THE COORDINATION OF SPEAKING AND LISTENING IN DIALOGUE

BLAIRE ELIZABETH MORGAN

A thesis submitted to the University of Birmingham for the degree of DOCTOR OF PHILSOPHY

School of Psychology

College of Life and Envrionmental Sciences

University of Birmingham

May 2013

ABSTRACT

The present research examined the coordination of speech production and speech comprehension. Whilst conversing with others we must coordinate the planning of our own speech output with the comprehension of our speech partner(s)'s utterances. However, very little is known about the coordination. Through examining participants' speech and eye-movements, this research questioned how people manage to coordinate their speaking and listening in dialogue. In 6 studies I demonstrate that tracking prediction processes offers an effective measure of online comprehension. When employing an appropriate dual-task, where two linguistic tasks competed for attention, cognitive resources had to be shared between the two tasks. Where necessary, participants actively engaged in strategies that helped to reduce processing demands including separation of production and comprehension, comparable processing of simple and complex syntactic structures, and using online prediction to preserve capacity. These capacity saving processes and strategy use enabled speech production to be effectively coordinated with speech comprehension in an experimental setting.

ACKNOWLEDGEMENTS

In the course of researching and writing this thesis I have been assisted by a number of friends and colleagues. I would like to offer my sincere thanks to my supervisors, Professor Antje Meyer, Director of Max Planck Institute for Psycholinguistics in Nijmegen, and Dr Andrew Olson, Senior Lecturer at the University of Birmingham, for their invaluable advice and support.

I wish to record my appreciation to my family, fiancée and friends for their continued emotional support and encouragement.

Contents

Chapter 1: Introduction		**1**
1	Motivation	2
1.1	Speech Production	2
1.2	Speech Comprehension	6
1.2.1	Parsing and Discourse Comprehension	8
1.3	Comprehension and the Visual World	11
1.4	Online Prediction and the Visual World	13
1.5	Dialogue	16
1.5.1	The Interactive Alignment Account of Dialogue	16
1.6	Language-based Dual-tasks	22
1.7	Working Memory and Language	24
1.8	Prediction and Working Memory Capacity	28
1.9	The Paradigm	30
1.10	Overview of thesis	33
Chapter 2: Predicting Online at the Verb		**35**
2	Experiment 1	36
2.1	Goals and Motivation	36
2.2	Hypotheses	38
2.3	Method	38
2.4	Linear Mixed Effects Models and 'R'	41
2.5	Experiment 1: Results	46

2.6	Discussion	48
Chapter 3: Effects of Digit-load on Online Prediction Processes		**50**
3	Experiment 2	51
3.1	Goals and Motivation	51
3.2	Hypotheses	55
3.3	Method	55
3.3.1	Lexical Decision Task	55
3.3.2	The Dual-task	57
3.4	Results	59
3.4.1	Digit Matching Task	59
3.4.2	Eye-movement Data	59
3.4.3	Linear Mixed Effects Model	60
3.5	Discussion	66
Chapter 4: Effects of Syntactic Complexity on Online Prediction Processes		**73**
4	Experiment 3	74
4.1	Goals and motivation	74
4.2	Hypotheses	79
4.3	Method	80
4.4	Results	85
4.4.1	Eye-movement Data	85
4.4.2	Linear Mixed Effects Model	86
4.5	Discussion	89

Chapter 5: Coordinating Speech Production and Speech Comprehension		**93**
5	Experiment 4	94
5.1	Goals and Motivation	94
5.1.1	The Capacity Demands of Speech Production	97
5.1.2	The Planning Process	101
5.1.3	Overview of experiment	103
5.2	Hypotheses	105
5.3	Method	106
5.4	Results	111
5.4.1	Eye-movement Data	111
5.4.2	Linear Mixed Effects Models	112
5.4.3	Participant Feedback	123
5.5	Discussion	123
Chapter 6: The Effect of a Cued Preview on Coordination		**130**
6	Experiment 5	131
6.1	Goals and Motivation	131
6.2	Hypotheses	135
6.3	Method	135
6.4	Results	137
6.4.1	Eye-movement Data	137
6.4.2	Linear Mixed Effects Models	138
6.4.3	Participant Feedback	146

6.4.4 Operation Span Measure 146

6.5 Discussion 147

Chapter 7: The Coordination of Speaking and Listening: Prediction as a Capacity Saving Process **155**

7 Experiment 6 156

7.1 Goals and Motivation 156

7.2 Hypotheses 156

7.3 Method 157

7.4 Results 157

7.4.1 Eye-movement Data 157

7.4.2 Linear Mixed Effects Models 158

7.5 Discussion 166

Chapter 8: General Discussion and Implications **169**

8.1 Summary of experiments and results 170

8.2 Discussion of the results 174

8.2.1 Online prediction as a capacity saving process 174

8.2.2 Processing demands and strategy use 176

8.2.3 How does this relate to natural dialogue? 179

8.2.4 How does this differ from natural dialogue? 184

8.2.5 The paradigm and potential applications 188

References **192**

Appendices **212**

Appendix 1

Appendix 1A: Sentence stimuli for Experiments 1 and 2 212

Appendix 1B: Script and output from Experiment 1's linear mixed effect model 213

Appendix 1C: The verb onsets, noun onsets and verb-noun distances for all sentences in Experiment 2 216

Appendix 1D: Stimuli for lexical decision task 217

Appendix 1E: Output from paired samples t-test on lexical decision task data 218

Appendix 1F: Experiment 2 cross-over effects 219

Appendix 1G: Sentence stimuli for Experiments 3-6 220

Appendix 1H: The verb onsets, noun onsets and verb-noun distances for all sentences in Experiments 3-6 227

Appendix 1I: Means and correlation from the Operation Span task (Experiment 5) 229

Appendix 1J: Number of disfluencies across production only and production with comprehension trial types in Experiments 4-6 230

Appendix 1K: The AIC values for each random effects model in the LMEMs 231

Tables

Table 1:	The number and percentage of valid fixations in the specific and general verb conditions	47
Table 2:	Mean eye-movement latencies to targets (in ms) across the specific and general conditions, relative to the verb and noun onsets	48
Table 3:	The mean and standard deviations of reaction times in the lexical decision task across specific and general verb types	57
Table 4:	The number and percentage of valid fixations in the specific and general verb conditions	60
Table 5:	The means and standard deviations of eye-movement latencies to the target across specific and general conditions, and both tasks	62
Table 6:	The means and standard deviations of eye-movement latencies to the target across specific and general verb conditions, the visual world only and dual tasks, and the first and second experimental halves	63
Table 7:	The average latencies to targets across the specific and general conditions, relative to the verb and noun onsets	65
Table 8:	The number and percentage of valid fixations in the specific and general verb conditions	86
Table 9:	The means and standard deviations of eye-movement latencies to the target across verb type and complexity	87
Table 10:	Mean eye-movement latencies to targets (in ms) across the specific, general and simple, complex conditions, relative to the verb and noun onsets.	88
Table 11:	The mean and standard deviations (ms) for eye-movement latencies to early and late verb positions across the specific and general conditions	89

Table 12:	The number and percentage of valid fixations across the specific and general verb conditions	111
Table 13:	The means and standard deviations of eye-movement latencies to the target across verb type and complexity	113
Table 14:	Mean eye-movement latencies to targets (ms) across the specific, general, simple, complex conditions, relative to verb and noun onsets.	114
Table 15:	The mean latencies to the target across the two verb type conditions when presented within the two different sentence types (embedded, other), and the two different verb positions (early, late)	115
Table 16:	The mean and standard deviations for eye-movement latencies in simple and complex sentences as shown across Experiments 3 (comprehension only) and Experiment 4 (production and comprehension)	119
Table 17:	The mean eye-movement latencies to the target across the specific and general verb conditions in Experiment 3 and Experiment 4	119
Table 18:	The number of trials where participants recalled, 0, 1, 2, or 3 items, as shown across PO and PC trials	121
Table 19:	The mean and standard deviations of speech onset latencies when 1, 2 or 3 items were recalled, as shown across PO and PC trials	122
Table 20:	The number and percentage of valid fixations across the specific and general verb conditions	137
Table 21:	The means and standard deviations of eye-movement latencies in the specific and general conditions and across simple and complex sentences	139
Table 22:	Mean eye-movement latencies (ms) across the specific, general and simple, complex conditions, relative to the verb and noun onsets.	139
Table 23:	The mean eye-movement latencies to the target across the specific and	142

general verb conditions in Experiment 3 and Experiment 5

Table 24: The mean and standard deviations for eye-movement latencies in simple and complex sentences as shown across Experiments 3 (comprehension only) and Experiment 5 (production and comprehension with cued preview) 143

Table 25: The mean and standard deviations for eye-movement latencies across verb type and complexity conditions in Experiments 4 (dual- task, no cued preview) and Experiment 5 (production and comprehension with cued preview) 144

Table 26: The mean and standard deviations of speech onset latencies when 1, 2 or 3 items were recalled, as shown across production only (PO) and production with comprehension (PC) trials 145

Table 27: The number and percentage of valid fixations across the specific and general verb conditions 158

Table 28: The means and standard deviations of eye-movement latencies in the specific and general verb conditions and across simple and complex sentences 159

Table 29: Mean eye-movement latencies to targets (in ms) across the specific, general and simple, complex conditions, relative to the verb and noun onsets. 159

Table 30: The means and standard deviations of eye-movement latencies to early and late verb positions in Experiment 6 160

Table 31: The average eye-movement latencies by the number of items recalled across the two verb types 162

Table 32: The mean eye-movement latencies to the target across the specific and general verb conditions in Experiments 3-6 163

Table 33: The mean and standard deviations of speech onset latencies when 1, 2 or 3 items were recalled, as shown across PO and PC trials 165

Table 34: Overview of main results 173

Figures

Figure 1:	A summary of the main speech production processes from Levelt (1989)	6
Figure 2:	The Cohort model of word recognition, proposed by Marslen-Wilson and Welsh (1978)	8
Figure 3:	A reconstruction of the interactive alignment account of dialogue as found in Pickering and Garrod, 2004	19
Figure 4:	Example of a visual display; the agent is the boy in the centre, the target is the cake, and the three remaining pictures are distractors	39
Figure 5:	The probability of fixating the target over time across the two types of verb condition (general, specific) and two tasks (comprehension (or visual world only), dual-task)	65
Figure 6:	Diagram showing the divisions and subdivisions of stimuli into the different structures (complex/simple), verb types (specific/general), sentence types (embedded clauses/other), and verb positions (early/late)	84
Figure 7:	A visual representation of the two types of trial; Speech Production Only and Speech Production and Comprehension	110

CHAPTER 1

INTRODUCTION

1. Motivation

The motivation for the present research is to gain some insight into how comprehension and production processes can be coordinated during speech. Whilst conversing with others we must coordinate the planning of our own speech output with the comprehension of our speech partner(s)'s utterances. This coordination system is extremely fast and efficient; turn-taking is achieved rapidly and utterances between partners often overlap. However, very little is known about the coordination. Here, my research questions how people manage to coordinate their speaking and listening in dialogue. Is it possible to plan and listen simultaneously? Do we wait to plan until comprehension has been completed? Or do we continuously switch been planning and listening? In psycholinguistic research, production and comprehension are often examined separately. Here I used a research paradigm that requires the two tasks to be dealt with concurrently. Through examining participants' speech and eye-movements I hoped to learn about the coordination process.

In the next sections of this chapter I motivate this work by briefly describing the processes involved in production, comprehension and dialogue. I also review the paradigm used in this research.

Background Information

1.1 Speech Production

Speech production is the process of translating an idea or intention into a linguistic representation. Production involves thinking of what one wants to say, choosing the most appropriate words for expressing this thought, and saying the words aloud. Researchers have argued that speech production can be broken down into a series of different stages. Levelt (1989) separates speech into three broad stages of i) Conceptualisation, ii) Formulation, and iii) Articulation.

Conceptualisation involves deciding how to best portray one's intention or ideas. This process starts with the speaker deciding on what to say in relation to whom they are speaking to, what has gone before, and the knowledge shared between interlocutors. Imagine, for example, that a speaker would like to convey the knowledge that their friend Bill has just washed the dishes alone in the kitchen. From this "intention" to speak, the speaker must select the appropriate information to fulfil this purpose, this is called "macroplanning". Macroplanning involves searching and selecting the information to be expressed in an utterance and deciding on the order of presentation. In explanation, speakers must activate suitable lexical concepts. For example, should the speaker to refer to their friend using their Christian name 'William', or 'Bill', a nickname, a pronoun ('he'), or a reflexive ('himself')?

The selection process is famously described by Levelt (1989) as a 'statisical mechanism', where activation spreads from conceptual nodes to lemma nodes, Activation increases until one lemma is selected. This search and selection process can be easy if the information has already been brought into focus or primed by previous speech, or more difficult when the information must be extrapolated from long-term memory (for example, when giving directions (Levelt, 1989)). It is thought that this process of searching and selecting is subject to capacity demands (see Chapter 5, section 5.1.1).

Speakers must also plan the informational perspective and assign referents; or "microplan". The process of microplanning involves ensuring that the interlocuters are on the same page. That is, any indication of referents (both animate and inanimate) must be jointly understood by all involved. Conversation fails if the speakers do not share the same representations. This is also true when introducing names or pronouns; if a speaker is to talk about Bill then they must be sure that their speech partner knows which Bill they are referring to. There are numerous ways of ensuring success within microplanning, for example ensuring that the focus of the utterance is most prominent in the sentence or creating sentence structures that clarify the topic at hand (for instance, conceptual prominence and topicalising), however we do not need to discuss these in detail here. Levelt (1989) refers to the outcome of conceptualising as a "preverbal message".

The preverbal message forms the basis for the linguistic formulation of the utterance. Formulation involves two levels of processing; "grammatical encoding" where lemmas (syntactic representations of words) and syntactic structure are retrieved, and "phonological encoding" where a phonetic plan is built for the utterance. The process of grammatical encoding involves building up a syntactic structure. Within this process lemma information and syntactic information must be organised alongside one another; the activation of certain lemmas will dictate what kind of syntactic structure can be used and conversely the activation of certain syntactic structures will dictate what kinds of lemmas can be utilised. Once the lemmas have been retrieved and the syntactic structure has been generated the speaker must generate the corresponding sound form. This involves retrieving the stored form representations of the words and combining them according to the phonological and phonetic rules of the language (see Levelt (1989) for details). This stage of speech production is more complex than just retrieving the phonetic information for consecutive words and producing one after the other; even retrieving the phonetic material for one word is thought to be made up of several levels (or tiers) of processing, when building a phonetic plan made up of multiple words the speaker has to also ensure that there is an appropriate rhythm, with the correct stress pattern over syllables in order to communicate the correct intonation for the speakers' goal (Levelt, 1989).

With reference to the aforementioned sentence "Bill wiped the plates", the process of encoding is as follows. The speaker begins at the message level with an intention to convey a particular idea or concept. This then activates the appropriate propositions, for example, the subject is "Bill"; the verb is "wash", the object is "dish"; the number (of objects) is many; and the timing is past tense. Retrieving the syntactic representations specifies the structure of the sentence; Noun^1, Verb [Past] (Determiner), Noun^2, [Plural]. The phonological structure of content words are accessed using semantic representation; /Bill/ /wash/ /dish/. These content words are added into the syntactic frame to give /Bill/ /wash/ + [Past] (Determiner) /dish/ + [Plural]. The phonological representations of these elements are then specified ready for articulation; /Bill/ /washed/ /the/ /dishes/ (see Harley, 2008 p. 402).

The final stage of speech production is articulation. The articulation process involves the coordination of multiple muscles, retrieval of motor programs and control of timing (as well as many more processes). The outcome of this last stage is the spoken utterance, "Bill washed the plates".

Psycholinguistic researchers tend to agree upon these steps of speech production but their time course is still a matter of some debate. Whilst in earlier research it was argued that information flows in a strictly serial fashion where the phonology of a word can only be activated once the lemma has been selected (Butterworth, 1992; Garrett, 1980, Levelt, Roelofs, & Meyer, 1999; Roelofs, 1992), most researchers now assume a cascading flow of information where, for instance, phonological processing of an utterance fragment can begin before grammatical encoding has been completed (Caramazza, 1997; Dell, 1986; Goldrick & Rapp, 2002; Harley, 1993; Humphreys, Riddoch, & Quinlan, 1988; Jescheniak & Schriefers, 1998; MacKay, 1987; Navarrete & Costa, 2005; Rapp & Goldrick, 2000 (see Caramazza & Miozzo,1997 and Morsella & Miozzo, 2002 for a review)).

Conceptualisation:

- Determining what to say
- Conceiving an intention
- Selecting relevant information in preparation for constructing the intended utterance
- End point: preverbal message

Formulation:

- Translating the conceptual representation into a linguistic form
- Lexicalisation: selecting the words that you want tos ay
- Syntactic planning: putting the words from the lexicalisation stage together to form a sentence
- Phonetic and articulatory planning
- Phonological encoding: words are turned into sounds

Execution:

- Detailed phonetic planning
- Motor planning
- Articulation

Figure 1. A summary of the main processes involved in speech production, split into the three stages of conceptualisation, formulation and execution. The stages and processes are originally taken from Levelt (1989), and adapted from Harley (2008).

1.2 Speech Comprehension

Although understanding our native tongue appears effortless, comprehension involves many levels of processing. For example, a stream of speech must be segmented into individual words, these individual words must be recognised from a huge host of items in the mental lexicon, and the listener must also retrieve the associated semantic and syntactic information.

The process of identifying speech from the sensory input involves three main stages, the first involves hearing the speech input, the next involves a search to identify the correct items from the mental lexicon and the final stage (in comprehension of a single word) is recognition. This is described in Frauenfelder and Tyler's (1987) stages of initial lexical contact; (lexical) selection; and word recognition. Initial lexical contact involves the sensory input making contact with the mental lexicon. After initial contact, the listeners will activate a set of possible lexical candidates (Frauenfelder & Peeters, 1998; Luce, 1986; McQueen, Cutler, Briscoe, & Norris, 1995; Tanenhaus, Magnuson, Dahan, & Chambers, 2000). The pathways to activation are still widely debated and theories include the well-known interactive TRACE model (McClelland & Elman, 1986) and the all-or-none Cohort model (Marslen-Wilson & Welsh, 1978). More recently, researchers have argued for a graded activation where phonemes, features and words are activated based on their similarity to the auditory input (Strauss, Harris, & Magnuson, 2007). This activation continues to accumulate until a threshold is reached and one lexical item can be chosen, at this point the selection process leads to word recognition (Norris, Cutler, McQueen, & Butterfield, 2006).

The process of spoken word comprehension does not end there however, the listener will then go on to acquire phonological, syntactic and semantic information regarding the lexical item, the end point of which is lexical access (McQueen, 2005).

Although this description might indicate that activation only runs in a serial manner from perception to acquisition of syntactic and semantic information, there is much evidence to suggest that activation may be bi-directional. For example, it has been claimed that lexical information can facilitate speech perception with lexical items having an effect on pre-lexical representations (McClelland, Mirman, & Holt, 2006). Regardless of the directions or interactions in identifying words however, the main stages of creating a set of lexical candidates, selecting a candidate, and integrating the item are largely agreed upon in the literature. These main stages of comprehension are highlighted in Figure 1 below which has been adapted from one of the original theories of speech comprehension, the Cohort model (Marslen-Wilson & Welsh, 1978).

