated with lexical processing, as the faster rate of articulation would put additional pressure on the processes of lexical access and retrieval, resulting in a higher incidence of associated hesitations and self-repairs.

4.2 Experimental Method

Three issues were borne in mind when selecting a methodology for use in the current thesis. First, the task used would have to allow the specific manipulation of factors that are known to affect the speed of selection and retrieval of words. Second, the task needed to allow speakers to produce relatively constrained speech within a spontaneous, naturalistic context. Finally, the task needed to maintain communicative intent, so that speakers believed that they were conveying information to someone else.

Oomen & Postma's (2001) version of the network task provides a useful framework that addresses these requirements. The task allows spontaneous description of the route the marker takes through the network, while properties associated with the pictures to be named can be experimentally manipulated in order to elicit naturalistic disfluency. In addition, the number of paths that connect one picture to the next can be varied to examine how increased choice in the complexity of descriptive options can influence associated disfluency. While speakers descriptions are spontaneous, they are relatively constrained in terms of the content words used to name the pictures presented and to describe the paths that connect them. It can also be presented as a communicative task in which speakers describe the networks to a listener.

4.2.1 Design

The experiments using the Network Task that are detailed in this thesis were designed as a naturalistic communication task between a speaker and a listener. In the Network Task, the speaker is presented with a series of networks of pictures connected by multiple paths. Their task is to describe to the listener the path that a marker takes as it traverses each network of pictures. They are told that the listener's job is to follow their instructions to fill in a blank network containing no pictures, and that the study investigates how easy it is for people to follow verbal instructions without the aid of eye-contact, gestures or body language. As disfluency has been argued to form part of discourse structure (e.g., Fox Tree, 2002; Swerts, 1998), the experiment was designed so that participants communicate information about each network to another party, who is in fact a confederate, in order to ensure goal-directed communicative intent on behalf of the speaker. As a result, the communicative setting was not actually a dialogue situation, but that of a communicative monologue produced by the speaker.

The purpose of using this task was to examine how disfluencies produced during spontaneous descriptions of each network would vary with properties associated either with the pictures contained within the networks or with the paths that connected the pictures. Speech rate was not systematically varied in the experiment, although speakers were encouraged to try and maintain their rate of speech to keep up with the position of the marker, which travelled at a constant speed through each network. Within each network, lexical properties associated with the picture names (such as their lexical frequency or picture name agreement) could be manipulated. It was also possible to manipulate the pictures in such a way as to affect pre-lexical processing (such as the ease of visual recognition through picture blurring).

Because of the incremental nature of speech planning processes, disfluencies are thought to reflect underlying difficulty with immediately upcoming speech, and so were considered to be directly related to the part of the network that the speaker was describing at the time (i.e., either the pictures or the path in between them).

Therefore, the variation in disfluency associated with each of the task manipulations was regarded as local to the part of the utterance in which it was produced. Prior studies, such as that of Oomen and Postma (2001), evaluated average disfluency and repair rates per 100 words, measured across all of the network descriptions in each condition. In the experiments presented here, the factors were varied within each network, and we were interested in disfluencies that were associated with (and hence local to) the picture description. As a result, network descriptions were broken into utterances that contained the description of the route from one picture to the next. Each of these utterances usually included the initiation of the route ("Then it goes ..."), the direction of the route ("... straight to the left"), and the naming of the object ("... to the gate"). Disfluencies occurring in each of these parts of the utterance were thought to reflect either initial utterance planning, formulation and planning of the path description, or selection and retrieval of the picture name, respectively. To analyse the relationship between different disfluencies and factors influencing these local production processes, each utterance was separated into *beginning*, *path* and *target* sections (as described in section 4.2.5, below), and analyses were performed on the proportion of utterances containing a disfluency, or a particular class of disfluency, in the relevant section. This analysis reflects the fact that multiple disfluencies often cluster together, such as in (1), but such conjoint disfluencies tend to reflect the accommodation and resolution of a single underlying difficulty. As a result, the experimental analyses presented in this thesis relate the likelihood of disfluency to different factors that affect production difficulty, as opposed to a measure of the rate or number of disfluencies produced during portions of speech.

(1) ".down to thee: uh .. plant thing"

While in Oomen and Postma's (2001) study, participants were told that their network descriptions would be recorded and played back to another listener, who would have to use them to fill in a blank network, in the present thesis, we decided to use a live and present confederate listener to give the task greater communicative

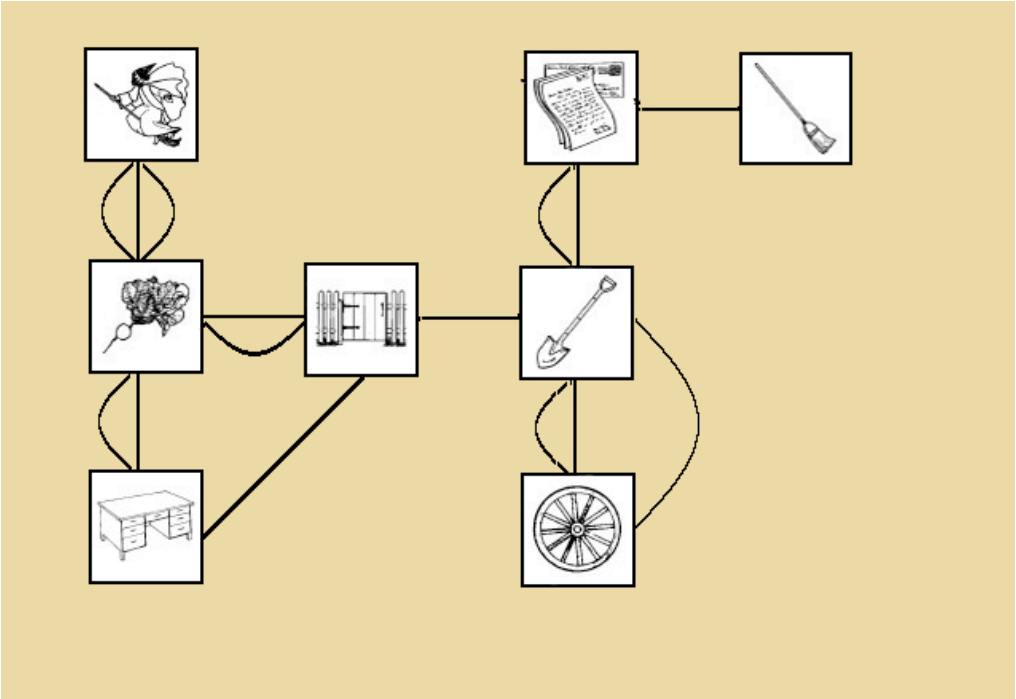


Figure 4.2: An example network used in the Network Task (Taken from Experiment 4).

validity. As we were primarily interested in speakers' production of disfluencies, and it has been argued that disfluencies may also contain metalinguistic communicative intent (Clark & Fox Tree, 2002), we decided that speakers would be more likely to elicit more natural patterns of disfluency if they were convinced that the task that they were engaged in was a communicative act. While the confederate listener was located behind a partition and did not engage directly with the participant during the course of their descriptions, participants were clearly aware of their presence and directed their network descriptions to the listener. It is possible that without a confederate listener in the room, speakers may have truncated their network descriptions and produced substantially fewer disfluencies, as Oviatt (1995) noted when examining disfluency rates in human-human and human-computer dialogues. Further, as the participants believed that the listener was filling in a blank network from their descriptions, this encouraged them to produce complete network descriptions, and try to avoid missing out or providing ambiguous references to items.

4.2.2 Materials

The networks used in all experiments presented here consisted of 8 pictures depicting objects, arranged in different configurations on a 4 x 3 grid and interconnected by one, two or three straight or curved lines. The lines could either connect pictures vertically, horizontally, or diagonally (see Figure 4.2 for an example). Each network was associated with a route through the objects, which was indicated by a red marker that moved at a constant pace along the lines connecting the pictures. Each route always consisted of nine steps, in which the marker started on a picture at the edge of the network, and passed through six of the pictures once along the route, and through two of the pictures twice, before stopping on the final picture of the network.

In their study, Oomen and Postma (2001) used two speeds for the marker to traverse the network. In their "normal speed" condition, the marker took 53 seconds to run through the network, while in their "fast" condition, the marker took 35 seconds. For the studies reported here, we determined in a pretest that 30 seconds was an optimum time to produce fast, yet errorful speech, while allowing speakers to produce complete descriptions of the networks. This traversal time is faster than the times used in Oomen and Postma (2001), as all pictures described in their study also included colour names, which were not required in the descriptions produced for the current experiments.

The pictures used in the Network Task varied between each experiment, and were selected based on the factors that were manipulated in that experiment. The picture sets used are detailed in the methodological sections of the experimental chapters. In addition, the numbers and types of paths connecting each picture in the networks differed, but always varied between one and three paths.

4.2.3 Apparatus

All experiments were performed on a Research Machines personal computer, connected to a 17-inch display. The software used to run the Network Task was spe-

cially designed, and was obtained courtesy of Albert Postma. Participants speech was recorded on a SONY TLD-D8 DAT Walkman recorder using a Senheiser C6 microphone. Recordings were then converted into .aiff files for subsequent transcription and acoustic analysis on the computer. All acoustic analyses were performed using the phonetics software, Praat (Boersma & Weenink, 2008).

4.2.4 Procedure

The procedure used in each of the Network Task studies was broadly similar, with exceptions detailed in the relevant experimental chapters. The speaker and a confederate listener sat on opposite sides of a partition. While both speaker and listener could hear each other clearly, the partition occluded any visual interaction. The speaker was informed that they would see a series of networks and that their task was to describe the route a marker took through each network so that the listener could fill in the route and picture names on a blank network. Participants were instructed to describe the path of the marker as it moved from one picture to the next, including the name of each picture, the direction the marker was taking, the shape of the line (straight or curved), and the position of the line in relation to any others. They were told to modulate their speech rate to try to keep up with the position of the marker as it traversed the network.

When a new network was revealed, the speaker wass instructed to immediately press the space bar to start the marker moving through the network. The marker appeared at the starting picture and began moving through the network along a predetermined route, taking approximately 30 seconds to traverse the entire network. Upon completion of their description, the confederate listener confirmed that they had marked down the description on their blank network, and the speaker pressed the spacebar to reveal the next network. No feedback was given by the confederate during the course of each network description.

4.2.5 Transcription and Coding

Network descriptions were not transcribed phonetically, however, care was taken to accurately transcribe all part-words and other speech sounds heard in the recordings. A system of symbols was devised to mark up disfluencies, in particular, prolongations, silent pauses and interruption points in a repair.

Disfluencies were coded into five classes based on Lickley's (1998) taxonomy: prolongations; filled pauses; silent pauses; overt repairs and repetitions (see section 2.3 for details of each disfluency class). Additionally, filled pause were coded into separate types (either *um* or *uh*), and repairs were coded as either insertions, deletions or substitutions. Prolongations, where possible, were coded as to whether they possessed either a reduced or non-reduced form, although this was only reliable for the function words *the* and *a*.

Prolongations and silent pauses were coded perceptually. This was primarily because of the problem of reliably measuring such hesitations, and determining an objective measure of pause length or prolongation duration, as these can vary from speaker to speaker based on factors such as speech rate (for further details, see sections 2.3.2, 2.3.3). Silent pauses were only coded when they occurred in isolation within a clause, and were not co-located with another disfluency, such as following a filler or prolongation, or at the interruption point within a repair. Similarly, when two adjoining words were both prolonged, such as the prolongation of both vowels in "to: the:...", this was only coded as a single instance of prolongation, because this conjoint hesitation is related to only a single underlying difficulty in the upcoming speech plan.

Each transcribed network description was broken up into 10 utterance moves, starting with the naming of the picture where the marker started. The following 9 utterances each described the path of the marker up to a particular picture, and included that picture's name. If an utterance did not explicitly include a picture name, we determined which part of the recording uniquely identified the path associated with the upcoming picture. If there was no clear description of either the path or the

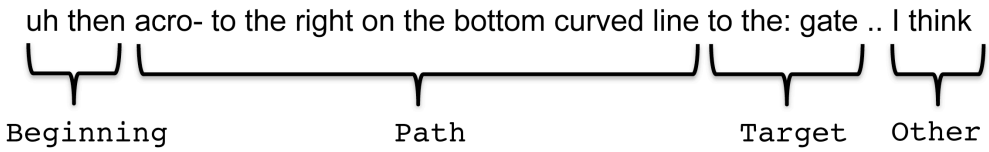


Figure 4.3: An example utterance broken up into coding sections

picture that it lead up to, the utterance corresponding to the upcoming picture was left blank, and any speech was appended to the beginning of the next distinctly coherent utterance. Each utterance was then separated into four sections (for an example see Figure 4.3):

1. **Beginning**: The first section contained any speech or utterance-initial disfluencies that occurred before speakers began describing the path to the next item.
2. **Path**: The description of the path between items, ending at the last content word of the path description
3. **Target**: All words up to and including the target name from the last preceding content word of the path description.
4. **Other**: Any additional speech following the naming of the target item that does not relate to the description of the path leading up to it, or the beginning of the next utterance.

As disfluencies were expected to be related to local production difficulties with upcoming material (i.e., disfluencies related to access and retrieval of the target picture name would be closely co-located with the target name), the focus of the present thesis is on disfluencies that occurred within the target region (with the exception of Chapter 7 which includes an analysis of all disfluencies produced by 12 participants).

The remaining regions constitute those parts of the utterances which do not refer to the target, as well as utterance-initial disfluencies. The latter were coded separately

as filled pauses in particular tend to occur most often in utterance-initial positions. These fillers often are thought to be associated with message level planning of the upcoming utterance, or as a continuation device between utterances (Barr, 2001), and so may not be due to specific difficulties related to the manipulations employed in the task.

To check for coding consistency, a portion of the network descriptions were independently coded by an experienced second rater from the Edinburgh Disfluency Group, and compared to the coding of the author. Second raters were blind to the item conditions used within each experiment and were given complete transcriptions of the network descriptions with all disfluency mark up removed. They were then asked to listen carefully to each network description and note exactly where they identified different classes of disfluency using the devised mark up scheme. While the identification of disfluencies such as filled pauses, repetitions and repairs was relatively straightforward and consistent across coders as these are discrete events, the identification of prolongations and silent pauses was more subjective as it involved a perceptual assessment of what was a "hesitant" delay. In order to ensure that coders were making such assessments of delay using similar perceptual criteria, a number of examples of each type of disfluency were worked through with both the author and coder prior to coding the network descriptions to ensure that identification was consistent between coders. While prolongations occurred frequently, and inter-coder reliability of identification of prolongations was high, isolated silent pauses occurring during the Target section that were not associated with another disfluency were relatively rare. Because inter-coder reliability for the identification of silent pauses was relatively low compared to other types of disfluency, in experiments where this was the case they were removed from the analysis altogether. Reliability comparisons for each experiment are provided in the experimental chapters.

4.2.6 Analysis

All disfluency data are presented in terms of the proportion of utterances within a given item condition that contain one or more occurrences of a particular type of disfluency. However, it is generally recognised that the use of ANOVA on proportionate data is inappropriate as the binomial distribution violates ANOVAs assumption of homogeneity of variance (e.g., Agresti, 2002). In binomially distributed data, the variance of values in the middle of the distribution will always be larger than variance observed for values at the ends of the distribution. Furthermore, as the range of confidence intervals calculated from ANOVA can extend beyond 0 or 1, beyond the range of the binomial distribution, it is possible that probability mass may be assigned to events that can never occur, increasing the likelihood of Type I errors, and leading to potentially spurious results. While the experiments in Chapter 5 used an arcsine transform to convert proportions into Rationalised Arcsine Units, this transform is only an approximation that increasingly breaks down at values close to 0 or 1. As the proportion of utterances containing certain classes of disfluency were often very low in these studies, this raises serious doubts about the reliability of the results obtained for infrequently occurring classes of disfluency.

In a recent article, Jaeger (2008) has made a cogent argument against the use of ANOVA, untransformed or not, for the analysis of proportionate data for the above reasons, and has argued for the use of logit mixed effects models (Breslow & Clayton, 1993; DebRoy & Bates, 2004) as a statistically valid alternative for the analysis of such data (see also Dixon, 2008). Logit mixed effects models are a form of generalised linear model that use a logit (or log-odds) transformation to convert binomially distributed data into a linear distribution. Such models confer several advantages: They are a statistically sound method of modelling binomial data across the full range of the binomial distribution. Additionally, they provide information about both the size and directionality of observed effects, and allow the inclusion of both by-participant and by-item random variation within a single model, removing the need for separate F1 and F2 analyses. Finally, they can be used to analyse data with both categorical and continuous predictor variables, and

are more robust to issues of missing data than are ANOVA (for a more complete overview of the utility of logit mixed models for categorical data analysis, see Jaeger, in press).

Logit mixed effects models require a degree of model fitting to ensure that the model fit to the data is optimal, and only contains predictors that add significant explanatory value to the model. Model fitting is performed through likelihood-ratio tests, which use a χ^2 distribution to determine a predictor variables significance in a model by comparing the data likelihood (i.e., the likelihood of the sample, given the model) of two models with and without the predictor variable. By adding predictor variables in turn, it is possible to determine the optimal model fit to the data while avoiding model over-fitting. All models described in this chapter were determined by iteratively adding each independent variable and interaction term to a null model containing only an intercept, and testing whether the variable significantly improved the model fit to the data. Additionally, all models included random subject and item effects.

Parameters for each fixed effect are fit to the data so that the model describes the data optimally. The output of the model (for an example, see Table 6.3) provides estimated coefficients of each fixed effect given by the model. These represent the strength and directionality of the effect in log-odds space, and so can be used to calculate an estimate of the change in the likelihood (in terms of its odds) for that condition. For categorical predictors, this is the change in log-odds between conditions; for continuous predictors, the coefficient represents the change in log-odds per unit of that continuous predictor. The standard error of the coefficient is also given, which is used to calculate Walds Z scores and corresponding significance values for each fixed effect.

All mixed effects models described within this thesis were implemented using the *lmer* function (*lme4* library, D. M. Bates & Sarkar, 2007) in the R statistical software package (R development core team, 2005).

4.3 Chapter Summary

In this thesis, the primary experimental methodology employed is the Network Task, an experimental paradigm that can be used to produce relatively constrained, yet spontaneous speech, in which properties associated with the content of what a speaker must say can be explicitly manipulated. In this way, the impact of factors thought to affect the ease of planning utterances, or the selection and retrieval of words can be manipulated to determine their affect on the likelihood of resulting in a disfluency. This approach contrasts with many prior studies examining disfluency production, that tend to use large corpora of unconstrained speech, and have focused how the rate of disfluency production varies with local and global factors thought to have a bearing on fluency and speech production processes. In this chapter I have detailed the overall methodology employed in the Network Task. In Chapters 5, 6 and 7 I will present experiments using this methodology to examine how lexical and pre-lexical factors that are know to affect the production of words in isolation affect disfluency likelihood during continuous speech.

Chapter 5

The Network Task: Exploratory investigations of disfluency production

While the distribution of disfluencies in spontaneous speech has been documented through corpus-based studies (e.g., Bortfield et al., 2001; Shriberg, 1996), and claims have been made as to the differing functions of different types of disfluency (e.g., Clark & Wasow, 1998; Clark & Fox Tree, 2002; Fox Tree & Clark, 1997), experimental studies to date have failed to fully address the underlying causes of disfluency. Take, for example, the finding that disfluencies are more likely to precede open-class words than closed class words (Maclay & Osgood, 1959). On current evidence, it is unclear what the underlying cause of these disfluencies might be. They could be a consequence of the relatively low frequencies (compared to closed-class words) with which open-class words are likely to occur, which Levelt (1983,1989) has attributed to lexical retrieval difficulties. Alternatively, disfluencies could be the result of increased planning demands associated with the greater choice of open-class words available to the speaker (Schachter et al., 1991), or because open class words possess greater uncertainty and so their retrieval tends to be less probable given the previous spoken context (Beattie & Butterworth, 1979). Indeed, they may be due to causes outwith the language system: if, for example, a speaker is trying to name an unfamiliar or ambiguous object (Siegman & Pope, 1966). Effectively, the difficulties signalled by disfluencies could occur at any stage of the speech production process: during conceptualisation, planning, formulation or articulation of the speech plan.

As it has been argued that different types of disfluency may be associated with different kinds of problems during production (Bortfield et al., 2001; Shriberg, 1994), manipulating the types of lexical and pre-lexical difficulty that speakers encounter while monitoring the disfluencies that they produce may provide more insight into this relationship. Clearly, a better understanding of the underlying causes of disfluency would provide an important contribution our understanding of language production in general.

In this chapter, we present three experiments that set out to explore how the disfluencies speakers produce are related to problems with what they are about to say. We used the Network Task (Oomen & Postma, 2001) to manipulate the content of what people say when describing a network of objects, in order to explore how factors known to influence production processes affect the production of disfluency in spontaneous speech. In Experiment 1, production difficulty associated with naming pictures was increased by varying the word frequency and name agreement of items in the networks; Experiment 2 extended this work by orthogonally manipulating frequency and name agreement to determine whether disfluencies may be attributed to one factor over the other. In Experiment 3, the visual accessibility of pictures was manipulated to assess the impact of difficulties that do not have their origin in the linguistic system. Subsequent naming studies were also performed on the items used in each of these experiments in order to investigate how the time-course of picture naming, which has been the focus of much research in the production literature, relates to delays in spontaneous speech that are accommodated through disfluency. These experiments allow us to begin to investigate the causes of disfluency directly, in an attempt to establish whether different disfluencies serve different purposes.

5.1 Experiment 1: Does lexical difficulty affect disfluency production?

The first experiment set out to explore how the disfluencies speakers produce are related to lexical difficulties associated with the words produced during speech

production. Previous research has established relationships between overall disfluency rate and factors such as utterance length (Shriberg, 1996) and planning load (Bortfield et al., 2001), however, this study aimed to take an approach to the study of disfluency that relates their production to processes occurring during lexical access, in order to investigate how they relate to difficulties associated with immediately upcoming words. There has been a long tradition of localist research into disfluency that has primarily focused on how the relative uncertainty of subsequent words is related to hesitation (e.g., Beattie & Butterworth, 1979; Butterworth, 1975; Goldman-Eisler, 1958b, 1961, 1968; J. G. Martin & Strange, 1968; Tannenbaum et al., 1965). Studies such as these have primarily focused on how the contextual probability of a word influences associated hesitation. This is effectively a measure of uncertainty that, given what has already been said, a particular word is more or less likely to be used. Yet this measure is not a property of the word itself, but a property of the context. A content word of high frequency may be just as likely to have a low contextual probability in a particular utterance as a word of low frequency.

Name agreement, in contrast, is a property of the concept itself. The concept associated with a particular picture may elicit several names that are more or less accessible to the speaker, and may also provide a better or worse lexical fit to the conceptual representation that the speaker is intending to convey. In a production model such as that of Levelt et al. (1999), activation of alternative lexical concepts would increase the time taken to correctly select a single one, as each of these lexical representations would engage in competition for selection. This process may take time to resolve, most likely longer than the selection of a lexical representation that has a unitary relationship with its associated concept.

In addition to a concept's name agreement, a word's lexical frequency is also known to influence the speed of access to its lexical and phonological representation (e.g., Oldfield & Wingfield, 1965; Wingfield, 1968). The locus of these frequency effects has been traditionally argued to reside in the processes of word-form retrieval and phonological encoding (Garrett, 1975, 1980; Jescheniak & Levelt, 1994; Levelt,

1989). Whether frequency effects are purely associated with word-form retrieval, or also impact other parts of the lexical system (Alario, Costa, & Caramazza, 2002; Almeida et al., 2007), it is evident that frequency effects on production appear to be lexical in nature (Jescheniak & Levelt, 1994; Navarette, Benedetta, Alario, & Costa, 2006).