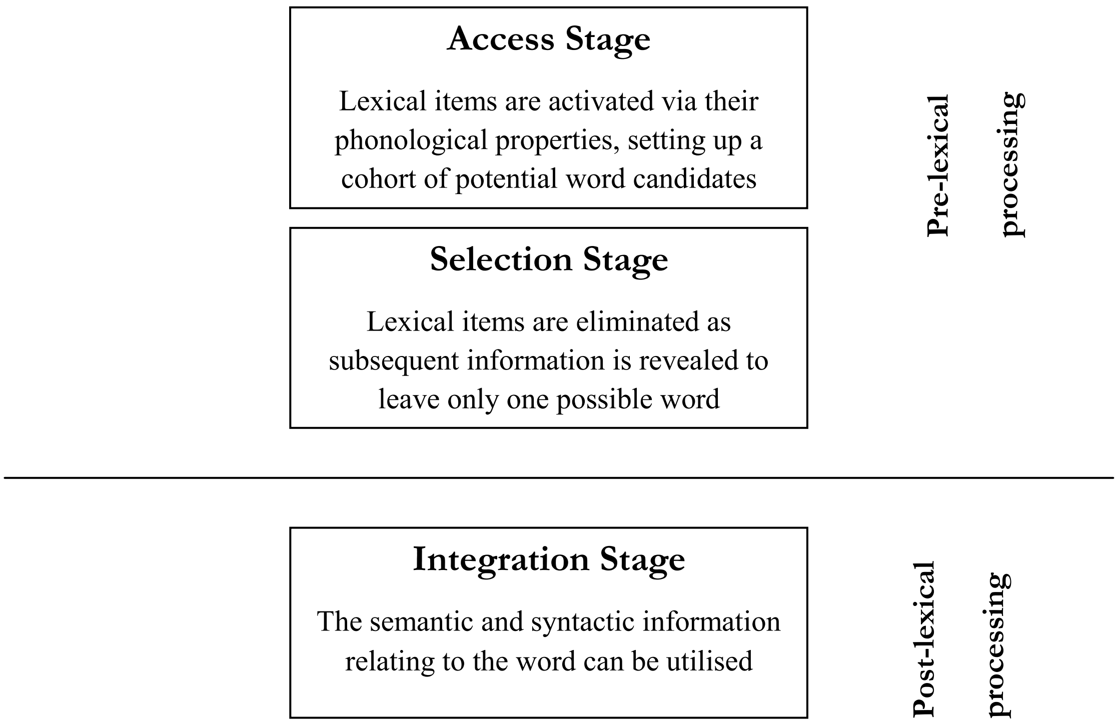

Figure 2. The Cohort model of word recognition, proposed by Marslen-Wilson and Welsh (1978). Word recognition begins with the creation of a cohort of possible lexical items (*access stage*) which is then reduced down until only one activated item remains (*selection stage*). This one item is recognised as the target word (*word recognition*). Next, the word's associated semantic and syntactic information can be accessed and put to use (*integration stage*) - this stage can be influenced by the context effects. Adapted from Harley, 2008.

1.2.1 Parsing and Discourse Comprehension

The above description describes the intricacies involved in processing a single word, however, we rarely hear only one spoken word; the majority of our spoken communications involve comprehending multiple sentences (or a "discourse"). This therefore necessitates additional processes including parsing the sentences to obtain the necessary syntactic and semantic information, integrating these aspects into a sentence interpretation, and assimilating this information with what has gone before.

Parsing

Parsing refers to analysing the syntactic structure of a sentence in order to be able to assign thematic roles (what is being done, who it is being done by and whom is it being done to).

This process is helped along by language-specific cues. For example, an extremely important cue in analysing the syntax of a sentence is word order; most languages have a specific word order rule and in English it is SVO (subject, verb, object) (Corrigan, 1986). Individuals can use this information as a guide to processing. One specific example is using the verb's argument structure (the possible themes that are associated with a given verb). This can greatly aid the parsing process. For example, hearing the verb "give" would activate the relevant thematic roles of an agent (who instigates the action), the theme (in this case, the action of giving something) and a recipient (who receives the theme) (Harley, 2008).

Numerous models of parsing have been proposed in the literature, for example, garden-path, interactive, and constraint-based models (Frazier, 1987; MacDonald, Pearlmutter, & Seidenberg, 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Trueswell, Tanenhaus, & Garnsey, 1994). Research in this area has highlighted how words are fitted together as they are encountered rather than waiting until the entire sentence has been presented. This incremental processing of language and the ability to use a verb's argument structure are highlighted in more detail in the following sections and are crucial to the design of my paradigm.

An important function to emphasise here, however, is the use of predictive processing in sentence comprehension. A predictive function of language has been included in numerous theories of sentence processing (Kamide, 2008), one such theory is the Syntactic Prediction Locality theory (Gibson, 1998) which describes the role prediction plays in parsing. Gibson suggests that whilst listening to speech input syntactic rules are activated/accessed and that this in turn activates representations that fit these syntactic requirements. For example, Gibson describes how hearing a noun phrase predicts that a verb will appear shortly. Or, as mentioned above, hearing "give" activates the thematic role of an agent, theme, and recipient.

The visual world paradigm (as described in more detail below) has offered evidence for prediction at numerous levels of sentence comprehension. For example, Kaiser and colleagues found that the interpretion of pronouns, demonstratives and reflexives is affected by several linguistic constraints including information structure, syntactic role and word order. Allopenna, Magnuson and Tanenhaus (1998) demonstrated that phonetic information can be used to make predictions, with acoustic information at the beginning of words driving comprehension and constraining lexical access. Other factors affecting language processing and predictions includes word frequency (Magnuson, Dixon, Tanenhaus, & Aslin, 2007); verb information (Altmann & Kamide, 1999); grammatical agents (Kamide, Altmann, & Haywood, 2003); and semantic similarity (Huettig, Quinlan, McDonald, & Altmann, 2006). For a full review of predictive processing in language comprehension, see Kamide (2008).

Context Effects

Even though the comprehension process becomes more complex when dealing with a string of words or multiple sentences (as is the case in conversation), speech recognition is extremely fast; we can recognise speech at a rate of 20 phonemes per second and identify spoken words in context after only 200ms (Marlsen-Wilson, 1984). Many researchers have demonstrated a processing advantage for speech, especially when presented within a relevant context. For example, it was demonstrated as far back as the 1960's that a semantic or grammatical context allows for more accurate word recognition (Lieberman, 1963). More recently, research has shown an advantage for processing words in context even in very young children (Fernald and Hurtado, 2006).

Therefore, one way in which speech comprehension may be sped up is by the presence of contextual information. A sentence's previous semantic or syntactic properties can restrict its possible upcoming themes and structures, and can therefore offer important cues for processing. One technique that has been used to examine different contextual effects on language processing is the visual world paradigm.

1.3 Comprehension and the Visual World

Researchers have begun exploring online language processing with eye-tracking techniques. Eye-movements and language are intrinsically linked, as first demonstrated by Cooper in 1974. Through presenting an illustration alongside linguistic input, where the input referred to aspects of the illustration, Cooper found that participants' eyes were drawn to objects as they were mentioned. The link between objects and eye-movements was also found to extend to associations; with eye-movements to a camera being initiated after hearing, "During a photographic safari..." (Cooper, 1974).

Such discoveries lead to the creation of the ‘Visual World’ paradigm (see Huettig, Rommers, & Meyer, 2011 for a comprehensive review). The paradigm involves presenting visual arrays alongside an auditory input. This set-up allows researchers to examine what is being processed and when, and also pin-point what parts of speech may be more difficult to access.

The first researchers to make use of this research tool were Tanenhaus, Spivey-Knowlton, Eberhard and Sedivy in 1995. They speculated that the visual world paradigm could help discover the events that happened when subjects were faced with syntactic ambiguity. For example, how participants process a sentence such as “Put the frog on the towel in the box”. Models suggest that “on the towel” is processed as a destination rather than a modifier, therefore, when participants reach the preposition ‘in’ they become confused.

Tanenhaus and colleagues recorded eye-movements to assess the online millisecond by millisecond processing of sentence comprehension. Alike to Cooper (1974), they demonstrated that subjects moved their gaze to objects shortly after they were named; this was taken as evidence of incremental processing of linguistic information. Furthermore, through presenting syntactically ambiguous sentences, such as “Put the frog on the towel in the box”, they could investigate how the visual context can affect how these syntactically ambiguous sentences are processed. They illustrated that when subjects are presented with only one referent (a frog) then “on the towel” will be taken as the destination of that frog; with eye gaze following from the frog to the empty towel. However, when two-referents are available (a frog on a towel and a frog on a napkin), “on the towel” is treated as a modifier in order to distinguish which frog is the subject of the sentence. These results indicate how visual context can be used in conjunction with linguistic information at very early stages of sentence processing, with subjects quickly looking to the visual scene to identify the intended referent.

The visual world paradigm has been commandeered to study several aspects of language processing including lexical access (Allopenna, Magnuson, & Tanenhaus, 1998); semantic processing (Huettig & Altmann, 2005); phonetic processing (McMurray, Clayards, Tanenhaus, & Aslin, 2008), and pragmatics (Chambers & San Juan, 2008; Sedivy, Tanenhaus, Chambers, & Carlson, 1999) to name but a few. The visual world paradigm has proved to be a sensitive and effective measure of linguistic processing.

Of most importance to us here, is how the visual world has been used to look at anticipatory language. The research I present in the subsequent chapters relies on the ability to predict upcoming words/concepts from the linguistic input. Therefore, the next section reviews how the visual world has been used to examine such prediction.

1.4 Online Prediction and the Visual World

The visual world paradigm has also been used to study online prediction. Anticipating what is to come in sentences or utterances has the advantage of reducing the amount of resources spent on processing, freeing up capacity for other cognitive activities. However, when an incorrect prediction is made this has the opposite effect, requiring additional resources to amend the erroneous construct (Kamide, 2008).

Altmann and Kamide (1999) demonstrated that predictions can be made online using information extracted at the verb. Participants were presented with a visual scene comprised of an agent and various objects; one target and three or four unrelated distractor items (for the example I use here the display is repectively made up of a boy, a cake, a ball, a toy car and a toy train). Alongside this visual scene participant would hear either sentence (1a) or sentence (1b).

(1a) The boy will eat the cake.

(1b) The boy will move the cake.

In the case of sentence (1a) the verb restricts the domain from which the noun can be chosen, and the visual scene restricts the referent to only one possible object (only edible objects can be eaten, and only the cake is edible). However, in sentence (1b) the domain and potential referents are not restricted (all objects are moveable and equally likely to be moved). Altmann and Kamide's tested whether the verb is used to anticipate the upcoming noun and hypothesised to see faster and more frequent eye-movements to the cake after hearing (1a) than after hearing (1b). Their results supported this hypothesis and demonstrated more frequent saccades to cake after hearing "eat" (54% of trials) relative to "move" trials (38% of trials). Fixations to the cake were also made significantly *faster* during "eat" trials relative to "move" trials (611ms and 838ms respectively). Importantly, in the "eat" condition (where predictions could be formed) saccades to the target were often initiated *before* noun onset (mean saccade onset -85ms), whilst saccades where made after noun onset in the "move" condition (mean saccade onset 127ms).

Kamide, Altmann and Haywood (2003) extended this work to explore various different manipulations. They demonstrated that predictions can be extended past a simple relationship between verb and subsequent noun, that predictions can be made based on preverbal information (such as an agents' identity), and anticipations can be made in other languages, such as Japanese. For example, Kamide et al. demonstrated that multiple linguistic sources can be used to drive predictions. They found that the grammatical agent together with the verb can be used to anticipate a future referent. For instance, participants were presented with a visual scene containing two agents; a man and a girl, and multiple objects including a motorbike, a carousel, a sweet jar and a beer. Alongside this scene participants could hear (2a), (2b), (2c), or (2d).

(2a) The man will ride the motorbike.

(2b) The girl will ride the carousel.

(2c) The man will taste the beer.

(2d) The girl will taste the sweets.

Kamide et al. demonstrated that world knowledge about the grammatical agent and the verb could be used to predict the upcoming noun, with increased looks to motorbike after hearing the first part of (2a), and increased saccades to carousel when hearing (2b), even though multiple objects could be combined with the verb. This experiment therefore demonstrates that it is not just verb information that can be used to make online predictions.

Altmann and Kamide (2007) also demonstrated that information regarding tense can be used to anticipate language. For example, when presented with an agent, an empty wine glass, a full pint of beer, and various other distractor items, participants would look to the intended referent based on the statement's tense. That is, after the verb in sentence (3a) participants fixated upon the pint of beer; as "will drink" signifies that something is about to be drank. Yet in sentence (3b) after hearing "has drunk" participants will move their gaze to the empty wine glass; as the tense signals that the action of drinking has already been completed.

(3a) The man will drink all of the beer.

(3b) The man has drunk all of the wine

Altmann and Kamide propose the reason for this gaze pattern, and for anticipatory eye-movements in general, is that subjects access the objects' affordances and this drives them to look at whichever object best fits the unfolding speech. My own research expands on Altmann and Kamide's (1999) design and focuses on the ability to use verb information to predict future referents; I will return to this topic in Chapter 2.

1.5 Dialogue

Dialogue (or a conversation involving more than two people) typically involves person A speaking whilst person B listens, then person B speaking as A listens, and the exchange continues in a similar manner. For conversation to flow successfully, all interlocutors must work together to ensure that what is spoken can be understood. It is crucial that speakers produce speech that is audience appropriate, referring only to shared knowledge or ideas and including inferences that can be easily deciphered. One way in which speakers can ensure successful communication is through adhering to Grice's conversational maxims of quantity, quality, relevance and manner (Grice, 1975). The degree to which people stick to these rules will largely depend on their motives, however, a "good conversationalist" will ensure that their speech partner is following at all times and that both interlocutors share the same aims and representations. The idea of interlocutors sharing the same representations has been coined as "alignment" by Pickering and Garrod (Garrod & Pickering 2004; Pickering & Garrod, 2004). These researchers have proposed a theory of how alignment between speakers aids dialogue; the *interactive alignment account*.

1.5.1 The Interactive Alignment Account of Dialogue

The majority of accounts of comprehension and production are based only on monologue. For example, Kulen, Allefeld and Haynes recently commented that "traditional approaches in cognitive psychology and cognitive neuroscience tend to focus on the isolated mind" (Kuhlen, Allefeld, & Haynes, 2012). Pickering and Garrod, however, are part of the minority of researchers who have attempted to describe a model of language processing that explicitly accounts for dialogue, demonstrating the importance of examining how individuals coordinate and adapt to one another.

Although there are a lot more processes being undertaken and coordinated during dialogue, conversing with others feels relatively effortless. Pickering and Garrod's model goes some way to explaining the "magic" behind conversation and why conversation is easy. They describe how comprehension and production become tightly coupled in dialogue due to a mechanism called "interactive alignment". Alignment is not simply coordinating a joint activity where two people are working together to achieve the same goal (as in playing tennis for example), but involves interlocutors sharing the same representations (Garrod & Pickering, 2004; Pickering & Garrod, 2004). It is thought to be an essential mechanism that helps to simplify conversational exchange.

The first process of interactive alignment is the alignment of situation models; this refers to interlocutors establishing a joint idea about of what one another is talking about. The success of a conversation largely relies on how well the interlocutor's situation models have been aligned. Evidence for this kind of alignment can be seen in Garrod and Anderson's work (Garrod & Anderson, 1987) where participants took part in a maze game. To perform well in the maze participants aligned their spatial and lexical representations; they used particular references to location that were created by the interlocutors themselves to navigate with one another. Aligned linguistic expressions included references to the "right indicator" or the "upside down T shape" (see p. 214 – 217 of Garrod & Anderson (1987) for examples).

One question posed here is how does alignment of situation models occur? It is clear from experience that we do not overtly negotiate with our speech partners on how to align speech, so how exactly is interactive alignment achieved? Pickering and Garrod argue that this "global" alignment of situation models arises from alignment at "local" levels of language processing. That is, interlocutors will align their lexical, semantic and syntactic presentations, which in turn lead to global alignment. Local alignment is thought to occur by mechanisms of priming; when interlocutors come across a specific word, phrase, or expression this activates particular representations, these representations subsequently become more accessible and are interlocutors are therefore more likely to use them again. This is thought to span across comprehension and production; hearing a specific word or phrase or expression primes representations and makes the interlocutor more likely to produce it in their own utterance. An example of this can be found in Branigan, Pickering and Cleland (2000) who demonstrate a case of syntactic priming. They had naive participants engage in dialogue with a confederate speaker. The researchers demonstrated that the syntactic structure of the confederate's speech became modelled in the participant's own utterance, with speakers often producing the same syntactic structure as the one they had just heard the confederate use.

Pickering and Garrod's model incorporates a bi-directional channel of alignment with alignment passing from comprehension to production, and from production to comprehension. They argue that as this alignment occurs due to priming the channels to alignment are direct and automatic; activation of specific representations in one individual automatically leads to activation of the same representations in the other individual. The pathways of alignment can be seen more clearly in Figure 2 below, adapted from Pickering and Garrod (2004). The horizontal lines indicate where alignment occurs across interlocutors, and as you willsee there are a lot of pathways to shared representations. By proposing shared representations and automatic alignment Pickering and Garrod are able to explain why dialogue is so repetitive, why it is easier than monologue (where there is no one to align with), and why it is often possible to complete one another's sentences.

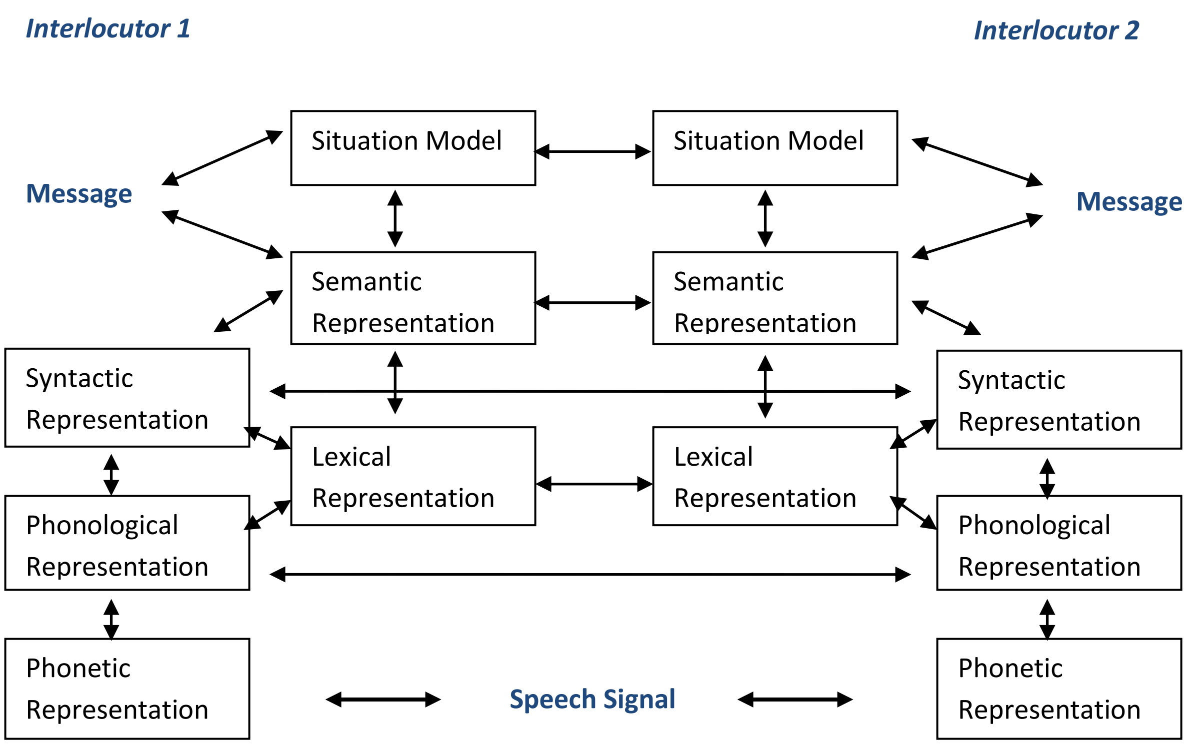

Figure 3. A reconstruction of the interactive alignment account of dialogue as found in Pickering and Garrod, 2004 (p. 176). This figure shows alignment (as depicted by arrows) across all levels of language processing from semantic to phonetic representations, and the alignment across interlocutors as illustrated by the horizontal lines.

Building upon claims by Pickering and Garrod, subsequent work has provided evidence for *why* dialogue is easier to understand than monologue. For example, Branigan, Catchpole and Pickering (2011) examined whether dialogue is easier to comprehend because the listener has access to multiple perspectives, in contrast to monologue where only the speaker's perspective is available. They demonstrate that an "overhearer" (a person that overhears someone else's utterances) can understand dialogue more easily and accurately than they can understand monologue. Branigan et al. suggest that interlocutors work together to create a perspective they both agree on; this perspective becomes “grounded” (Branigan et al., 2011; Clark, 1996) and easier to understand.

Pickering and Garrod suggest that this kind of alignment may not be restricted tospeech, indeed there is gathering evidence for non-linguistic alignment, for example imitation of others’ body movements (Chartrand & Bargh, 1999); mimicry of facial expressions (Bavela, Black, Lemery, & Mullett, 1986); imitation through mirror neurons (Rizzolatti & Arbib, 1998); and alignment of gesture between interlocutors (Mol, Krahmer, Maes, & Swets, 2012; Parrill & Kambara, 2006). Some of these alignment processes are also present in dialogue. Furthermore, research has also demonstrated that alignment in conversation may not be constrained to human interactions; Branigan et al. (2010) highlight how alignment even occurs in human-computer interactions (or HCIs), with interlocutors adapting their speech in order to ensure communicative success. Examples of adaption in this contextinclude adapting speech rate to align with a computer (Bell, Gustafson, & Heldner, 2003) and changing the length of utterances to align with that of the computer (Brennan, 1991). The influence of alignment can even be strong enough to invoke the production of ungrammatical utterances (Ivanova, Pickering, McLean, Costa, & Branigan, 2012).

Coincidentally, this increasing focus on coordination between two speakers can also be seen in cognitive neuroscience. A recent study by researchers in Germany examined how listeners’ and speakers’ electroencephalogram (EEG) are coordinated with one another (Kuhlen, et al., 2012). Although this coordination only involved a “unidirectional” exchange rather than dialogue the researchers were able to examine where speakers’ and listeners’ EEG correlate with one another. The EEG of speakers were recorded while they narrated short stories, these stories were later played to a set of listeners. Two stories were superimposed on top of one another and listeners had to attend to one of the stories whilst ignoring the other; this allowed the researchers to separate out the processing of the story from the sensory input. The researchers suggest that in social interactions (in this case a listener engaging (unidirectionally) with a speaker) individuals not only coordinate their behaviour but also coordinate their "mental states". For example, the listeners' EEG correlated with that of the speakers yet with a time delay of around 12.5 seconds. Kuhlen et al. suggest that these time delays correspond to where situation models are being developed. And as we know, building these situation models is crucial, as failure to align situation models could lead to failures in communication (Pickering & Garrod, 2004).

To summarise, although coordination between two interlocutors has begun to attract more attention, dialogue has for the most part been avoided in the psycholinguistic literature with researchers basing a huge amount of their work on monologue. Yet it is clear from the information presented in this section that a lot changes in dialogue; we have to coordinate our own input with that of our speech partners, we have to keep a record of what has gone before and use this information to make decisions about what comes next. Although this sounds like a complex intricate task, we deal with this process extremely well and communicating with others appears relatively effortless. Pickering and Garrod's interactive alignment account has helped to explain our excellent performance and highlight how dialogue benefits from features absent in monologue.

Pickering and Garrod's theory of interactive alignment helps to explain how it is possible to coordinate our speaking and listening, and with such apparent ease. In the situations described above the interlocutors benefit from dialogue situations; the alignment mechanisms serve to put them at an advantage. However, what happens when the speakers do not have everything going for them? For example, what happens when participants have to keep track of two unrelated strands of conversation? What about when a new speech partner joins in on a conversation and has missed the previous utterances? Do these situations cause breakdowns in coordination, or are interlocutors still able to piece together speech and successfully coordinate production and comprehension? Coordinating speech production and speech comprehension entails that two highly related linguistic processes must be organised and synchronised with one another. This thereby creates a situation in which two processes must compete for available cognitive resources. The interactive alignment theory described above helps to explain how the demand for these cognitive resources can be minimised, for example, through re-using the same syntactic structures. However, it should be noted that these alignment processes might not always be available to speakers. The experiments in this thesis examine the coordination process (and in a context where alignment is reduced). To do this, I employ both a speech production and a speech comprehension task. These two tasks will inevitably compete for resources, afterall they require access to the same language processors and working memory capacity (see section 5.1). The experiments create what is known as a 'dual-task scenario'.

Therefore, also of relevance to us here is previous work that has employed a dual-task that is linguistic in nature.

1.6 Language-based dual-tasks

Dual-task studies involve employing two tasks which participants have to complete simultaneously. Examining the extent to which we can do two things at once has recently received much attention; this is not surprising due to both the practical implications and scientific insight that this issue conveys. For example, when performance of one task hinders the performance of another it would suggest that at least some aspects of these tasks require processing by the same system or mechanism. However, when certain tasks can be performed simultaneously without any obvious damaging effects to either one, such as eating and watching TV, it appears as though each task is processed by a separate dedicated system or mechanism (Ferreira & Pashler, 2002; Pashler, 1994).

In a dual-task paradigm subjects are asked to make speeded responses to two tasks, and the reaction times (RTs) and accuracy rates are measured. Although there are various types of dual-task scenarios, the underlying assumption is that if two tasks tap into the same resource pool then capacity must be shared between them resulting in postponement of one or both of the tasks. That is, task postponements are due to capacity demanding aspects of the two tasks competing for the same cognitive resources.

Dual-tasks have been widely implemented to examine the processing of auditory or visually presented words. For example, one issue that is widely debated between researchers is whether lexical processing (or more specifically lexical access) is an automatic or attention limited process. An example of this research comes from McCann, Remington and Van Selst (2000). They used a pitch discrimination task (task 1) with a concurrent lexical decision or naming task (task 2). Through manipulating the stimulus onset asynchrony (SOA, or the time between the two tasks), the latencies to the lexical-decision and naming tasks increased dramatically. They further demonstrated that word frequency could not be processed whilst attention was devoted to the pitch discrimination task. The authors suggest that lexical processing and pitch discrimination tasks tap into the same resource pool and that some aspects of access cannot be processed while attention is directed elsewhere (however see Cleland, Gaskell, Quinlan, & Tamminen (2006) for a conflicting view).

The dual-task paradigm has also been used to study production processes. For example, Ferreira and Pashler (2002) manipulated different stages of word production to examine which stages are resource demanding. Using this dual-task logic, Ferreira and Pashler found evidence that lemma selection and phonological word-form selection are capacity demanding processes, whereas phoneme selection on the other hand is a “peripheral” process that can be completed without delay even when attention is split between two tasks (cf. Cook & Meyer, 2008).

To summarise, dual-task paradigms are useful tools for examining capacity demands and forcing participants to coordinate two tasks simultaneously. The above mentioned research also highlights that dual-tasks are a sensitive measure of assessing how two tasks can be coordinated and the effects of this coordination. This suggests that a dual-task scenario is well-suited to examining the research question of how speech production and speech comprehension can be coordinated.

1.7 Working Memory and Language

Baddeley and Hitch (1974) argue for working memory as a distinct subset of short-term memory that is involved in the temporary processing and storage of information. That is, working memory deals with the information that is active in memory at a given time. The Baddeley and Hitch model describes four components of working memory; the Central Executive, the Phonological Loop, the Visuo-spatial Sketchpad; and the Episodic buffer. I describe these components in more detail in section 3.1. The focus of this section, however, is how working memory relates to language processing.

Working memory is argued to be involved in many complex activities, including language processing (Gathercole & Baddeley, 1993). Much recent has focused on relationship between working memory and language, with researchers highlighting how components of working memory are involved in language processing. It should be noted, however, that definitions of working memory differ drastically in the literature. I take the view here that working memory is subject to capacity constraints and involves several subcomponents that interact with one another (see section 3.1).

There has been much evidence to suggest that the central executive component of working memory plays a vital role in language comprehension. Gathercole and Baddeley (1993) neatly describe three main principles that underlie the relationship between the central executive and language comprehension;. Firstly, language comprehension involves both processing and storage of information (i.e., in order to access the surface, semantic and syntactic stuctures the linguistic utterance must be processed, this processing occurs incrementally with new chunks of information being retrieved as the sentence unfolds, this thereby necessitates that 'intermediate representations' are stored as the subsequent information is processed). Secondly, the processing and storage of information is served by the same 'pool' of cognitive resources, and resources must be shared between these two tasks. The final premise, is that individuals differ in their ability to store or process information; what is known as their 'working memory capacity'. These individual differences in working memory capacity can be due to limitations in the total amount of resources available, or to the efficiency with which they carry out cognitive processes.

Evidence to support these claims comes from Daneman and Carpenter (1980). They created what is now an infamous measure of working memory capacity; the reading span task. This task requires participants to store and process information simultaneously (i.e., it engages their central executive). Individuals who are better able to coordinate processing and storage processes have a better 'working memory span'. These estimations of working memory capacity (or spans) are tightly related to measures of language comprehension, for example there is a high correlation between working memory span and verbal SAT scores. Daneman and Carpenter propose that an individuals working memory span can directly determine their ability to process language (for meaning).

There has also been many studies that have focused in on working memory capacity and syntactic processing. The majority of this work is located within the realm of written comprehension (i.e., reading studies). For example, it has been proposed that individuals with lower working memory capacities struggle more when processing a complex syntactic structure (Daneman & Carpenter, 1980; King & Just, 1991). The reason for this is thought to lie with the amount of intermediate representations that require storing as the sentence unfolds. These individuals' lack of cognitive resources or their inefficiency in processing leads to poorer performance than their high working memory opposites.

This relationship between working memory and syntactic structure has been accommodated into various theories of comprehension, for example the Causal Inference Process Model (van den Broek, 1994; van den Broek, Fletcher, & Risden, 1993), and the Capacity Constrained Comprehension (CCC) theory (Carpenter, Miyake, & Just, 1994; Just & Carpenter, 1992). The former theory describes how inferences in comprehension rely on activation spreading from long-term memory to working memory. The transference of concepts from long-term to working memory is thought to require cognitive resources. Therefore, the degree to which (predictive) inferences are generated during comprehension relies on the working memory capacity of the individual. Relatedly, the CCC theory posits that individual differences in inference ability is due to the amount of activation/cognitive resources available to store and process information. That is, individuals with low working memory capacity have less resources available to transfer concepts or representations from long-term memory to working memory; this thereby means that these individuals are less able to keep several predictive inferences in mind at one time. Individuals with high working memory capacity, however, are able to consider several possible inferences in comprehension simultaneously. These individuals are better able to deal with complex syntactic structures that entail storage and processing of several intermediate representations, to deal with syntactic ambiguity more efficiently, and make more accurate predictions about what is coming up in speech by activating multiple alternative representations (Macdonald, Just, & Carpenter, 1992). This distinction between high and low working memory capacity and syntactic processing will become more crucial in Experiments 3-6 of this thesis.

So far, I have focused on how working memory relates to language comprehension. I now turn to working memory and speech production. It has been suggested, for example, that working memory plays a role in storing or buffering speech output (Gathercole & Baddeley, 1993). Akin to language comprehension, working memory is thought to be involved in storing intermediate representations as an utterance is built (i.e., storing concepts whilst the formulation process proceeds, holding onto the syntactic structure as the phonetic plan is processed etc). Further to this, it has been argued that working memory aids the cognitive processes involved in speech production, namely, moving from one level of processing to the next. For example, working memory could be involved in accessing the mental lexicon, creating the syntactic or semantic structure, or building the phonological structure of the speech output (Gathercole & Baddeley, 1993).

Indeed, there is some evidence that the central executive component of working memory plays a role in planning speech outputs. For instance, Power (1985) asked participants to engage in a dual-task; one task involved producing a plausible sentence from two set words, the second involved a three- or six-digit load task where participants had to repeat the digits in the correct sequence. Power hypothesised that if the central executive is involved in creating syntactically and semantically plausible sentences then sentence production will be disrupted by the presence of the digit-task. The results demonstrated that under the six-digit load task participants' speech became predictable and stereotyped. This effect was not found in the three-digit or no-digit load conditions. The results were taken to suggest that the central executive is involved in early speech production processes, namely, construction of the semantic content of speech utterances. Research into the role of working memory in language continues to be a strongly developing area. However, it is clear that at least some components of working memory (i.e., central executive specifically) play an important role in language processing. I now move on to one specific part of language processing and its relation to working memory capacity; prediction. This subject area has important implications for the design of my research paradigm (as will become clear section 1.9).

1.8 Prediction and Working Memory Capacity

As alluded to in some of the previous sections, our comprehension measure relies on measuring prediction processes. My design is based upon Altmann and Kamide's (1999) study where participants must use the verb information to predict what noun will appear in the following fragment of speech. If, like Altmann and Kamide, I show that participants are able to predict using the verb then I have shown evidence of online prediction (and thus online comprehension).

In order to examine the coordination of speaking and listening I must also incorporate a production task alongside the comprehension measure, thereby creating a dual-task scenario. As highlighted in the section above, dual-tasks examine whether the tasks employed are capacity demanding and whether both tasks rely on the same resource pool. Therefore, with these points of interest in mind, also of relevance here is any research that examines whether prediction processes rely on working memory capacity and demand cognitive resources. Research in this area *could* indicate whether we should expect online prediction at the verb to rely on working memory, and consequently whether a secondary task should reduce the amount of capacity available for prediction.

There is little research that examines prediction and working memory, however some researchers have investigated the link between prediction in written comprehension and working memory capacity. For example that by Gibson and Thomas (1999). They argue that when a sentence contains many referents, verb or noun-phrases and sentence heads as is the case in “The apartment that the maid who the service had sent over was cleaning every week was well decorated” (taken from Frazier, 1985) individuals can no longer make predictions about all of the verb phrases present as the demand on memory is too high. In the above sentence participants would not be able to make predictions based on the middle verb phrase as storage and processing costs are too great. This has subsequently been named the “VP (Verb-Phrase) forgetting hypothesis” and has been replicated in previous studies, such as that by Vasishth and colleagues (Vasishth, Suckow, Lewis, & Kern, 2010). This research therefore supports the notion that predictions in comprehension aredependent on working memory capacity.

Related research comes from Estevez and Calvo (2000). They carried out a naming task where the probe word to be named was either a predictable continuation to the sentence (as set up by the context) or an inconsistent continuation. Individual differences in working memory capacity were also measured using areading span task (Daneman & Carpenter, 1980). Results indicate that high working-memory capacity speeds up the time course of prediction, however prediction does not become automatic. Similar findings are reported by Linderholm (2002), who speculates that prediction in reading is a demanding process and that in some instances only those with high working memory capacities are able to make predictive inferences. Although this research links prediction with written rather than spoken comprehension, it suggests that working memory capacity plays an important role in predicting upcoming themes.

Although there is little research linking prediction and working memory capacity, and no research that I know of that investigates this link in spoken comprehension, the evidence that is present indicates that prediction *is* dependent upon working memory capacity. This could therefore suggest that predicting at the verb in the comprehension task may be disrupted by the presence of a secondary task. However, there are vast differences between written and spoken comprehension, and individuals are extremely well practiced in coordinating their speaking and listening, therefore the participants in this study could be resilient to the presence of a dual-task.

1.9 The Paradigm

This section introduces the design of the dual-task I employ in this research. The design incorporates techniques described both in the dual-task section outlined above (section 1.6) and the two sections on the visual world paradigm (sections 1.3 and 1.4). The first task presented within the dual-task scenario is a speech production task. It involves presenting participants with three object pictures and asking them to create a sentence that incorporates these set objects. There is no set format or syntactic structure that must be employed however participants must try and include as many of these objects as possible within their utterance. On half of the trials participants are asked to produce their utterances immediately, and, on the other half of the trials, participants must coordinate the planning of their utterance with a secondary speech comprehension task (therefore allowing a comparison of a single-task condition with a dual-task condition).

The secondary comprehension task utilises the visual world paradigm. My design is based on the study by Altmann and Kamide (1999) who demonstrated how verb information can be used online to predict upcoming referents. For example, when presented with a visual display containing one edible item and three inedible items, hearing the (specific) verb "eat" cues participants to pick out the only edible item as the upcoming referent. This is contrasted with the case where participants hear the (general) verb "move" which could signal any of the four objects in the visual scene. If you recall, this prediction effect manifested itself in faster eye-latencies when the verb highlighted one particular referent (what we have named specific verbs) compared to when the verb could signify any of the four objects in the visual display (general verbs).

Crucially, in the visual world paradigm participants' eye-movements reflect where their visual attention is focused. It is thought that the link between speech input and eye-gaze arises because the "verbal information affects the listeners' allocation of attention, which in turn governs the direction of their gaze" (Huettig et al., 2011, p. 166). Furthermore, visual attention is thought to demand central capacity (Broadbent, 1958, 1982; Bundesen, 1990), and visual or linguistic information that requires more processing entails more attention, which is mirrored in longer or delayed fixation times in the visual world. This link between visual attention and eye-movements allows researchers to examine fixationtime and duration in order to learn about the processing demands involved in a particular linguistic structure or task. Because verbal information directs visual attention, and eye-gaze is a measure of where visual attention is focused, I can thereby examine the demands imposed by various linguistic inputs and tasks. If the structures or tasks I employ reduce the amount of capacity available for comprehension and visual attention, then this should be mirrored in participants' eye-movements.

The link between visual attention and eye-gaze and the proven effect of predicting with verbs will be exploited in this research to specifically examine whether comprehension still proceeds online when capacity must be shared between two tasks. If comprehension still occurs online we should see a significant difference in eye-movement latencies after hearing the specific verbs compared to after hearing general verbs. If, however, comprehension is delayed in the presence of the secondary production task the difference between hearing a specific or a general verb may be decreased or completely extinguished. Using dual-task logic, if production and comprehension processes tap into the same pool of cognitive resources then capacity will have to be shared between the two tasks and, therefore, we should see delayed responses to one or both tasks. Therefore, this design allows me to 1) examine whether prediction at the verb is a capacity demanding process, and 2) gain some insight into how production and comprehension processes are coordinated (especially in cases where the interlocutors do not have access to processes of alignment). More details on the exact nature of the dual-task I employ can be found in Experiments 4 to 6.

Dual-task methodology has not yet been used alongside the visual world paradigm however this allows the opportunity to uniquely examine production and comprehension alongside one another whilst monitoring the processes using an online, time-sensitive technique.

1.10 Overview of thesis

In Experiment 1, I report the results of a replication of Altmann and Kamide's (1999) study into the incremental nature of speech comprehension and the ability to use verb information online. I demonstrate how, when using an extended set of verb pairs and a highly controlled visual world display, we obtain very similar eye-movement latencies to those documented in the original study.

Experiment 2 examines how this online prediction process is affected by the presence of a secondary task. I implement a dual-task design whereby participants must coordinate speech comprehension (the visual world task) with a concurrent digit-matching task. I demonstrate that online prediction is preserved in this dual-task setting and that a secondary phonological task does not compete for the same cognitive resources as speech comprehension.

Experiment 3 extends the stimulus set to 56 verb pairings that are used to build both simple and complex syntactic structures. Previous research has indicated that different processing demands are imposed in simple and complex sentences, and that manipulating syntactic complexity is an effective way of influencing resource-availability. I aimed to observe whether the ability to use verb information online to predict upcoming referents is maintained in complex syntactic structures, or whether this ability is reduced relative to its simple syntactic counterparts. I replicate the finding that complex syntactic structures require more processing effect, as evident in an overall increase in processing time. However, I also show that online prediction (at the verb) is resilient to these increased capacity demands, with the strength of the prediction effect being equivalent across the two levels of complexity.

Chapter 5 introduces a new element to our experimental design; speech production. Experiments 4, 5 and 6 (Chapters 5-7) examine how a speech comprehension and a speech production task are coordinated. Eye-tracking in the visual world allows us to monitor whether participants continue to comprehend online when comprehension is coordinated with a production task. Through measuring onsets in speech production I am also able to examine whether speech production is affected by the presence of a comprehension task. Throughout these three experiments I illustrate how efficiently and effectively two complex speech tasks can be coordinated. I demonstrate how capacity must be shared between the two tasks and, as the overlap between comprehension and production increases, the ability to predict online decreases. However, I also show that the ability to predict online is extremely resilient to capacity demands and that although prediction is reduced the effect always remains highly significant. I discuss how the ability to predict online aids comprehension and can help to conserve resources

Finally, Chapter 8 reiterates the main findings of this research project and discusses these in relation to previous research in speech production, speech comprehension and dialogue. I also discuss the implications of this work, important methodological issues and the potential progression of this research area.

CHAPTER 2

PREDICTING ONLINE AT THE VERB

2 Experiment 1

2.1 Goals and Motivation.

This first experiment examines online spoken comprehension through the manipulation of verb information. Following from Altmann and Kamide (1999), I aimed to replicate that information extracted from the verb can be used to predict which of four objects will be referred to in upcoming speech. Within this replication I hoped to demonstrate that prediction at the verb with a wider set of stimuli (32 verb pairs compared to Altmann and Kamide's 16 pairs) and with a more standardised visual world display. In Altmann and Kamide's original study, the visual displays did not appear to be controlled. Sometimes the agent's eye-gaze is directed at the target item and sometimes it cues a distractor item; sometimes the target is closer to the agent than the distractors, and on other occasions it is further away. Therefore, in order to ensure standardisation across visual displays, the displays I implemented here contained no cues from the agent's eye-gaze and all objects are equidistant from the agent. Replicating the use of verb information in this study would allow the new stimuli and displays to be used in subsequent experiments, and crucially, enable online prediction to be utilised as a sensitive measure of comprehension which can be tracked during various task manipulations.

Altmann and Kamide (1999) presented participants with a visual display containing one agent, one target item and either 3 or 4 distractor items. Alongside the display, participants heard a sentence that either contained a verb that could refer to the target only, or a verb that could refer to the target or the distractors. For example, in the visual world context of a boy (agent), cake (target), toy car, toy train and ball (distractors), participants could be presented with either

(1) The boy will *eat* the cake (verb refers to target item only)

(2) The boy will *move* the cake (verb can refer to target or distractors)

Altmann and Kamide either asked participants to perform a meta-linguistic task which involved making judgements about whether the auditory sentence could apply to the visual scene (Experiment 1) or just recorded their eye-movements with no additional language task (Experiment 2).