Both of these factors associated with the concept or word to be expressed impact the speed of picture naming in isolation, although their effects could be interactive rather than purely additive depending on what assumptions are made about the architecture of the production system (Griffin & Bock, 1998; Shatzman & Schiller, 2004). It is also likely that they will increase the time taken to produce picture names when the naming task is presented within a demanding spontaneous speech context, such as the Network Task. If this is the case, then we would anticipate finding a direct relationship between lexical difficulty, as indexed by a picture's frequency and name agreement, and local hesitations or other disfluencies that allow for the resolution of these processes.

5.1.1 Method

The experiment set out to examine how lexical difficulty associated with the naming of pictures during a spontaneous production task influenced the likelihood of disfluency in the same region of speech as the picture name. The experiment was presented as a communication task between a speaker and a listener. Participants described the route taken by a marker through a network of pictures to a listener situated behind a screen.

Participants

Twenty students from the University of Edinburgh participated in the experiment (8 male, 12 female). All were native British English speakers and had no speech or hearing problems. Two undergraduate Edinburgh University Psychology students who were involved in the study acted as confederates throughout the experiment.

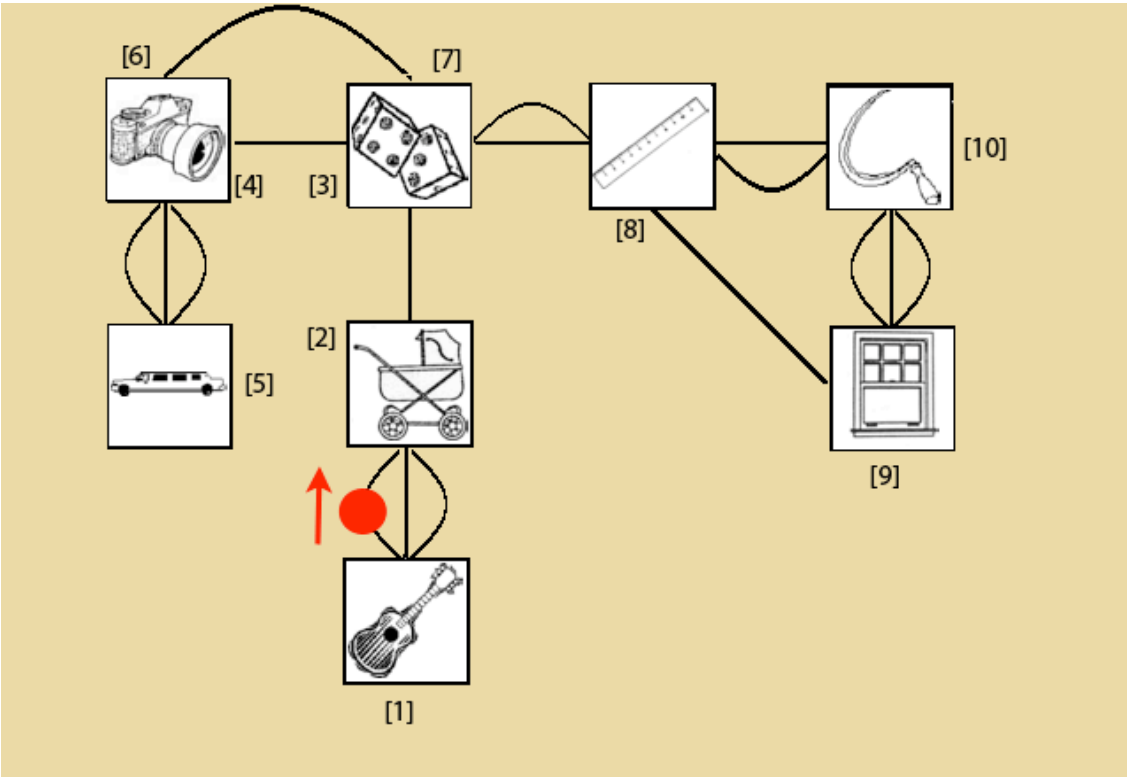


Figure 5.1: A sample network used in Experiment 1. The red marker progresses through the network at a constant pace, passing along one of the paths that connects the pictures. The numbers mark the order of the pictures that that the marker passes through in the network.

Materials

The Network Task consisted of 6 visually presented networks based on those used by Oomen and Postma (2001). Each network contained 8 black and white line drawings, arranged in different configurations and connected by one, two or three straight or curved lines. Each network was associated with a pre-determined route through the pictures that was indicated by a red marker that moved along the lines connecting each picture. The route started on one picture and made nine further steps through the network, so that six of the pictures occurred once in the route, and two of the pictures occurred twice. Figure 5.1 shows an example network. The numbers beside each picture indicate the direction of the route that the marker takes through the network.

To construct the networks, 36 experimental and 12 filler pictures were chosen out of 366 pictures from the two sets of pictures used to generate the Beckman Spoken Picture Naming Norms (Griffin & Huitema, 1999): those of Huitema (1996) and Snodgrass and Vanderwart (1980). Selection of pictures was based on two factors: the picture's name agreement and the lexical frequency of the preferred name. 18 pictures with high name agreement and high frequency dominant names (designated HF items), and 18 pictures with low name agreement and low frequency dominant names (designated LF items) were selected. Frequency and name agreement statistics for all experimental pictures are provided in Appendix A. All pictures possessed consonant-initial dominant names, as determiners such as *the* and *a* are typically given a full vowel sound before words with vowel onsets. Inclusion of pictures with a name possessing a vowel onset could interfere with correctly identifying non-reduced prolongations before target names in the transcription. Twelve filler pictures were also selected, possessing high name agreement and medium frequency relative to the experimental items (mean BNC frequency = 23.19 counts per million, hereafter cpm; s.d. = 9.55).

Name agreement: The pictures used in Griffin and Huitema's (1999) study were categorised based on the consistency with which participants ascribed different names to them. Pictures were classified as high name agreement if more than 90% of the participants in Griffin and Huitema's study provided the same name for it, while pictures possessing less than 50% agreement with target name were classified as low name agreement. Pictures could result in alternative names that varied *intrinsically*, i.e., they were valid alternative names for a target picture produced as a result of access to alternative lexical representations associated with the same underlying concept. Otherwise, variation in picture names could be due to *extrinsic* naming differences, i.e., alternative names that corresponded to different concepts to the intended referent, possibly due to picture ambiguity or conceptual mis-selection. For example, a picture of a stove could also be given the intrinsically varying names *oven*, *hob* or *cooker*, while a picture of an orange may be given extrinsically varying names such as *ball* or *circle*. Only pictures whose two most frequently used

alternative names were intrinsic variants of the preferred picture name were considered. Pictures that were given alternative names that varied extrinsically were not included. In total, high name agreement pictures had dominant names which were produced by 96.7% (s.d. = 3.4%) of Griffin and Huitema's (1999) respondents, while low name agreement pictures given the dominant name of 35.1% (s.d. = 9.1%) of the time.

Word Frequency: Lemmatised word frequencies were taken the British National Corpus (BNC) frequency database (Kilgariff, 1995). High frequency words that were selected for use in this experiment possessed frequencies ranging from 25 (cow) to 532 (hand) cpm (mean = 143.5 cpm ; s.d. = 147.3), while low frequency words had frequencies ranging from 0.5 (gavel) to 18 (molecule) cpm (mean = 4.7 cpm; s.d. = 5.2).

Networks: Each network comprised eight pictures: 3 HF pictures, 3 LF pictures and 2 filler pictures. The fillers were included in each network at the 2 nodes which were always passed through twice by the marker, as including experimental items at these positions may lead to practice effects on these items. In each network the picture sequences were alternated between HF and LF items. Each network was constructed so that semantically and phonologically related targets were not adjacent. For example, dog and cat were not in sequence due to their semantic similarity and dice and desk were separated due to phonologically similar onsets and rhymes. An example of the order of pictures in a pathway is shown in Figure 5.1.

Procedure

The experimental procedure used in this experiment followed the description given in section 4.2.4. The speaker and confederate listener (who the speaker believed to be another participant) were led into the testing room and asked to sit on opposite sides of a screen. Both were given instruction sheets to read. The participant heard the experimenter ask if the confederate understood the instructions for their part in the experiment and heard the confederate say yes, as if they were a genuine

participant. While both speaker and listener could hear each other clearly, the screen occluded any visual interaction between them.

The speaker was informed that they would see a series of networks and it was their task to describe the route the red marker took through each network so the listener could fill in the route and picture names on a blank network. They were instructed to describe the path of the marker, making sure to include the name of each picture, the shape of the line the marker was moving along (straight or curved), the position of the line in relation to the others (right, left or middle), the angle of the line (horizontal, vertical or diagonal) and the direction the marker was taking. The speaker was instructed to talk in complete sentences and to modulate their speech rate to keep up with the position of the marker (as in Oomen & Postma, 2001), while doing their best to be as accurate as possible as the listener could not ask for any repeated descriptions.

The order of network presentation was randomized across participants. The marker started moving through the first network when the space bar on the keyboard was pressed. It took 30 seconds to pass through the whole network, which was found to be an appropriate speed for eliciting fast connected spontaneous speech containing sufficient numbers of disfluencies. Upon completion, participants pressed the space bar again to reveal the next network.

Transcription and coding

Transcription and coding was carried out by the author. 20% of the transcriptions were independently verified and coded by another researcher from the Edinburgh Disfluency Group. The two coders agreed on 84% of identified disfluencies. Items for which transcription or coding differed were re-examined by both researchers until a consensus was reached.

The description of each network was separated into 10 utterances, where each utterance named the upcoming item and described the path the marker took to reach it. The description of the starting item was included as a separate utterance to

the description of the transition to the second item. Where an item or a path was not named (e.g., the first item in each network had no path description leading up to it), the next utterance was considered to begin at the onset of the next path description, after the naming of the previous item, as in (1). In situations where an item the marker was moving from was not named until after the description of the path to the next item (as in (2), where duck is the starting item of the network), any disfluencies produced immediately preceding the naming of the first item were attributed to the first utterance, and the path description and second item description were attributed to the utterance of the item that the path was leading to.

(1) it goes from the boot ‖ straight down to the monkey

(2) the dot goes along to the right the straight line from the duck to the leaf

While disfluencies associated with the target item were of primary interest, disfluencies occurring throughout each utterance utterance were also coded. Disfluencies were assigned one of 4 positions within an utterance, either *beginning*, *path*, *target* or *other*, as described in section 4.2.5.

Four classes of disfluencies were coded: filled pauses, prolongations and repetitions and repairs. Silent pauses, however, were not explicitly coded in this study because of difficulties with reliable perceptual identification. Hesitations were further broken down into individual disfluency types for additional analysis. *Um* and *uh* were coded as separate types of filled pause; prolongation of the determiners *the*, *a*, and *to*, as well as the non-reduced forms of *the* and *a*, were also coded separately. Repetitions and repairs were both coded as a single classes in this study. Examples of each type of disfluency are illustrated in Table 5.1.

Table 5.1: Examples of each class of coded disfluency.

Disfluency Class	Type	Example (in bold)
Filled Pause	*Um*	along the straight line to the **um** trophy.
	Uh	its going up to the **uh** house
Prolongation	*a:*	... moving left on a line to **ay:** flag
	the:	down to **the:** lion badge
	to:	along **to:** a hammer.
Repetition Repair		towards **the–the** sickle um along the and then left towards the **squ–rectangle**

5.1.2 Results

Out of a total of 720 items, 29 items were not named in participants descriptions (7 HF and 22 LF items), and were removed from further analysis. Disfluency analysis was performed using the remaining 691 data points.

Picture Name Agreement

On average, HF pictures resulted in 94.6% name agreement, while LF pictures resulted in 40.6% dominant name agreement. The average number of different names used for each picture (including the dominant name) was 1.56 (s.d. = .85) for HF items and 4.83 (s.d. = 2.12) for LF items. Despite the selection of items which only had intrinsically varying alternative names in the Beckman picture naming norms, a number of pictures still elicited extrinsic alternative names (e.g., the LF item *sickle* elicited the intrinsic alternative name "scythe," as well as the extrinsic alternative names "hook" and "rope").

To assess the extent to which intrinsic and extrinsic alternative names were produced for each picture, a by-items ANOVA was performed on the number of alternative names, with item type as a between items factor, and alternative name type as a within-items factor. A main effect of item type was found [$F(1, 34) = 45.97$, $p < .001$], demonstrating that LF items were given significantly more alternative names than HF items, as detailed above. A main effect of alternative name type was also observed [$F(1, 34) = 7.98$, $p < .01$], as well as an item type x alternative

Table 5.2: Proportion of utterances in Experiment 1 that contained different classes of disfluency in each of the four identified utterance locations: the beginning of the utterance, during the path description, immediately prior to naming the target item, and elsewhere in the utterance (i.e., after naming the target item).

Disfluency Class	Utterance Location			
	Beginning	Path	Target	Other
Any Disfluency	5.4%	14.6%	30.1%	3.5%
Filled Pause	5.1%	3.9%	3.9%	1.3%
um	3.9%	2.5%	2.5%	0.9%
uh	1.2%	1.6%	1.4%	0.4%
Prolongation	0.0%	3.5%	22.4%	0.9%
Repetition	0.0%	1.0%	1.6%	0.1%
Repair	0.3%	6.4%	1.6%	0.6%

name interaction [$F(1, 34) = 7.61$, $p < .01$]: Items were given, on average, significantly more intrinsic (1.53) than extrinsic (.67) alternative names, and this effect was driven by LF items.

Distribution and rates of disfluency production

As all disfluency data are presented in terms of the percentage of utterances in a given condition, performing analysis of variance (ANOVA) on untransformed proportions would violate the homogeneity of variance assumptions inherent in ANOVA (Winer, 1971; Agresti, 2002). Therefore, all analyses of disfluency rates were performed using logit mixed effects models including item category as a fixed effect and random subject and item effects. All descriptive statistics are reported as untransformed percentages.

Overall distribution: Disfluencies were classified as either occurring utterance-initially, during the description of the path between items, prior to the naming of the target item within each utterance, or elsewhere in the utterance, which normally corresponded to after the naming of the target picture, but before the beginning of the next utterance. Table 5.2 presents the proportion of utterances that contained disfluencies of different types in each of the four pre-specified locations within an utterance. Table 5.3 details where different classes of disfluency tended to occur within

Table 5.3: The proportion of observations of each disfluency class in different locations within an utterance. Percentages for each disfluency class sum to 100%.

Disfluency Class	Utterance Location			
	Beginning	Path	Target	Other
Any Disfluency	8.4%	26.9%	58.4%	6.3%
Filled Pause	35.0%	29.0%	27.0%	9.0%
um	39.7%	26.5%	25.0%	8.8%
uh	25.0%	34.4%	31.3%	9.4%
Prolongation	0.0%	12.4%	83.9%	3.6%
Repetition	0.0%	36.8%	57.9%	5.3%
Repair	3.3%	72.1%	18.0%	6.6%

utterances, in terms of the percentage of total observations of each class of disfluency. Overall, disfluencies were observed more frequently in the Target location than elsewhere in the utterance: 58% of all disfluencies were produced immediately prior to the naming of target items. Most of these disfluencies were prolongations, which primarily occurred during the target description. Other classes of disfluency were observed in the Target location, but relatively infrequently. Almost all utterance initial disfluencies were filled pauses, and while the number of observed fillers was relatively evenly spread between the utterance onset, the path and the target descriptions, slightly more fillers were observed at the beginning of utterances than in in other positions. Repetitions were relatively infrequently observed, but tended to occur in prior to target items, while repairs were produced more often during the path description than elsewhere, and were the most common class of disfluency in this location.

While the general patterns of occurrence of different classes of disfluency may help to characterise their differing roles in production, the focus of this experiment was on how lexical difficulty associated with naming the target item influenced the production of associated disfluency. One question is whether lexical difficulty only has a local influence on production, or whether difficulty naming an upcoming picture can influence disfluency associated with planning processes earlier on in an utterance. To evaluate this possibility, disfluency rates in other parts of an utterance were examined with respect to the item manipulation. No effects of the

Table 5.4: Proportion of utterances in Experiment 1 that contained a disfluency in the Target location overall, and by disfluency class.

Disfluency Class	Item Type	
	HF	LF
Any Disfluency	15.1%	37.7%
Filled Pause	2.0%	5.9%
um	1.7%	3.4%
uh	0.3%	2.6%
Prolongation	9.4%	28.9%
reduced *the* or *a*	5.2%	17.5%
non-reduced *thiy* or *ay*	3.1%	4.2%
Repetition	1.2%	2.1%
Repair	1.5%	1.9%

item manipulation were found on the proportion of utterances containing a disfluency, either at the beginning of an utterance [HF = 6.6%, LF = 4.5%; coefficient $= 0.11, SE = .308; p > .5$] or during the path description [HF = 12%, LF = 13.4%; coefficient $= 0.24, SE = .342; p > .5$], indicating that any effects associated with the difficulty of naming items only had an influence on disfluencies immediately local to the picture name.

Disfluency prior to picture names: Table 5.4 presents the proportion of utterances in each condition that contained a disfluency prior to the picture name, broken down by class and type of disfluency.[1] A logit mixed effects model of overall disfluencies immediately preceding the target item showed a significant effect of item type (LF vs. HF) on the total number of disfluencies produced [coefficient= $1.177, SE = .244; p < .001$]. Prolongations were the most frequently occurring class of disfluency, and were produced significantly more often prior to LF target names than before HF items [coefficient= $1.246, SE = .242; p < .001$]. Filled pauses were also found to occur more often prior to LF items [coefficient= $1.22, SE = .557; p < .05$]. Repetitions and repairs associated with the item name were each produced in less than 2% of utterances. Neither repetitions nor repairs occurred significantly more often before LF than HF items [$p > .5$].

[1] In some utterances, more than one disfluency type preceded a target item. Therefore, the sum of percentiles across disfluency types and classes are not equal to the total percentage of items preceded by a disfluency.

Differences in the frequency of individual types of hesitation were also examined. The filler *um* was produced more often than the filler *uh* before picture names, but only the production of *uh*s resulted in an effect of the item manipulation that approached significance [coefficient= $2.578, SE = 1.491; p = .08$]. However, it is difficult to make strong conclusions based on the results for different types of filled pause, as they were produced relatively infrequently in the Target position (only 17 *um*s and 10 *uh*s were observed prior to picture names across the experiment). When examining the occurrence of reduced and non-reduced prolongations of the determiners *the* and *a*, over three times as many reduced prolongations were observed prior to picture names. Reduced forms occurred significantly more often before LF items than HF items [coefficient= $1.684, SE = .368; p < .001$], however no effect was observed for non-reduced prolongations [$p > .5$]. Prolongations of the determiner *to* were also observed significantly more often prior to LF items [coefficient= $0.99, SE = .329; p < .01$], but because it was not possible to identify reduced and non-reduced forms, an analysis of vowel type was not performed (although they are included in overall rates of prolongation).

5.1.3 Discussion

The purpose of this experiment was to provide an initial characterisation of the occurrence of disfluency during spontaneous utterances produced when speakers described each network, and to examine how increased difficulty associated with the selection and retrieval of the picture names influenced the rate of disfluency production. As we assumed that disfluency is a local phenomenon that reflects production difficulties as they arise, we were not interested in overall rates of disfluency *per-word* during each utterance, but whether manipulation of the lexical difficulty of picture names would influence the likelihood of disfluencies produced in the immediate vicinity of the picture name.

Supporting the hypothesis that disfluencies are local phenomena, the picture manipulation had no effect on the likelihood of a disfluency occurring earlier in an

utterance. Disfluencies were observed during the path description and at the beginning of utterances, but the proportion of utterances in which they occurred in these locations did not vary with the manipulation of difficulty associated with the picture names. Utterance initial fillers may have been related to accommodating overall utterance planning processes, while disfluencies occurring during the path description may be due to local problems of choice in formulation a description of the path. Both of these possibilities will be addressed in later parts of this thesis.

The manipulation of picture naming difficulty did, however, have a strong effect on the production of disfluencies local to the picture name within the descriptive utterances. Utterances relating to LF pictures contained more disfluencies overall, and specifically more prolongations and filled pauses prior to the item name, than utterances relating to HF pictures. This result suggests that the production of hesitations is directly related to local difficulty with the selection and retrieval of words. However, because the two factors that were used to determine naming difficulty in this experiment (i.e., the picture's name agreement and the frequency of its dominant name) were not manipulated independently of each other, no direct conclusions can be made about the separate effects of these factors on disfluency likelihood, or whether different types of disfluency reflect difficulty at different stages of production. Further, some of the pictures were given extrinsic alternative names, despite attempts to select pictures that would provide only intrinsic variation in naming. This suggests that in some cases speakers may have had difficulties recognising pictures and determining appropriate conceptual representations, and hence that some of the observed disfluencies may have been associated with conceptualisation difficulties, rather than delays in selecting and retrieving an appropriate name. As a result, it is not possible to claim that the effects observed were purely lexical in nature.

Another issue that arises from this study is whether the different types of hesitation produced tend to be associated with different problems encountered during production. Most of the disfluencies observed prior to picture names were prolongations, a class of hesitation that has not been consistently addressed in previous research

on disfluency. Fox Tree and Clark (1997) have argued that the production of the non-reduced form of *the* (pronounced "thiy") is used by speakers to signal upcoming difficulty, and is associated with a longer subsequent delay in speaking. While we did not examine the relationship between vowel non-reduction and upcoming delay, we found no evidence to support the assertion that non-reduced forms are directly related to upcoming production problems, as the production of non-reduced forms did not increase significantly prior to naming difficult items. In contrast, reduced forms of the determiners *the* and *a* were produced more often than non-reduced forms, and were prolonged more often prior to naming LF pictures. This suggests that it is the prolongation of a prior vowel final word, rather than quality of the vowel, that is related to subsequent production difficulty. The difference between our results and the assertions of Fox Tree and Clark (1997) could be due to dialectical differences: British English speakers may be more likely to produce non-reduced prolongations than American English speakers. However, this would also suggest that Fox Tree and Clark's arguments about the use of *thiy* as a signal of difficulty are the result of dialectical variations associated with American English speech, and may not apply universally to different dialects of English or to other languages.

According to other research (e.g., Barr, 2001; Clark & Fox Tree, 2002; Shriberg, 1996), the filled pauses *um* and *uh* would also be expected to display different patterns of distribution within an utterance. For example, Shriberg (1996) found that the filler *um* tended to occur more often at the beginning of utterances, while *uh*s were more likely to occur in the middle of an utterance. This study found mixed evidence for these distributional claims. While more *um*s occurred utterance-initially than at other locations within the utterance, overall, more fillers were observed within an utterance than at the beginning. Filled pauses that were produced prior to item names were affected by naming difficulty. However when examined separately, only the filler *uh* was produced significantly more often prior to LF items, despite more *um*s being produced overall. This may suggest that the production of *uh* is more closely associated with the resolution of lexical difficulties, but because relatively few fillers of either type were observed in this position (17 *um*s and 10 *uh*s) across the entire experiment, such claim s require further corroboration.

In summary, this experiment demonstrated that the manipulation of factors that influence the difficulty of picture naming can impact the production of local disfluencies when the pictures are named in the context of spontaneous speech. This effect appears to be restricted to the production of hesitations, rather than repetitions or repairs. Picture naming difficulty was varied both through picture name agreement and the frequency of its dominant name, factors that are thought to influence the speed of selection and retrieval of picture names. However, this experiment did not distinguish between the effects of these two factors, nor did it establish whether the hesitations observed were produced as a result of purely lexical rather than pre-lexical difficulties. These are issues that will be investigated further in subsequent experiments.