Altmann and Kamide found that participants looked to the target significantly more than the other distractors, and that the onset of the first saccade to the target post verb onset was significantly faster in the eat condition compared to the move condition (what I will subsequently refer to as specific verb and general verb conditions). These first post-verb onset saccades were made 85ms *before* noun onset in the eat/specific condition, and 127ms *after* noun onset in the move/general condition (Experiment 1). This demonstrated that participants' eye-movements were mediated by the verb, and that verb information could be accessed online and used to predict upcoming nouns.

The experiment I document here is modelled on Altmann and Kamide's second experiment which does not require any meta-linguistic judgements to be made about the material presented. I present participants with a visual scene comprised of an agent, a target item, and three distractor items. Alongside the visual scene participants heard a sentence containing a verb that can refer to the target only ("Specific verb condition") or can refer to any of the items on screen ("General verb condition"). Some of these verb pairings have been taken from Altmann and Kamide's study; however I have extended the stimuli set to include an additional 16 verb pairs. The visual scenes were newly created and additional factors such as item saliency, gaze direction of the agent, and distance between agent and items were controlled; all items were line drawings, the agent's gaze was toward the participant and not directed at any of the items, and the agent was equidistant from all items (an example can be seen in Figure 3). All sentences contain the same syntactic structure; ‘The (*agent*) will (*specific/general verb*) the (*target*)’. Examples of the sentences include:

(1) The man will drive/admire the car (church, television, watch)

(2) The boy will smash/pass the plate (slipper, cushion, balloon)

(3) The woman will sharpen/break the pencil (bottle, computer, telephone)

2.2 Hypotheses.

I expect to replicate an effect of verb type, with participants looking to the target items significantly faster in the specific verb condition compared to the general verb condition.

2.3 Method.

Participants.

34 Undergraduate students from the University of Birmingham (28 female: 6 male, ages 19-26) took part in return for either course credits or payment of £4.00. All participants were native English speakers and had normal to corrected vision and hearing.

Materials and apparatus.

32 sets of stimuli were used. Each set was comprised of one visual scene and two associated auditory sentences. Each visual scene incorporated an image of a person (man, woman, boy, or girl) and four object pictures, one of which was the target. The visual scene was presented on a 15’’ computer monitor at a resolution of 1024x768pixels. The picture stimuli were black and white line drawings, 178x142 pixels in size. The pictures were taken from Snodgrass and Vanderwart (1980); Szekely et al. (2004); Clip Art; created using Paint; or hand-drawn.

The auditory sentences were recorded at a slightly slower than normal speaking rate, by an English native speaker with a neutral accent (mono, 64 kHz sampling rate, 16 bit sampling resolution).

64 auditory sentences were created in total; 32 were comprised of a specific verb that could refer to the target item only (e.g., eat), and the other 32 contained a general verb that could refer to any of the four items (e.g., move). The sentence pairs were identical apart from the verb.

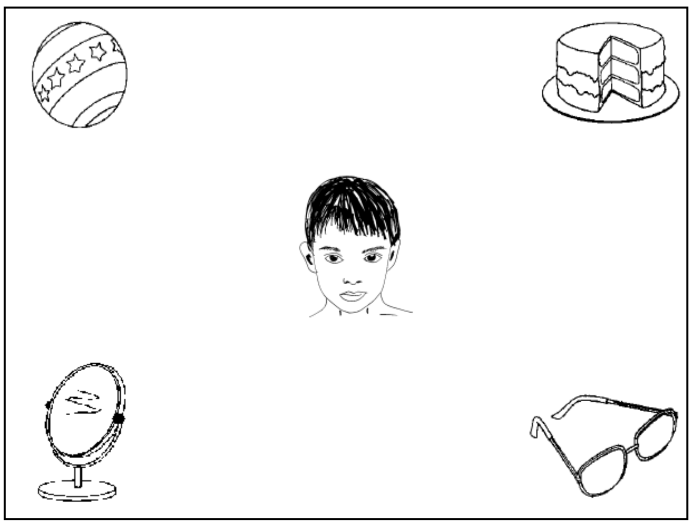

Figure 4. Example of a visual display; the agent is the boy in the centre, the target is the cake, and the three remaining pictures are distractors.

The specific and general verbs were matched for frequency. 26 of the 64 sentences were taken from Altmann and Kamide (1999) and the remaining 38 sentences were newly created. A list of the stimuli used in this experiment can be found in Appendix 1A. (Please note, Appendix 1 appears at the end of this thesis and includes information on the materials used, Appendix 2 appears in the accompanying CD and contains the scripts and outputs from all linear mixed effects model ran throughout this research).

Eye-movements were recorded using a desktop-mounted EyeLink 1000 eye-tracking system. Recordings were taken from the right eye, sampling at 1000Hz.

Design.

The experiment used a repeated measures design; each visual scene was viewed twice, once with the specific sentence and once with the general sentence. The independent variable was which type of verb (specific or general) was heard, and the dependent variable was the latency of eye movements to the target picture.

The location of the target picture was counterbalanced across trials, appearing in each corner of the screen an equal number of times. The order of sentences was randomised, but programmed to ensure that no corresponding specific and general sentences were heard adjacently. Half of the sentences were presented in the specific verb condition first, half in the general verb condition first.

Procedure.

Participants were sat in front of a 15" computer screen with their head placed on the chin rest of the desktop eye-tracker. Participants were seated approximately 57cm away from the display. Participants were instructed that the experiment measured changes in pupil dilation as they viewed a visual scene; an approach originally implemented in early visual world paradigms, such as that by Cooper (1974). The instructions in this study made clear that participants did not have to make any manual response to the stimuli presented, but that they had to keep their eyes on the screen at all times. These instructions were chosen to minimise participants' attempts to uncover the true motive of the experiment.

At the start of the experiment I conducted a 9 point calibration of participants' eye-movements. All calibrations were validated by the EyeLink machine and any poor calibrations were repeated by the experimenter. The calibration took approximately 30s and the experiment itself ran for 15 minutes.

Participants were presented with a preview of the target and distractor pictures for 3000ms, followed by the onset of the agent's face that appeared in the centre of the screen. The onset of the sentence was dependent upon fixation on this central image.

2.4 Linear Mixed Effects Models and 'R'.

All the eye-movement data presented in this thesis have been analysed using linear mixed effects models. This statistical technique offers several advantageous features; it is flexible and can deal well with large complex data sets, missing data, and non-spherical error variance across items and participants (Barr, Levy, Scheepers, & Tily, 2013; Winter, 2013). Furthermore, alternate F1 and F2 analyses have recently received much criticism (Baayen, Davidson, & Bates, 2008; Brysbaert, 2007; Raaijmakers, 2003; Raaijmakers, Schrijnemakers, & Gremmen, 1999).

Bates (2005) describes linear mixed effects model as characterising "the dependence of a response on one or more covariates" whilst also characterising "the 'unexplained' variance in the response" (Bates, 2005 p. 27). For example, in Experiment 1 I am interested in examining the fixed effect verb type whilst also considering the unexplained (or random) variance that comes from participants and items. Within a linear mixed model, we examine how the dependent variable changes across different levels of the fixed factors (for instance across the two verb types), and also examine the variation in response to random factors. Dealing with subjects and items as crossed random effects in this way allows us to model all variance across subjects and items and ensures that we retrieve an accurate significance value for our fixed effects (Baayen et al., 2008).

I performed linear mixed effects models using the open source program 'R', version 2.14.2 (R Development Core Team, 2012). In each experiment I analysed the latency to targets by examining the onsets of first gaze (to targets) made after the point of verb onset. The eye-movement latencies to the target could be directly compared across the two conditions; the latency after hearing a specific verb compared to the latency after hearing the general verb. If the specific verb allows for prediction of the upcoming verb we should see faster eye-movement latencies in this condition. This is an effective way of examining prediction without having to calculate cumulative likelihoods of fixations across the targets and all distractors.

Methodology for linear mixed effects models.

Before describing the results I will explain the methodology used to fit the linear mixed effects models. (The procedure for fitting the models is repeated across all experiments).

I began by modelling the random effects; this involved examining the variance across participants and items. For example, participants may differ considerably in their overall eye-movement latencies (this is a random effect of participants on the intercept, or the *by-participant intercept*); participants' eye-movements may also differ considerably from one item to the next (a random effect of items on the intercept, or *by-item intercept).* Alternatively each participant may show different levels of response to the factors. For example, in Experiment 1, some participants may show large differences in eye-movement latencies across the specific and general verb conditions whilst other participants may only show a very small difference; this is called a random effect of the *by-participant "slope"*. If the design includes more than one fixed effect than there will be multiple by-participant and by-item slopes; this will become more obvious in the subsequent chapters when more complex experimental designs are implemented.

It is important to examine variance on both the intercept and slopes in order for results to be generalised to the population from which the participants are drawn and to other items. According to Barr et al. (2013) only examining intercepts and ignoring slopes is in fact worse than only conducting an F1 analysis in an ANOVA, as this can drastically inflate Type I errors and lead to anti-conservative p-values.

In the model analyses documented here, I explore all possible random effect models (tested with a maximal model for fixed effects) and examine which model describes our data best. That is, I compare a “maximal” random effects model that includes all possible intercepts and slopes and compare this with simpler models where only some intercepts or slopes are included. The best random effects model is chosen by testing all the different combinations of parameters present. I chose the best model based on its AIC value; the better the model fits the data the lower the AIC value. This technique has been described as an “information-theoretic approach” (Burnham & Anderson, 2002). The AIC, or Akaike Information Criterion, is a measure which balances “quality of fit” against the number of parameters in the model (Bates, 2005). In essence, the AIC value is a calculation of how well a model fits (or describes) your data set. The AIC value can be used to rank several models from the best (i.e., a structure that models the data best) to the worst, and has been described as “a simple, compelling concept” which is “easy to compute and effective in a wide variety of applications” (Burnham & Anderson, 2004, p.10 – 11).

If two models share the same AIC value or are not significantly different from one another then the most parsimonious model is chosen to avoid over-fitting the model. Over-fitting the model refers to a Type II error, where the p-value is underestimated as variance is allocated to random effects rather than fixed effects.

The parameters of the random effects models are always tested using the lmer function in 'R' (Bates, Maechler, & Bolker, 2012) and likelihood ratio tests (see Bates (2005) for a description of the lmer procedure). Other researchers, such as Barr et al. (2013), argue that a maximal structure for random effects should always be used; this means that all parameters where random effects could appear in your data should be modelled in the random effects structure. However, the examples they offer in their recent 2013 paper involve more simple experimental designs with only a single factor and two levels. Whilst this 1x2 design mirrors that of Experiment 1, as the experiments continue we include several more factors each with more than one treatment level. As the design becomes more convoluted and the number of fixed effects increase, the number of random effects also increases; this means that the number of parameters in the random model increases very rapidly. Likewise, as the number of parameters increase and the random model becomes more complex, the potential for over-fitting the model becomes more likely, thus increasing the chance of making Type II errors and missing a significant effect where one does exist. Therefore, in the experiments I describe in this thesis (excluding Experiment 1), using a maximal structure could entail an increased risk of under-estimating the fixed effects.

Consequently, here I use what I believe to be a happy median; I test all different random effects parameters to ensure that I am incorporating all the random variance into my model, and I choose which structure models this random variance the best as to avoid the issue of over-fitting. It should be noted that the maximal model of random effects is included in this comparison wherever possible, however, in several cases these models fail to converge as there are too many parameters to fit. Where this happens, the model is simplified by examining slopes for each factor but no longer estimating different covariances for each by-participant or by-item slope. The best model is chosen based on its AIC value. Indeed, when exploring the different parameters of a random effects model the AIC value is used a calculation of which model is doing the best job of describing the empirical data (Burnham & Anderson, 2004).

Once the best random effects model has been chosen I then go on to examine the fixed effects. In Experiment 1, for instance, I need to test whether the fixed effect of verb type (otherwise known as the independent variable) has a significant impact on the model. A significant impact on the model is akin to a significant main effect in an analysis of variance (ANOVA). In this case, a significant impact of verb type on the model indicates a significant difference between the specific and general verb conditions. Again, the linear mixed effects model for fixed effects is always fit using the lmer function in 'R' and likelihood ratio tests.

When there are multiple fixed effects (as is the case in Experiments 2-6) a backwards selection to narrow down the parameters to be included in the model. This involves starting with all parameters included in the model, that is, all fixed effects and all interactions between the fixed effects, and removing parameters one by one to examine whether the model becomes better or worse as a result. The effect or interaction that is shown to be contributing least to the model (as shown by the AIC and p value) is removed first and this procedure is continued until the factor on the last step is significant at the 0.05 level. If there is a significant interaction in the data, the data set is split to examine where this interaction stems from.

This procedure of first selecting the random effects model and then fitting the fixed effects model is repeated across all experiments. Therefore, in each results section the linear mixed effects models will be split into the subsections of "Random effects model" and "Fixed effects model".

2.5 Experiment 1 Results.

2.5.1 Eye-movement data.

Eye-movement variables.

I was only interested in analysing the eye-movements made within specific areas of interest; namely, the area surrounding the central face and the target picture. I used EyeLink Data Viewer to create an interest area report for the eye-movement data. The report included information on whether the target was fixated and when it was fixated in relation to the onset of the verb. Other information contained in the report included the type of verb encountered and timings relative to each particular trial, for example, when the sentence started playing.

Exclusion criteria.

Any trials in which participants did not fixate the target were removed from analysis (522 trials were removed from the general verb condition, 439 from the specific condition). It is thought that we take around 200ms to program an eye-movement (Matin, Shao & Boff, 1993), therefore, I was only interested in eye-movements made 200ms or more after verb onset; our rationale here being that any fixations to targets prior to this time point could not possibly be mediated by information gleaned from the verb. Twenty-six trials were excluded due to fixating before the 200ms time-point in the general condition, forty-three trials were excluded from the specific condition. Any eye-movements that were over 2.5 standard deviations above the participant mean were also removed from analysis. However in this experiment no trials fitted this criteria.

In total in Experiment 1, participants fixated on the target in 56% of the trials. They fixated the target on 52% of trials in the general verb condition, and on 60% of the trials in the specific verb condition.

Table 1. The number and percentage of valid fixations in the specific and general verb conditions. [Experiment 1].

	Number of Valid Fixations	Total Number of Trials	% of Trials with Valid Fixations
Specific Verb Condition	649	1088	60
General Verb Condition	566	1088	52
Total	1215	2176	56

2.5.2 Linear mixed effects model.

Random effects model.

First, I chose the best random effects model. As described previously, our data could be modelled with random effects on the intercept, on slopes, or both and on items, on participants, or both. Here, the preferred random effects structure here includes participants and items as random effects on the intercept; (1 | PptNum) + (1 | Item). As mentioned earlier each random model is chosen based on the best (i.e., lowest) AIC value. The random model chosen always has a lower AIC than the maximal (or fullest tested) model (see Appendix K).

Fixed effects model.

The next step involves working with the fixed effects. Verb type is shown to add significant weight to the model (χ^2= 77.938, $p < .001$). In other words, we see a significant main effect of verb type. (For comparisons sake, the corresponding main effect in an ANOVA here would be $F(1, 33) = 95.89$, $p < .001$). Table 2 (below) compares the mean latencies to the target across the two conditions, and demonstrates significantly faster latencies in the specific verb condition compared to the general verb condition, with a massive 202ms advantage for specific verbs. The scripts for running this linear mixed effects model can be found in Appendix 1B and 2A.

Table 2. Mean eye-movement latencies to targets (in ms) across the Specific and General conditions, relative to the verb and noun onsets. [Experiment 1].

	Specific Condition (Eat)	General Condition (Move)	Difference between conditions (eat-move)
Verb Onset	1019 (SD 370)	1221 (SD 432)	202
Noun Onset	218	385	167

Note: The standard deviation of eye-movement latencies is shown in parentheses.

2.6 Discussion.

Replicating Altmann and Kamide's findings, I have demonstrated the online use of verb information. Eye-movement latencies to targets are shorter when preceded by a specific verb that complies with only one of the object pictures depicted. These results suggest that (1) the simple active sentences, such as "the boy will eat the cake" are interpreted online, on a word-by-word basis, and (2) the amendments to the visual scenes and the extension of the stimulus set was successful; my new verb pairs and line drawings still allow for online prediction at the verb. This, therefore, enabled me to use this extended stimulus set in all subsequent visual world experiments.

Although the latencies documented here could be construed as slightly delayed, I believe this can be explained by the distinction between tasks that incorporate a metalinguistic task and those that include little to no instruction. Altmann and Kamide (1999) demonstrated this difference through giving participants both specific instructions (Experiment 1) and no instruction (Experiment 2). When given no instruction or meta-linguistic task, reaction times similar to those reported here were obtained (1246ms and 988ms, Altmann & Kamide; 1221ms and 1019ms, here). However, when subjects had to make a response to the presented material (judge whether the target was present in the display) their responses become significantly faster (611ms specific; 838ms general). Altmann and Kamide propose that this delay in the former context is due to participants being less inclined to form an association between the sentence and the visual display. Therefore, the relatively long latencies I report here are not out of place considering participants had no specific instruction. Regardless of the delay in latencies, however, Altmann and Kamide were able to demonstrate that verb information can help to guide participants' eye-movements even in the absence of a meta-linguistic task.

Whilst it is most likely that our results reflect online prediction, there is an alternative explanation for these findings. Analysis of the distance between the verb and the noun across the two conditions revealed that the sentences containing specific verbs had a significantly shorter verb-noun distance ($M=740.4$, $SE=15.9$) than the sentences containing general verbs ($M=903.9$, $SE=18.9$); $t\ (62) = 6.627$, $p < .001$). Therefore, in the specific condition the noun information was heard earlier and for a longer duration compared to in the general condition. This methodological error could account for the earlier eye-movements to targets in the specific condition; it is possible than subjects' eye-movements were mediated by noun information rather than verb information. In order to rule out this possibility the task was repeated with the verb-noun (V-N) distance controlled (Experiment 2).

CHAPTER 3

EFFECTS OF DIGIT-LOAD ON ONLINE PREDICTION

3 Experiment 2

3.1 Goals and Motivation.

Experiment 2 aims to remove the possibility of that eye-movements are faster in the specific condition due to shorter verb durations and nouns information being heard sooner. To ensure that it is the *verb* information that mediates eye-movements, the verb-noun (V-N) distance for each specific-general sentence pair was altered to be within 10ms of one another with no significant difference across conditions (specific, M= 715ms, SD= 85ms; general, M= 716ms, SD= 84ms, $t(31) = -.898$, $p = .38$; see Appendix 1C). Furthermore, to ensure that other confounding variables, such as uniqueness point and age of acquisition, do not affect the ability to process verbs across the two conditions I performed a lexical decision task that compared the time taken to recognise all the verbs utilised in this research (see section 2.3).

A second objective of this experiment was to examine whether the ability to anticipate a sentence's upcoming theme is resource limited or can proceed in a relatively automatic manner. Similar questions concerning capacity have targeted other areas of language processing. For example, McCann et al. (2000) used a dual task study to examine whether spoken word recognition demands cognitive resources. Tasks that tap into a limited resource pool are thought to reduce the capacity available to other resource demanding processes. As a result, performance on one or both tasks will be diminished. As described in the introduction, McCann et al. demonstrated that lexical processing cannot be performed in conjunction with another capacity demanding task and therefore conclude that word processing must be resource limited.

Whole sentence comprehension involves several processing levels before a complete representation is formed. Whilst dual-task studies have suggested that even the individual processing levels are capacity demanding (e.g., Hohlfeld, Sangaks, & Sommer, 2007; McCann et al., 2000), online measures of comprehension, such as the visual world paradigm, have indicated that these processes are dealt with incrementally as the sentence unfolds (Allopenna et al., 1998; Altmann & Kamide, 1999; Huettig & Altmann, 2005; Kamide, Altmann & Haywood, 2003; Tanenhaus et al., 1998). Indeed, the results of Experiment 1 provide evidence for this incremental processing. Such visual world studies, however, have not manipulated resource availability; participants' only requirement is language comprehension. It is, therefore, unclear whether prediction in comprehension can still be achieved online when participants have to share capacity between multiple tasks.

This experiment examines whole sentence comprehension and, through implementing a secondary task, directly assesses whether prediction in comprehension is a resource limited process. Resource limited processes are thought to require access to a system known as 'working memory'.

What is working memory?

Working memory, as I describe here, is based upon the Baddeley and Hitch model (1974). Baddeley and Hitch argue for working memory as a distinct subset of short-term memory that is involved in the temporary processing and storage of information. That is, working memory deals with the information that is active in memory at a given time.