5.2 Experiment 2: Separating effects of frequency and name agreement on disfluency production

Experiment 1 demonstrated that specific difficulty associated with an upcoming word can affect the production of associated disfluencies, however, in this experiment the two factors that were employed to manipulate lexical difficulty were not varied independently. There is reason to believe that both lexical frequency and name agreement could have separate effects on the production of disfluency. Word frequency effects on picture naming latencies are well documented (e.g., Griffin & Bock, 1998; Jescheniak & Levelt, 1994; Oldfield & Wingfield, 1965; Wingfield, 1968), and a word's frequency has been considered to have a key influence on the speed of lexical access. Given reliable effects of frequency on picture naming latencies in isolation, it would be reasonable to assume that this factor could also influence the likelihood of overt delays and hesitations during spontaneous speech. Similarly, the relationship between name agreement and picture naming latencies also suggests that this factor could impact the production of hesitations prior to naming a picture in the Network Task. Lachman (1973) observed strong effects of codability on picture naming latencies that persisted, even when the effects of frequency and age of acquisition were controlled for. These factors have been argued

to have different loci of effect within models of production such as that of Levelt et al. (1999), which suggests that they may have additive, or potentially interactive effects on disfluency production. However, there may not be a direct relationship between longer naming latencies and the production of disfluency. Disfluency may only result in instances in which the production system encounters a difficulty that is severe enough to require an overt delay in the continuation of speech. Minor delays in production may be accommodated without adversely impacting the speech plan if, for example, there is enough material in the articulatory buffer to enable the resolution of lexical access while maintaining speech fluency.

Experiment 2 set out to further examine the relationship between disfluency and lexical difficulty by orthogonally manipulating both picture name agreement and the frequency of picture names, in order to determine whether, and if so how, effects of these factors observed in isolated picture naming translate into hesitations and disfluencies in spontaneous speech.

5.2.1 Method

This experiment used the Network Task to further investigate how difficulty naming pictures in the context of spontaneous utterances influenced the rate and type of associated disfluency produced. The design extended from Experiment 1 by orthogonally manipulating the name agreement and name frequency of the pictures used in the Network Task.

Participants

24 students from Edinburgh University (16 women and 8 men) volunteered to participate in the experiment. All were native English speakers and had no speech or hearing disorders

Table 5.5: Mean frequency (Freq) and percent name agreement (NA) statistics for items in Experiment 2. Frequency is presented in counts per million, % NA is the proportion of speakers providing the dominant name for each item. Standard deviations are presented in brackets.

Freq Class	NA Class	CELEX Freq	% NA
High	High	160.2 (102.6)	0.97 (0.02)
High	Low	364.8 (455.5)	0.53 (0.18)
Low	High	11.5 (9.9)	0.99 (0.01)
Low	Low	23.4 (29.0)	0.39 (0.26)

Materials

In order to obtain name agreement statistics that were derived from the same population as the participant pool, 80 object names were selected from the International Picture Naming Project (IPNP) across a range of values corresponding to high and low frequency and and high and low dominant name agreement (calculated as the proportion of speakers that provided the dominant name for a picture). Colour pictures corresponding to these object names were obtained primarily from two sources: Rossion and Portois's (2004) set of "Snodgrass and Vanderwart-like" pictures and a set of public domain clip-art drawings of objects. To determine name agreement statistics for these images, an online pretest was conducted with 86 volunteer participants from the University of Edinburgh. This pre-test was an online survey in which demographic data was recorded (age, gender and first language) and participants were shown each of the 80 pictures in random order and asked to type the name that would ordinarily be used for the object shown. From the responses provided to the pre-test survey, percentage name agreement for each item was calculated. Percentage name agreement scores generated during the pre-test correlated highly with those obtained by E. Bates et al. (2003) on a similar set of pictures with the same names [Pearson's $r = 0.679, p < .001$]. Frequency statistics were obtained for all pre-test items form the CELEX lexical database (Baayen, Piepenbrock, & van Rijn, 1993).

From the pre-test items, a subset of 48 pictures were selected to form an orthogonal matrix of high and low frequency, and high and low name agreement items, with 12

pictures in each condition cell. Table 5.5 provides mean CELEX frequency (in cpm) and percent name agreement scores for the items in each of the cells in the design. A list of all experimental items used and their frequency and name agreement statistics is provided in Appendix A.

Networks: 6 networks were constructed using the 48 experimental pictures. Each network contained 2 pictures from each of the 4 frequency x name agreement conditions in the design, placed so that items that were semantically related or had similar onsets did not follow each other in the path description. The route that the marker took through the network was pre-specified as before, but always passed through 6 items once and two items twice. Unlike Experiment 1, filler pictures were not included in the networks, as descriptions leading to items that the marker passed through twice could be included as experimental utterances on the first pass, when speakers had no prior experience of them. Utterances produced when speakers re-encountered those items were not included in the analysis.

Procedure

The experiment was presented as a non-visual communication task between the participant and a confederate listener, in a similar way to Experiment 1. The experimental procedure followed that detailed in section 4.2.4 of the Methodology chapter.

Transcription and coding

Each network description was separated into ten utterances, where each utterance described the route to the upcoming picture, and included its name, in the same way as described in Experiment 1. The two utterances from each network that contained descriptions relating to pictures that the marker had already passed through were not included in the analysis. Each network description therefore contained eight experimental utterances relating to two items from each item condition.

Transcription and coding was performed by the first author. 20% of transcriptions were check by a second coder who was blind to the experimental manipulations. The first and second coders agreed on 78% of disfluent classifications. Of the 23 discrepancies in classification, 16 related to the presence or absence of a silent pause, while 7 related to a prolongation. After discussion, the author's coding decisions were used in the subsequent analysis. Disfluencies were classified as either filled pauses, silent pauses, prolongations, repetitions or repairs. In addition, the fillers *uh* and *um* were coded separately, as were three types of repair: insertions, substitutions and deletions. Silent pauses were included in the coding scheme, however, only when they did not occur at natural prosodic boundaries and were not co-located with another disfluency, such as following a filled pause. The location of disfluencies was coded into one of three positions, as before: at the beginning of an utterance (i.e., prior to any content words in the utterance), during the path description, or immediately prior to naming the target item (i.e., the Target location).

5.2.2 Results

In total, the 24 speakers produced 1152 experimental (i.e., non-filler) utterances. In 44 (3.5%) utterances, the target item was not named, and these utterances were removed from further analysis. The following analysis was performed on the remaining 1108 utterances.

Picture Name Agreement

On average, high name agreement pictures resulted in 98.4% (s.d. = 3.7%) dominant name agreement, while pictures with low name agreement resulted in 46% (s.d. = 26.3%) dominant name agreement. High name agreement items were given 1.25 (s.d. = 0.45) names on average (including the dominant name), while low name agreement items were given 5.75 (s.d. = 2.25) different names. When low name agreement items were not given the dominant name, intrinsically varying alternative names relating to the same underlying concept were produced significantly more often (in 36.9% of utterances) than names that related to a different concept

Table 5.6: Proportion of utterances in Experiment 2 that contained a disfluency in each of the four identified utterance locations: *Beginning*, *Path*, or *Target*.

Disfluency Class	Utterance Location		
	Beginning	Path	Target
Any Disfluency	10.9%	13.5%	24.5%
Filled Pause	10.7%	3.8%	3.6%
um	6.6%	0.7%	1.1%
uh	4.2%	3.1%	2.5%
Prolongation	0.0%	2.4%	18.5%
Silent Pause	0.1%	0.4%	2.2%
Repetition	0.1%	1.5%	1.7%
Repair	0.1%	7.0%	3.2%

(16.1% of utterances, $t(23) = 2.27, p < .05$). On average, low name agreement pictures were given 3.0 (s.d. = 2.3) intrinsic alternative names and 1.9 (s.d. = 2.3) extrinsic alternative names, however this difference in type of alternative name did not reach significance [$t(23) = 1.51, p = .14$].

Distribution and rates of disfluency production

Overall distribution: Overall, 44% of produced utterances contained one or more disfluencies. Table 5.6 details the proportion of utterances that contained different classes of disfluency either utterance initially, during the path description or prior to the target name. In addition, the proportion of each class of disfluency that occurred in each of the pre-specified utterance locations is given in Table 5.7. The overall distribution of disfluency observed in Experiment 2 was similar to Experiment 1. Just over half of all observed disfluencies were associated with the naming of pictures, and this was primarily due to prolongations, which overwhelmingly tended to be produced prior to picture names. Silent pauses that were not associated with other disfluency were relatively infrequent, but also occurred almost entirely in the Target location. Most filled pauses, in contrast, tended to occur at the beginning of utterances, however the distribution of *um*s and *uh*s appeared to differ. While slightly more *uh*s were observed overall than *um*s (112 *uh*s in total, compared to 92 *um*s), 78% of *um*s were produced utterance initially, whereas *uh*s tended to be more evenly distributed throughout utterances, and were associated

Table 5.7: The proportion of observations of each disfluency class in different locations within an utterance in Experiment 2. Percentages for each disfluency class sum to 100%.

Disfluency Class	Utterance Location		
	Beginning	Path	Target
Any Disfluency	19.7%	27.9%	52.2%
Filled Pause	58.5%	21.5%	20.0%
um	78.5%	8.6%	12.9%
uh	42.0%	32.1%	25.9%
Prolongation	0.0%	11.6%	88.0%
Repetition	2.7%	45.9%	51.4%
Repair	0.9%	69.2%	29.9%

with both descriptions of the path and picture name. Repetitions were relatively infrequent, but were produced with similar frequency during the path description and target name. Repairs, once again, tended to occur during the path description, and were the most common class of disfluency at this location. The majority of repairs during the path description were substitutions, which tended to reflect either lexical or syntactic error repair.

To assess whether properties of the pictures influenced macro-planning and formulation processes, the effects of frequency and name agreement of upcoming pictures were evaluated with respect to disfluencies occurring earlier in the utterance. Neither frequency or name agreement affected disfluency rates at the beginning of utterances [frequency: coefficient $= 0.19, SE = .24; p > .5$ name agreement: coefficient $= 0.69, SE = .45; p > .1$], or during the path description [frequency: coefficient $= 0.27, SE = .36; p > .5$ name agreement: coefficient $= 0.72, SE = .51; p > .1$], providing further evidence that difficulties associated with selecting and retrieving picture names appear to be local, and do not affect earlier planning and formulatory processes.

Differences in type of response: The analyses above described overall disfluency rates across all experimental utterances, regardless of the type of response given to a particular picture. Yet as the name agreement data shows, over 50% of pictures possessing low name agreement were given names that varied from the dominant

name. This was anticipated as an expected consequence of the design. However, because the lexical frequency of the dominant picture name was also manipulated, this presents a potential confound for pictures given other names. The frequency of alternative names produced for pictures may vary substantially from the frequency of the expected picture name. For example, a low name agreement picture with a low frequency name, such as *trophy*, may be given a high frequency alternative name, such as *cup*, which may be retrieved more easily. Therefore, to evaluate the effect of the lexical frequency manipulation on disfluencies produced prior to target names, the analysis of disfluency data was restricted to utterances in which the target was given the dominant (or expected) name.

It should be noted that this restriction also potentially removes the utterances in which speakers would be anticipated to have the most difficulty. Instances where speakers used a subordinate but valid alternative, or an incorrect name for an item would be expected to be stronger candidates for potential difficulty during conceptual or lexical processing. Indeed, when disfluency rates were examined by response type, 47% of intrinsically varying names and 45% of extrinsically varying names were preceded by a disfluency, compared to 21% of utterances containing the dominant name. However, effects of response type could not be fully evaluated as there were not enough utterances containing extrinsic or intrinsic alternative names to allow post-hoc analyses to be performed.

Disfluency prior to picture names: Table 5.8 presents the proportion of utterances by condition that contained different classes of disfluency in the Target location. A logit mixed effects model containing frequency and name agreement as fixed effects were performed on disfluencies occurring immediately before the target item [log likelihood = -349.9]. A main effect of name agreement was found on overall disfluency production [coefficient = $1.35, SE = .308; p < .001$], however the frequency effect was non significant [coefficient = $0.25, SE = ..322; p > .5$]. Adding the frequency x name agreement interaction did not improve the overall model fit [log likelihood = $-349.9, \chi^2(1) = 0.12, p > .5$].

Table 5.8: Experiment 2: Proportion of utterances in each condition that contained a disfluency prior to dominant picture names.

Disfluency Class	Overall Percent (%)	Name Agreement	Frequency	
			High	Low
			Percent (%)	Percent (%)
Total Disfluency	20.9	High	11.0	15.4
		Low	30.1	27.4
Prolongation	16.6	High	7.7	12.6
		Low	24.9	21.3
Filled Pause (um, uh)	2.2	High	1.8	2.0
		Low	2.5	2.4
Um	0.6	High	0	1.0
		Low	0.5	0.8
Uh	1.6	High	1.8	1.0
		Low	2.0	1.5
Silent Pause	1.8	High	0.7	1.0
		Low	2.0	3.5
Repetition	1.9	High	1.9	0
		Low	3.0	2.6
Repair	3.1	High	0.7	1.7
		Low	3.1	7.1

When each disfluency class was analysed, effects of name agreement were found for some classes of disfluency, but no other significant effects of frequency or interactions were observed. Prolongations were produced more often before low name agreement items [coefficient $= 1.27, SE = .325; p < .001$], but not before low frequency names [coefficient $= 0.27, SE = .325; p > .5$]. In addition, a significant effect of name agreement on the likelihood of repairs was also found [coefficient $= 1.80, SE = .789; p < .05$], however no significant effects of name agreement were observed for other disfluency classes [$p > .1$]. In all cases, frequency effects were found to be non-significant [$p > .1$]. It should be noted that for all disfluency classes with the exception of prolongations, overall disfluency rates were very low (i.e., 3% or less, representing less than 25 occurrences of each disfluency class across utterances containing dominant names).

5.2.3 Discussion

Following on from Experiment 1, this experiment sought to dissociate separable effects of picture name agreement and the lexical frequency of picture names on the production of disfluencies associated with the picture name. While the previous experiment indicated that lexical difficulty associated with naming pictures had local effects on disfluency production, this experiment made a clearer distinction between the influence of the two factors used to manipulate naming difficulty. The results of this experiment indicate that picture name agreement had a significant local effect on disfluencies produced, however, the frequency of the word used did not have affect the likelihood of an associated disfluency. Similarly to Experiment 1, the majority of hesitations observed prior to picture names were prolongations. The general pattern of disfluency observed across the entire utterance suggests that hesitation phenomena such as prolongations and silent pauses are much more likely to be associated with local difficulty in the selection and retrieval lexical representations than other disfluencies such as filled pauses and repairs. Filled pauses tend to occur at the beginning of utterances, and could be related to planning or may possibly have a communicative function, while repairs appear to be more closely related to syntactic or lexical reformulation that has less to do with difficulties retrieving words as it does with the mis-selection of words at structural choice points within an utterance.

While significant effects of name agreement was observed for prolongations, the lack of observations of disfluency classes other than prolongations prior to the target name makes it difficult to draw reliable conclusions about different classes of disfluency. In 810 utterances, 117 prolongations were observed prior to correctly named items, yet only 18 filled pauses, 12 silent pauses and 11 repetitions were observed. Therefore, while effects of name agreement on the production of prolongations appear to be reliable, greater power or a more sensitive method of analysis may be required to provide further support for the relationships observed for other classes of disfluency.

The results of this experiment suggest that the increase in hesitations prior to low name agreement picture names may be attributable to greater difficulty associated with the selection of names for an identified concept. However, it is not clear whether the effects of name agreement observed in this experiment are exclusively lexical in nature. As Vitkovitch and Tyrrell (1995) demonstrated, name agreement effects can be attributable to two sources: they may associated with the selection of an appropriate lexical representation for an identified concept; or they may be pre-lexical, due to difficulties in object recognition and conceptualisation. The present study restricted analysed utterances to those in which speakers provided the expected dominant name for the item. This reduces, but does not eliminate the likelihood of speakers encountering pre-lexical difficulty during picture identification in these utterances. Speakers may have still had trouble identifying a picture, but once identified, they produced the expected name for it. In any case, the locus of difficulty did not appear a differential effect on the likelihood of a disfluency prior to alternative picture names. While pictures were given a valid alternative name more often than an inappropriate or invalid alternative, disfluency rates for utterances containing these two types of naming variants were similar.

The lack of a frequency effect on the production of disfluency is perhaps surprising, given the reliability of observed frequency effects on naming latencies in picture naming studies. However, there are a few possible accounts for this null result. First, the frequency of the pictures used in the LF condition may have been too high to generate observable increases in disfluency rates. In this study, the mean CELEX frequency of low frequency picture names was 17.5 cpm. This is low relative to the high frequency pictures (mean CELEX =262.5 cpm), but other studies have used items with much lower frequency names. For example, in Jescheniak and Levelt's (1994) study, their low frequency pictures had a mean name frequency of 6 cpm. This difference could have a substantial effect: as response latencies have been shown to vary with the logarithm of frequency (Oldfield & Wingfield, 1965), a relatively small numerical difference in the frequency of low frequency items can have a significant impact on the time taken to name them. A second, related explanation for this result is that with the exception of extreme states of retrieval difficulty,

such as tip-of-the-tongue states (e.g., Burke, MacKaay, Worthley, & Wade, 1991), a word's frequency may simply not have enough of an impact on the time taken to select and retrieve a name to warrant the production of a disfluency. For example, Jescheniak and Levelt (1994) found a 64ms difference in average response latencies between high and low frequency pictures. If disfluencies are produced as a result of significant delays in production processes, the amount of delay introduced by the retrieval of a low frequency picture name may be easily accommodated within the ongoing processes of production.

5.3 Experiment 3: Visual accessibility and disfluency production

Effects of naming difficulty on disfluencies observed in the previous experiments appear to be due to a picture's name agreement rather than the lexical frequency of the picture name. However, effects of name agreement could be attributable to difficulties during lexical selection or due to earlier problems correctly identifying objects. In Vitkovitch and Tyrrell's (1995) study, both pictures with multiple names and those that were often mis-named took longer to name than high name agreement controls. However, in an object decision task, only pictures that tended to elicit incorrect names resulted in longer recognition times. They suggested that increased naming latencies for these items was likely to be the result of difficulties encountered at the stage of object recognition. If pre-lexical difficulties can increase the time taken to name pictures, then they may also increase the likelihood of disfluency in a spontaneous context. This raises the question of whether disfluency is primarily a result of difficulties associated with language production processes, or whether disfluency may be a more general way for speakers to accommodate any cognitive problems they are engaged in while speaking.

To investigate this possibility, Experiment 3 set out to examine whether disfluency is influenced by pre-lexical processes by manipulating the ease of picture recognition. This was achieved by blurring the images of objects used in the Network Task, which we expected would increase the difficulty associated with identification of the items and access to their related concepts.

5.3.1 Method

The experimental method was identical to that of Experiment 1, with the exception of the materials used in the networks, as described below.

Participants

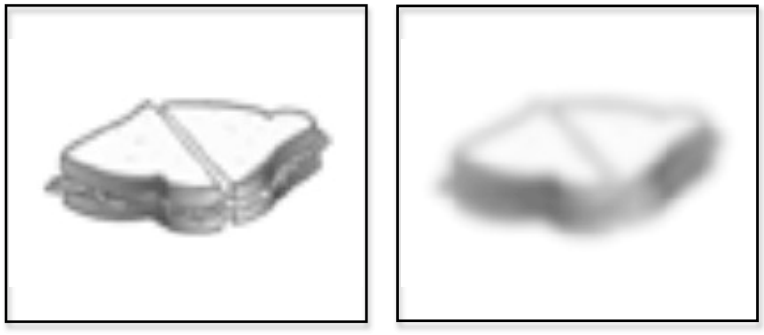
20 students (7 male, 13 female) from the University of Edinburgh participated in the experiment. All were native British English speakers.

Materials

8 visually presented networks were used, based on the networks used by Oomen and Postma (2001), each containing 8 pictures of objects. 64 pictures of objects were selected from Rossion and Portois's (2004) set of 260 of "Snodgrass and Vanderwart-like" objects. The images used were grayscale textured images, which were found to be more suitable for blurring.

All selected images had BNC frequency counts (Kilgariff, 1995) of 10-30 per million (mean = 19 cpm), and a CELEX (Baayen et al., 1993) frequency within the range of 8-45 cpm (mean = 19.4). Image names were high name agreement, having been rated 4 or above on a scale of 1-5 in a rated picture name agreement measure (Barry, Morrison, & Ellis, 1997). Frequency and name agreement statistics for all experimental pictures are provided in Appendix A. The test images were blurred using a Gaussian blur with a 1.5 pixel radius to create a set of 48 clear images and a set of 48 blurred images that had a similar degree of visual naming difficulty. 16 filler items were also selected from the same item set. Figure 5.2 shows an example of the effect of the blurring manipulation on a sample picture.

Each network contained 8 pictures (6 test items and 2 fillers), as described in Experiment 1. 2 sets of networks were created with alternating clear and blurred items, so that the use of clear and blurred images was counterbalanced across participants. Items were positioned so that semantically and phonologically related items were not adjacent to each other.



Clear Blurred

Figure 5.2: An example of the blurring manipulation used on all items. A 1.5 pixel radius blur was applied to the picture of a sandwich, above right.

5.3.2 Procedure

The testing procedure used was the same as for the preceding experiments and as described in section 4.2.4. Each participant was assigned to one of 2 counterbalanced conditions and presented with 8 networks, ordered randomly. As in the previous experiments, the marker took 30 seconds to pass through each network. Upon completion of the experiment, participants were debriefed about the true nature of the experiment and the confederate's part in it.

Transcription and Coding

Descriptions of each network were broken up into 10 utterances, a explained before. Within each utterance disfluencies were identified as occurring either at the beginning of an utterance, during the path description, prior to target items, or elsewhere (as before). Transcription and coding was carried out by the author. 20% of transcriptions were independently verified and coded as in Experiment 1. The mean inter-rater coding agreement was 87%.

5.3.3 Results

Disfluency rates were analysed immediately preceding (and including) the target name, with no other content word intervening between the disfluency and the target. 23 items (7 clear and 16 blurred) were not named and the analysis was performed

Table 5.9: The proportion of utterances in Experiment 3 that contained different classes of disfluency in each of the four identified utterance locations: *Beginning*, *Path*, *Target* or *Other*.

Disfluency Class	Utterance Location			
	Beginning	Path	Target	Other
Any Disfluency	8.4%	18.1%	26.1%	1.6%
Filled Pause	7.4%	5.1%	2.8%	0.5%
um	4.9%	3.0%	1.7%	0.5%
uh	2.5%	2.3%	1.1%	0%
Prolongation	0.3%	3.0%	18.0%	0.3%
Repetition	0.5%	1.3%	0.4%	0.1%
Repair	0.3%	6.3%	1.1%	0.2%

from the remaining 937 utterances. There was no significant difference between numbers of removed clear and blurred items $[F(1, 19) = 2.21, p > .15]$.

Name agreement

In contrast to the previous experiments, all pictures in this experiment possessed high name agreement. This was reflected in the percentage name agreement scores for both clear and blurred items. 87.9% of blurred pictures were given the dominant name for the picture, while 89.2% of clear pictures were given their expected name. There was no significant difference in name agreement between clear and blurred items $[t(47) = .97, p > .3]$. On average, 0.79 alternative names were produced for each clear item, and 0.65 alternative names were produced per blurred item. This difference was also non-significant $[t(47) = .96, p > .3]$

Distribution and rates of disfluency production

Overall distribution: Overall, 45% of utterances contained one or more disfluencies at some stage of the utterance. Table 5.9 presents the proportion of utterances that contained a disfluency in each location. In addition, Table 5.10 details the proportion of each type of disfluency that occurred in each location. General patterns of distribution were similar to those noted in experiments 1 and 2. Overall disfluency rates were higher prior to the target item than elsewhere in the utterance. Prolongations tended to occur predominantly prior to target items, while filled pauses

Table 5.10: The proportion of observations of each disfluency class in different locations within an utterance. Percentages for each disfluency class sum to 100%.

Disfluency Class	Utterance Location			
	Beginning	Path	Target	Other
Any Disfluency	13.3%	34.2%	49.8%	2.7%
Filled Pause	45.7%	33.8%	17.2%	3.3%
um	47.9%	30.2%	16.7%	5.2%
uh	41.8%	40.0%	18.2%	0.0%
Prolongation	1.3%	12.6%	84.8%	1.3%
Repetition	13.9%	63.9%	19.4%	2.8%
Repair	6.3%	78.8%	12.5%	2.5%

occurred most often at the beginning of utterances. Repetitions and repairs were produced most frequently during the path description, however, repetitions were the least common class of disfluency and were very infrequently observed overall. The patterns of occurrence of different disfluency classes suggest that they tend to be associated with different types of production processes, however there is still clearly some flexibility as to their role when they occur in different locations within an utterance.

Disfluency prior to target items: Table 5.11 details the variation by item condition of the percentage of utterances that contained different classes of disfluency prior to the target name. A logit mixed effects model containing the blurring condition as a fixed effect and random subject and items effects demonstrated a main effect of item condition [coefficient $= .372, SE = .183; p < .05$]. Significantly more utterances contained a disfluency when the item was blurred than when it was clear. When individual disfluency classes were examined, blurred items were preceded by significantly more prolongations than clear items [coefficient $= .442, SE = .217; p < .05$]. For the classes of filled pause, repetition and repair, no significant effects of picture blurring were observed either (all $p > .1$).

5.3.4 Discussion

The results of Experiment 3 demonstrated that manipulating speakers' access to object representations through image blurring significantly increased the overall rate

Table 5.11: Proportion of utterances in Experiment 3 that contained a disfluency in the Target location overall, and by disfluency class.

Disfluency Class	Item Condition	
	Clear	Blurred
Any Disfluency	18.1%	25.2%
Filled Pause	2.3%	1.5%
um	1.2%	1.0%
uh	0.3%	1.2%
Prolongation	12.4%	20.1%
reduced *the* or *a*	1.5%	3.8%
non-reduced *thiy* or *ay*	4.5%	5.5%
Repetition	0.4%	0.4%
Repair	1.5%	1.9%

of disfluency production associated with naming these objects. The blurring manipulation did not appear to affect conceptualisation or lexical selection processes, as the proportion of items given the expected name remained consistent across both item conditions. However, the picture blurring manipulation only increased the likelihood of a prolongation prior to the target name: it did not affect the incidence of other classes of disfluency, which all occurred infrequently in the Target location. The data for individual disfluency classes showed similar patterns to those observed in Experiments 1 & 2: prolongations comprised the majority of disfluencies prior to the target name, while filled pauses were relatively infrequent in the Target location and tended to occur at the beginning of utterances. Repairs and repetitions were observed in less than 2% of utterances and showed no effects of picture blurring.

The results of this experiment indicate that pre-lexical difficulties associated with object recognition can also result in disfluency: the production of disfluency is not purely associated with problems of lexical processing. Moreover, this experiment suggests that where a difficulty is encountered within an utterance may have a stronger bearing on the type of disfluency produced than the stage of production at which a difficulty arises. Prolongations appear to be more generally associated with the accommodation of difficulties producing content words, regardless of the source of the problem. One possible reason this is that prolongations may be a more parsimonious way of accommodating difficulty within the context of a noun phrase.

Noun phrases generally contain vowel-final determiners that are easily prolonged. This would allow for the resolution of problems without interrupting the ongoing speech, while other hesitations, such as fillers, would require the overt cessation of speaking.

However, it should also be noted that in both Experiments 1 and 2, while filled pauses were less frequent prior to picture names, observable effects of name agreement were still found. If fillers are considered to reflect more "severe" underlying difficulty, the lack of an effect on fillers in this experiment would suggest that difficulties associated with object recognition introduced by the blurring manipulation may be more easily resolved than those associated with selection and retrieval of words. But such a hypothesis is tentative at best: it relies on a null result that may also be due to any number of alternative factors, and would require further experimental investigation.

5.4 Experiment 4: Disfluency and the speed of picture naming

When a speaker encounters a production difficulty that takes some time to resolve (whether this is a difficulty associated with planning, formulation or lexical access), they may often accommodate it through the production of a disfluency: either a hesitation such as a filled pause or prolongation if they can anticipate the problem before articulation, or a repair (such as a substitution or even a complete restart) if the problem is detected after articulation of the problematic portion of the utterance. Such disfluencies tend to be produced as soon as a difficulty is encountered, and so a disfluency can often be considered a marker of some kind difficulty associated with what a speaker is about to say. The experiments described in this chapter have focused on this issue by manipulating properties of a speakers upcoming speech, such as the frequency or name agreement of to-be-produced words, while examining the pattern of disfluencies that speakers produce. In this way, it may be possible to gain an idea of exactly what factors and their associated underlying processes influence speaker difficulty during speech, and whether there is a

qualitative relationship between types of production difficulty encountered and the types of disfluency that speakers use to accommodate them.

Fundamental to these ideas of speaker difficulty is the notion of delay. When a speaker encounters a production difficulty, it takes time for this to be resolved before the speaker can resume fluent speech. If a delay is short, it may be relatively easy to accommodate without interrupting the flow of speech. If a delay is longer, it is more likely that speech will have to be suspended, potentially through a hesitation, while the underlying problem is resolved. Therefore it is reasonable to hypothesise that there would be a direct relationship between the length of a delay and the likelihood of a disfluency. Indeed, Clark and Fox Tree (2002, also Fox Tree & Clark, 1997) have gone further, suggesting that the production of different types of hesitation are used by speakers to signal different lengths of upcoming delay. Delays in production processes that have an influence on the production of hesitations should also be directly observable in isolated naming studies, and this raises the question of whether there is a demonstrable relationship between the likelihood of disfluency associated with naming a picture in spontaneous speech and the speed of producing that name in isolation.

5.4.1 Method

This naming study was performed to obtain naming latencies for the picture sets used in Experiments 1-3, in order to determine whether the observed effects of lexical and pre-lexical processing difficulty on disfluency likelihood were also evident in the speed of picture naming. If disfluency production is thought to directly reflect underlying production difficulty, we would expect to see a relationship between the effects observed on picture naming and those found during spontaneous speech.

Participants

20 Edinburgh University students participated in the experiment. All were native British English speakers and were drawn from the same population that participated in the original experiments. All were paid for their participation in the experiment.

Materials

The experimental materials consisted of the pictures used in experimental trials in Experiments 1-3. Pictures from each experiment constituted 1 block of trials in the naming study (see the materials sections (5.1.1, 5.2.1, and 5.3.1 for further details about the pictures used). In addition to these three blocks of trials, a practice block consisting of 16 items was presented before the experimental blocks. These practice items were all imageable and nameable objects, however factors such as frequency or codability were not controlled for. As each image had originally been presented as a 70 x 70 pixel image in the original experiments, they were all scaled up to 200 x 200 pixel images to be clearly visible on the computer screen for the sake of the naming task. An unsharpen mask was applied to each picture to smooth lines that appeared excessively pixelated as a result of the image scaling.

Apparatus

The experiment was designed and run using E-Prime on a Research Machines PC. Vocal timing data was recorded through a microphone connected to a button box with millisecond sensitivity. In addition, all sessions were recorded on a Sony Walkman DAT recorder for later coding.

Procedure

Participants were informed that they would see pictures of objects presented on the screen and that their task was to name each picture as quickly and accurately as possible. It was emphasised that they needed to use a name that they felt best described each picture, and to attempt to name all pictures, even if they were not sure what they were. Participants were encouraged to only produce the picture name, and to try not to produce disfluencies or other vocalisations prior to the picture name.

Each trial consisted of a fixation cross presented in the centre of the screen for 1000ms, followed by the presentation of an image and a short concurrent auditory

tone. Each picture remained on the screen for 1000ms after the onset of the participant's response (as triggered by the vocal response to the button box) and was then replaced by the fixation cross in preparation for the next trial. If the voice trigger failed to trigger when the participant named an object and the picture remained on the screen for more than one second after their response, they were instructed to name it again louder.

Participants completed the block of practice trials with the experimenter in the room, to ensure they were speaking loudly enough to reliably activate the voice trigger, and to give them the opportunity to ask any questions before starting the experimental blocks. After this, they were presented with the 3 blocks of experimental trials, with a break in between each. The experiment took approximately 10 minutes to complete.

Coding and errors

All participants naming responses were coded based upon the dominant names for each object used in Experiments 1-3. Items for which the name given was the same as the dominant item name were coded as correct. Items for which participants produced a name that was synonymous with the target name, or could be considered a valid alternative name (including names that expanded on the target name, e.g., water bottle for the item *bottle*) were coded as intrinsic variants. Items for which participants produced a name that could not be considered a valid alternative, or were misidentified (e.g., using the name "shoe" for a picture of a *cap*) were coded as extrinsic variants. Items for which participants did not produce a name, or stated that they did not know were coded as errors and removed from further analysis. All recordings were also checked for trials in which participants produced a vocalisation prior to naming the item, or repeated the name suggesting a failure of the voice trigger. These trials were treated as invalid and removed from further analysis. Finally, responses shorter than 400ms and longer than 4000ms were treated as outliers and removed from the analysis.

Table 5.12: Mean naming latency, error rates and name agreement statistics for HF and LF items in Block 1. The items are the same as those used in Experiment 1.

Item	RT (ms)	Errors	% Correct	% Intrinsic	% Extrinsic
HF	870	18	96.3%	3.1%	0.6%
LF	1407	32	30.8%	45.6%	23.7%

5.4.2 Results

As each block consisted of items drawn from separate experiments they will be considered separately here, and reported as Blocks 1, 2 and 3. There was no intention to compare directly between blocks as the items within each block were drawn from separate sources and were of differing standards of visual quality.

Block 1

The items presented in Block 1 were drawn from Experiment 1, detailed above. In total, there were 36 pictures in this block. 18 pictures were classified as possessing high frequency and high name agreement (HF pictures) and 18 with low frequency and low name agreement (LF pictures).

Errors and Name Agreement: Errors were classed as trials in which either the participant produced a vocalisation prior to naming the item, failed to name the item, there was a timing error, or the response time was outside of the pre-specified range. Table 5.12 details the number of error trials for each item condition. The proportion of utterances given a dominant, intrinsically varying and extrinsically varying name are also provided in Table 5.12. On average, participants used significantly more alternative names for LF pictures (6.44 s.d. = 3.43) than they did for HF items 1.56 (SD = .86; $t(34) = 5.861, p < .001$).

Reaction Times: The naming latency analysis was carried out using a linear mixed effects model on response times for items that participants named correctly or used intrinsic variant of the dominant name. Any trials in which naming latencies fell more than 2.5 standard deviations from a participant's mean were trimmed. On average participants took 537ms longer to name LF items than they did to name HF

pictures. This difference was highly significant [log likelihood = −3824;coefficient = 361.9, $SE = 56.3; p < .001$]. A by-item correlation of observed naming latencies for each picture and average disfluency rates from Experiment 1 demonstrated that average naming latencies for each item correlated highly with disfluency rates in the Target location [Pearson's $r(36) = .739, p < .001$].

Block 2

The pictures in Block 2 were taken from Experiment 2. In total, there were 48 experimental items: 12 items in each of the cells orthogonal in an orthogonal matrix of high and low frequency and high and low name agreement (see section 5.2.1 for details). Because the frequency of picture names was explicitly manipulated as a separate factor, only pictures given the dominant name were included in the response time analysis.

Errors and name agreement: Table 5.13 shows the number of errors per condition as well as the percentage of trials that were given correct, intrinsically varying or extrinsically varying names. When name agreement was examined as a factor, each participant produced significantly more errors for low name agreement items (mean = 8.21) than for high name agreement items (mean = .46) [$F1(1, 19) = 496.5, p < .001, F2(1, 44) = 52.0, p < .001$]. A 2-way ANOVA (frequency x name agreement) performed on the number of names given to each item showed no significant difference between the average number of names for low frequency items (mean = 3.33 names) and high frequency items (mean = 3.45 names): F(1, 44) = .028, however low name agreement items were given on average 4.95 different names, significantly more than the average for high name agreement items (mean = 1.83) [$F(1, 44) = 17.27, p < .001$].

Reaction Times: A linear mixed effects model containing name agreement and frequency as categorical fixed effects was performed on RTs for correctly named pictures demonstrated highly significant effect of name agreement on mean response time. Participants took 315 ms longer to name low name agreement items (mean RT = 1194ms, SD = 302ms) than they did to name high name agreement items (mean

Table 5.13: Mean naming latency, error rates and name agreement statistics for items varying by frequency (Freq) and by name agreement (NA) in Block 2. The items are the same as those used in Experiment 2.

Condition		RT (ms)	Errors	% Correct	% Intrinsic	% Extrinsic
Freq	NA					
High	High	855	2	97.1%	0.9%	0.6%
High	Low	1183	9	66.2%	24.2%	9.5%
Low	High	903	4	98.3%	0.8%	0.8%
Low	Low	1185	13	46.7%	45.8%	7.5%

RT = 879ms, SD = 166ms): [log likelihood = −4900;coefficient = $315.9, SE = 52.1; p < .001$]. However, response times to low frequency items (mean RT = 1044ms, SD = 304ms) were not significantly longer than for high frequency items (mean RT = 1020ms, SD = 271ms): [coefficient = $51.93, SE = 51.96; p > .1$]. Mean response time for items in this study was found to correlate highly with average disfluency rates associated with items observed in Experiment 2 [Pearson's $r(48) = .628, p < .001$.

Block 3

The pictures used in Block 3 were taken from experiment 3, as presentations of blurred and clear versions of each picture were counterbalanced across participants in the original experiment, the same design was used here. Blurred and clear versions of each picture were randomly assigned to one of two counterbalancing conditions, so that each participant saw an equal number of clear and blurred pictures and half the participants named all pictures in each item condition. Pictures that were given correct names or intrinsic variants were included in the response latency measure. Pictures that were mis-named were excluded.

Errors and name agreement: Table 5.14 details number of errors per condition as well as the percentage of trials that were given correct, intrinsically varying or extrinsically varying names. On average, 2.2 names were used for each blurred item, compared to an average of 1.6 names for clear items. A paired samples t-test found this difference to be significant [$t(47) = 2.86, p < .01$]. It should be noted that because of the counterbalancing of items across participants in block 2, each item

Table 5.14: Mean naming latency, error rates and name agreement statistics for clear and blurred items in Block 3. The items are the same as those used in Experiment 3.

Item	RT (ms)	Errors	% Correct	% Intrinsic	% Extrinsic
Clear	893	10	87.2%	10.5%	2.4%
Blurred	974	12	79.6%	11.1%	9.4%

was only presented to 10 participants in each condition, as opposed to 20 for items in blocks 1 and 3.

Reaction Times: On average, naming latencies were 84 ms longer for blurred pictures compared to clear pictures. This difference was found to be significant [log likelihood $= -5967$;coefficient $= 71.9, SE = 16.9; p < .01$]. As the blurring manipulation was counterbalanced across subjects, this 84ms difference must be directly attributable to increased difficulty associated with identifying blurred items. Mean response latencies for clear and blurred pictures were found to correlate with average disfluency rates for the same items from Experiment 3 [Pearson's $r(48) = .334, p < .05$], however this correlation was noticeably weaker than those observed in the first two blocks.

5.4.3 Discussion

The results of the naming studies demonstrate that variations in lexical and prelexical properties associated with the items used in Experiments 1-3 had significant effects on picture naming latencies when they were named in isolation. In Block 1, a large difference in mean naming latencies was observed between HF and LF items. As increases in picture naming latencies associated with low frequency items are generally in the order of 60-100ms (e.g., Jescheniak & Levelt, 1994; Griffin & Bock, 1998), this would suggest that most of the difficulty associated with this manipulation is attributable to differences in picture name agreement. Previous studies that have explicitly examined the effects of picture name agreement on response times have also found similar lengths of delay (e.g., Lachman, 1973; Lachman et al., 1974).

The naming latencies for items in Block 2 also support this conclusion. Pictures with low name agreement resulted in significantly longer naming times, although this effect was smaller in magnitude than observed in Block 1. This difference in magnitude is attributable to differences in name agreement between low name agreement items in the two blocks. LF items in Block 1 were only given their dominant name 30% of the time, while low name agreement pictures in Block 2 possessed on average 56% dominant name agreement. More significant may be the proportions of extrinsic alternative names provided. On average, 23.7% of LF items in Block 1 were given an extrinsic alternative name, while low name agreement items in Block 2 were given an extrinsic alternative 8.5% of the time. While these items were not included in the calculations of naming latencies for either group of items, it suggests that participants may have had greater conceptual difficulties with LF items from Block 1.

While Block 2 exhibited a strong effect of name agreement on response latencies, no significant effects of frequency were observed on response times. However, no effect of frequency on disfluency rates was found for these experimental items either. The similarity between the pattern of effects on picture naming and disfluency suggests that processing delays evident in the picture naming data may be the result of the same processes that are affecting disfluency likelihood, namely the resolution of lexical selection. This also suggests that the lack of an observable frequency effect on disfluency production is likely to be the result of the pictures used in this task, and that disfluency may show a relationship with the frequency of picture names in other item sets that do generate observable frequency effects on picture naming.

However, the lack of a frequency effect in the picture naming data is surprising considering the general robustness of frequency effects in naming studies. However, assuming that frequency effects on naming would be discrete from, and additive to, effects of name agreement also assumes a discrete staged model of production in which these factors operate independently. There is evidence that frequency effects interact with other factors involved in lexical selection, such as contextual constraint (Griffin & Bock, 1998) or contextual probability (Beattie & Butterworth,

1979), and so it is possible that effects of frequency and name agreement may also interact.

The picture naming results for Block 3 provided evidence that object recognition processes can influence picture naming latencies. Blurred pictures took, on average, 81 ms longer to be named than counterbalanced clear versions of the same pictures. This demonstrates a clear effect of early object recognition processes on picture naming. While a greater proportion of blurred items were mis-named, dominant name agreement remained high for blurred items, suggesting that the item blurring slowed object recognition, but did not dramatically affect conceptualisation processes. The magnitude of the blurring effect supports the suggestion made based on the disfluency data, that difficulties introduced by the blurring manipulation were not as difficult to resolve as those relating to the resolution of lexical selection.

The data from the naming studies suggest that the length of delay associated with picture naming is closely related to the likelihood of producing a disfluency. Across all three sets of naming data, significant by-items correlations were observed between naming latencies and disfluency rates observed in the previous experiments. Notably, the correlations between disfluency rates and naming latencies for Experiments 1 & 2, where items varied in terms of their lexical properties, were much higher than when production difficulty varied as a result of pre-lexical processing. This indicates that there is a close link between the time taken to resolve lexical selection processes and the likelihood of an associated disfluency.

5.5 Conclusions

The experiments presented in this chapter set out to use the Network Task to provide an initial characterisation of when different classes of disfluency tend to occur, and more importantly, to investigate how difficulties associated with the selection and retrieval of picture names during spontaneous speech influence the the production of associated disfluency. Experiment 1 demonstrated that lexical difficulties with upcoming words tend to result in the production of hesitations

immediately prior to to the naming of target items. Experiment 2 extended this result, showing that disfluencies were produced more often prior to naming low name agreement pictures. This result suggests that hesitations were produced to allow the resolution of lexical selection problems. However, observed disfluencies may have also been due to problems during conceptualisation, rather than exclusively lexical processes. This experiment could not discount this possibility, despite the fact that the analysis was restricted to correctly named items. Experiment 3 lent support to to this possibility, showing that disfluencies may also occur as a result of pre-lexical processes of object recognition.

Yet mid-utterance hesitations do appear to be directly related to the time course of selection and retrieval of words. Whether different types of hesitation are related to different lengths of anticipated upcoming delay, or whether the type of disfluency produced is simply a function of the point at which a difficulty is identified is an issue that will be investigated further in the following chapter.

Chapter 6

Lexical Influences on Disfluency Production

6.1 Introduction

In the previous chapter, Experiments 1 and 2 focused on how factors that are associated with lexical access affected the likelihood of disfluency production immediately prior to the picture name. These experiments demonstrated that mid-utterance hesitations reflect local difficulties associated with the production of individual words or phrases. The subsequent naming studies suggested that disfluencies produced are related to the length of delay associated with the resolution of underlying production processes. While a strong effect of name agreement was observed both on naming latencies and disfluency rates, surprisingly, no effects of frequency were observed. As frequency effects are consistently obtained in picture naming studies, delays associated with naming low frequency items might also be expected to elicit hesitations. It is possible that the null effect of picture name frequency on disfluency likelihood occurred because the frequency manipulation employed may not have been strong enough to affect disfluency rates. An alternative explanation may be that as frequency effects observed on picture naming latencies tend not to be very large in magnitude (for example, Jescheniak & Levelt, 1994, found a 62ms benefit for pictures with high frequency names), such short delays may be easily accommodated during normal fluent production. As a result, low frequency names may have to be extremely uncommon to generate long enough delays to result in disfluency. Yet the fact that no frequency effect was found in the naming study

suggests that this may have been due to properties of the pictures themselves. For example, the fact that pictures were selected from multiple sources may have introduced unintended variability in factors such as the image quality associated with pictures used in different conditions. This suggests the need to carefully control the selection of pictures, not only to ensure that they are matched for frequency and name agreement, but to ensure that they reliably elicit effects of both of these variables during picture naming.

Another issue associated with the conditions used in the design is the measure of name agreement employed. In the previous experiments, the percentage of pictures assigned the dominant name was used both in item selection, and in the reporting of naming consistency. While percent name agreement provides a measure of how consistently the dominant name is produced, it does not take into account the variety of alternative responses given to a picture. There is potentially a large difference in naming uncertainty between a picture that is given one of two possible names half the time, and another picture also possessing 50% dominant name agreement, but that is given many other names less frequently. In this case, one would expect a greater degree of overall uncertainty associated with the picture possessing more names. An alternative, and more robust measure of name agreement that encapsulates this measure of naming uncertainty is the H-statistic. H is an information statistic computed from speakers naming responses by the formula:

$$H = \sum_{i=1}^{k} p_i log_2(1/p_i) \tag{6.1}$$

where k refers to the number of different names given to each picture and p_i is the proportion of participants that provide each name. Increasing H values reflect decreasing name agreement, and, generally, a decreasing proportion of subjects who provide the same name. A picture with perfect name agreement (i.e., that elicited the same name from every participant) would have an H score of 0, while a picture that is given two names equally often will have an H score of 1. However, H is unbounded and scores higher than 1 are possible if a picture is given several

names relatively infrequently. Unlike metrics such as percentage name agreement, the H-statistic captures a measure of both the consistency with which a particular name is given and the number of alternative names provided. It therefore provides a continuous, and more reliable measure of naming uncertainty. The H-statistic (sometimes also termed the U statistic) has been used in previous studies that have addressed name agreement as a factor in picture naming. For example, Lachman (1973) found a very high correlation ($r = .82$) between H-statistic scores calculated from participants' responses and their response times to a set of pictures that varied from H = 0 to H = 3.79.