The original working memory model is split into three parts, the first of which is the *Central Executive*. Argued to be the most important part of working memory, the central executive is involved in controlling the flow of information, processing and storing information, and accessing information from the long-term memory store (Gathercole & Baddeley, 1993). The central executive has a limited processing capacity and functions most efficiently when all resources are allocated to the executive system. Tasks that are thought to require the central executive include mental arithmetic (Hitch, 1980), language comprehension (Gathercole & Baddeley, 1993) and reasoning (Baddeley & Hitch, 1974). It has also been argued that the central executive may even form the basis of general intelligence (Kyllonen & Christal, 1990). Baddely (2003, pp. 835) neatly describes the executive as a "homunculus, the little man taking all the decisions".

The central executive feeds input into its two sub- (or "slave") systems; the *Phonological Loop* and the *Visuo-Spatial Sketchpad*. The phonological loop is specialised for the storage of verbal material and is further divided into the phonological store and articulatory rehearsal. The former subcomponent is involved in storing phonological input, and in order to prevent this information from decaying and being lost, articulatory rehearsal is required to refresh the phonological store. Rehearsal also works to recode printed words and pictures and other non-phonological information into a phonological representation that can be maintained in the phonological store. The second slave system is the Visuo-Spatial Sketchpad. This system is dedicated to processing and storing visual and spatial information and generating images.

Recently, researchers began to notice that the original model outlined above could not account for some of the findings in the literature, for example, the benefit of chunking information which allows individuals to immediately recall a sentence as long as 15 words, but only 5 or 6 unrelated words (Baddeley, 2003). Another crucial flaw was that the original model did not allow for the two sub-systems to interact with one another. The model has now been extended to include an additional structure; the episodic buffer.

The episodic buffer is assumed to be controlled by the central executive and, unlike the other components, it can be brought into conscious awareness. It offers a way for the phonological loop and the visuo-spatial sketchpad to interact and for different types of information to be integrated together. According to Baars (2002) the buffer can be thought of as the central executive's storage component.

One thing that is agreed upon in the working memory literature is that the components work most effectively when attention can be allocated to one task only. However, when multiple tasks compete for working memory resources the components must work harder to coordinate the necessary computational processes. These multiple tasks are, therefore, competing for available resources. The outcome of this competition will be the diminished performance of one of the tasks, or, possibly, both tasks, compared to when the tasks are completed without competition. This is the 'dual-task interference effect' (see section 1.6).

In relation to our research question here, if prediction is capacity demanding, and thus requires the involvement of working memory, resources must be shared between prediction processes and processing of the secondary task. As a result, we should observe diminished performance in one or both tasks. However, if prediction proceeds independently of working memory, in a more automatic manner, performance of the two concurrent tasks should proceed without disruption.

The secondary task employed in this experiment is a digit matching task; digit tasks are widely assumed to require working memory and commonly employed to examine working memory capacity. In order to compare or recall digit strings, written stimuli must be recoded into phonological form, temporarily held in the phonological store, and refreshed via articulatory rehearsal to prevent decay or loss; these processes require resources (Klapp, Marshburn & Lester, 1983). Here, participants see two sets of four digits, before and after the visual world task, and are asked to distinguish whether these two digits sets are identical or different. The accuracy rate of digit-matching was recorded.

3.2 Hypotheses.

If accessing and using verb information to predict is a resource limited process then we will see reduced performance in the digit task, a reduction or absence in the amount of prediction employed (as measured by the difference in eye-movement latencies across the specific and general conditions), or both. If, however, this prediction can occur in a relatively automatic manner then we will see preserved performance in both tasks.

I also expect to see some dual-task interference, with delayed latencies in the dual-task condition relative to when comprehension is completed on its own.

3.3 Method.

Prior to running the dual-task study I collected a separate sample of participants to take part in the lexical decision task mentioned above. This task was designed to test whether participants process the specific and general verbs presented in this study (and in Experiments 3-6) at a similar rate. If the two types of verbs are accessed at similar rates then there is no processing advantage for one verb type over the other and the differences in eye-movement latencies can be put down to prediction effects.

3.3.1 Lexical decision task.

Participants.

17 native English speaking participants took part (16 female, 1 male, ages 18-21, mean age 19 years, 2 months). None of the participants had taken part in Experiment 1 and all were subsequently prohibited from taking part in Experiment 2 (to avoid any disruptive memory effects). All participants were from the University of Birmingham, had normal to corrected vision and hearing and took part in return for course credits.

Materials, design and procedure.

The lexical decision task was run using the computer program E-prime. The task required participants to decide if a given item was a word or non-word, pressing the “m” key on a computer keyboard for a word, and “n” for a non-word. Each word was presented visually and remained on screen for 1000ms or until the participant had responded by pressing the keyboard. The next trial then began after a 1000ms delay, during which time the visual word was replaced with a fixation mark (+). Participants completed 10 practice trials, followed by two experimental blocks; there were 224 experimental trials in total, 112 containing real words (56 specific verbs, 56 general verbs[1]) and 112 containing non-word items. The non-words were created using the online pseudo-word generator ‘Wuggy’, recently put forward by Keuleers and Brysbaert (2010), which uses numerous criteria including subsyllabic structure, letter lengths, neighbours, and transition frequencies to choose the best possible pseudo-word matches. The task took approximately 10 minutes to complete. For a detailed list of the stimuli used in this lexical decision task see Appendix 1D.

[1] Please note, 112 verbs were tested in this lexical decision in preparation for future experiments. However, only 64 verbs were tested as part of Experiment 2 to match those presented in Experiment 1.

Results.

To ensure that general and specific verbs are equally easy to identify, I compared reaction time responses to general and specific verbs. The lexical decision task revealed no significant difference in word recognition across the two conditions (t (16) = -.507, p = .62), see Appendix 1E. Therefore, any future differences between the two types of verb can be safely interpreted as evidence of online comprehension effects. Table 3 below shows the means and standard deviations of reaction times in the lexical decision task.

Table 3. The mean and standard deviations of reaction times in the lexical decision task across specific and general verb types. [Experiment 2, Lexical Decision Task].

	Mean Reaction Time (ms)	**Standard Deviation (ms)**
Specific Verbs	633	195
General Verbs	627	184

3.3.2 The dual-task.

Participants.

34 students from the University of Birmingham (25 females, ages 18-38, mean age 20 years) took part in this study in return for course credits or payment of £5. All participants were native English speakers with normal to corrected vision and hearing. Five additional participants were tested but removed from the analysis due to equipment malfunction (n=1), not being a native English speaker (n=3) and an astigmatism on right eye (n=1).

Materials and apparatus.

All visual displays were taken from Experiment 1, and tested alongside the newly altered auditory sentences. These distance betweent the verb and noun in these auditory sentences were matched by splicing the sentences using the program Praat and removing segments of silence between words. For the working memory task, 32 matching digits pairs (6547-6547), and 32 mismatching pairs (6547-6587) were presented. The mismatching digit pairs always differed by one digit. The order of presentation of matching and mismatching pairs was randomised for each participant. The same apparatus was used as in Experiment 1 with the addition of a button box. Participants used the button box to respond to the digit matching task, the left-hand red button was used if the digit sets matched, and the right-hand green button was used if the digits mismatched.

Design.

Participants carried out both the visual world only task (as in Experiment 1) and the visual world with working memory task (from now on referred to as the "dual-task"). The order in which participants completed each task was counterbalanced. This allowed both a direct comparison with Experiment 1, and also an assessment of the impact of working memory load.

Procedure.

Again, the procedure is predominantly the same as in Experiment 1 with a couple of minor amendments. The instructions given to participants were more explicit; they were asked to first familiarise themselves with all object pictures in the preview stage, and then to look at whichever object was mentioned in the auditory sentence. The reasoning here was that participants did not appear to take in the whole visual scene in Experiment 1, with eye-gaze often restricted to one or two of the object pictures. As a result, when encountering the verb participants may not have known where to direct their attention, in fact many participants were searching the screen for the referent during the critical stages of the trial. Latencies, therefore, might not represent a purposeful fixation motivated by verb information but a saccade made whilst searching around the screen.

During the dual-task, participants initially observed a set of four digits for 1500ms, followed by the preview scene, central face, and auditory sentence (i.e., the visual world task). A second set of digits then appeared and remained present until the participant responds. Participants responded to the working memory task via a button press, as described above. Participants' eye-movements were calibrated at the beginning of each task. The experiment ran for approximately 30 minutes and participants were offered a break in between the two tasks. Participants were not given feedback on their performance.

3.4 Results.

3.4.1 Digit matching task.

It should be noted here that our aim was not to look at how the comprehension task affected digit matching, but how digit matching affected comprehension. Therefore, the digit matching data was analysed only for accuracy ratings. Participants who performed very poorly in the digit-matching should be removed from the eye-movement analysis (as these participants are not accurately coordinating two tasks). However, all 34 participants performed extremely well on the digit matching task with over an 80% accuracy level, and therefore no participant data was eliminated from the mixed effect model.

3.4.2 Eye-movement data.

Eye-movement variables.

The eye-movement variables used in analysis are the same as those described in Experiment 1.

Exclusion criteria.

The exclusion criteria were also described in Experiment 1. 880 trials where participants did not fixate on the target were removed from analysis (629 trials were removed from the general verb condition, 251 from the specific condition). 55 trials were excluded due to fixating before the 200ms time-point; 29 trials were excluded from the general verb condition and 26 were removed from the specific verb condition. Any eye-movements that were over 2.5 standard deviations above the participant mean were also removed from analysis. 16 trials were removed in total; 14 from the general condition and 2 from the specific condition.

In total, participants fixated on the target in 77% of the trials. They fixated the target on 67% of trials in the general verb condition, and on 87% of the trials in the specific verb condition.

Table 4. The number and percentage of valid fixations in the specific and general verb conditions. [Experiment 2, Comprehension Task].

	Number of Valid Fixations	Total Number of Trials	% of Trials with Valid Fixations
Specific Verb Condition	1669	1920	87[2]
General Verb Condition	1291	1920	67
Total	2960	3840	77

3.4.3 Linear mixed effects model.

The independent variables that I was interested in here are the effects of task (will eye-movement latencies be affected by the presence of a second task), and the effects of verb type (will the online prediction effect still be present when attention must also be allocated to a

[2] It should be noted that there are always more valid trials included from the specific verb condition. This highlights, in its own right, that the ability to comprehend online is greater after hearing specific verbs compared to general verbs.

secondary task). The dependent variable was the eye-movement latencies to the target under these different conditions.

Random effects model.

In this experiment there are more areas where variance may occur, this is due to the presence of more fixed factors. The additional factors here are i) experimental half (participants complete the visual task twice, one in the presence of a secondary task and once on its own, and participants could show different levels of variance across the two experimental halves), and ii) task (participants may show more or less variance in their eye-movements when a secondary task is present). In light of these additional factors there are extra parameters to test in random effects model. Not only could there be a random effect of participants and/or items on the intercept, participants could show different levels of variance across verb type, experimental half, task or a combination of the fixed effects.

The random effects model was ran as described in section 2.4 and the preferred structure included random effects of participants and items on the intercept (1 | PptNum) (1 | Item); a by-participant slope for verbtype and experimental half (1 | GeneralSpecific:ExpHalf: PptNum); and a by-item slope for experimental half (1 | ExpHalf: Item). (In explanation, 1 | GeneralSpecific:ExpHalf: PptNum fits a different term for each verb type x experimental half x participant condition but assumes that covariances are equal. A model that contained 0 + | GeneralSpecific:ExpHalf: PptNum would be doing the same yet estimating different covariances).

Fixed effects model.

After selecting the most appropriate random effect structure I then compared models with different fixed effect combinations. When analysing the full data set the model comparison preferred a simple no interaction model, and also excluded Task (χ^2= 12.601, p > .05) The two remaining factors were included in the model, indicating that Verb Type (χ^2= 81.746, p < .001) and Experimental half (χ^2= 31.094, p < .001) add significant weight to the model.

Table 5. The means and standard deviations of eye-movement latencies to the target across specific and general verb conditions, and the visual world only and dual tasks. [Experiment 2, Full data set].

<table>
<tr><th>Task</th><th colspan="2">Specific Verb</th><th colspan="2">General Verb</th><th colspan="2">Prediction Effect</th></tr>
<tr><td></td><td>Latency to Target (ms)</td><td>SD (ms)</td><td>Latency to Target (ms)</td><td>SD (ms)</td><td>Specific-General Verb Difference (ms)</td><td>Task Difference (ms)</td></tr>
<tr><td>Comprehension Only</td><td>962</td><td>232</td><td>1178</td><td>215</td><td>216</td><td rowspan="2">48</td></tr>
<tr><td>Dual Task</td><td>989</td><td>241</td><td>1157</td><td>255</td><td>168</td></tr>
</table>

Note: The final column shows the strength of prediction, which is the difference between specific and general verb latencies, and highlights how the strength in the prediction effect changes across the two tasks.

The inclusion of Experimental half in the model suggests that there was a significant difference in responses over each experimental half; completing a given task first or second has an effect on eye-movement latencies. The means demonstrate that this difference occurs due to faster eye-movement latencies in the second experimental half where participants have already completed the visual world task once before (either in the presence or absence of the secondary task), compared to the first experimental half when the visual world information is initially encountered (see Table 6).

Table 6. The means and standard deviations of eye-movement latencies to the target across specific and general verb conditions, the visual world only and dual tasks, and the first and second experimental halves. [Experiment 2, Full data set].

Task	**Experimental Half** *($1^{st}/2^{nd}$)*	**Specific Verb Condition** *Mean latency to target*	*SD*	**General Verb Condition** *Mean latency to target*	*SD*
Comprehension Only	1	962	232	1178	215
	2	840	276	1004	315
Dual-task	1	989	241	1157	255
	2	921	252	1106	240

I believe that in the second experimental half participants could remember the visual world displays and use this knowledge to quickly locate the correct referent. These carry over effects indicate that the differences in eye-movements across conditions are not only down to prediction but memory effects. Therefore, to ensure the effects relate to prediction I now modelled the data from the first experimental half only.

Random effects model.

The best random effects structure included participants and items on the intercept (1|PptNum) (1|Item); a by-participant slope for verb type (1|GeneralSpecific:PptNum); and a by-item slope for task and verb type (1|Task:GeneralSpecific:Item).

Fixed effects model.

The preferred fixed effects model included verb type (χ^2= 47.003, p < .001), however the model revealed that there was no significant effect of task (0.111, p = .74). (The script for running this linear mixed effects model can be found in Appendix 2B).

A difference for general and specific verbs is maintained in a dual-task situation even when attention must also be allocated to a secondary digit task. In fact, when examining the first experimental half only (where there are no confounding carry-over effects) we notice that there is no effect of task; eye-movement latencies are equivalent whether or not a secondary digit task is present, with similar differences between specific and general verbs across both tasks.

For comparison purposes, I also explored whether our analysis of first looks to the target mirrors an assessment of cumulative probabilities over time (a popular method with this type of data). The equivalent evaluation can be seen in Figure 4 where the probability of fixating the target across time is illustrated across conditions. The results mirror those of my LMEM; the equal distance between each pair of grey and black lines demonstrates that latencies to specific and general verbs are similar across the two tasks, whilst the close proximity of the straight and dotted lines illustrates how latencies are almost equivalent in the comprehension (or visual world) only task and the dual-task.

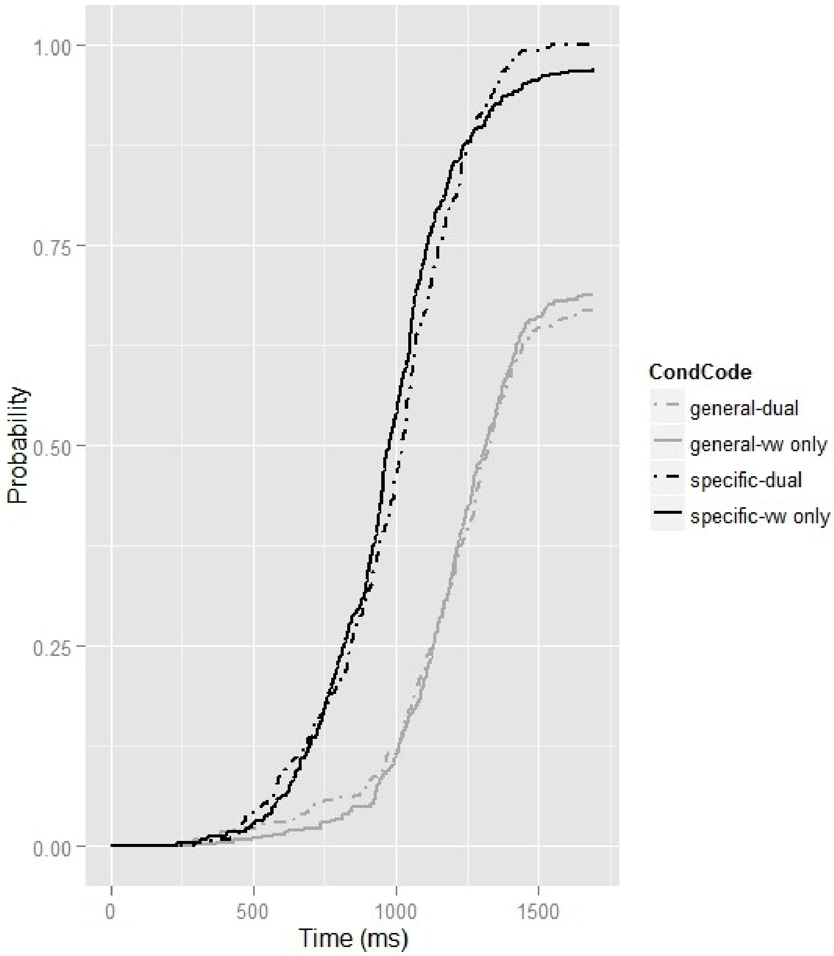

Figure 5. The probability of fixating the target over time across the two types of verb condition (general, specific) and two tasks (comprehension (or visual world only), dual-task).

Table 7: The average latencies to targets across the Specific and General conditions, relative to the verb and noun onsets. [Experiment 2, First Experimental Half only].

	Specific Condition (Eat)	General Condition (Move)	Difference between conditions (eat-move)
Verb Onset	962	1178	216
Noun Onset	247	462	215

3.5 Discussion.

The presence of a working memory task does not appear to prevent the ability to predict the upcoming theme of a sentence; in the dual-task condition there is still a significant difference between latencies after hearing specific verbs and those after hearing general verbs. I also suggest that fixations to targets within the specific condition here are initiated prior to processing the noun. Although the latencies documented in the bottom row of Table 7 show that the mean latency is 247ms after noun onset, previous research suggests that it around 200ms to initiate an eye-movement (Matin et al. , 1993) and approximately 200ms to recognise a word in context (Marslen-Wilson, 1984). This would suggest that any fixations made within 400ms of noun onset are in fact initiated prior to the when the noun information can be used. Therefore I argue here that fixations made during the specific condition are (on average) initiated prior to the noun information being of use to participants. On the other hand, eye-movements in the general condition rely on the noun being heard; only at this point can the correct referent to be distinguished. This is evident in latencies over 400ms in the general condition. This corroborates the claim that *verb* information is mediating eye-movements and suggests that we have successfully eliminated any doubts about this that were raised in Experiment 1.

We also find that the presence of the secondary task does *not* impact upon the strength of the prediction effect: the ability to predict is not diminished at all in the presence of the digit-matching task. One possible assertion could be that predicting at the verb is not a resource-limited process but proceeds automatically. Indeed, adults can comprehend their native language with apparent ease and accuracy, when listening to speech it is impossible to prevent comprehension. Similar arguments for automaticity have targeted processes in dialogue, for example the aforementioned mechanism of interactive alignment proposed by Pickering and Garrod (2004; Garrod & Pickering, 2004). Furthermore, dual-task studies have offered evidence for certain aspects of comprehension, such as semantic processing, as automatic (Fischer, Miller, & Schubert, 2007). However, an abundance of individual difference research has indicated a strong link between working memory capacity and language proficiency. For example, Daneman and Carpenter (1980) developed the reading span as a measure of individual differences. Participants read or hear a series of sentences and then must later recall the final word of each sentence. The resultant reading span score has been correlated with several measures of language comprehension in both children and adults, with higher reading span scores predicting greater proficiency in comprehension. Therefore, this would contradict the notion that (predictive) comprehension is independent from working memory and takes place automatically.

Relatedly, dual-task studies looking at connected speech, rather than single words, suggest that comprehension disrupts secondary tasks, such as driving, lexical decisions, and probe detection (Kubose, Bock, Dell, Garnsey, Kramer & Mayhugh, 2005; Shapiro, Zurif, & Grimshaw, 1987; Nicol, Love, & Swinney, & Hald, 2006). Kubose et al. (2006) examined whether production and comprehension place different demands on attention, or working memory. They used a dual-task that involved participants hearing or producing connected speech whilst in a driving simulator. Both speech tasks were found to have a detrimental effect on driving skill, with comprehension resulting in just as much degradation as production.