Snodgrass and Vanderwart (1980) also used H as a measure of name agreement in their large picture norming study. They found that picture name agreement correlated with both image agreement and visual complexity, but was independent of other factors, such as frequency and familiarity. Snodgrass and Vanderwart (1980) distinguished between what they considered *concept* as opposed to *name* agreement, reflecting differences based on conceptual, rather than linguistic ambiguity. This is essentially the same distinction that we have made in this thesis between *intrinsic* and *extrinsic* variation in naming, and is important as it reflects different sources of disagreement within or outwith the language production system. In the experiments in Chapter 5, attempts were made to restrict the item sets to those that only possessed linguistic variation in name agreement, and disfluency analyses were only performed on utterances that elicited correct or intrinsically varying names. Yet pictures still elicited names reflecting conceptual ambiguity, and hence it was not possible to attribute effects of name agreement to purely lexical processes.

In addition, two other factors that have been shown to impact picture naming latencies, and therefore may have an influence on the production of disfluency were included as *post-hoc* factors within the experimental analyses. These were included to evaluate alternative possibilities about the primary locus of of delays in lexical processing that could result in hesitation or disfluency. Age of acquisition (AoA) effects have been shown to be a significant predictor of picture naming latencies in many studies dating back to the 1970s (e.g., Carroll & White, 1973b; Lachman,

1973; Lachman et al., 1974; Snodgrass & Vanderwart, 1980; Morrison, Ellis, & Quinlan, 1992; Morrison, Chappell, & Ellis, 1997; Barry et al., 1997; Pérez, 2007; Bonin, Chalard, Meot, & Fayol, 2002). Many of these studies have also found AoA effects to be significantly correlated with observed frequency effects (Carroll & White, 1973b; Snodgrass & Vanderwart, 1980; Snodgrass & Yuditsky, 1996; Barry et al., 1997), leading some researchers to claim that most, if not all observed frequency effects in lexical tasks are confounded with AoA, and are really effects of age of acquisition in disguise (Carroll & White, 1973b; Morrison & Ellis, 1995). However, in addition to a main effect of word frequency, Barry et al. (1997) found an interaction between frequency and AoA, suggesting that AoA effects on picture naming may be primarily associated with low frequency names. Other studies have used regression techniques to demonstrate significant, independent effects of both frequency and AoA on picture naming latencies (Lachman et al., 1974; Snodgrass & Yuditsky, 1996), suggesting that these factors may be related but independent contributors to picture naming times. What is clear though, is that AoA does have a strong effect on picture naming: All studies that have examined AoA effects on picture naming have found significant independent effects on picture naming latencies.

More recently, on the basis of differences in the frequency-AoA relationship observed between picture naming and word naming tasks, Brysbaert and colleagues (Belke, Brysbaert, Meyer, & Ghyselinck, 2005; Brysbaert, Van Wijnendaele, & De Deyne, 2000; Brysbaert & Ghyselinck, 2006; Ghyselinck, Lewis, & Brysbaert, 2004) have argued for both a frequency related and frequency independent AoA effect. They found that AoA effects in picture naming tasks were much larger than those that would be expected on the basis of results from tasks focusing on word reading and lexical decision, while frequency effects remain relatively consistent across tasks. While AoA effects have traditionally been thought to arise during phonological encoding (Gerhand & Barry, 1998, 1999), and may result because early acquired words have "more complete", and hence, more easily accessible phonological representations (Brown & Watson, 1987), Brysbaert and Ghyselinck (2006) proposed that frequency independent AoA effects are more likely to be related to processes of

competition at the conceptual level and during lexical selection, as early acquired words appear to be stronger competitors than late acquired words. This proposal would suggest that frequency independent AoA effects are in fact more closely related to the effects of name agreement, which would have a similar proposed locus of effect.

The second *post-hoc* factor that was included was the length (in terms of the number of syllables) of picture names. Klapp, Anderson, and Berrian (1973) originally suggested that the phonological length of a word could impact production latencies, supported by evidence of a small but significant effect of word length on naming latencies (∼15ms) observed in a picture naming task. However, this result has been subsequently called into question. In a similar study, Bachoud-Levi, Dupoux, Cohen, and Mehler (1998) did not find any evidence of a word length effect on picture naming, and argued that Klapp et al. (1973) failed to control for picture familiarity in their original study, which they suggested may account for the benefit that they observed. In contrast, Meyer, Roelofs, and Levelt (2003) demonstrated that word length effects do arise when pictures with monosyllabic and bisyllabic names are placed in separate, rather than intermixed blocks, and that this difference may be due to how speakers modulate their response criterion in such speeded picture naming tasks. However, large scale picture naming studies such as those of Snodgrass and Yuditsky (1996); Barry et al. (1997) have not found word length to be a significant predictor of naming latencies. Effects of word length would be expected to arise during the encoding of phonological forms and the subsequent generation and buffering of the phonetic and articulatory plan, but any delay associated with producing longer picture names in spontaneous contexts may be insufficient to reliably influence disfluency likelihood.

This chapter presents two experiments that investigated how lexical effects related to variations in picture name agreement and frequency impact the production of local hesitation phenomena during continuous speech, using a design that improves both the treatment of frequency and name agreement. In addition, both age of acquisition and word length are included as *post-hoc* factors in the analyses. The aim

was to evaluate how disfluency production is affected by the difficulty of selection and retrieval of names, by manipulating both the frequency and name agreement of high concept agreement pictures, and participants prior exposure to pictures used in the task.

In Experiment 5, participants described networks of pictures which they had not encountered before. In this case, we anticipated that observed disfluencies would reflect delays in the selection and retrieval of words. However, because participants had no prior exposure to the items, we cannot rule out pre-lexical difficulties associated with object recognition or conceptualisation. In Experiment 6, the pictures described in the network task were familiarised through a prior naming task in which participants also received feedback about the dominant name of each picture, priming subsequent production. We expected that the pre-exposure would minimise prior pre-lexical difficulties, but would also facilitate subsequent retrieval of words through repetition priming, resulting in a reduction in disfluency related to lexical retrieval processes. Any differences in the patterns of disfluency observed between experiments could cast light on the ways in which speakers accommodate different types of lexical difficulty while producing an ongoing utterance.

6.2 Experiment 5

Experiment 5 used the same version of Oomen and Postmas (2001) Network Task that was used in the experiments described in the previous chapters. Participants were required to describe networks that contained pictures that varied in name agreement and in the frequency of their dominant name, and we analysed the occurrence of disfluencies immediately preceding the name of each picture.

As name agreement and frequency have both been found to affect naming latencies and error rates in previous picture naming studies (e.g., Jescheniak & Levelt, 1994; Lachman, 1973; Snodgrass & Vanderwart, 1980; Vitkovitch & Tyrrell, 1995),

we expected to observe an increase in disfluency immediately prior to the naming of pictures with low name agreement or low frequency names, relative to high frequency, high name agreement items.

6.2.1 Method

Participants

Twenty four students from the University of Edinburgh were paid for their participation (8 male, 16 female). All were native British English speakers with no speech or hearing difficulties and had normal or corrected-to-normal vision.

Materials

9 visually presented networks (8 experimental and 1 practice network) based on those used by Oomen and Postma (2001) were used in the Network Task. Each network contained 8 black and white line drawings, arranged in different configurations and connected by one, two or three straight or curved lines.

To construct the networks, 56 experimental and 16 filler pictures were chosen from the set of 520 images used in the IPNP (Szekely et al., 2004, see Appendix A.4 for a list of items). Pictures were selected to form a 2x2 matrix of high and low name agreement images with high and low frequency dominant names, with 14 pictures in each condition (for details of all items used in the experiment, see Appendix A). All pictures possessed either monosyllabic or bisyllabic dominant names. We used the H-statistic (Lachman, 1973; Snodgrass & Vanderwart, 1980) as a measure of name agreement. Items with high name agreement had a mean H score of 0.019 (s.d. = 0.027) and 100% dominant name agreement, while items with low name agreement had a mean H score of 1.57 (s.d. = 0.38) and 63% dominant name agreement (as calculated from the IPNP norms). In order to avoid naming difficulties due to conceptual ambiguity, only pictures with high concept agreement were selected. On average, low name agreement pictures possessed 86% concept agreement.

Lemmatized word frequencies were taken from the CELEX English database (Baayen et al., 1993). The mean CELEX frequency for high frequency names was 98.2 cpm (s.d. = 50.5 cpm), with frequencies ranging from 41 cpm (chicken) to 210 cpm (wall). The mean CELEX frequency for low frequency names was 3.9 cpm (s.d. = 2.5 cpm), with frequencies ranging from 1 cpm (radish) to 9 cpm (monk).

Age of acquisition norms for all picture names were taken from the IPNP norms (E. Bates et al., 2003), which used a 3-point scale based on data taken from the MacArthur Communicative Development Inventories (Dale & Fenson, 1996). They presented AoA on a simple 3-point scale, where 1 = words acquired (on average) between 8 and 16 months; 2 = words acquired (on average) between 17 and 30 months; 3 = words that are not acquired in infancy (¿ 30 months). It should be noted, however, that this scale is limited in scope as it only differentiates between words acquired during infancy. As a result it may not be directly comparable to other studies including AoA that have used more extensive rated and objective scales and measurements.

Both AoA and the length of picture names (in terms of their number of syllables) were included in the analyses as *post-hoc* measures. Therefore, while conditions were matched for frequency and name agreement, no explicit controlling of AoA or word length was performed when selecting items. Therefore groupings of items by AoA or number of syllables were inherently unbalanced. However, the use of mixed-effects models allows for the inclusion of unbalanced independent variables within a design, and so both of these factors were included in the experimental analyses.

Individual networks contained 7 experimental pictures and one filler picture in the starting square of the network. In the experiments described in Chapter 5, the starting picture in each network was included as an experimental item. However, across these experiments, over 50% of the pictures that were not named occurred in the first square of the network. As there was no path leading up to these items, participants also did not reliably provide complete descriptions of them. Therefore, pictures used in the starting square of each network were treated as fillers. Pictures

from each item condition were interspersed throughout the networks so that pictures from the same condition, or that were semantically or phonologically similar, would not follow each other in participants descriptions. Four sets of counterbalanced networks were constructed so that, across participants, items from each condition were presented in every square in a network. One practice network was constructed out of non-experimental items that had high name agreement (mean H = 0.03, s.d. = 0.06), and frequencies between the ranges of the two experimental conditions (mean CELEX frequency = 32.9 cpm, s.d. = 7.7).

Each network was associated with a route through the objects, which was indicated by a marker that moved at a constant pace along the lines connecting the pictures. A route always consisted of nine steps, in which the marker passed through six of the pictures once along the route, and passed through two of the pictures twice. Utterances referring to these items on the first pass were included as experimental utterances, while the second pass was designated a filler utterance.

Procedure

The experimental procedure followed that described in Chapter 4. No significant changes were made in terms of the procedural methodology or the instructions given to participants.

Transcription and coding

Each network description was broken up into 10 utterances, where an utterance described the path of the marker up to a particular picture, and included the picture name. *Beginning*, *Path* and *Target* locations were identified within each utterance as before.

Transcription and coding was carried out by the first author. 15% of network descriptions (240 utterances) were independently transcribed and coded by an experienced second rater. The raters coded 97% of the target descriptions identically.

In cases where there was disagreement, the coding decisions of the first author were used in the subsequent analysis.

Analysis

The experiment was designed to examine differences in the occurrence of disfluency associated with pictures assigned to categories of high and low name agreement and high and low frequency. As both frequency and the H-statistic are continuous predictor variables, the use of logit mixed models allows us to model variation along these continuous predictors, as opposed to modelling based upon categorical condition assignment. While items in all 4 experimental conditions were discrete from each other (see Appendix A), the use of H and log frequency as continuous predictor variables can provide a more accurate representation of the influence of these two predictors across the full range of the distribution.

6.2.2 Results

Name Agreement: Table 6.1 presents H-statistic scores and percent dominant name agreement, calculated from total responses made to each item in that task. A 2-way by-items ANOVA on calculated H scores for Experiment 5 demonstrated a significant effect of name agreement [$F(1, 52) = 35.49, p < .001$]: low name agreement items produced higher H scores, confirming that more names were used for low name agreement items than for high name agreement items. No significant effect of frequency on H scores was observed [$F(1, 52) = 3.28, p > .05$]. When alternative name type was examined, there was no difference between the proportion of utterances in which pictures were either mis-named or given valid alternative names [9.2% vs. 12.6%; $t(55) = 0.98, p > .3$].

Removed items: Utterances were removed from the analysis if speakers failed to name the target item. Additionally, because frequency was a continuous variable in the analysis and word frequencies related only to the expected name for a given item, only items where speakers used the dominant item name were included. Out of 1344 experimental utterances, participants failed to name the target item in 34 (2.5%)

Table 6.1: Mean H-statistic scores and percent dominant name agreement calculated from responses given in Experiments 5, 6 and the naming study. The mean proportion of un-named items are also presented for Experiments 5 & 6. Standard deviations are presented in brackets.

Experiment	Name Agreement	Frequency					
		High			Low		
		H -Statistic	% Dom NA	% Removed	*H* -Statistic	% Dom NA	% Removed
Experiment 5	High	0.29 (.57)	95.8 (13.3)	5.0 (4.4)	0.41 (.49)	84.9 (27.5)	16.5 (7.3)
	Low	1.27 (.65)	74.4 (16.9)	27.2 (9.9)	1.91 (1.07)	57.3 (27.8)	45.0 (11.9)
Experiment 6	High	0.06 (.11)	99.4 (1.7)	2.0 (3.8)	0.22 (.41)	92.6 (19.2)	8.5 (5.4)
	Low	0.53 (.37)	93.2 (10.5)	7.6 (6.6)	0.60 (.45)	89.8 (14.0)	11.2 (9.2)
Naming Task	High	0.06 (.22)	99.0 (3.6)		0.21 (.30)	85.1 (28.6)	
	Low	0.84 (.52)	81.2 (16.3)		1.41 (.73)	64.1 (21.5)	

of the utterances. In 281 (20.9%) utterances, speakers used a name other than the dominant name for the target item. 165 (12.3%) of these names were classed as valid alternative names for the target item, while 114 (8.5%) were considered incorrect names. The following disfluency analyses were based on the remaining 1029 correctly named utterances. Table 6.1 details the proportion of utterances removed from each item condition. A logit mixed effects model with frequency and name agreement conditions as fixed effects was fitted to the proportions of removed items [log likelihood = -535.2]. Adding the frequency x name agreement interaction did not significantly improve the predictive power of the model [log likelihood $= -534.8, \chi^2(1) = 0.63, p > .1$], indicating that there was no significant interaction between these factors. More items were removed from low name agreement conditions [coefficient $= 3.11, SE = .78, p < .001$], and low frequency conditions [coefficient $= 1.72, SE = .81, p < .05$].

Disfluency Analysis: Table 6.2 displays the proportion of utterances in each condition containing a disfluency preceding the target name, and the proportion of utterances containing a prolongation, filled pause, repetition, repair or silent pause. In addition, filled pause rates are broken down by type, detailing separate rates for the fillers *um* and *uh*. While some utterances contained more than one class of disfluency, or more than one instance of a particular disfluency prior to the target name, analyses were performed on the likelihood of the presence of a particular class of disfluency, rather than counts of individual disfluencies.

Table 6.2: Proportion of utterances (and standard deviation, in brackets) in Experiment 5 that contained a disfluency overall and for each disfluency class. Filled pause rates are also separated by type (*um* and *uh*).

Disfluency Class	Overall Percent (%)	Name Agreement	Frequency: High Percent (%)	Frequency: Low Percent (%)
Total Disfluency	15.07 (13.77)	High	8.21 (8.95)	11.29 (11.39)
		Low	14.38 (12.48)	26.42 (15.19)
Prolongation	10.52 (11.60)	High	4.13 (5.53)	9.25 (9.70)
		Low	10.33 (12.14)	18.38 (13.2)
Filled Pause	2.61 (6.63)	High	1.54 (3.79)	0.67 (2.26)
		Low	2.50 (4.47)	5.75 (11.25)
Um	1.49 (5.60)	High	0.62 (2.12)	0.33 (1.63)
		Low	1.13 (3.05)	3.88 (10.22)
Uh	1.35 (4.09)	High	0.96 (2.60)	0.33 (1.63)
		Low	1.79 (4.11)	2.33 (6.36)
Silent Pause	1.15 (3.32)	High	0.67 (2.26)	0.71 (2.40)
		Low	0.67 (2.26)	2.54 (5.19)
Repair	1.75 (1.75)	High	2.25 (4.76)	0.67 (2.26)
		Low	1.29 (3.51)	2.79 (7.28)
Repetition	0.74 (2.51)	High	0.92 (2.48)	0.33 (2.36)
		Low	1.21 (3.27)	0.50 (2.45)

Logit mixed effects models were fitted to the proportion of disfluent utterances, overall and for each disfluency class. In all models, H, log frequency were applied as continuous predictors, while word length (in terms of number of syllables) and AoA were applied as categorical predictors of disfluency likelihood. Best-fit model parameters for all models containing significant predictors are summarised in Table 6.3. In each case, interactions between these four fixed effects were tested to see if they significantly improved the base model fit using a χ^2 test, and if so, were included in the model as interaction terms. For the classes of repetitions, repairs and silent pauses, an examination of models including H, log frequency number of syllables and AoA as predictors revealed no significant effects of either predictor ($p > .1$), and are therefore not addressed further.

The proportion of utterances containing a disfluency was fit by the base model model containing H, log frequency, AoA and number of syllables as predictors. Adding the H x log frequency interaction term did not significantly improve the model fit [log

Table 6.3: Best-fit logit mixed effects models containing significant predictors for Experiment 5.

							Random Effects	
Disfluency Class	Fixed Effect	Coefficient	*SE*	Wald *Z*	*p*		Variance	*SD*
Any Disfluency	Intercept	-2.598	0.683	-3.805	< .001	*Item*	0.234	0.483
log likelihood = -383.4	*H*	0.447	0.156	2.861	< .005	*Subject*	0.166	0.400
	Log Frequency	-0.171	0.077	-2.226	< .05			
	No. syllables	-0.202	0.292	-0.689	n.s.			
	AoA	0.468	0.145	3.227	< .005			
Prolongation	Intercept	-3.410	0.793	-4.301	< .001	*Item*	0.245	0.495
log likelihood = -295.7	*H*	0.394	0.179	2.198	< .05	*Subject*	0.269	0.519
	Log Frequency	-0.195	0.087	-2.243	< .05			
	No. syllables	-0.119	0.332	-0.361	n.s.			
	AoA	0.584	0.174	3.357	< .001			
Filled Pause (um, uh)	Intercept	-5.362	1.753	-3.059	< .005	*Item*	0.079	0.281
log likelihood = -100.8	*H*	1.986	0.686	2.894	< .005	*Subject*	0.331	0.575
	Log Frequency	0.274	0.289	0.945	n.s.			
	No. syllables	-0.516	0.614	-0.841	n.s.			
	AoA	0.164	0.300	0.545	n.s.			
	H x Log Frequency	-0.372	0.181	-2.050	< .05			
Um	Intercept	-6.504	2.774	-2.345	< .05	*Item*	0.000	0.000
log likelihood = -60.1	*H*	2.692	1.088	2.473	< .05	*Subject*	1.241	1.143
	Log Frequency	0.382	0.469	0.815	n.s.			
	No. syllables	-0.579	0.859	-0.674	n.s.			
	AoA	-0.043	0.426	-0.101	n.s.			
	H x Log Frequency	-0.5231	0.280	-1.895	0.058			
Uh	Intercept	-6.287	2.097	-2.998	< .005	*Item*	0.446	0.668
log likelihood = -62.4	*H*	0.594	0.469	1.268	n.s.	*Subject*	0.622	0.788
	Log Frequency	-0.032	0.226	-0.142	n.s.			
	No. syllables	0.087	0.853	0.102	n.s.			
	AoA	0.352	0.456	0.772	n.s.			

likelihood $= -382.6, \chi^2(1) = 1.55, p > .1$], demonstrating that these two factors did not interact. Speakers were more disfluent preceding items with low name agreement (i.e. high H) and low frequency names. The coefficient estimates indicate that the odds of a speaker producing a disfluency were 1.56 ($e^{0.447}$) times higher for each unit increase in H, while the odds of a speaker producing a disfluency were .84 ($e^{-0.17}$) times as high for each unit increase in log frequency. In addition, AoA was found to have a strong effect on disfluency likelihood. Pairwise comparisons of different levels of AoA found that this effect was significant between early (i.e. <16 months) and late (i.e. > 30 months) acquired words [coefficient $= 0.95, SE =$

$0.29, z = 3.22, p < .005$], while effects between these levels and the intermediate level were non-significant [$p > .1$]. In line with the results of Barry et al. (1997), we also tested whether this AoA effect interacted with log frequency, however, adding this interaction term to the model did not significantly improve the model fit [log likelihood = $-382.3, \chi^2(1) = 2.12, p > .1$], suggesting that these factors had non-interactive effects on disfluency likelihood. Adding other interaction terms also did not further improve the model fit. Finally, no significant effect of word length in terms of the number of syllables was found. It should be noted that these overall disfluency effects were driven almost entirely by the production of prolongations and filed pauses.

When individual disfluency classes were examined, the incidence of prolongations was also fit using the base model: adding an H x log frequency interaction term (or other interaction terms) did not improve the model fit further [log likelihood = $-298.7, \chi^2(1) = 0.15, p > .1$]. Speakers produced more prolongations preceding low name agreement and low frequency words, as well as words possessing a late AoA.

When the occurrence of filled pauses was examined, including the H x log frequency interaction significantly improved the model fit over the base model [log likelihood = -100.8, $\chi^2(1) = 4.94, p < .05$]. While frequency was not found to be a significant independent predictor of filled pause likelihood, a main effect of name agreement and a significant frequency by name agreement interaction were observed. No effects of word length or AoA were observed on filled pause likelihood. When the fillers *um* and *uh* were examined separately, the occurrence of *um* was also best predicted by a model with fixed effects including an H x log frequency interaction [log likelihood = -60.3]. However, only H was a significant predictor in this model. While the main effect of frequency was non-significant, the frequency x name agreement interaction approached significance, suggesting that effects on *um*s were driving those found for overall filled pause production. In contrast, no factors were found to be reliable predictors of the occurrence of *uh*.

6.2.3 Discussion

The results of Experiment 5 demonstrated effects of both name agreement and frequency of the target name on immediately preceding disfluency. Disfluencies occurred more often before the naming of low name agreement pictures as well as before those with low frequency dominant names, demonstrating effects of the frequency of picture names that were not previously observed in Experiment 2, where frequency was also manipulated.

In addition, while no effects of word length were observed, a significant effect of AoA on disfluency likelihood was found. This effect was independent of, and in addition to effects arising from picture name agreement and the frequency of picture names. It should be noted, however, that this AoA effect was a *post-hoc* observation that was based on a simplified scale that differentiated between words learnt at stages of infancy ranging from less than 16 to over 30 months. In contrast, other studies, such as that of Barry et al. (1997), have used a much wider scale, ranging from less than 2 years up to more than 13 years. Thus, while the scale employed here differentiates between words acquired during and outside of infancy, it provides no further clarity as to when later acquired words were learned, and how variations in later word learning might influence disfluency likelihood. While it is suggestive that AoA effects that have been previously observed on picture naming latencies may also influence the likelihood of disfluency in spontaneous contexts, further work explicitly examining AoA on a more complete scale would need to be performed before strong conclusions could be drawn about the strength and locus of its effect on disfluency production. These results do, however, provide further evidence of a relationship between lexical processing and disfluency: as the demands of selecting and retrieving words during spontaneous speech rise, it becomes more likely that speakers will produce a disfluency to allow for the resolution of these processes.