Previous research has also speculated that there is a link between prediction and working memory. The aforementioned study by Gibson and Thomas (1999) demonstrates that when a sentence contains many referents, verb/noun-phrases and sentence heads as is the case in "The apartment that the maid who the service had sent over was cleaning every week was well decorated", individuals can no longer make predictions about all of the verb phrases present as the demand on memory is too high. This research therefore supports the notion that predictive comprehension *is* dependent on working memory capacity.

Moreover, researchers have previously looked into prediction in written comprehension (i.e., reading), and examined whether prediction in text is affected by working memory capacity. For example, the aforementioned studies by Estevez and Calvo (2000) and Linderholm (2002) demonstrated that working-memory capacity affected the ability and time course of predicting. Therefore this research further indicates that predicting upcoming themes is reliant on working memory capacity.

The above claims therefore indicate that rather than prediction processes being automatic perhaps they don't tap into the same resource pool as the digit-matching task. If you recall, in the introduction to this chapter I explained how working memory capacity (and therefore performance in the comprehension and digit-matching task) will only be affected if the two tasks draw upon the same pool of resources. If both tasks require access to the same resources then the available working memory capacity must be shared between the two tasks. Here, I explore whether this is the case. Digit tasks are presumed to engage phonological working memory; is it possible that our sentence comprehension task employed here does not require a great amount of phonological working memory capacity?

Evidence to support this claim comes from neuropsychological studies. In several cases, brain damaged patients with severely reduced phonological memory exhibit largely preserved comprehension skills. For example, patient IL, who acquired a phonological loop deficit and had poor short-term memory skills, could not repeat sentences verbatim yet could correctly paraphrase the meaning (Saffran & Marin, 1975). Therefore, the phonological loop is clearly not necessary for all types of language comprehension.

Moreover, models of language comprehension agree that phonological working memory is not employed in the processing of single words, clauses, and sentences containing simple syntactic and semantic structures (like those presented in this study) (Gathercole & Baddeley, 1993). Such processing is assumed to proceed in real time without delay (i.e., online), and does not necessitate reference to working memory. In contrast, sentences that are syntactically or semantically complex *do* require the use of phonological working memory. Here, analysis of the complex syntax and semantics is thought to occur “offline”, as indicated by evident lags in time between when the sensory input is encountered and when its interpretation becomes complete. (Please note, subsequent uses of ‘offline’ processing simply refer to situations where there is a time lag in incremental processing). Such complex forms are thought to include passive sentences, embedded clauses, sentences containing numerous content words and sentences in which word order is vital for understanding (Gathercole & Baddeley, 1993).

It is argued that in order to interpret the entire structure and meaning of these sentences the phonological representation of each item must be held in working memory, and consequently requires greater processing capacity (Baddeley et al., 1987; Martin, 1990). Studies disrupting the use of the phonological loop in comprehension have provided evidence for this distinction between syntactically simple and complex sentences. Articulatory suppression techniques, where subjects must produce irrelevant speech, impair the rehearsal process of the phonological loop and lead to the decay of stored items. Using this technique, research has shown that suppressing articulation does not affect the verification of simple active sentences, such as “Canaries have wings”, however does affect the verification of more complex sentence structures such as “She doesn’t mind going to the dentist to have fillings, but doesn’t like the pain when he gives her the injection in the beginning” (Baddeley, 1987; Baddeley, Eldridge & Lewis, 1981; Gathercole & Baddeley, 1993).

Additionally, the aforementioned patient, IL, demonstrates a pattern of performance that supports the simple-complex sentence distinction. IL could correctly paraphrase simple sentences yet often incorrectly interpreted more syntactically complex structures including passives and embedded clauses. Another patient showing a similar response pattern is EA, whose accuracy of sentence interpretation declined as the complexity of syntactic analysis increased (Friedrich, Martin, & Kemper, 1985). And finally, patient PV, reported by Vallar & Baddeley (1984, 1987), whose impairment of phonological working memory resulted in problems with comprehension for syntactically complex structures, but was preserved for simpler structures.

Such evidence suggests that the active sentences implemented here were too simple to necessitate a great amount of phonological working memory. Processing of sentences such as “the boy will eat the cake” are simple active forms that can be dealt with online. Consequently, if I were to reproduce these experiments using complex sentence structures we would presumably find that prediction processes impose greater demand on the phonological loop and should therefore interfere with, or receive interference from, the storage and processing of the digit task.

One more important distinction to be made here is that between phonological and semantic working memory. Recently, research has indicated the presence of separate capacities for phonological and semantic information (Hanten & Martin, 2000; Martin, 2005; Martin & He, 2004; Martin, Shelton, & Yaffe, 1994). This proposal is based upon the dissociation between retention of semantic and phonological material in working memory. For example, Martin, et al. (1994) reported patient A.B, who had a very restricted working memory span and as a result impaired retention of semantic material. However, patient A.B showed preserved phonological memory, and could perform well on measures of phonological retention, such as digit span and digit matching tasks. In sharp contrast however, another patient, E.A, demonstrated the opposite pattern; preserved semantic retention in the presence of a phonological memory deficit. When these two patients were studied in relation to language proficiency, results revealed that patient A.B's semantic retention deficit manifested itself in poor sentence *comprehension*. Whereas patient E.A's phonological retention deficit manifested itself in poor sentence *repetition,* with comprehension skills preserved. These findings imply separate capacities for semantic and phonological information and, moreover, suggest that these two capacities play different roles in language. A notion supported by the fact that semantic and phonological short-term memory are located in separate brain regions (Martin, 2005). (See also, Yudes, Macizo, & Bajo, 2012, who demonstrate that phonological working memory does not underlie the ability to coordinate comprehension and production processes).

Given such evidence, one could propose that the ability to comprehend and predict sentences relies not on the phonological loop but upon semantic working memory capacity, with comprehension (and thus most likely prediction) significantly impaired with a reduced semantic memory capacity. Therefore, if the working memory task implemented in this study had been semantic in nature we may have seen that occupying semantic capacity would result in poor performance in prediction, the working memory task, or both. I return to this issue in Experiment 4 (Chapter 5).

To conclude here, I believe that either the two tasks were too easy to interefere with one another, or the premise that both tasks in this study tap into the same pool of resources is missing. It appears as though both tasks can be processed with little interference because they do not require working memory capacity to be shared across the two tasks. Instead, the digit-matching task appears to have had uninterrupted access to phonological working memory, and the comprehension task uninterrupted access to semantic working memory. This would explain why there is no significant effect of task found in this experiment. In future studies where coordination is examined it is important to choose two tasks that require access to the same working memory capacity.

CHAPTER 4

EFFECTS OF SYNTACTIC COMPLEXITY ON ONLINE PREDICTION PROCESSES

4 Experiment 3

4.1 Goals and Motivation.

The main aim of this study is to extend the stimuli set and create more variability in the visual world materials. Until now the auditory sentences presented to participants contain an identical structure: “The [agent] will [verb] the [noun]”. This repetition gives a strong cue as to when the crucial target word is about to appear. After only a small number of trials participants will be aware that the same structure is repeated and could choose to focus in on the target region only.

In the subsequent experiments it is extremely important that comprehension and planning overlap in order to mimic a normal conversation pattern. Therefore, I need to ensure that participants are listening to the whole sentence rather than minimising their comprehension; to do this the repetitive and predictive syntax of the sentences must be changed to include a variety of different structures, where the verb and noun information could be heard at different time-points. In this study I examine online prediction at the verb using a variety of sentence structures, including both simple and complex sentences.

During spoken sentence comprehension, listeners must rapidly map acoustic information onto lexical representations, parse the sentence into correct syntactic structures, and decipher the meaning of both individual words and the construct as a whole. This task inevitably becomes harder when the sentences we must decode become more complex. As mentioned in Experiment 2’s discussion certain linguistic analyses can take place online as the sentence unfolds, whilst others require more processing time before analyses are complete and occur “offline”, with semantic and syntactic analyses lagging behind sensory input. The time taken to process linguistic information has been found to vary with the complexity of the sentence, with simple syntactic structures being dealt with online and complex structures being dealt with offline (Baddeley et al., 1978; Baddeley et al., 1987; Baddeley et al., 1981; Friedrich et al., 1985; Gathercole & Baddeley, 1993; Martin, 1990). The online nature of simple sentence structures is thought to be possible as they do not need access to phonological working memory. In Experiment 2, I implemented simple syntactic structures whilst attempting to examine the effects of competition for working memory resources, yet the above argument suggests that these linguistic analyses do not rely greatly on working memory capacity.

In this experiment I examine whether prediction in comprehension can occur online when presented within a more complex structure. Possible outcomes here include: 1) as syntactic complexity increases so do the latencies to the target, 2) the syntactic complexity affects online prediction with higher levels of processing affecting the resources available for prediction, 3) prediction at the verb can still occur online and at an equivalent speed across simple and complex sentences, as the verb information can be dealt with incrementally and independent of syntactic complexity.

Simple versus complex sentence processing.

Research has shown that different sentence structures impose different cognitive demands. For example, using an auditory moving windows task (self-paced listening) Fallon, Peele and Wingfield (2006) noted that pause durations increased in line with syntactic complexity. Participants were presented with simpler subject-relative centre-embedded clause sentences (“The author that insulted the critic hired a lawyer”) and more complex object-relative centre-embedded clause sentences ("The author that the critic insulted hired a lawyer"). Listening times were significantly delayed in the more complex sentence structures relative to the more simple structures.

Similarly, Tun, Benichov and Wingfield (2010) examined the effects of syntactic complexity alongside different perceptual inputs. Younger and older adult participants listened to sentences that varied in syntactic complexity and sound level, and were asked to make true/false judgements to comprehension questions. Tun et al. found significantly slower responses to syntactically complex sentences relative to simpler sentence structures. Furthermore, their comprehension accuracy, although near ceiling, was higher for syntactically simpler sentences (0.93 correct) than more complex forms (0.84 correct). An interesting finding here is the interaction between perceptual input and complexity of sentences; Tun et al. found that older adults had problems with sentences presented at low sound levels, but only when the sentences were syntactically complex. They still performed well with simpler sentences produced at low sound levels. This further reiterates the extra processing demands that are present in comprehending complex sentence structures.

Other researchers have manipulated both syntactic complexity and memory load. Gordon, Hendrick and Levine (2002) used an online self-paced reading technique to compare reading times in subject-cleft sentences (simple condition) and object-cleft sentences (complex condition). They manipulated memory load via similarity between word lists and the sentence content; the memory load was either matched or unmatched to the sentence content. The matched condition comprises a greater memory load due to interference between items.

Example.

Condition	Memory Load	Sentence
Matched, Subject Cleft	Joel-Greg-Andy	It was Tony that liked Joey before the argument began.
Matched, Object Cleft	Joel-Greg-Andy	It was Tony that Joey liked before the argument began.
Unmatched, Subject Cleft	Joel-Greg-Andy	It was the dancer that liked the fireman before the argument began.
Unmatched, Object Cleft	Joel-Greg-Andy	It was the dancer that the fireman liked before the argument began.

The results of this experiment showed that there were longer reading times and more errors made in complex object-cleft sentences compared to the simpler subject-cleft sentences. Analysis of recall revealed that the number of items recalled was affected by both syntactic complexity and match, with recall higher for unmatched items, and for subject-cleft sentences (Matched, Subject-Cleft, 92.5%; Matched, Object-Cleft, 90.8%; Unmatched, Subject-Cleft, 93.8%; Unmatched, Object-Cleft, 92.4%). These findings reiterate that sentences differing in syntactic complexity are subject to different capacity demands. More importantly, these results also suggest that a secondary memory load can affect comprehension.

However, other research suggests that task demands may account for some of the differences accredited to syntactic complexity. For example, Love, Nicol and Swinney (2006) looked at whether the differing activation levels in Broca's region are in fact down to different comprehension demands posed by simple and complex sentence structures, or whether this is due to differences in task demands. Through testing simple and complex sentences under three different task demands (passive listening, probe verification, and theme judgements) Love et al. demonstrated that Broca's region was in fact recruited to the same degree in simple and complex sentences, however, recruited to differing degrees depending on the task demand condition. They propose that it is what task that is asked of participants that influences the activation of Broca's area rather than syntactic complexity. This highlights the importance of choosing appropriate tasks, such as Altmann & Kamide's, that strike a balance between passive listening and tasks that require manipulation and consideration of the linguistic material.

Although syntactic complexity has been found to affect processing times, this does not necessarily mean it will impact upon prediction. Other research has found that whilst interpretation of the overall meaning and syntactic structure may be delayed, the processing of individual constituents (e.g., a noun or verb phrase or a clause) can occur online. That is, there is a distinction between local and global processing. Marslen-Wilson and Tyler (1980) describe how local processing refers to the processing that occurs within constituents. For example, comprehension of a word, phrase or clause involves local processing. Global processing, on the other hand, refers to processing across-constituents and interpreting the overall meaning and structure of a sentence. In their study, Marlsen-Wilson and Tyler examined the time-course of local processing via word-recognition processes and global processing via the presence and absence of context. They demonstrated that words within a constituent can be processed online independent of their position in the sentence.

To summarise here, research suggests that syntactic complexity does affect online measures of comprehension, such as listening times (Fallon et al., 2006; Gordon, et al., 2002; Tun et al., 2010). More complex sentence structures appear to be harder to process, taking longer to access and requiring greater resource allocation. Language comprehension is also affected by concurrent memory load, as demonstrated via manipulations in semantic similarity in Gordon et al.'s 2002 study. These findings also suggest that including different levels of syntactic complexity is a good way of manipulating processing demands.

More recent online techniques such as the visual world paradigm have examined the incrementality of language comprehension, yet neglected to use these research techniques to examine whether such online processes may be affected by different processing demands. In this study, I examined whether the ability to predict online at the verb (as demonstrated in Altmann & Kamide, 1999 and experiments 1 and 2) is preserved or diminished by increasing the processing demands required. Processing demands were manipulated here via the level of syntactic complexity; the vital verb information was presented either in a simple or complex syntactic structure. I aimed to observe whether the ability to use specific verb information ('eat') is maintained in complex syntactic structures, or whether this ability was reduced relative to its simple sentence counterpart.

4.2 Hypotheses.

If the complex structures are sufficiently difficult to require some offline processing, then we could see a general slowing of processing in complex sentence structures compared to its simple sentence counterparts.

However, research has also highlighted a distinction between local and global processing; which refers to the difference between parsing words inside the same constituent, and the processing involved in making a full syntactic interpretation of the whole sentence. That is, words within a phrase can be put together online without having a whole sentence interpretation ready. This was certainly the case in Experiment 2 where the verb mediated fixations to the target prior to the noun being heard. Therefore, it is likely that our vital verb information will be processed online within its individual constituent (the verb phrase) prior to full sentence interpretation; this would mean that the complex sentence structure will have little effect on anticipatory eye-movement latencies. If this is the case, the strong effect of verb type should still be present, with an advantage for specific verbs, and this difference between specific and general verbs should be comparable across simple and complex sentences.

However, in the very unlikely case that prediction depends on a full sentence interpretation, eye-movements to the target will be delayed until after interpretive processes are complete, resulting in slowed latencies and potentially a reduction or loss of the verb effect. This would manifest itself in an interaction between sentence complexity and verb type.

4.3 Method.

Participants.

28 students from the University of Birmingham (25 females, ages 18-23, mean age 19 years, 6 months) took part in the study in return for course credits. All participants were English Native Speakers and had normal to corrected vision and hearing.

Materials and apparatus.

56 sets of stimuli were used in this experiment. Each set is comprised of one visual scene and *four* associated auditory sentences. The visual scenes are still made up of an agent's face (a man, woman, boy or girl) and four object pictures (one target picture and three distractors). The

target pictures are also repeated as distractors in other visual scenes. The visual scene was presented on a 15'' computer monitor at a resolution of 1024x768pixels. The picture stimuli were black and white line drawings, 178x142 pixels in size. The pictures were taken from Snodgrass and Vanderwart (1980); Szekely et al. (2004); Clip Art; created using Paint; or hand-drawn.

The four types of sentences presented alongside a visual scene are; i) Simple sentences containing a specific verb, e.g., 'The boy will eat the cake'; ii) Simple sentences containing a general verb, 'The boy will move the cake'; iii) Complex sentences containing a specific verb, e.g., 'The boy who is the captain of the football team at school will eat the cake'; and finally, iv) Complex sentences containing a general verb, 'The boy who is the captain of the football team at school will move the cake'. There are 224 sentences in total; 112 Complex structures (56 of which contain a specific verb, 56 general verbs), and 112 Simple sentence structures (56 containing specific verbs, 56 general verbs). The vital region of each sentence, between the verb and noun, was identical across each simple/complex pair, for example "The boy will *eat the cake"*, and "The boy who is the captain of the football team at school will *eat the cake*". The frequencies of specific and general verbs are matched across items, and the distance between the verb and noun are matched across all four sentence types for each item. A full list of the verb-noun distances in Experiments 3-6 can be found in Appendix 1H.

The complex sentences are further subdivided into different types of sentence structure. Half of the complex sentences contain an embedded clause, such as in the aforementioned, "The boy *who is the captain of the football team at school* will eat/move the cake". (Please note, 'embedded' sentences here refer to sentences that contain a relative subject clause, as seen above. See Appendix 1G for more examples of these 'embedded' structures). The other half are other complex structures; this category is comprised of a variety of complex forms including sentences containing complex conjunctions (such as whilst, because, that, e.g., “The woman remembers she needs to go shopping whilst she is playing/dusting the piano”) and sentences containing many content words (e.g., “The girl in the pink stripy dress and shiny black stilettos will call/dodge the policeman”). Using a variety of sentences allows us to examine whether the online interpretation of verbs is mediated by sentence type, for instance, whether online verb processing is preserved in sentences containing many content words but not in embedded clauses.

The verb appears in two different forms; in the future tense, will eat/will move, and present tense, is playing/is dusting. This is designed to add more variety to our sentence structures and avoid the auxiliary “will” from being used as an indication that the verb is coming up. As a result participants have to listen carefully to the entire sentence.

To add further variety to the stimuli and to examine whether processing is affected by the verb’s location, the verb may now appear either at the beginning or the end of the sentence. In half of the sentences containing an embedded clause, the verb is presented inside the embedded clause; “The boy *who is riding/stroking the horse* has to do his maths, geography, and chemistry homework tonight”. However, in the other half of this sentence type the verb appears outside of the embedded clause; “The boy who is captain of the football team at school *will eat/move the cake”*. These two verb positions are referred to as “early” (occurring in the first half of the sentence) or “late” (occurring in the second half). The “other” complex sentences also have the two different verb positions, for example, “The man is sailing/watching the boat because it is a lovely clear day outside” (early verb position), and, “The woman remembers she needs to go shopping whilst she is dusting/playing the piano” (late verb position). The variety of sentences can be seen more clearly in the diagram in Figure 2, and a full list of the sentences used in Experiments 3-6 can be found in Appendix 1G.

Participants' eye-movements in response to these sentences were recorded using a desktop-mounted EyeLink 1000 eye-tracking system. Recordings were taken from the right eye and measured at 1000Hz.

Design.

This experiment used a within subjects design. Items were presented between-subjects, whereby each visual scene was seen only once by each participant and for each item (e.g., eat or move) participants heard only one of the four possible associated sentences. However, every participant heard a selection of all sentence variations. For example, each participant heard 14 embedded clauses with an early verb position (7 of which contained a specific verb, the other 7 a general verb); 14 embedded clauses with a late verb position (7 specific, 7 general); 14 other sentences with an early verb position (7 specific, 7 general); and 14 other sentences with a late verb position (7 specific, 7 general).

The independent variables (IVs) here were; a) Complexity (whether the sentence is simple or complex in structure), b) Verb Type (whether the sentence contains a specific or general verb), c) Sentence Type (whether the sentence has an embedded clauses or is an other complex sentence), and d) Verb Position (whether the verb is positioned early or late on in the sentence). The dependent variable (DV) was the latency of eye-movements to the target object.

The location of the target was counterbalanced across trials, appearing in each corner of the visual scene an equal number of times. Each participant heard the same number of specific and general verbs, the same number of embedded clauses and other sentences, the same number of future tense (will) and present tense (-ing) verb forms, and the same amount of early and late verb positions.

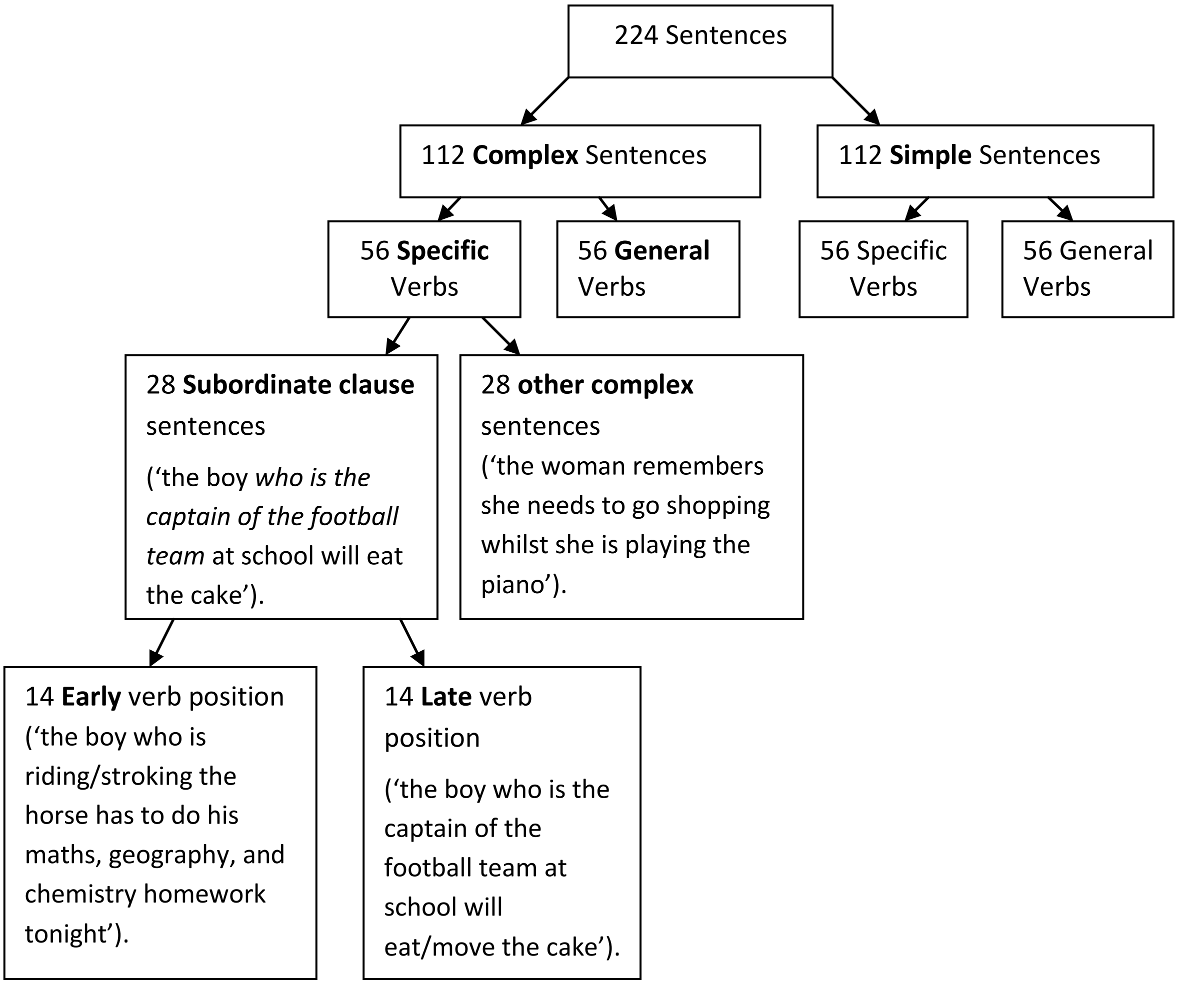

Figure 6. Diagram showing the divisions and subdivisions of stimuli into the different structures (complex or simple), verb types (specific or general), sentence types (embedded or other), and verb positions (early or late).

Procedure.

Alike to Experiments 1 and 2, participants sit approximately 57cm away from the display with their head placed on the chin rest. Their eye-movements are calibrated at the start of experiment (this process is described in Experiment 1's procedure section).

Participants were told that the experiment looked at eye-movements as they viewed a visual scene. On each trial, the object pictures appeared on screen for 3000ms, participants were asked to use this time to look at all four items. They were asked to then look at the agents' face in the centre of the screen. The onset of the sentence was gaze-dependent upon this central image. The experiment took approximately 30 minutes to complete.

4.4 Results.

4.4.1 Eye-movement data.

Eye-movement variables.

All eye-movement variables included in Experiments 1 and 2 are similarly used here. Additionally, two new variables are created using information in the stimuli set, these are verb position (early, late) and sentence type (embedded, other).

Exclusion criteria.

The exclusion criteria used here are described in Experiment 1. 434 trials where participants did not fixate on the target were removed from analysis (296 trials were removed from the general verb condition, 138 from the specific condition). 21 trials were excluded due to fixating before the 200ms time-point (12 trials were excluded from in the general condition, and 9 from the specific condition). Any eye-movements that were over 2.5 standard deviations above the participant mean were also removed from analysis; 22 trials were removed in total (13 from the general condition, 9 from the specific condition).

In total, participants fixated on the target in 72% of the trials. They fixated the target on 67% of trials in the general verb condition, and on 87% of the trials in the specific verb condition.

Table 8. The number and percentage of valid fixations in the specific and general verb conditions. [Experiment 3].

	Number of Valid Fixations	Total Number of Trials	% of Trials with Valid Fixations
Specific Verb Condition	646	784	82
General Verb Condition	488	784	62
Total	1134	1568	72

4.4.2 Linear mixed effects model.

The independent variables that I was interested in here are the effects of verb type (will the online prediction effect still be present within this varied set of stimuli), and complexity (will eye-movement latencies be slower when participants are presented with a complex sentence). I also want to examine whether there is a significant interaction between verb type and complexity; a significant interaction would suggest that prediction processes are dealt with differently depending on whether the verb is presented within a simple or a complex sentence. The dependent variable was the eye-movement latencies to the target under these different conditions.

When considering the complex sentences only (or the ‘complex data set’) there are two additional factors that must be analysed; verb position (whether eye-movements are affected by the verb appearing early or late in the sentence) and sentence type (whether eye-movements are affected by the verb appearing in an embedded clause or other sentence structure). When analysing the complex data I am looking at whether there is a three-way interaction between verb type, verb position and sentence type; whether there are any two-way interactions; and whether there is a main effect of verb type, verb position or sentence type.

This linear mixed model section below is separated into analyses of the "full data set" and "complex data set". Each analysis follows the pre-defined format of first testing the random effects models followed by fitting the fixed effects model.

Full data-set.

Random effects model.

The preferred random effects structure included participants as random effects on the intercept (1| PptNum), and verb type and complexity as random effects on the by-item slope (0+ GeneralSpecific:SimpComplex|Item).

Fixed effects model.

I then compared models with different fixed effect combinations. The model comparison preferred a simple no interaction model, with the inclusion of the two fixed effects (Verb Type: χ^2= 53.679, p < .001; and Complexity: χ^2= 11.159, p < .001) but no interaction (Verb Type*Complexity: χ^2= .002, p = .96). From the means table below you can see that latencies in the specific condition are significantly faster than in the general condition, and, similarly, latencies in the simple condition are significantly faster than in the complex condition.

Table 9. The means and standard deviations of eye-movement latencies to the target across verb type (specific, general) and complexity (simple, complex). [Experiment 3, full data set].

Complexity	Verb Type			
	Specific Verb Latency (ms)	SD (ms)	General Verb Latency (ms)	SD (ms)
Simple	867	245	1086	224
Complex	961	303	1173	318

Table 10: Mean eye-movement latencies to targets (in ms) across the Specific, General and Simple, Complex conditions, relative to the verb and noun onsets. [Experiment 3, full data set].

	Specific Simple (ms)	**General Simple (ms)**	***Difference between conditions***	**Specific Complex (ms)**	**General Complex (ms)**	***Difference between conditions***
Verb onset	867	1086	219	961	1173	212
Noun onset	148	364	216	321	533	212

Note: The specific latencies are under 400ms which means that these eye-movements are mediated by the verb information not the noun (see Chapter 3 discussion).

Complex data set.

The data was further analysed to examine the effects of verb position (early, late) and sentence type (embedded, other), this analysis therefore includes eye-movements when listening to the complex structures only. The factors analysed in this model included verb type, verb position, and sentence type.

Random effects model.

The best random effects structure modelled participants and items as random effects on the intercept (1| PptNum) (1| Item), and a by-item slope for verb type (1| Item: GeneralSpecific).

Fixed effects model.

Again, I used a backward selection method to select the necessary parameters. The model comparison preferred a simple no interaction model, with no significant three-way or two-way interactions. The preferred model further excluded sentence type (χ^2= .697, p =.40), however found a significant main effect of both verb type (χ^2= 32.473, p < .001) and verb position (χ^2= 34.439, p < .001). The means reveal significantly faster latencies after hearing specific compared to general verbs, and significantly faster latencies when the verbs are presented late on in the sentence (see means table below). The full linear mixed effect model can be found in Appendix 2C.

Table 11. The mean and standard deviations (in ms) for eye-movement latencies to early and late verb positions across the specific and general conditions. [Experiment 3, complex data set].

	Early Verb Position		Late Verb Position	
	Mean Latency (ms)	SD (ms)	Mean Latency (ms)	SD (ms)
Specific Condition	973	303	858	244
General Condition	1203	294	1038	237

4.5 Discussion.

I have demonstrated that verb information is processed online as evident in faster eye-movement latencies to the target following specific verbs. In support of previous experiments reported earlier in this thesis, I have shown that verb information can be processed and used online to predict a future referent. Importantly, I have shown that syntactic complexity affects processing, with more complex sentence structures requiring more processing time, an indicator of being subject to great processing demands. This finding is supported by previous work that has also highlighted that processing demands vary with syntactic complexity (Fallon et al., 2006; Gordon et al., 2002; Tun et al., 2010). For example, the aforementioned study by Gordon et al. (2002) demonstrated longer reading times and more errors made within the more complex object-cleft sentences compared to the simpler subject-cleft structures (see also Waters, Caplan & Yampolsky, 2003; Waters & Caplan, 2001; Fallon et al., 2006; and Tun et al., 2010 for similar findings).

The findings here suggest that some types of sentence may be dealt with in a similar way, as illustrated in similar latencies across embedded clause sentences and other structures. It appears as though these two types of sentence structures that I have chosen to include in our complex syntactic group require similar processing effort.

Interestingly, although I note that complex sentence structures resulted in increased processing time I also demonstrate that the effect of verb type (i.e., the difference between specific and general verbs) remains unaffected by the increased processing demands. That is, the effect of verb type is equivalent across simple and complex sentences. Therefore, this suggests that although the overall processing times increased with task difficulty, the ability to anticipate language remains intact. The fact that the strength of prediction remains unaffected by increased syntactic complexity would suggest that participants did not have to wait until interpretation processes were complete before predicting. Instead, this preserved prediction effect indicates that the verb information presented here was processed locally and local processing was not affected by any processing overhead introduced by more complex sentences. In explanation, the vital verb information was presented within a constituent, and participants were able to process this constituent incrementally.

Previous research has highlighted a distinction between local and global processing. For example, Baum (1989) demonstrated that in cases of Broca's aphasia (a language production problem that also results in difficulties with syntactic aspects of language comprehension) grammatical violations within clauses (requiring local processing) could be identified online however across-clause violations (requiring global processing) were not detected. These results were later replicated in Haarman and Kolk (1994) and thus support the idea that local syntactic processing is less demanding than processing of global syntax. This research suggests that if verb information were to be presented across constituents then we may have seen a disruption to the verb effect.

Such evidence may also help to explain why we observed similar processing across the different sentence types. Here I have demonstrated that our different types of sentence structure are subject to equivalent processing demands, as evident within the similar processing times across embedded and other structures. Conversely, previous research has demonstrated that different sentence structures are subject to different processing demands; for instance, the aforementioned distinction between subject-cleft and object-cleft sentences. Obviously, the sentences here cannot be directly compared to those used in other studies as I didn't employ the contrast of subject-cleft and object-cleft structures. Yet, it appears that the difference in results here can be explained in terms of local versus global processing. The level of global processing is what differs between object-cleft and subject-cleft pairs. Object-cleft sentences are designed to be more complex due to information being split across more constituents; these sentences must be dealt with globally in order to successfully interpret the meaning. In our study, it is important that the vital verb information is presented in the same way across sentences in order to allow comparisons of prediction. Therefore, in both embedded and other structures although the surrounding material differed the vital information was identical. This, therefore, allowed participants to process information locally and in a similar manner across the various complex sentence structures.

Another interesting finding is that participants were faster to process and use verbs when they appeared towards the end of the sentence. This was true for both embedded clause and other sentences. In regards to our embedded clause sentences, this increased processing speed at the end of utterances fits nicely with previous research which has demonstrated that parsing information *inside* of an embedded clause is more difficult than parsing information surrounding the clause (e.g., Miller, 1974). In regards to other sentence structures, however, the reasoning for increased processing speed towards the end of sentences is less clear. Other research, for example the aforementioned study by Fallon et al. (2006), has conversely demonstrated online wrap up effects with comprehension processes *slowing down* towards the end of a sentence. In sharp contrast, this study shows an increase, rather than delay, in processing speed at the end of a sentence. Participants became faster at processing and utilising verb information as the sentence unfolded. From this one could postulate that decoding the sentence becomes easier with the more information available. Indeed, a similar claim was offered by Holmes & Forster (1970) in their early study of auditory detection. They proposed that when a sentence starts participants have very little linguistic material to work with, and the ways in which the sentence can unfold are endless. However, towards the end of the sentence they have already received the majority of linguistic material and as a result the decoding load decreases. Therefore, in our case, at the end of the sentence participants have been given the majority of information and thus the possibilities of the upcoming material are more greatly restricted. In fact, the nature of our experiment allows the possible future material to be further restricted as participants are aware that the information must include one of the four items displayed in the visual scene in front of them. The visual world acts as a contextual environment which participants can use to help influence lower level (word) processing.

In conclusion, I have successfully demonstrated the online interpretation and utilisation of verb information in complex sentences. I have shown that online prediction is resilient to the effects of syntactic complexity, with the strength of the prediction effect being equivalent across simple and complex sentences. This suggests that anticipation in language remains intact at least under some processing demands.

CHAPTER 5

COORDINATING SPEECH PRODUCTION AND SPEECH COMPREHENSION

5 Experiment 4

5.1 Goals and Motivation.

Experiment 2, which forced coordination between speech comprehension and a secondary phonological task, revealed no reduction in the ability to predict online at the verb in a dual-task condition. Similarly, Experiment 3 revealed how online prediction is maintained in complex syntactic structures, and that increasing processing demands through syntactic complexity does not affect prediction processes. Experiment 4 further examined how speech comprehension is affected by various levels of processing demands. I reverted back to examining comprehension within a dual-task setting (like Experiment 2). However, I now began to investigate my overall research question: the coordination of speech production and speech comprehension. This was achieved through implementing the visual world comprehension task alongside a concurrent speech production task.

The goals here were largely two-fold: 1) to examine whether, and how, speech production and speech comprehension can be coordinated efficiently in an experimental setting, and 2) to examine the prediction processes in the presence of an appropriate secondary task. In the discussion of Experiment 2, I considered the distinction between phonological and semantic working memory and the importance of choosing a task that tapped into the relevant memory stores. By examining speech comprehension alongside a concurrent speech production task we can be confident that both tasks employ language processors and that the two must compete for available resources. Indeed, evidence from dual-task studies have demonstrated that coordinating two language-based tasks creates higher levels of interference. For example, Hohlfeld et al. (2004) looked at whether semantic processing (as measured by amplitude of the N400 effect) can be performed alongside another concurrent task. The additional task was either also language-related (response to letters) or non-linguistic (a spatial task). Results indicated that the amplitude of the N400 effect was delayed when an additional task had to be performed and, furthermore, this delay was lengthened when the additional task was linguistic in nature. This finding, therefore, supports the idea that two language based tasks will strongly compete for the same cognitive resources.

Moreover, previous research has suggested that, when using an appropriate memory load, a concurrent memory task affects online language processing (Hohlfeld et al., 2004; Nicol et al., 2006; Mattys & Wiget, 2011). For example, Nicol et al. (2006) examined sentence comprehension alongside the presentation of externally presented visual word probes. They hypothesised that when an auditory sentence is being produced fluently, without interruption, the visual probe will not be integrated into the sentence. Results indicated that a visually presented probe will not be integrated into the fluent ongoing sentence, even if the probe is a better continuation of the sentence than the subsequent auditory word. This can be interpreted as evidence that sentence comprehension saturates capacity and therefore the visual probe cannot be simultaneously processed and integrated. Further evidence to support this idea is that when the same sentences were presented at slower rates the probe could now be integrated into the sentence. This is interpreted by Nicol et al. as evidence that when the processor is made available (through interruption or time waiting for the next word) there are now sufficient resources available for effectively processing the probe word.

Research into prediction and memory has highlighted that, at least in written language processing, memory capacity may impact online prediction (Estevez & Calvo, 2000; Linderholm, 2002), and could also affect how prediction processes are dealt with (Otten & Van Berkum, 2009).

Very little is known about how speech production and speech comprehension are coordinated in dialogue; is it possible to plan and listen simultaneously? Do we rapidly switch between tasks? Does this task switching affect how much speech is processed? Do we prioritise one task over the other? By implementing an experimental design which forces coordination between the two tasks I hoped to gain some insight into the coordination process. The design allowed me to look at the cost of doing both tasks at once, and through using a sensitive online measure (eye-tracking and the visual world) begin pinpointing where speech planning occurs relative to speech comprehension.

The secondary task I implemented was speech production. To recap, speech production involves transforming an intention into an utterance, commencing with the intention to speak and what to say; a conceptual task. This secondary task creates a more ecologically valid task-setting; we coordinate our speech production and comprehension daily and must store concepts (or ideas) of what we want to say in memory as we listen to others. The similarities between the dual-task and real dialogue are explained below in section 5.1.3.

In order to create a successful dual-task setting our secondary task, speech production, must be subject to resource availability. This is necessary in order to examine any postponements of the comprehension task and to pinpoint any interference between Task 1 (production) and Task 2 (comprehension). In the following section I examine the cognitive demands involved at each stage of speech production.

5.1.1 The Capacity Demands of Speech Production.

The processes involved in producing speech were briefly reviewed in Chapter 1, section 1.4, however, to recap here there are three broad stages of i) Conceptualisation, ii) Formulation, and iii) Articulation. Conceptualisation involves deciding how to best portray one's intention or ideas. After conceptualisation comes the formulation stage of speech production. Here, the concept produced during conceptualisation is transformed into a linguistic structure. The final stage of speech production is articulation where the utterance is translated into covert speech.

Previous research has examined the extent to which these speech production processes are subject to resource availability, for example, Levelt (1989). He describes automatic processes as those that do not require conscious awareness or intention, and suggests that they do not share computational resources with any other processes. Therefore, it is not surprising that conceptualisation, a process that is by definition an intentional activity, is labelled as a controlled process. Indeed, speakers must have an intention of what they wish to say, be aware of how they want to portray this intention and control the selection of relevant information. These tasks all require processing capacity. More specifically, constructing the preverbal message through retrieving and selecting relevant information (macroplanning) is thought to occupy most of the speaker's attention, with macroplanning often being subject to great amounts of memory search and planning (Levelt, 1989). Levelt also suggests that self-monitoring requires attentional resources, with speakers constantly keeping tabs on both their internal and overt speech. Conversely, it is suggested that all other stages take place automatically, with formulation and articulation not being subject to executive control.

Garrod and Pickering (in Meyer, Wheeldon & Krott, 2007) looked at Levelt's stages in relation to Bargh's "four horsemen of automaticity". A process is said to be automatic when it satisfies some or all four criteria, or "horsemen", of automaticity (Bargh, 1994). The first horseman is *awareness*; Bargh suggests that a process is automatic when you are not aware of it. The second horseman is *intentionality*; an automatic process is initiated involuntarily. The third horseman is *efficiency*, with automatic processes being faster and more efficient than one that demands greater processing. The final horseman of automaticity is controllability, or *interruptibility*; automatic processes cannot be easily modified or halted once they have been started.

Garrod and Pickering support the claim that conceptualisation is a capacity demanding process, and suggest that it is controlled with respect to all four horsemen; people are obviously *aware* of what they want to and do say; speakers *intend* to speak; speech production is *not efficient* (in fact thinking of the right thing to say or how best to say it can be very difficult); and finally, speech is *interruptible* with speakers choosing to modify or discard speech during planning or even mid utterance.

The automaticity of Levelt's formulation stage is, according to Garrod and Pickering, more complex to identify than conceptualisation. Unlike Levelt, who in his book states that formulating is an automatic process, Garrod and Pickering argue that some formulation processes have a degree of controllability. For instance, they suggest that lexical selection is at least to some extent controlled, as speakers must choose between words and decide on the degree of explanation or description needed to portray a given picture. They further suggest that accessing lexical items is not always efficient, as demonstrated in the tip-of-the-tongue phenomenon.