The majority of disfluencies recorded were prolongations, which were over four times as common as filled pauses. The likelihood of occurrence of prolongations was affected both by the name agreement and frequency of the upcoming target item.

Because participants were describing a route, target names were normally preceded by a preposition and determiner (such as "...to the"), both of which are short vowel-final words that are easily prolonged. Given the speeded nature of the task, this may have favoured the use of prolongations to accommodate short-term production difficulties. When filled pauses were examined, *um*s were primarily associated with low name agreement items, and it is possible that they were produced as a result of difficulties associated with picture recognition or the selection of names. Speakers had particular difficulty with low frequency, low name agreement items, as shown by the significant frequency by name agreement interaction, as well as the high proportion of un-named or mis-named items of this type. In contrast, no effects were observed for *uh*, or for other types of disfluency. This finding suggests that, in contrast with *um*, the production of *uh* in this population is not directly associated with difficulties during lexical access of picture names. However, it should be noted that filled pauses were produced very infrequently prior to target items, and despite stronger methods of analysis, it may not be possible to draw strong conclusions about their underlying cause.

Whereas increased difficulty in assigning a name to a picture clearly affects the likelihood of disfluency, speakers would have also engaged pre-lexical processing as they identified the pictures and formulated their speech plan. Because each picture in the Network Task was encountered for the first time as participants selected, retrieved, and uttered its name, the time taken to select a name may also have been affected by processes involved in object recognition and conceptualisation, creating additional cognitive demand on the production system. For this reason, the observed disfluencies may not be attributable to purely lexical processes.

There are two potential sources of this effect. Participants may correctly identify a picture, but give different correct or synonymous names to it (such as hen for a picture of a chicken), or they may use incorrect names reflecting uncertainty about the concept depicted (e.g., spider for beetle). Despite the fact that items with high concept agreement were used in the networks, pictures still elicited alternative

names that reflected conceptual ambiguity, and therefore the possibility that disfluencies were produced as a result of pre-lexical difficulties associated with object recognition and conceptualisation could not be ruled out. Experiment 6 addresses this issue by attempting to resolve pre-lexical problems in advance of the Network Task.

6.3 Experiment 6

Experiment 6 was designed to reduce conceptual demands on speakers as they performed the Network Task by minimising any pre-lexical difficulties encountered when identifying target pictures. The same networks were used as in Experiment 5, however, before describing the networks, participants performed a picture naming task to familiarise themselves with the pictures used in the networks. In this task, pictures were presented on the screen and after they were named the dominant name for the picture was displayed underneath. By reducing the likelihood of difficulties associated with identifying the pictures, we anticipated that disfluencies produced during the Network Task would reflect difficulties primarily associated with the selection and retrieval of picture names.

The naming task was also expected to have an effect on subsequent lexical processing of pictures during the network task. Lexical priming of picture names during the naming task would also be expected to facilitate subsequent retrieval of names for those items, and hence we would expect to observe a reduction in disfluency rates associated with lexical retrieval processes. However, if participants gave a non-dominant name for an item during the naming task, exposure to the dominant name could actually result in increased lexical competition when that picture is encountered again. It was hypothesised that pictures that were given the dominant name in the Network Task, but had previously elicited an alternative name would result in higher rates of disfluency relative to items that were given the dominant name in both tasks.

The naming task had a subsidiary purpose: It also allowed us to measure naming latencies for the item set. As observed in Chapter 5, if disfluencies accommodate difficulties associated with picture naming, we would expect pictures that elicited longer naming latencies to also be associated with a higher likelihood of disfluency.

6.3.1 Method

Participants

Twenty four University of Edinburgh students were paid for their participation (11 male, 13 female). All were native British English speakers with no speech or hearing difficulties and had normal or corrected-to-normal vision.

Materials

Experiment 6 consisted of a naming task, followed immediately by the Network Task. The pictures used in both were identical to those used in Experiment 5, with the exception that the pictures presented during the naming task were enlarged to 300 x 300 pixels in size, so that they were clearly visible on the computer screen.

Procedure

During the naming task, the participant and confederate were isolated in separate rooms from each other. Participants were informed that during the first part of the study they would see pictures appear on the computer screen that they had to name as quickly and accurately as possible. Once they named the picture, they were given feedback about the dominant name for that picture. They were also told that this was not necessarily the "correct" or only name as some pictures could be given multiple valid names. They were instructed to name pictures as quickly as possible and to use the first name that came to mind for each picture.

Each trial consisted of a 500ms fixation cross, followed by a picture presentation. One second after participants named the picture (recorded from voice onset) the dominant name for that picture was presented on the computer screen, underneath

the picture. Both the picture and name remained on-screen for 2 seconds, followed by a 500ms inter-stimulus interval and presentation of the fixation cross signalling the start of the next trial. Participants were given a practice block of 12 items before moving on to the main block of 72 trials. A pseudo-random picture presentation order was used, so that phonologically or semantically similar items did not follow each other. Filler items used in the network task were interspersed with the experimental items, so that participants named all pictures that they subsequently encountered during the Network Task. Following the completion of this task, the confederate was brought into the room and the Network Task proceeded as described in Chapter 4.

Transcription and coding

Transcription and coding of recordings from the Network Task was carried out in the same way as for Experiment 5. 15% of network descriptions (240 utterances) were independently transcribed and coded by an experienced second rater. The raters coded 94% of the target descriptions identically. In cases where there was disagreement, the coding decisions of the first author were used in the subsequent analysis.

6.3.2 Results

Naming Study

One participant's data were lost due to a recording error. The subsequent analysis was performed on data from the remaining 23 participants.

Name Agreement: Mean H-statistic scores were calculated for each item condition from names given during the naming study (see Table 6.1). A 2-way by items ANOVA found a significant effect of name agreement [$F(1, 52) = 58.6, p < .001, CI = \pm 0.26$], confirming that items ascribed to the low name agreement conditions were given more names than those in the high name agreement conditions.

Table 6.4: Mean response latencies (RT) for each item condition in the naming task. Standard deviations are presented in brackets.

	Frequency	
	High	Low
Name Agreement	RT (ms)	RT (ms)
High	783 (92)	908 (118)
Low	1082 (168)	1467 (389)

A significant effect of frequency was also observed [$F(1, 52) = 7.86, p < .01, CI = \pm 0.26$]. There was no interaction [$F(1, 52) = 2.62, p > .1$].

Naming Latencies: Trials were removed from the latency analysis if the voice key was not triggered immediately upon first response, if participants produced a filled pause before responding, or if they repaired their initial response (75, or 5.8% of trials). In addition, responses more than 2.5 standard deviations outside of a participants mean response time were treated as outliers and removed (36, or 2.8% of trials). Finally, items were removed if participants used a name other than the dominant name for an item (170, or 13.2% of trials). In total, 281 trials (21.8%) were disregarded, and the naming analysis was performed on data from the remaining 1007 trials.

Mean naming latencies for correctly named pictures in each item condition are presented in Table 6.4. Observed naming latencies were broadly in line with those found for the same items in the IPNP study (Szekely et al., 2004). Overall, items with high name agreement were named faster than those with low name agreement [845ms vs. 1275ms, $CI = \pm 81$ms], and high frequency picture names were produced faster than low frequency picture names [932ms vs. 1188ms, $CI = \pm 86$ms]. Naming latency data was fit using a linear mixed effects model containing H and log frequency as continuous linear predictors (log likelihood = -7039.4). Adding an additional interaction term did not significantly improve the fit of the model, [log likelihood $= -7038.2, 2(1) = 2.39, p > .1$]. t values for each fixed effect were determined from this model, along with probabilities based on 10,000 Markov Chain Monte Carlo (MCMC) samples. This model demonstrated a significant effect of H

[coefficient = 194.4, SE = 25.7; t = 7.54, $p < .001$], and a significant effect of log frequency [coefficient = −60.4, SE = 11.9; t = −5.09, $p < .001$] on mean response latencies.

In a separate logit mixed effects analysis, a significant by-items relationship was found between the mean naming latency for each picture and the likelihood of a disfluency, taken from Experiment 5 [log likelihood = -360.1, coefficient = 0.0017, SE = 0.0003, $p < .0001$]. A significant by-items correlation between naming latency and average disfluency rate was also found [$r(56) = .80, p < .001$].

Network Task

Name Agreement: Table 6.1 details H-statistic scores by condition, calculated from total responses made to each item. A 2-way by-items ANOVA demonstrated a significant effect of name agreement [$F(1, 52) = 19.7, p < .001$], with low name agreement items producing higher H scores. There was no effect of frequency condition on H scores [$F(1, 52) = 1.3$]. However, no effects of frequency or codability were observed on percent dominant name agreement [Frequency: $F(1, 52) = 2.12, p > .1$; Name agreement: $F(1, 52) = 1.65, p > .1$] Yet when alternative name type was examined, pictures were mis-named less often than they were given valid alternative names [5% vs. 1.2%; $t(55) = 2.15, p > .05$].

However, prior item exposure during the naming task did result in more consistent naming of items when they were subsequently encountered during the network task. A 3-way by-items ANOVA (experiment x frequency x name agreement) performed on item H-statistic scores between Experiment 5 and Experiment 6 found a main effect of experiment [$F(1, 52) = 42.6, p < .001$], with higher mean H scores in Experiment 5 (mean H = 0.96 vs. 0.35) and an experiment by name agreement interaction [$F(1, 52) = 18.35, p < .001$]. This confirms that the naming task improved naming consistency for low name agreement items in Experiment 6, relative to those observed in Experiment 5.

Table 6.5: Proportion of utterances (and standard deviation, in brackets) in Experiment 6 containing a disfluency overall and for each disfluency class.

Disfluency Class	Overall Percent (%)	Name Agreement	Frequency: High Percent (%)	Frequency: Low Percent (%)
Total Disfluency	16.65 (14.99)	High	12.04 (12.12)	13.46 (13.57)
		Low	16.50 (14.73)	24.58 (16.78)
Prolongation	12.58 (13.16)	High	8.96 (9.66)	9.12 (12.21)
		Low	12.75 (12.43)	19.50 (15.58)
Filled Pause	2.19 (4.82)	High	1.46 (3.56)	2.21 (4.16)
		Low	1.83 (5.31)	3.25 (5.97)
Um	0.92 (3.72)	High	0.88 (3.14)	0.29 (1.43)
		Low	0.87 (4.29)	1.63 (5.08)
Uh	1.34 (3.47)	High	0.58 (1.98)	1.92 (4.05)
		Low	1.25 (3.61)	1.63 (3.92)
Silent Pause	1.81 (3.97)	High	0.88 (2.36)	2.00 (3.54)
		Low	1.79 (3.79)	2.83 (5.54)
Repair	1.58 (3.28)	High	0.88 (2.36)	1.58 (3.16)
		Low	1.87 3.92)	2.00 (3.55)
Repetition	1.20 (3.56)	High	1.17 (3.37)	0.67 (2.26)
		Low	1.00 (2.72)	1.96 (1.96)

Removed items: Out of a total of 1344 experimental (i.e. non-filler) utterances, in 14 (1%) of the utterances the target item was not named. In 86 (6.4%) of the utterances, speakers used a name other than the expected name for the target item. 69 (5.1%) of these names were classed as valid alternative names for the target item, while 17 (1.3 %) were mis-named. Disfluency analyses were based on the remaining 1244 (92.5%) correctly named utterances. Table 6.1 details the proportion of utterances removed from each item condition for Experiment 6. A logit mixed effects model with frequency and name agreement conditions as fixed effects, and subject and item as random effects was fitted to the proportions of removed items (log likelihood = -294.0). Adding the frequency by name agreement interaction did not significantly improve the predictive power of the model, [log likelihood $= -293.8, \chi^2(1) = 0.459, p > .1$]. More items were removed from low name agreement conditions [coefficient = 0.69, SE = .32, $p < .05$]. Effects of frequency on removed items were non-significant [coefficient = -0.25, SE = 0.14, $p = .08$].

Disfluency Analysis: Table 6.5 lists the mean disfluency rates overall, and for each condition, calculated in the same way as for Experiment 5. Logit mixed effects models with H and log frequency as continuous fixed effects, and AoA and number of syllables as categorical fixed effects were fitted to the proportion of utterances containing a disfluency overall, and for each disfluency class. Only models of overall disfluency and of prolongations were found to contain significant predictors at the $p < .05$ level. In all other models, adding fixed effects terms did not significantly improve the model fit over the null model.

Table 6.6 presents logit models that contained significant predictors of disfluency. For overall disfluency rates, adding the H x log frequency interaction term significantly improved the model fit over the base model, demonstrating an increase in the likelihood of disfluency as item H scores increased, in addition to a small but significant interaction between frequency and name agreement. These observed effects were driven by the occurrence of prolongations, which exhibited the same pattern of effects, while other types of disfluency were not reliably predicted by either fixed effect. In addition, no effects of either AoA or number of syllables on disfluency likelihood were observed.

To test whether the naming task reduced disfluency rates between experiments, a model was fit to combined disfluency data from Experiments 5 and 6, including experiment, H and log frequency as fixed effects. For overall disfluency, adding experiment as a predictor did not significantly improve the model compared to one with only H and log frequency as predictors [log likelihood $= 1848.6, \chi^2(1) = 0.187, p > .5$]. However, for the class of prolongations, a fully saturated model including experiment provided an improved fit over a model only containing H and log frequency as predictors [log likelihood $= -959.1, \chi^2(1) = 12.51, p < .05$]. This model contained H and log frequency as significant predictors, and crucially, a significant interaction between experiment and log frequency [coefficient = 0.26, SE = .117, $p < .05$]. This demonstrates a reduction of frequency effects on the production of prolongations between Experiment 5 and Experiment 6.

Table 6.6: Best-fit logit mixed effects models containing significant predictors for Experiment 6.

							Random Effects	
Disfluency Class	Fixed Effect	Coefficient	*SE*	Wald *Z*	*p*		Variance	*SD*
Any Disfluency	Intercept	-1.767	0.520	-3.397	< .001	*Item*	0.031	0.176
log likelihood = -522.4	*H*	0.742	0.193	3.842	< .001	*Subject*	0.489	0.699
	Log Frequency	-0.030	0.077	-0.389	n.s.			
	No. syllables	-0.077	0.205	-0.380	n.s.			
	AoA	-0.073	0.102	-0.725	n.s.			
	H x Log Frequency	-0.127	0.056	-2.259	< .05			
Prolongation	Intercept	-2.729	0.618	-4.416	< .001	*Item*	0.245	0.495
log likelihood = -298.8	*H*	0.801	0.228	3.505	< .001	*Subject*	0.269	0.519
	Log Frequency	0.044	0.093	0.478	n.s.			
	No. syllables	0.075	0.240	0.311	n.s.			
	AoA	0.013	0.121	0.105	n.s.			
	H x Log Frequency	-0.161	0.066	-2.426	< .05			
Filled Pause (um, uh)	Intercept	-3.703	1.188	-3.114	< .005	*Item*	0.001	0.001
log likelihood = -122.6	*H*	0.275	0.272	1.007	n.s.	*Subject*	1.705	1.306
	Log Frequency	-0.188	0.136	-1.378	n.s.			
	No. syllables	0.415	0.502	-0.827	n.s.			
	AoA	0.011	0.250	-0.046	n.s.			

"Switch" items

One question of interest was whether changes in the name that speakers gave to an item between the naming study and the network task had an influence on the likelihood of disfluency when they encountered that item again. During the naming task, participants named each picture and then were presented with the picture's dominant name. When they encountered the pictures again in the Network task, some speakers who had originally given an alternative name "switched" to the dominant name, suggesting that the feedback provided after each picture presentation primed speakers to subsequently use the dominant name when they re-encountered the pictures. In these cases the production of the dominant name would be affected by lexical competition with the name that the speaker produced earlier during the picture naming phase. It was hypothesised that as a result of increased competition in utterances in which speakers switched names, a higher likelihood of disfluency would be predicted during the production of network descriptions, compared to utterances where speakers produced the same name as they had in the naming task.

Compared to a fully saturated model of disfluency likelihood containing H and log frequency as predictors [log likelihood = -626.2], adding "switch" items as a predictive factor significantly increased the model fit [log likelihood $= -621.2, \chi^2(1) = 9.95, p < .005$]. "Switch" items were more than twice as likely to be preceded by a disfluency as other correctly named items, over and above effects associated with frequency and name agreement [coefficient $= 0.719, SE = .21; p < .001$]. For the class of prolongations, adding "switch" items also improved the model fit over the fully saturated model including H and log frequency [log likelihood $= -527.5, \chi^2(1) = 6.71, p < .01$]: prolongations were 1.9 times more likely prior to switched items [coefficient $= 0.646, SE = .24; p < .01$]. However, no significant improvement in model fit was observed for filled pauses [log likelihood $= -147.2, \chi^2(1) = 3.8, p > .05$], indicating that there was no significant increase in filled pause production associated with "switch" items.

6.3.3 Discussion

The results of Experiment 6 highlight a number of issues that provide compelling evidence for a relationship between local hesitation and specific aspects of lexical processing. Most notably, given that the naming study was intended to minimise potential picture recognition difficulties while also priming production of the dominant picture name, the persistent effects of name agreement observed support the earlier findings of an association between disfluencies produced in the vicinity of picture names and the resolution of lexical selection processes.

According to current models that detail the process of picture naming (e.g., Levelt et al., 1999), the naming task would be expected to facilitate both visual-structural processes of object recognition and conceptualisation as well as lexical access during production. However, as other authors have noted (Ellis, Flude, Young, & Burton, 1996; Barry, Hirsh, Johnstone, & Williams, 2001), because of the length of time (more than 5 minutes) and the number of intervening items between presentation in the naming and network tasks, it is unlikely that the observed reduction in disfluency rates is due to purely pre-lexical priming processes. Instead, facilitation

of production is likely to be the result of speakers accessing a picture's name on two separate occasions. This faciliation of lexical access through repetition priming could either affect the process of lexical selection, by increasing the activation strength of the dominant picture name relative to other alternatives, or it could facilitate the phonological retrieval by strengthening the association between an item's lemma and its phonological form.

The effects of name agreement on disfluency production observed in this experiment were consistent with those observed in Experiment 5 and other previous experiments. However, effects of both frequency and AoA that were observed in Experiment 5 were attenuated as a result of repetition priming during the naming task. As both frequency and AoA are thought to influence the speed of lexical retrieval, either by increasing the resting activation of lexical representations for high frequency words (Levelt et al., 1999), or as a result of more complete phonological representations for early-acquired words (Brown & Watson, 1987), this result suggests that the priming during the naming task facilitated subsequent lexical retrieval, rather than lexical selection processes.

Despite priming of dominant picture names, overall disfluency rates were not reduced compared to Experiment 5. Indeed, slightly more prolongations were produced than in Experiment 5, and these continued to be affected by picture name agreement. However, there was no evidence of the previously observed effect of name agreement on the production of filled pauses, and a comparison of filled pauses produced in experiments 5 and 6 showed a reduction in the occurrence of fillers associated with low name agreement items. It is possible that such fillers produced in Experiment 5 were the result of difficulties in object recognition under the pressurised constraints of the task, as filled pauses would be expected to result from more severe processing delays than the production of prolongations. In this case, as all pictures had already been identified in the naming task, one would anticipate a uniform facilitation of subsequent recognition during the network task in Experiment 6. However, it is not possible to make a clear distinction between a lexical and pre-lexical accounts of the reduction of filled pauses.

So far it has been assumed that the naming task would facilitate the selection and retrieval of pictures names during the network task, and the reduction in *um*s suggests that it eased some of the more severe problems speakers occasionally encountered. While priming of the dominant picture name during the naming study may have facilitated retrieval if the speaker was already predisposed to use the primed name, if the speaker had an alternative preferred name for the picture, priming of the dominant name may have actually increased the likelihood of disfluency by increasing lexical competition between the speaker's preferred name for the item and the primed name.

There is evidence for both of these processes in effect: After priming of the dominant name, frequency effects on prolongations that were observed in Experiment 5 were reduced significantly (as shown by the experiment by frequency interaction), indicating that speakers encountered less difficulties associated with the retrieval of picture names. As participants produced the dominant name in the naming task in over almost 75% of responses to low name agreement items, it is assumed that lexical priming reinforced the activation of the associated lexical information, facilitating subsequent retrieval when the pictures were encountered for a second time. However, not all pictures were given their dominant names in the naming task. In some instances, participants produced an alternative name for pictures prior to obtaining feedback about the dominant picture name. When pictures were encountered again in the Network Task, some participants who had previously produced an alternative name "switched" their response to the dominant name. In these cases, participants were twice as likely to produce prolongations before the picture names as in cases where the dominant name was used throughout. These disfluencies may reflect delays resulting from increased lexical competition. If a speaker produced the dominant name during the naming task, the displayed name would increase the activation of that lexical representation, facilitating its selection and retrieval when the picture is re-encountered in the Network Task. However, if a speaker produced an alternative name for the item during the naming study, the displayed name would prime the dominant competitor to the name that they used. As a result, when the picture was encountered again in the Network Task, both the

previously used name and the dominant name would be highly activated, resulting in increased competition for selection.

In the present chapter, we have shown that disfluencies are associated with local lexical difficulty in spontaneous speaking. One possibility is that they are "used" by the speaker to *signal* a difficulty, as claimed by Clark and Fox Tree (2002). In the following chapter, we turn our attention to this issue and present an analysis of the timings of disfluencies recorded during Experiment 5 and their attendant silences.

Chapter 7

Temporal analysis of disfluencies

7.1 Introduction

Throughout this thesis, the focus has primarily been on determining how the likelihood of disfluency is influenced by different lexical processes in language production. Disfluencies have been treated essentially as symptoms of underlying production difficulty. The studies detailed in the previous chapters have demonstrated a relationship between the length of delay associated with lexical production processes and the likelihood that a disfluency will be produced. But so far, disfluencies have been treated in a categorical fashion: we have focused on the presence (or absence) of disfluency. If disfluencies are produced to accommodate delays in speech, then it seems reasonable to assume that the length of the anticipated delay may have an impact on both the type and length of disfluency produced.

In a series of papers, Herbert Clark and colleagues (Clark, 1994; Clark & Fox Tree, 2002; Clark, 2002; Clark & Wasow, 1998; Fox Tree & Clark, 1997) have proposed that speakers *use* hesitations and disfluencies to signal varying length of upcoming delay. They argued that hesitations such as *uh* and *um* are planned by speakers in a similar fashion to other conventional words, and therefore possess meta-cognitive meaning, similar to other interjections, such as *oh* or *ah*. The meaning inherent in a hesitation such as a filled pause is to announce the cessation of speaking and the initiation of a delay in speech. Fundamental to this hypothesis is the idea that speakers can use different types of hesitations to signal different lengths of upcoming

delay, and hence they must be planned and formulated in a similar way to other conventional words.

To support their argument, Clark and Fox Tree (2002) examined the occurrence of different types of filled pause in the London-Lund (LL) corpus of spontaneous conversation. They found that, on average, *um*s are followed by a delay in speaking 60% of the time, while *uh*s are followed by a delay in about 30% of instances. Clark and Fox Tree also measured the length of pauses, measured in terms of *prosodic units*, and found that the delays following *um*s were significantly longer than those following *uh*s (on average, 0.68 vs. 0.29 prosodic units). On the basis of this analysis, they concluded that filled pauses consistently mark a hiatus in speech that displays a regular relationship with subsequent pauses before speech is resumed. However, both of these measures appear to be debatable in their consistency. For example, saying that 30% of *uh*s are followed by a pause suggests that 70% are not. Further, if the reason to produce an *uh* is to signal to a listener to prepare for a delay in speaking, one would expect listeners to need a more regular relationship than this in order to reliably extract the intended meaning from the filler. The second issue is with their measure of delay. In the LL corpus, pauses were marked my professional transcribers as lasted for perceptually coded "prosodic units". The assumption inherent in their analysis is that the length of these prosodic units was consistent across both transcribers and disfluent locations. What Clark and Fox Tree (2002) were actually measuring (and they readily admit it) was the unitised perception of delay.

A similar argument can be made about the claim by Fox Tree and Clark (1997) concerning prolongations. As prolongations of *the* and *a* are the most frequently-observed disfluency in the present thesis, this claim is of particular relevance. Fox Tree and Clark (1997) note that prolongations are sometimes articulated with a full vowel (*the* pronounced "thee"). Because this alternation exists, speakers are assumed to have "chosen" the full vowel, again to signal an upcoming suspension in speech. Once again, the evidence presented in support of this argument comes from

a transcribed corpus in which suspensions of speech are determined perceptually by the transcribers (Bell et al., 2003, present similar evidence).

To investigate the claims concerning the differences between fillers, this chapter presents a subsidiary analysis focusing on the time taken by the disfluencies recorded in Experiment 5 and their attendant pauses.

7.2 Temporal analyses of disfluency

Recordings from 12 participants in Experiment 5 were analysed and transcribed using Praat (Boersma & Weenink, 2008). In contrast to earlier analyses, all disfluencies in any part of each utterance were transcribed and measured. In total, 548 hesitation-type disfluencies were identified (comprising fillers, prolongations, and silent pauses). Except in the case of silent pauses, any pre- or post-disfluency silence was also identified. The duration of each disfluency, together with any associated silence, was recorded for further analysis.

Silence durations were analysed using linear mixed effects models. In each case, models including predictors of interest were compared to a base model including an intercept and per-participant and per-image random variation. Where model fit was reliably improved by the addition of predictors, t values for each effect were determined, along with probabilities based on 10,000 Markov Chain Monte Carlo (MCMC) samples. Table 7.1 details the number and average length of each type of disfluency, as well as the proportion of disfluencies that were followed by a silent pause and their average duration, when observed.

7.3 Fillers

In an explicit treatment of the nature of um and uh, Clark and Fox Tree (2002) claimed that these fillers were part of the *collateral message* in which the speaker is commenting on his or her performances (Clark, 1994, 2002). Their argument that *um* and *uh* serve different functions is in part based on a finding that *um*

Table 7.1: Average duration of disfluencies and subsequent pauses in analysed speech.

Disfluency Type	Count	Average Disfluency Length (ms)	% Pause	Average Pause Length (ms)
um	91	420	35.2%	354
uh	78	251	21.8%	300
the	71	343	38.0%	362
thee	25	381	64.0%	411
a	7	389	28.6%	482
ay	25	406	36.0%	343

tends to be followed by a longer pause than *uh*. However, this finding is based on analyses of three corpora, two of which were written corpora in which the length of post-disfluency pauses was estimated by transcribers using a number of dots. Although some evidence exists to support Clark and Fox Tree's (2002) finding (e.g., Barr, 2001; Fox Tree, 2001), it has recently been called into question (O'Connell & Kowal, 2005).

We identified 169 fillers in the analysed speech, 78 of which were *uh*s. *um*s took significantly longer to utter than *uh*s (excluding silence, log likelihood = -998.3, coefficient = 141.9, $p < .001$; including all silence, log likelihood = -1181.0, coefficient = 167.8, $p < .001$). However, a base model of post-disfluency silence was not improved by adding disfluency type as a predictor ($\chi^2(1) = 0.14$, $p > .1$), showing that silences following *um*s were not significantly longer than those following *uh*s. Thus in the present study there is no evidence to support Clark and Fox Tree (2002).

7.4 Prolongations

Fox Tree and Clark (1997) argued that prolongations with full vowels (e.g., *the* pronounced "thee") are produced as an alternative to the reduced form with a schwa vowel by speakers to signal an upcoming suspension of speech. They showed that "thee" was more likely to be followed by a suspension of speech than its

reduced equivalent. Once again, this finding was based on a written corpus in which suspensions had been hand-transcribed.

In the sample we analysed there were 128 prolongations of *the* or *a*.[1] Of these, 54 were followed by silence (25 following a full vowel). A logistic mixed effects model including random participant and item variation to predict the likelihood of a post-prolongation silence was not improved by the inclusion of vowel quality as a predictor ($\chi^2(1) = 2.00$, $p > .1$). Similarly, vowel quality did not improve a linear mixed effects model predicting the length of the post-prolongation silence ($\chi^2(1) = 0$, $p > .1$). In other words, there was no evidence in our sample to suggest that a full-vowel prolongation signalled an upcoming suspension of speech.

As would be predicted, however, the duration of the prolonged words themselves were predicted by vowel quality, with full vowels taking longer than reduced vowels (log likelihood $= -777.4$, coefficient $= 53.3$, $p = .019$). This finding replicates a standard phonetic effect, improving our estimate of the reliability of the other findings reported.

7.5 Discussion

In the analyses presented here, there was no evidence that different disfluencies signalled differing upcoming delays: the silences following *um* were no longer than those following *uh*, and similarly there were no differences in the likelihood or length of silences following full vs. reduced-vowel prolongations. On the other hand, our analyses established (uncontroversially) that *um* takes longer to articulate than *uh*, and full vowels last longer than their reduced equivalents. One possibility is that it is this difference that drives the perception of a post-disfluency pause in transcribed corpora (e.g., transcribers miscategorise a lengthened vowel as a lengthened post-vowel pause). Alternatively, it may be that differences in post-disfluency silence are only found in certain circumstances (for example, Barr, 2001 only considered utterance-initial *um*s and *uh*s). What is clear is that in the present analysis we found

[1] Although we report prolonged *to* elsewhere, for dialectal reasons reduced and unreduced vowels were not clearly distinguishable in our sample, and *to* was not included in this analysis.

no evidence to support the claims of either Clark and Fox Tree (2002) or Fox Tree and Clark (1997). Given the more sophisticated analyses used in this chapter, our findings add to those of O'Connell and Kowal (2005) in posing a challenge to those claims.

Chapter 8

General Discussion

This thesis presented a series of experiments that set out to explore how difficulties encountered during spontaneous speech production influence the production of hesitations and other forms of disfluency. In particular, the aim was to relate delays associated with particular production processes to the likelihood of occurrence of different types of disfluency. While prior research has primarily focused on the use of corpus studies to examine the relationship between disfluency and speech production, this thesis took an experimental approach, using the Network Task methodology to generate spontaneous utterances that were sufficiently constrained to allow the manipulation of lexical properties associated with the words that speakers used. Given that this task placed the process of picture naming within a context of spontaneous production, the series of experiments presented here set out to investigate whether factors known to influence lexical access also affect the likelihood of disfluency.

8.1 Summary of main findings

The main experiments reported in this thesis investigated how lexical frequency and name agreement, two factors known to affect the speed of selection and retrieval of pictures names (Lachman, 1973; Lachman et al., 1974; Oldfield & Wingfield, 1965; Wingfield, 1968), influenced the production of disfluency when pictures were presented in the context of a spontaneous network description task. Experiment 1

provided an initial characterisation of the relationship between lexical difficulty and the production of mid-utterance hesitations. This experiment demonstrated that increased lexical difficulty associated with an upcoming picture name resulted in an increase in the likelihood of hesitations, and in particular the prolongation of a preceding function word.

Experiment 2 explored this relationship further, by orthogonally manipulating the frequency and name agreement of lexical items. The results of this experiment suggested that the consistency with which a picture is given a particular name has a strong influence on the likelihood of disfluency. This interpretation is supported by a large body of work that has linked disfluency to various measures of uncertainty and choice in speech production. For example, hesitations in speech are more likely to occur prior to words of low contextual probability (Beattie & Butterworth, 1979; Goldman-Eisler, 1961; J. G. Martin & Strange, 1968) or in situations where the speaker has to make choices about what they have to say next (Christenfeld, 1995). In contrast, the lexical frequency of picture names was found to have no effect on disfluency likelihood. This is surprising given the reliability of frequency effects that have been observed in picture naming studies (e.g., Oldfield & Wingfield, 1965; Jescheniak & Levelt, 1994), and their central importance to current models of lexical access. It was hypothesised that disfluencies may be insensitive to the relatively small variations in naming times resulting from changes in the frequency of words. However, a subsequent picture naming study also failed to find a reliable frequency effect, suggesting that any underlying effect may have been outweighed by non-lexical production difficulties associated with the picture set used. These experiments did establish an important relationship between hesitation and the resolution of lexical choice: disfluencies are more likely to occur when speakers have multiple descriptive options to choose from, and it is the resolution of this process that is likely to introduce delays in production. It is possible that this relationship may also extend beyond the resolution of lexical selection to other other aspects of conceptual, syntactic or structural planning decisions.

Experiment 3 explored the relationship between disfluencies and difficulty outwith the language production system, demonstrating an increase in hesitations associated with blurred over clear pictures. Therefore it appears that hesitations in speech are a more general response to difficulty: disfluencies can be produced as a result of any kind of difficulty that introduces a delay in the production of speech, and are not tied to exclusively lexical processes.

Throughout these initial experiments different classes of disfluency tended to be primarily associated with different locations within an utterance, and the location of different disfluency classes was indicative of the role that they play in managing difficulties during ongoing production processes: filled pauses (in particular *um*s) tended to occur most often utterance-initially, suggesting that in these instances they may be related to the macro-planning of upcoming speech, while repairs tended to occur within the path descriptions at points where speakers had to resolve conceptual and syntactic choices relating to competing path descriptions. Prolongations, on the other hand, consistently occurred prior to picture names, suggesting that they are related to the accommodation of short term delays in the selection and retrieval of words. However, it could also be argued that prolongations are most common here simply because content words tend to be closely associated with short vowel final words such as *the* and *a* that are easily prolonged, and that this is the most parsimonious way of introducing a delay into the speech plan without creating an overt interruption to ongoing speech. Such a proposal would suggest that the production of one disfluency over another may be less closely related the type or severity of an underlying difficulty than it is to the location at which it is encountered.

In addition to examining the likelihood of disfluency associated with naming pictures in spontaneous contexts, an isolated naming study using the same experimental materials demonstrated similar effects of naming difficulty, name agreement and visual blurring on picture naming latencies. Significant correlations suggested that these were mapping onto the same processes, and therefore that there is a close relationship between naming difficulty, as measured by the response latencies to picture names, and disfluency likelihood.

The experiments reported in Chapter 6 extended these studies by investigating how lexical frequency and name agreement influenced the production of disfluency using more carefully controlled measures of name agreement and analyses that were more appropriate to the data. In addition, effects of word length and age of acquisition were factored into the analyses. These experiments sought to evaluate how lexical factors influenced disfluency in two different contexts: when speakers must also engage recognition and conceptualisation processes, and when these pre-lexical processes are minimised.

Experiment 5 examined how frequency and name agreement affected disfluencies produced immediately preceding correctly named items when speakers had no prior experience of the items presented in the networks. In contrast to Experiment 2, effects of both frequency and name agreement were observed on the the likelihood of disfluency production. Disfluencies occurred more often prior to naming pictures of low name agreement and with low frequency dominant names. Additionally, an independent effect of age of acquisition was also found on disfluency likelihood, however, no effects of word length were observed. These results indicate that processing delays during both lexical selection and word-form encoding could result in hesitation or disfluency. However, as speakers had no prior exposure to the pictures used, difficulties could have arisen due to pre-lexical problems during object recognition, and so observed disfluencies could not be attributed to exclusively lexical processes associated with the retrieval of their names.

Therefore, Experiment 6 sought to minimise pre-lexical difficulties by familiarising participants with items through a prior naming task, in which they were given feedback about the dominant picture name. Despite improved naming consistency as a result of the naming task, effects of name agreement remained consistent with those observed in Experiment 5, providing convincing evidence that previously observed effects of name agreement arise during the resolution of lexical selection. However, no effects of frequency or AoA were observed in this study, possibly because naming the pictures during the prior task facilitated their production through repetition

priming, which would be expected to benefit phonological encoding processes when the items were encountered again.

8.2 Discussion

The general pattern of occurrence of different disfluency classes varied substantially from other corpus-based studies of disfluency production (such as those of Bortfield et al., 2001; Shriberg, 1994). Most likely, this was due to particular aspects of the methodology that may have influenced speakers' speech strategies. First, it should be noted that while the majority of disfluencies discussed in the experiments in this thesis were prolongations and filled pauses, in all experiments, we were restricting our analyses to disfluencies that occurred immediately prior to a target name. When the whole utterances were examined (as detailed in Chapter 5) other classes of disfluency occurred more frequently elsewhere in the utterances. For example, while we observed a low incidence of repetitions and repairs immediately prior to target names, these disfluencies were observed more frequently during the path description, suggesting they were more closely related to mis-selection of common descriptors for path related terms. Similarly, filled pauses were observed much more frequently at the beginning of an utterance than they were immediately prior to target names. However, the nature of the Network Task did add substantial additional pressure to speakers' formulation and production processes, which was clearly challenging for some participants, and this may have had an influence on the types of disfluency observed. In comparison with the unconstrained spontaneous speech used in other corpus studies, speakers were under significant time pressure to plan, formulate and produce their descriptions in order to keep up with the marker. This may have forced speakers to adopt a more incremental approach to the planning and production of speech units than would normally be the case in conversational speech, while also reducing available resources for self-monitoring. Such a production strategy would be likely to favour short-term hesitations such as prolongations during an utterance, as they do not require speakers to interrupt their flow of speech. Furthermore, while other studies have often used spoken dialogues

as their data source, the Network Task essentially resulted in a communicative monologue from a single speaker. Therefore, observed disfluencies were most likely to have been due solely to difficulties that the speaker was encountering, rather than disfluencies that may be related to conversational aspects of turn taking or addressee monitoring in dialogue (Clark, 1994; Clark & Krych, 2004).

The majority of disfluencies observed in the present experiments were prolongations, although these forms of hesitation have often been overlooked in studies of speech error and disfluency (e.g., Beattie & Butterworth, 1979; Blackmer & Mitton, 1991; Bortfield et al., 2001; Maclay & Osgood, 1959; Oomen & Postma, 2001; Shriberg, 1996), although some recent studies have highlighted their association with planning problems (e.g., Bell et al., 2003; Fox Tree & Clark, 1997). While Fox Tree and Clark (1997) have made a distinction between the use of reduced and non-reduced forms of *the* to signal upcoming delays, in our investigation we showed that they did not differentially signal a following silence. For this and other (dialectical) reasons, our analyses of Experiments 1-6 did not differentiate between reduced and non-reduced prolongations. But our findings do suggest that whether reduced or non-reduced, they play an important role in the accommodation of short-term production difficulties, particularly as they do not require a speaker to halt or interrupt their flow of speech. In contrast, fillers occurred relatively rarely in our experiments. This may be because in normal conversational speech they are typically found elsewhere within an utterance.

It is important however to point out that *um*s and prolongations were observed in all conditions; that is, they cannot be considered to provide unequivocal evidence that a certain type of difficulty has been encountered, and the speaker may of course produce them for a variety of reasons (for example, utterance-initial fillers have been associated with uncertainty: Brennan & Williams, 1995). Indeed, other types of hesitation and disfluency, such as the filler *uh*, repetitions, and repairs were also observed, but showed no significant relationships with the lexical factors manipulated in our experiments. This may be a matter of power, as these types of disfluency occurred relatively infrequently prior to picture names, or their production could

be associated with other processes, as suggested by their preponderance in other parts of the utterances.

The fact that frequency and name agreement effects appeared to have different influences on prolongations and filled pauses throughout the thesis suggests that the production of prolongations and of the fillers could be associated with different types of production difficulty. In Experiments 5 and 6, name agreement affected the likelihood of a prolongation prior to the picture name, despite the more consistent naming of low name agreement items in Experiment 6. We propose that this increase in prolongations associated with low name agreement items is due to increased delays in lexical selection processes. While prior exposure to each picture and its dominant name would be expected to facilitate the subsequent selection when the picture is encountered again, alternative candidate names would still be activated for low name agreement items, impacting the time taken to resolve selection, and resulting in a delay before retrieval and articulation can commence. In isolated naming studies (e.g., Lachman, 1973; Lachman et al., 1974; Snodgrass & Yuditsky, 1996; Vitkovitch & Tyrrell, 1995), this delay is silent; in continuous speech, this delay must be accommodated where it is encountered in the context of ongoing speech. The filler *um*, in contrast, was only affected by name agreement in Experiment 5, when speakers would also have had to resolve any pre-lexical difficulties associated with identifying pictures and accessing related concepts prior to lexical access. It is possible that the higher rate of production of *um*s is due to difficulties with items possessing low concept name agreement. In a similar way to the object decision latencies observed by Vitkovitch and Tyrrell (1995), deciding what a picture is takes time, which increases the likelihood of a longer hesitation in continuous speech. Where the object has already been identified in Experiment 6, the delays are minimised, resulting in a commensurate reduction in the likelihood of producing an *um*.

The only direct and isolated effect of frequency was on the likelihood of a prolongation in Experiment 5. The fact that following the naming task no effects

of frequency and AoA on prolongations were observed in Experiment 6, while effects of name agreement on prolongations were unaffected provides support to the proposal that frequency and name agreement are influencing different processes in production. Furthermore, it is likely that the effects of both frequency and AoA were attenuated as a result of repetition priming of picture names from the naming task, suggesting that these factors both operate at the level of word-form encoding. While effects of name agreement are thought to impact lexical selection processes, it is possible that prolongations observed in Experiment 5 prior to low frequency picture names were the result of delays in word-form retrieval in situations where the speech plan was under pressure to keep up with articulation. Frequency may have had no effect in Experiment 6 because prior activation of lexical representations during the preceding familiarisation task may have overridden any underlying differences in resting activation and speed of retrieval between low and high frequency word-forms. This is congruent with picture naming studies that have demonstrated a reduction in frequency effects on response latencies over multiple presentations (e.g., Oldfield & Wingfield, 1965; Griffin & Bock, 1998). Additionally, we note that in the naming study low frequency items took on average 256ms longer to name, compared to 430ms longer for low name agreement items. Hence, the effects of exposure to the dominant name may have been insufficient to override any differences between low and high name agreement items.

There was also clear evidence that the distribution of prolongations differed from that of fillers across experiments. One possible account for this difference is in terms of the anticipated time needed to resolve a difficulty. If the type of disfluency produced is related to the length of the anticipated delay required to resolve a problem in production (Clark & Fox Tree, 2002; Smith & Clark, 1993), then the time required to resolve lexical selection difficulties may not only be shorter, but more predictable than for earlier pre-lexical difficulties. We should point out that post-disfluency silences do not differ in our data, contrary to existing claims (Clark & Fox Tree, 2002; Fox Tree & Clark, 1997). Moreover, in the data examined in Chapter 7, prolongations lasted longer than fillers (395 vs. 341 ms; coefficient $= 46.1$, CI $= 24.1 - 69.3$, $p < .001$), giving at least some credence to the view

that the time required may be anticipated. In fact, the extra time afforded by a prolongation (of an existing word) may still be less than that afforded by a new filler included into the utterance.

According to Roelofs (1992), lexical selection proceeds through a process of spreading activation between lexical concepts in which the time-course of selection can be predicted by Luce's choice ratio (Luce, 1959). Given the resting activation and relative activation strengths of alternative lexical concepts, this could allow the point in time at which a single lexical candidate is selected to be predictable in advance. If the production system utilises a similar process, it may be possible to determine whether lexical selection can be fluently resolved in the time available before the commencement of articulation, and if not, estimate the delay required. Indeed, spreading activation accounts of speech planning processes often also include some kind of simple monitoring mechanism that identifies situations there is a high amount of competition between alternative nodes of the same class (Postma, 2000 Nov 16), and such a monitor could also signal the need for accommodation of a delay in selection processes. In these cases, a prolongation may result if it provides sufficient additional time to resolve any selection process. Clearly, other factors will enter into this calculation: For example, the time taken to resolve selection and create an articulation plan will additionally depend on the time-course of phonological encoding, in addition to the cognitive load that the production system is under at the time. In contrast, the resolution of conceptual difficulty (here, deciding what a picture represents) may not be amenable to such a calculation, and may therefore engender the production of a different type of disfluency, such as a filler.

Alternatively, the production of filled pauses could be under a greater degree of strategic control than the production of prolongations and other hesitations. While we would not go as far as Clark and Fox Tree (2002) in considering filled pauses as genuine words that need to be explicitly planned, speakers may have some awareness of the difficulties they encounter, particularly during conceptualisation and formulation of upcoming speech. This could be due to internal monitoring of conceptual and formulatory processes, to check, for example, the validity or appropriateness

of a planned constituent. Internal conceptual monitors (e.g., in the monitoring accounts of Levelt, 1983; Blackmer & Mitton, 1991) have been been proposed as a checking mechanism to compare the appropriateness of the planned speech output relative to the intended message. But as the conceptual monitor would be active as soon as the intention to speak is initiated, it is also possible that it is sensitive to difficulties or delays enountered in the pre-lexical formulation of the message-level plan. As message-level formulation is a more centrally mediated process, such difficulties may be dealt with in a more strategic fashion than later stage lexical difficulties, and hence be more open to accommodation through a filled pause. In comparison, hesitations such as prolongations would be considered to be the result of more automatic processes during lexical access that result from an immediate requirement to suspend speech to accommodate a production delay.

8.3 Conclusion

This thesis establishes a relationship between speakers' use of disfluency and the different problems of lexicalisation they are attempting to accommodate, using an experimental paradigm which allows us to manipulate the words that speakers are likely to use during spontaneous, unplanned speech. In particular, it points to the important role played by the prolongation of function words such as *the* in order to resolve lexical difficulty. Moreover, it suggests that issues of choice and competition (what to call a picture), as opposed to frequency (how easy a name is to retrieve) are among the primary causes of within-utterance disfluency. However, it leaves questions about what it is that causes speakers to use different types of disfluency to resolve different production problems open to further research. Integrating other approaches examining the incremental nature of spontaneous speech, such as the eye tracking paradigm of Griffin (2001) with the Network Task could provide further information about both the time-course of disfluencies and of associated production processes.

Appendix A

A.1 Items used in Experiment 1

Lemmatised frequency counts were taken from the British National Corpus (Kilgariff, 1995). Percent dominant name agreement data was obtained from the Beckman Spoken Picture Naming norms (Griffin & Huitema, 1999).

High Frequency, High Name Agreement			Low Frequency, Low Name Agreement		
Item	Frequency (BNC cpm)	% Dom NA	Item	Frequency (BNC cpm)	% Dom NA
bird	90	91%	chalice	1	20%
camera	38	100%	cradle	4	43%
cat	54	98%	cylinder	12	38%
computer	170	96%	dresser	3	43%
cow	25	91%	dynamite	1	32%
desk	49	91%	flasks	4	33%
dog	124	96%	gavel	0	43%
guitar	34	98%	limousine	2	38%
gun	55	91%	metranome	0	43%
hand	532	100%	molecules	19	31%
heart	152	98%	pram	3	48%
house	490	98%	rectangle	4	31%
pencil	14	100%	rosary	1	33%
star	91	100%	scythe	1	11%
sun	95	98%	sheild	14	39%
table	231	96%	suitcase	8	37%
tree	147	98%	tomahawk	0	28%
window	193	100%	weights	9	41%
Mean	143.5	96.7%	Mean	4.7	35.1%
SD	147.3	3.4%	SD	5.4	9.1%

A.2 Items used in Experiment 2

Lemmatised frequency counts were taken from the CELEX lexical database (Baayen et al., 1993). Percent dominant name agreement data was obtained from the International Picture Naming Project (E. Bates et al., 2003). In addition, percent name agreement data that was obtained forom the pre-test survey is also provided.

High Frequency, High Name Agreement				Low Frequency, High Name Agreement			
Item	Frequency (cpm)	% Dom NA (survey)	% Dom NA (IPNP)	Item	Frequency (cpm)	% Dom NA (survey)	% Dom NA (IPNP)
bed	269	98%	100%	banana	8	100%	100%
bone	69	99%	100%	bat	14	100%	100%
book	434	99%	100%	bra	6	98%	100%
chair	136	99%	100%	butterfly	10	99%	100%
dog	115	94%	100%	comb	5	100%	100%
dress	87	97%	100%	ghost	31	98%	100%
fish	163	98%	100%	kite	5	99%	100%
flower	93	100%	100%	pear	6	100%	100%
horse	132	100%	100%	skeleton	12	99%	100%
sun	152	94%	100%	spider	7	97%	100%
train	81	95%	100%	witch	32	98%	100%
tree	191	95%	100%	zebra	2	100%	100%
Mean	160.2	97.3%	100.0%	Mean	11.5	98.8%	100.0%
SD	102.6	2.1%	-	SD	9.9	1.2%	-

High Frequency, Low Name Agreement				Low Frequency, Low Name Agreement			
Item	Frequency (cpm)	% Dom NA (survey)	% Dom NA (IPNP)	Item	Frequency (cpm)	% Dom NA (survey)	% Dom NA (IPNP)
bag	80	67%	84%	beetle	8	16%	44%
block	54	35%	55%	crackers	2	26%	84%
bottle	116	42%	90%	fire hydrant	1	34%	71%
boy	349	26%	90%	hamburger	5	16%	84%
city	257	22%	85%	ice cream	-	64%	52%
girl	438	63%	92%	pot	36	72%	73%
glass	145	80%	71%	priest	49	30%	43%
letter	206	59%	68%	safety pin	17	79%	53%
man	1629	49%	94%	tape recorder	9	13%	75%
picture	174	65%	83%	trophy	4	72%	50%
wine	79	71%	67%	wheat	29	35%	58%
woman	850	51%	69%	wood	97	11%	55%
Mean	364.8	52.5%	79.0%	Mean	23.4	39.0%	61.8%
SD	455.5	18.4%	12.5%	SD	29.0	25.7%	14.8%

A.3 Items used in Experiment 3

Frequency counts were obtained from the CELEX lexical database (Baayen et al., 1993). Percent dominant name agreement data was obtained from Barry et al. (1997).

Item	Frequency (cpm)	% Dom NA	Item	Frequency (cpm)	% Dom NA
balloon	7	100%	jug	10	88%
barrel	21	91%	ladder	16	96%
basket	24	96%	lamp	35	96%
bell	42	100%	leaf	81	100%
belt	27	96%	lemon	15	100%
boot	39	96%	lion	25	100%
bow	13	82%	lock	15	88%
bowl	33	100%	mouse	18	82%
button	26	100%	pen	26	96%
cake	34	100%	pipe	31	100%
candle	16	100%	pot	36	81%
cap	37	91%	rabbit	19	96%
cat	67	100%	refrigerator	10	93%
clock	39	100%	sandwich	10	100%
cow	40	100%	seal	14	88%
crown	24	100%	sheep	40	96%
doll	25	71%	shirt	61	100%
fence	30	91%	shoe	79	100%
flag	26	100%	skirt	29	100%
flower	93	100%	sofa	6	67%
fork	15	100%	suitcase	19	77%
fox	15	100%	sweater	15	83%
glasses	32	86%	swing	18	96%
guitar	8	98%	tie	34	100%

A.4 Items used in Experiments 5 & 6

Frequency counts were taken from CELEX (Baayen et al., 1993). *H*-scores and % dominant name agreement were obtained from the IPNP (E. Bates et al., 2003).

High Frequency, High Name Agreement					Low Frequency, High Name Agreement				
Item	Frequency (cpm)	*H*-Score	% Dom NA	AoA	Item	Frequency (cpm)	*H*-Score	% Dom NA	AoA
ball	111	0.00	100%	1	bra	6	0.00	100%	3
bell	42	0.00	100%	3	broom	8	0.00	100%	1
bone	69	0.00	100%	3	cactus	3	0.06	100%	3
chair	136	0.00	100%	1	comb	5	0.00	100%	1
desk	91	0.00	100%	3	giraffe	2	0.03	100%	1
dog	115	0.00	100%	1	kite	5	0.00	100%	3
fish	163	0.03	100%	1	pear	6	0.00	100%	3
flower	93	0.00	100%	1	pumpkin	2	0.03	100%	2
horse	132	0.00	100%	1	saw	1	0.03	100%	3
leaf	81	0.06	100%	3	scissors	4	0.08	100%	1
moon	59	0.08	100%	1	shovel	4	0.03	100%	1
ring	49	0.00	100%	3	turtle	4	0.00	100%	1
tree	191	0.03	100%	1	zebra	2	0.03	100%	2
wheel	44	0.00	100%	3	zip	2	0.06	100%	1
Mean	98.3	0.01	100.0%	1.86	Mean	3.9	0.03	100.0%	1.86
SD	45.7	0.03	-	1.03	SD	2.0	0.03	-	0.95

High Frequency, Low Name Agreement					Low Frequency, Low Name Agreement				
Item	Frequency (cpm)	*H*-Score	% Dom NA	AoA	Item	Frequency (cpm)	*H*-Score	% Dom NA	AoA
bird	103	2.42	80%	1	beetle	8	2.01	44%	1
boat	76	1.20	71%	2	clamp	2	1.97	50%	3
branch	94	1.48	68%	3	cork	5	1.03	85%	3
chicken	41	1.25	72%	1	corkscrew	1	1.75	50%	3
cloud	56	1.06	81%	2	dustpan	1	1.57	69%	3
coat	61	1.22	56%	1	hoe	3	1.10	77%	3
floor	176	1.76	52%	3	mixer	2	2.09	39%	3
gate	69	1.10	60%	3	monk	9	1.87	43%	3
letter	206	1.66	68%	3	plank	7	1.79	55%	3
present	55	1.52	67%	2	pliers	1	1.22	60%	3
scale	82	2.02	56%	3	radish	1	1.77	58%	3
shirt	61	1.16	76%	1	syringe	2	1.45	63%	3
soldier	83	1.85	69%	3	trophy	4	1.72	79%	3
wall	210	1.90	38%	3	trumpet	8	1.11	96%	1
Mean	98.1	1.54	65.3%	2.21	Mean	3.9	1.60	62.0%	2.71
SD	56.7	0.41	11.8%	0.89	SD	3.0	0.36	17.1%	0.73

References

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley-Interscience.

Alario, F. X., Costa, A., & Caramazza, A. (2002). Frequency effects in noun phrase production: Implications for models of lexical access. *Language and Cognitive Processes*, *17*, 299-312.

Alho, K. (1995). Cerebral generators of mismatch negativity (mmn) and its magnetic counterpart (mmnm) elicited by sound changes. *Ear and Hearing*, *16*, 38–51.

Almeida, J., Knobel, M., Finkbeiner, M., & Caramazza, A. (2007). The locus of the frequency effect in picture naming: When recognizing is not enough. *Psychonomic Bulletin & Review*, *14*, 1177-1182.

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

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 comprehension. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *33*, 914-930.

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

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database.* Centre for Lexical Information.

Bachoud-Levi, A., Dupoux, E., Cohen, L., & Mehler, J. (1998). Where is the length effect? a cross-linguistic study of speech production. *Journal of Memory and Language*, *39*, 331-346.

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

Barry, C., Hirsh, K., Johnstone, R., & Williams, C. (2001). Age of acquisition, word frequency, and the locus of repetition priming of picture naming. *Journal of Memory and Language*, *44*, 350-375.

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency and name agreement. *Quarterly Journal of Experimental Psychology*, *50A*, 560-585.

Bates, D. M., & Sarkar, D. (2007). *lme4: Linear mixed-effects models using 24 classes* (R package version 0.9975-12).

Bates, E., D'Amico, S., Jacobsen, T., Szekely, A., Andonova, E., Devescovi, A., et al. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, *10*, 344-380.

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.

Belke, E., Brysbaert, M., Meyer, A. S., & Ghyselinck, M. (2005). Age of acquisition effects in picture naming: evidence for a lexical-semantic competition hypothesis. *Cognition*, *96*, B45-B54.

Bell, A., Jurafsky, D., Foster-Lussler, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in english conversation. *Journal of the Acoustic Society of America*, *113*, 1001-1024.

Blackmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, *39*, 173-194.

Boersma, P., & Weenink, D. (2008). *Praat: doing phonetics by computer (version 5.0.2) [computer program] heep://www.praat.org.*

Bonin, P., Chalard, M., Meot, A., & Fayol, M. (2002). The determinants of spoken and written picture naming latencies. *Br J Psychol*, *93*(Pt 1), 89–114.

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

Branigan, H. P., Lickley, R. J., & McKelvie, D. (1999). Non-linguistic influences on rates of disfluency in spontaneous speech. In *Proceedings of the 14th International Congress of Phonetic Sciences.* Berkely, CA.

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 states of speakers. *Journal of Memory and Language*, *34*, 383-398.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Society*, *88*, 9-25.

Brown, G. D., & Watson, F. L. (1987). First in, first out: word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Mem Cognit*, *15*(3), 208–216.

Brutten, E. (1963). Palmar sweat investigation of disfluency and expectancy adaptation. *Journal of Speech and Hearing Research*, *6*(40-48).

Brysbaert, M., & Ghyselinck, M. (2006). The effect of age of acquisition: Partly frequency related, partly frequency independent. *Visual Cognition*, *13*, 992-1011.

Brysbaert, M., Van Wijnendaele, I., & De Deyne, S. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychol (Amst)*, *104*(2), 215–226.

Burke, D. M., MacKaay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip ot the tonge: what causes word finding failures in young and older adults. *Journal of Memory and Language*, *30*, 542-579.

Butterworth, B. L. (1975). Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research*, *4*, 75-87.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*, 177-208.

Carroll, J. B., & White, M. N. (1973b). The effects of age of acquisition on an object classification task. *Quarterly Journal of Experimental Psychology*, *25*, 85-95.

Chafe, W. L. (1980). *The pear stories: cognitive, cultural, and linguistic aspects of narrative production*. Norwood, N.J.: Ablex Pub. Corp.

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

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

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press. Available from `http://www.loc.gov/catdir/toc/cam023/95038401.html`

Clark, H. H. (2002). Speaking in time. *Speech Communication*, *36*, 5-13.

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

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*, 62-81.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, *37*, 201-242.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39.

Collard, P., Corley, M., MacGregor, L. J., & Donaldson, D. I. (2008). Attention orienting effects of hesitations in speech: Evidence from erps. *Journal of Experimental Psychology-Learning Memory and Cognition*, *34*, 696-702.

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.

Culatta, R., & Leeper, L. (1988). Dysfluency isn't always stuttering. *Journal of Speech and Hearing Disorders*, *53*, 486-488.

Cutler, A. (1982). *Slips of the tongue and language production.* Berlin: Mouton Publishers.

Dale, P., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods Instruments & Computers*, *28*, 125-127.

DebRoy, S., & Bates, D. M. (2004). Linear mixed models and penalized least squares. *Journal of Multivariate Analysis*, *91*, 1-17.

Deese, J. (1984). *Thought into speech: The psychology of a language.* Englewood Cliffs, NJ: Prentice Hall.

Dell, G. S. (1986). A spreading activation theory of retrieval in language production. *Psychological Review*, *93*, 283-321.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and non-aphasic speakers. *Psychological Review*, *104*, 801-838.

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language*, *59*, 447-456.

Eklund, R. (2001). Prolongations: A dark horse in the disfluency stable. In *Proceedings of DiSS '01: Disfluency in Spontaneous Speech Workshop* (p. 5-8). Edinburgh, UK.

Eklund, R. (2004). *Disfluency in Swedish human-human and human-machine travel booking dialogues.* Unpublished doctoral dissertation, Linköping University.

Eklund, R., & Shriberg, E. E. (1998). Crosslinguistic disfluency modelling: A comparitive analysis of Swedish and American English human and human-machine dialogue. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP '98)* (Vol. 6, p. 2631-2634). Sydney, Australia.

Ellis, A. W., Flude, B., Young, A., & Burton, A. (1996). Two loci of repetition priming in the recognition of familiar faces. *Journal of Experimental Psychology-Learning Memory and Cognition*, *22*, 295-308.

Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, *100*, 233-253.

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

Ferreira, F., & Swets, B. (2002). How incremental is language production? evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*, 57-84.

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 use um and uh in spontaneous 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. (2002). Basic meanings of you know and i mean. *Journal of Pragmatics*, *34*, 727-747.

Fromkin, V. (1973). *Speech errors as linguistic evidence* (Vol. 77). The Hague: Mouton.

Garrett, M. F. (1975). The analysis of sentence production. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 9, p. 133-177). New York: Academic Press.

Garrett, M. F. (1980). Levels of processing in sentence production. In B. L. Butterworth (Ed.), *Language production: Vol. 1.* London, UK: Academic Press.

Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, *411-458*.

Gerhand, S., & Barry, C. (1998). Word frequency effects in oral reading are not merely age-of-acquisition effects in disguise. *Journal of Experimental Psychology-Learning Memory and Cognition*, *24*, 267-283.

Gerhand, S., & Barry, C. (1999). Age-of-acquisition and frequency effects in speeded word naming. *Cognition*, *73*(2), B27-36.

Ghyselinck, M., Lewis, M. B., & Brysbaert, M. (2004). Age of acquisition and the cumulative-frequency hypothesis: a review of the literature and a new multi-task investigation. *Acta Psychol (Amst)*, *115*(1), 43–67.

Godfrey, J. J., Holliman, E. G., & McDaniel, J. (1992). SWITCHBOARD: telephone speech corpus for research and development. In *Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing.* San Francisco: IEEE: IEEE.

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*.

Goldman-Eisler, F. (1961). The distribution of pause duration in speech. *Language and Speech*, *4*, 232-237.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech.* New York: Academic Press.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers.* New York, NY: Academic Press.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, *82*, B1-B14.

Griffin, Z. M., & Bock, J. K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language*, *38*, 313-338.

Griffin, Z. M., & Bock, J. K. (2000). What the eyes say about speaking. *Psychological Science*, *11*, 274-279.

Griffin, Z. M., & Huitema, J. (1999). *Beckman spoken picture naming norms.* [Online]. Available: http://langprod.cogsci.uiuc.edu/norms/.

Hawkins, P. 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).

Heike, A. E., Kowal, S., & O'Connell, D. C. (1983). The trouble with "articulatory pauses". *Language and Speech*, *26*, 207-216.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*, 91–117.

Huitema, J. (1996). *The huitema picture collection.* [Electronic database].

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

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434-446.

Jarvella, R. J. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior*, *10*, 409-416.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*, 824-843.

Johnson, W. (1955). *Stuttering in children and adults. thirty years of research at the University of Iowa*. Minneapolis, MN: University of Iowa Press.

Johnson, W. (1961). Measurement of oral reading and speaking rate and disfluency of adult male and female stutterers and nonstutterers. *Journal of Speech and Hearing Disorders*, *7*, 1-20.

Johnson, W., & Associates. (1959). *The onset of stuttering: Research findings and implications*. Minneapolis, MN: University of Minnesota Press.

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. *J Psycholinguist Res*, *32*, 37–55.

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, *11*, 201-258.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: the role of mutual knowledge in comprehension. *Psychol Sci*, *11*, 32–38.

Kilgariff, A. (1995). *BNC database and word frequency lists.* [Online]. Available: http://www.itri.brighton.ac.uk.

Kircher, T. T. J., Brammer, M. J., Levelt, W. J. M., Bartels, M., & McGuire, P. K. (2004). Pausing for thought: engagement of left temporal cortex during pauses in speech. *Neuroimage*, *21*, 84-90.

Kircher, T. T. J., Brammer, M. J., Williams, S. C., & McGuire, P. K. (2000). Lexical retrieval during fluent speech production: an fMRI study. *Neuroreport*, *11*, 4093–4096.

Klapp, S., Anderson, W., & Berrian, R. (1973). Implicit reading in speech, reconsidered. *Journal of Experimental Psychology*, *100*, 368-374.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, *3*, 129-140.

Kowal, S., & O'Connell, D. C. (2000). Psycholinguistische aspekte der transkription: Zur notation von pausen in gespr achstranskripten. *Linguistische Berichte*, *183*, 355-380.

Kutas, M., & Hillyard, S. A. (1980). Reading between the lines: event-related brain potentials during natural sentence processing. *Brain and Language*, *11*, 354–373.

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

Lachman, R. (1973). Uncertainty effects on time to access the internal lexicon. *Journal of Experimental Psychology*, *99*, 199-208.

Lachman, R., Shaffer, J. P., & Hennrikus, D. (1974). Language and cognition: Effects of stiimulus codability, name-word frequency and age of acquisition on lexical reaction time. *Journal of Verbal Learning and Verbal Behavior*, *13*, 613-625.

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; London: MIT Press.

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

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 Conference on Phonteic Sciences* (Vol. 4, p. 192-195). Stockholm, Sweden.

Lickley, R. J. (1996). Juncture cues to disfluency. In *Proceedings of the 4th International Conference on Spoken Language Processing*. Philedelphia, PA.

Lickley, R. J. (1998). *HCRC disfluency coding manual* (HCRC Technical Report No. 100). Edinburgh Univesity.

Lickley, R. J., & Bard, E. G. (1998). When can listeners detect disfluency in spontaneous speech? *Language and Speech*, *41*, 203-226.

Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, *15*.

Mahl, G. (1957). Disturbances and silences in the patient's speech in psychotherapy. *Journal of Abnormal and Social Psychology*, *42*(3-32).

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*, 1–71.

Martin, J. G., & Strange, W. (1968). Determinants of hesitations in spontaneous speech. *Journal of Experimental Psychology*, *76*, 474-&.

Martin, N., Weisburg, R. W., & Saffran, E. M. (1989). Variables influencing the occurrence of naming errors: Implications for models of lexical retrieval. *Journal of Memory and Language*, *24*, 462485.

Meyer, A. S., Roelofs, A., & Levelt, W. J. M. (2003). Word length effects in object naming: The role of a response criterion. *Journal of Memory and Language*, *48*, 131-147.

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: eye movements during noun phrase production. *Cognition*, *66*, 825-833.

Miller, G. R., & Hewgill, M. A. (1964). The effects of variations in nonfluency on audience ratings of source credibility. *Quarterly Journal of Speech*, *50*, 36-44.

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology*, *50A*, 528-559.

Morrison, C. M., & Ellis, A. W. (1995). Roles of word-frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology-Learning Memory and Cognition*, *21*, 116-133.

Morrison, C. M., Ellis, A. W., & Quinlan, P. T. (1992). Age of acquisition, not word frequency affects object naming, not object recognition. *Memory and Cognition*, *20*, 705-714.

Navarette, E., Benedetta, B., Alario, F. X., & Costa, A. (2006). Does word frequency affect lexical selection in speech production? *Quarterly Journal of Experimental Psychology*, *59*, 1681-1690.

Obler, L. K., & Albert, M. L. (1984). Language in aging. In M. L. Albert (Ed.), *Clinical neurology of aging* (p. 245-253). New York, NY: Oxford University Press.

O'Connell, D. C. (1988). *Critical essays on language use and psychology.* New York: Springer.

O'Connell, D. C., & Kowal, S. (2005). Uh and um revisited: Are they interjections for signaling delay? *Journal of Psycholinguistic Research*, *34*, 555-576.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*, 273-281.

Oomen, C. C. E., & Postma, A. (2001). Effects of time pressure on mechanisms of speech production and self-monitoring. *Journal of Psycholinguistic Research*, *30*, 163-184.

O'Shaughnessy, D. (1992). Recognition of hesitations in spontaneous speech. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (p. 521-524).

Oviatt, S. (1995). Predicting spoken disfluencies during human-computer interaction. *Computer Speech and Language*, *9*, 19-35.

Pérez, M. A. (2007). Age of acquisition persists as the main factor in picture naming when cumulative frequency and frequency trajectory are controlled. *Quarterly Journal of Experimental Psychology*, *60*, 32-42.

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology-Learning Memory and Cognition*, *24*, 539-557.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends Cogn Sci*, *11*, 105–110.

Plauché, M. C., & Shriberg, E. E. (1999). Data-driven subclassification of disfluent repetitions based on prosodic features. In *Proceedings of the International Conference of Phonetic Sciences* (Vol. 2, p. 1513-1516). San Fransisco, CA.

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.

Postma, A. (2000 Nov 16). Detection of errors during speech production: a review of speech monitoring models. *Cognition*, *77*(2), 97–132.

Postma, A., & Kolk, H. (1993). The Covert Repair Hypothesis: prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech and Hearing Research*, *36*, 472-487.

Postma, A., Kolk, H., & Povel, D. J. (1990). On the relation among speech errors, disfluencies, and self-repairs. *Language and Speech*, *33*, 19-29.

R development core team. (2005). *R: A language and environment for statistical computing.* Available: http://www.R-project.org. R Foundation for statistical computing.

Rahman, R. A., & Melinger, A. (2007). When bees hamper the production of honey: Lexical interference from associates in speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *33*, 604-614.

Roelofs, A. (1992). A spreading activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107-142.

Rossion, B., & Portois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, *33*, 217-236.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, *50*(696-735).

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.

Schiffrin, D. (1987). *Discourse markers* (Vol. 5). Cambridge: Cambridge University Press. Available from `http://www.loc.gov/catdir/toc/cam031/86018846.html`

Schriefers, H. J., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, *29*, 86-102.

Shallice, T., & Butterworth, B. (1977). Short-term memory impairment and spontaneous speech. *Neuropsychologia*, *15*(6), 729–735.

Shatzman, K. B., & Schiller, N. O. (2004). The word frequency effect in picture naming: Contrasting two hypotheses using homonym pictures. *Brain and Language*, *90*, 160-169.

Shriberg, E. E. (1994). *Preliminaries to a theory of speech disfluencies*. Unpublished doctoral dissertation, University of California at Berkeley.

Shriberg, E. E. (1996). Disfluencies in switchboard. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP '96)* (p. 11-14). Philedephia, PA: ICSLP.

Siegel, G., Lenske, J., & Broen, P. (1969). Suppression of normal speech disfluencies through response cost. *Journal of Applied Behavioural Analalysis*, *2*, 265–276.

Siegman, A. W., & Pope, B. (1966). Ambiguity and verbal disfluencies in the TAT. *Journal of Consulting Psychology*, *30*, 239-245.

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, *32*, 25-38.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *6*, 174-215.

Snodgrass, J. G., & Yuditsky, T. (1996). Naming times for the Snodgrass and Venderwart pictures. *Behavior Research Methods Instruments & Computers*, *28*, 516-536.

Svartvik, J., & Quirk, R. (1980). *A corpus of English conversation.* Lund, Sweden: Gleerup.

Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics*, *30*, 485-496.

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., et al. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language*, *51*, 247-250.

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.

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.

Vitkovitch, M., & Tyrrell, L. (1995). Sources of disagreement in object naming. *The Quarterly Journal of Experimental Psychology Section A*, *48*, 822 - 848.

Wade, E., Shriberg, E. E., & Price, P. J. (1992). User behaviors affecting speech recognition. In *Proceedings of the International Conference on Spoken Language Processing* (p. 995-998). Banff, Canada.

Wheeldon, L. R., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, *37*, 356-381.

Winer, B. J. (1971). *Statistical principles in experimental design* (2d ed ed.). New York: McGraw-Hill.

Wingfield, A. (1968). Effects of frequency on identification and naming of objects. *American Journal of Psychology*, *81*, 226-234.