There are more factors to consider with the automaticity of grammatical encoding. Is it not clear as to whether speakers are aware of the grammatical structure they employ. Garrod and Pickering suggest that awareness could be related to one's literate abilities, with more literate speakers being aware of when they use certain grammatical structures, and less literate speakers having less awareness of their grammar use. Yet, grammatical encoding must involve some choice between the multiple possible structures that could be employed. Furthermore, there is evidence that more complex sentence structures are thought to require more processing power in both comprehension and production than simple sentence structures (Fallon et al., 2006; Ferreira, 1991; Gordon et al., 2002). This would suggest that grammatical encoding is not entirely efficient.

Similar claims about formulation processes have been made in studies using dual-task logic. For example, Ferreira and Pashler (2002) examined whether the three stages of single word production, lemma selection, phonological word-form selection, and phoneme selection, are subject to central or modular processing. Central processing refers to where multiple processes share the same resource pool, with capacity being distributed between them. Modular processing, on the other hand, refers to processing carried out by a dedicated (modular) system and are thought to be carried out automatically. As mentioned in the Introduction, dual-task studies try to determine which processes are centrally processed (i.e., share processing capacity with other tasks) by manipulating the duration of given processes and examining whether this in turn affects performance of another unrelated task. Ferreira and Pashler manipulated the processes of lemma selection, word-form selection, and phoneme selection, and the unrelated task was tone discrimination. If processing capacity is shared between lemma selection (or equally word-form or phoneme selection) and tone discrimination then increasing the duration of lemma selection should increase the time taken to discriminate between tones to the same degree. The duration of lemma selection was manipulated by using high or low constraint cloze sentences, for example, a highly constrained close sentence where lemma selection would be extremely fast would be “Bob was tired so went to (picture of bed)”, whilst a low constrained cloze sentence here would be “She saw a picture of a (picture of a bed)”. Ferreira & Pashler found that participants named the picture slower after the low constraint cloze sentences and, importantly, this propagated to tone discrimination latencies. This, therefore, suggests that lemma selection is a centrally processed, non-automatic, process. Ferreira & Pashler also found evidence that phonological word-form selection is centrally processed.

Conversely, the results suggested that phoneme selection does not impose central processing demands and can be processed alongside another task. Thus, Ferreira and Pashler demonstrated that some formulation processes are controlled and capacity demanding, whilst others take place independently of central capacity demands. Corroborating evidence comes from Meyer and van der Meulen (2000). They used eye-tracking to examine viewing time for an array of objects, with the intention of finding out how far participants got with word production before moving on to the next item. It is thought that participants will attend to an object up until all the capacity demanding processes are complete. Through manipulating phonological relatedness Meyer & van der Meulen found that participants move onto the next object only once the phonological word-form had been accessed, yet before the necessary phonemes had been selected (cf. Cook and Meyer, 2008).

Overall, it is assumed that the earlier stages of speech production are the most resource limited, with conceptualisation processes thought to be highly demanding along with some formulation processes. The later processes, however, such as late formulation processes and articulation, are considered to proceed fairly automatically. Within my design participants were required to engage, at the least, in conceptualisation processes; a resource limited stage of production.

Also of relevance here, is how far individuals plan before they start speaking. That is, do they have an idea for the whole sentence structure or do they only plan the first phrase or clause? It is important to be aware of how much information is typically stored in memory prior to articulation, and also have an estimate of how much planning should overlap with comprehension processes.

5.1.2 The Planning Process.

Retrieval of speech must be fast and efficient as speakers are generally fluent (Levelt, 1989). Because production is so fast and efficient it has been argued that production is incremental and that the different stages can be achieved in parallel, for example we can be accessing the lemma information for the second word as we are articulating the first word. In this way, a speaker could begin articulating their utterance before knowing their subsequent speech (Ferreira & Swets, 2002). Evidence to support this idea comes from the examination of disfluencies during speech production. Speech production tasks have shown that participants pause before producing major constituents (Goldman-Eisler, 1967; Henderson, Goldman-Eisler & Skarbek, 1966). These pauses are thought to demonstrate planning stages of production. It has also been noted that speakers will alternate between periods of fluent speech and periods of pause-filled speech. These findings suggest that the whole utterance has not been planned prior to articulation starting (Ferreira, 1991; Henderson et al., 1966).

Although there is agreement that some aspects of language production occur online, there is research to suggest that some advance planning of speech takes place. For example, Meyer (1996) examined what linguistic information is retrieved before articulation begins. This experiment tested Dutch speakers using a picture-word interference task. Participants were presented with two objects and asked to produce a noun-phrase, such as “the arrow and the bag”, or a sentence, “the arrow is next to the bag”. Alongside the pictures participants heard an auditory distractor word which could be semantically or phonologically related to either the subject noun (arrow) or object noun (bag). Meyer found that semantically related distractors increased the mean speech onset for both types of nouns, and phonologically related distractors decreased the mean onset time for subject nouns. However, there was no phonological facilitation in naming the object nouns. These results indicate that before speech onset participants have retrieved semantic (and possibly syntactic) information for the utterance, and have also retrieved the phonological word-form for the first (subject) noun.

Some researchers claim that speakers plan on a ‘word-by-word’ basis. That is, speech is put together only one word at a time to prevent the processing system becoming overwhelmed or the content or sequence of information being disrupted. Evidence that supports this claim comes from Griffin and colleagues. In an eye-tracking study, Griffin and Bock (2000) demonstrated that within visual scenes the gaze durations on objects and time between fixations and word onsets were similar no matter where the object came in the produced sentence. They suggest that if multiple lexical items were activated at once then the gaze duration and time between fixation and onset would instead be faster towards the end of the sentence. Therefore, they interpret these findings as evidence of word-by-word processing.

Other researchers, however, argue for clausal units of planning (Allum & Wheeldon, 2007; Martin, Miller & Vu, 2004; Smith & Wheeldon, 1999). Such researchers propose that in a given sentence, the lexical representations (including semantic and syntactic information) of all content words are activated before speech onset. More recent research has also suggested that speakers may even access the phonology of all content words before speaking (Alario, Costa & Caramazza, 2002; Costa & Caramazza, 2002; Damien & Dumay, 2007; Schriefers, 1992). For example, Damien & Dumay (2007) demonstrated that presentation of a phonologically related distractor speeded up the production of noun phrases.

Overall, there appears to be a consensus that although some language production processes are achieved incrementally some advance planning must take place before we start to speak. However, the debate still remains as to how far we plan before articulation begins. What is important to note here, however, is that at least some degree of planning should overlap with comprehension processes in the dual-task, and at the very least participants should be storing and processing the semantic information for their utterance. The importance of this overlap between planning and comprehension will now be explained in more depth as I describe what the experiment entails.

5.1.3 Overview of Experiment.

This experiment examines how speech production and speech comprehension are coordinated using a dual-task paradigm. Through measuring prediction at the verb I am able to observe whether participants can still comprehend online whilst having to also complete a concurrent speech production task. The design should allow me to examine the costs of doing both tasks at once through comparisons to where the same speech comprehension and production tasks are dealt with in isolation: a single- versus dual-task comparison.

The comprehension measure is identical to that described in Experiment 3, participants are presented with a visual array and must use the auditory information to target which of the objects in the array will be heard. If participants are able to predict online at the verb, as demonstrated in Experiments 1-3, then we should see faster latencies to the target in the specific verb condition compared to the general verb condition. The verb information will be presented in both simple and complex sentence structures.

The production task requires participants to produce a sentence that incorporates three given objects. For example, participants may be presented with a display that depicts a woman, a cup and a piano, and would produce a sentence such as ‘The woman puts the cup on the piano’. In half of the experimental trials participants produced this sentence immediately, on the other half of the trials this production task was coordinated with the visual world comprehension task.

In the dual-task trials participants are presented with the three objects which they must use to produce an utterance, however they are prevented from producing this sentence immediately and instead must complete the comprehension task first. Once the comprehension task has ended participants are then immediately cued to produce their utterance. During these trials participants must store the three objects in memory. This storage is meant to replicate the storing of concepts that happens in normal conversation. Whilst conversing with others one must keep track of concepts that have been mentioned and where applicable must also keep hold of ideas or concepts that one would like to introduce.

Because the cue to produce their sentence occurs immediately after comprehension there is a possibility that participants will attempt to plan their utterance during the comprehension task. This overlap between planning and comprehension should be highlighted in participants' eye-movement latencies or speech onsets. Those who prioritise production and plan during the visual world should have slower eye-movement latencies and a reduction in online prediction (a smaller difference between the specific and general conditions). Similarly, participants who attempt to swap back and forth between tasks are likely to miss vital information in the comprehension task and could demonstrate slower eye-movement latencies and/or a reduction in online prediction. Conversely, those who wait to plan until after the comprehension task has finished will have longer speech onset latencies.

As highlighted in section 5.1.2 some advanced planning of speech is necessary before articulation can start, and many studies, for example Meyer (1996), have demonstrated that an overall semantic concept of the utterance is in place before production begins. Therefore, at least some conceptualisation and formulation processes must be completed at the point of speech onsets. As described above, where this advanced planning is achieved should be located in the dual-task design.

To summarise, through a combination of measuring eye-movement and speech onset latencies and gathering feedback from participants it should be possible to pinpoint the strategies used to coordinate the speech production and speech comprehension tasks, whether that be rapidly switching between tasks, planning and listening simultaneously, deferring one task until the other is complete or prioritising one task over the other.

5.2 Hypotheses.

If prediction processes are not affected by the presence of a second linguistic task then we should see similar prediction effects to those noted in Experiment 3 where comprehension was tested alone. If however, the concurrent production task slows comprehension then we should see delayed latencies to the target, and perhaps an elimination of the prediction effect (specific-general verb difference).

The production task is presented both alone and alongside the comprehension task; this allows us to also examine whether the comprehension task has an effect on speech production. If production is slowed down by the concurrent comprehension task then we should see longer speech onset latencies in the production with comprehension trials compared to the production only trials.

5.3 Method.

Participants.

29 University of Birmingham students took part in this study in return for course credits (20 female, ages 18-28, mean age 19 years, 5 months). All were native English speakers with normal to corrected vision and hearing. Two additional participants were tested but removed from the analyses due to extremely large standard deviations in eye-movement latencies.

Materials.

52 sets of stimuli were used. Each set consisted of a production display, visual world display, four associated auditory sentences, and a production cue. The visual world displays and associated sentences were taken from Experiment 3. Each visual world appears alongside a specific and a general verb, and a simple and complex sentence (see section 4.3). Four stimuli sets were removed after Experiment 3 due to negative feedback about the clarity of the sound which was mirrored in very small numbers of fixations on the target (see Appendix G).

The production displays were comprised of three object pictures. Two pictures were presented adjacently 6.5 inches (16.51cm) apart, and the third picture was presented in a centred position 5 inches (12.7cm) underneath the other two objects. The production cue consisted of a question mark presented in the centre of the screen. This question mark cued participants to start speaking (see Figure 6).

Design and procedure.

This experiment used a within subjects design. Items were presented between-subjects, whereby each visual scene was seen only once by each participant, and for each item (e.g., eat/move) participants heard only one of the four possible associated sentences. However, every participant heard a selection of all sentence variations. The main independent variables (IVs) here were; a) Complexity (whether the sentence is simple or complex in structure), b) Verb Type (whether the sentence contains a specific or general verb); c) Sentence Type (whether the sentence has an embedded clauses or is an other complex sentence); d) Verb Position (whether the verb is positioned early or late on in the sentence); e) Trial type (whether the participant is taking part in a speech production only or speech production and comprehension trial); and f) Task (whether the participant is taking part in the production and comprehension dual-task or comprehension only task). The dependent variables (DVs) were the latency of eye-movements to the target object, and the speech onset latencies. Each participant heard the same number of specific and general verbs, the same number of embedded clauses and other sentences, the same number of future tense (will) and present tense (-ing) verb forms, and the same amount of early and late verb positions.

Participants took part in 5 practice trials and 104 experimental trials. Half of the experimental trials consisted of “Speech Production Only” trials, and the remaining 52 trials were “Speech Production and Comprehension” trials.

‘Speech Production Only’ trials.

In the Speech Production Only trials, the production display was presnted for 1500ms and immediately followed by the production cue. Participants had to combine the three objects from the production display into one sentence; this sentence had to be produced as quickly as possible after the production cue appeared. The onset of participants’ speech was measured using a voice key. The production cue remained on screen until the participant had completed their sentence, at which point the experimenter initiated a key press to continue on to the next trial.

‘Speech Production and Comprehension’ trials.

These trials require both speech production and speech comprehension, and coordination between the two tasks. First, the production display appears (for 1500ms) and then participants see a “stop signal” presented for 500ms. This stop signal signified that participants must wait to produce their sentence until after they had completed a comprehension task. The comprehension task involves using the visual world. Participants are presented with the visual scene (comprised of four object pictures and an agent’s face) and are asked to familiarise themselves with the object pictures before looking to the face which appears (after 3000ms) in the centre of the screen. As the participants fixated on the face the auditory sentence began to play, participants had to look to the target (the object mentioned within that sentence) as quickly as possible. The eye-movement latencies to the target were recorded. At the end of the sentence the production cue appeared and participants had to produce their sentence using the objects from the production display (see Figure 6).

The order of Speech Production Only and Speech Production and Comprehension trials were randomised, however, programmed so that no more than four of the same trial type were presented consecutively.

All participants were asked for feedback regarding how they coordinated the two tasks. The experiment took approximately 30 minutes to complete.

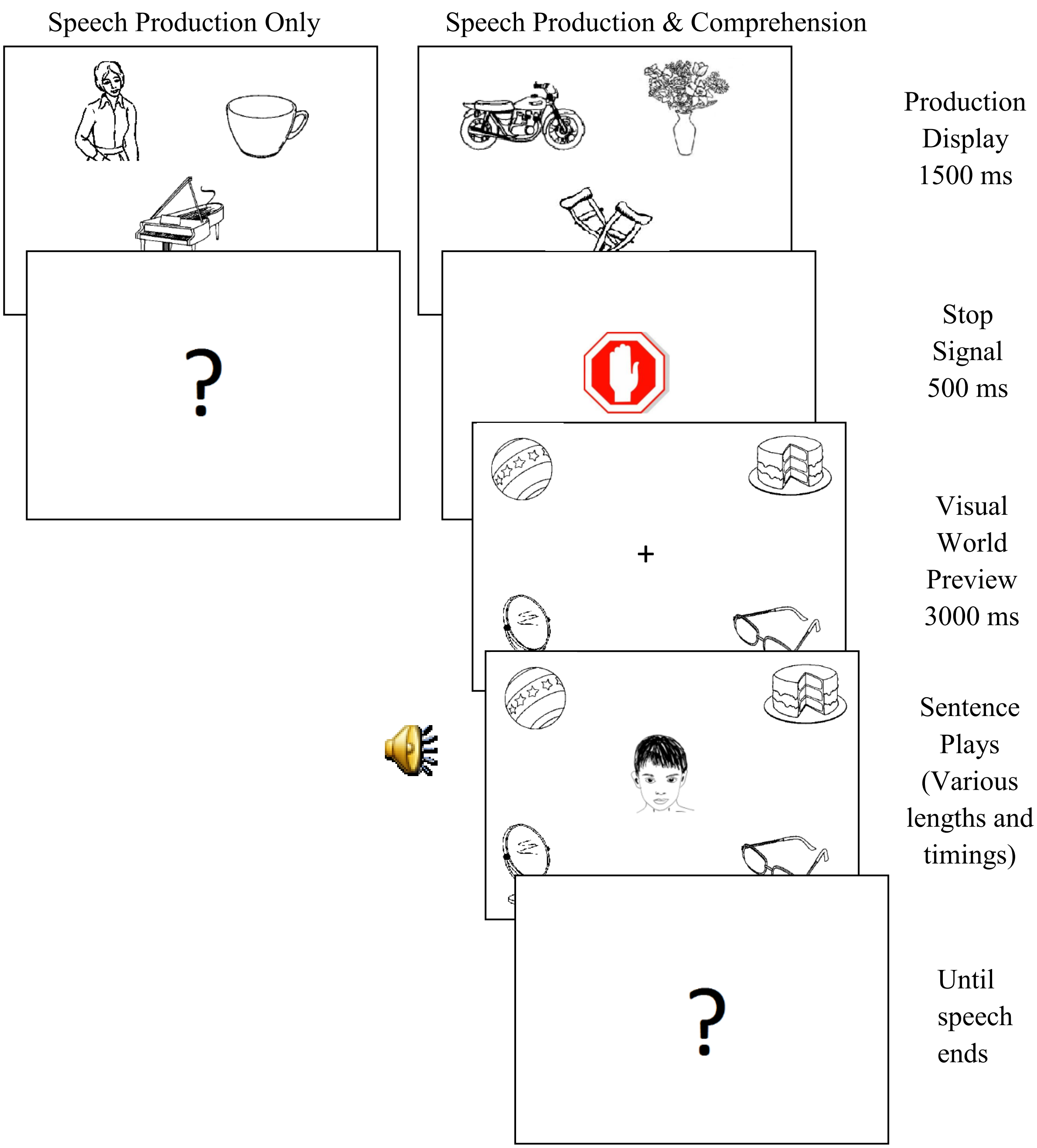

Figure 7. A visual representation of the two types of trial; Speech Production only (left) and Speech Production and Comprehension (right). Speech production only trials involve a production display followed immediately by a production cue display. Speech Production and Comprehension trials involve a production display, followed by a stop signal display, a preview display, a gaze-dependent and auditory sentence display, and lastly, a production cue display. The right-most column gives the onscreen timings for each type of display.

5.4 Results

5.4.1 Eye-movement data.

Eye-movement variables.

All eye-movement variables described in Experiments 1 and 3 are similarly used here.

Exclusion criteria.

The exclusion criteria are also described in Experiment 1. 593 trials where participants did not fixate on the target were removed from analysis (350 trials were removed from the general verb condition, 243 from the specific condition).

31 trials were excluded due to fixating before the 200ms time-point (11 trials were removed from the general condition, 20 trials were excluded from the specific condition). 16 trials where eye-movement latencies were over 2.5 standard deviations above the participant mean were also removed from analysis (9 trials were excluded from the general condition, 7 from the specific condition).

In total, participants fixated on the target in 61% of the trials. They fixated the target on 56% of trials in the general verb condition, and on 68% of the trials in the specific verb condition.

Table 12. The number and percentage of valid fixations across the specific and general verb conditions. [Experiment 4].

	Number of Valid Fixations	Total Number of Trials	% of Trials with Valid Fixations
Specific Verb Condition	511	754	68
General Verb Condition	404	754	56
Total	915	1508	61

5.4.2 Linear mixed effects model

In Experiments 4-6 there are many more factors that need to be considered during data analysis. Not only must the factors analysed in Experiment 3 be included in the mixed model analysis, but additional factors from the speech production task must also be included. Further to this, I must also examine any differences between the single and dual-task conditions to assess whether comprehension and/or production are affected by the presence of a concurrent task.

To make these analyses and comparisons clear this section is split into "Comprehension Task", "Cross-Experiment Comparison", and "Production Task". The comprehension task subsection examines the effects of verb type (differences between the specific and general conditions) and sentence complexity (differences between simple and complex sentence structures). This subsection also assesses whether the production task impacts upon comprehension by examining the effects of speech onsets and the number of items recalled and included in speech on eye-movement latencies.

The second subsection, the cross-experiment comparison, compares whether the ability to predict online is affected by the secondary task. Here the strength of the prediction effect in this experiment is contrasted with that found in Experiment 3 where there was no secondary production task.

Finally, the production task subsection conversely examines whether the comprehension task impacts upon production. Here comparisons are made across the production only (PO) and production with comprehension (PC) trials to examine whether the speed of speech onsets and/or the amount of items recalled in speech is affected by the presence of the comprehension task.

Comprehension task.

The variables that I was interested in here are the effects of verb type and complexity and whether there is a significant interaction between the two. I also examined the complex sentences further to explore any differences across verb positions (early versus late) and sentence types (embedded clause versus other).

Full data-set

Random effects model

The preferred random effects structure included participants as random effects on the intercept (1|PptNum), and a by-item slope for complexity (with different estimations of covariance) (0 + SimpComplex | Item).

Fixed effects model

I then compared models with different fixed effect combinations. A no interaction model was preferred (χ^2 = 2.327, p = .13), with a main effect of verb type (χ^2 = 24.445, p < .001) and complexity (χ^2 = 5.701, p < .05) only. The analysis demonstrates shorter latencies to the target after hearing a specific verb compared to a general verb, and shorter latencies when listening to a simple sentence compared to a complex sentence (see Table 13). Furthermore, eye-movements to the targets were initiated prior to processing the noun information (see Table 14).

Table 13. The means and standard deviations of eye-movement latencies to the target across verb type (specific, general) and complexity (simple, complex). [Experiment 4, full data set].

Complexity	Verb Type			
	Specific Verb Latency (ms)	SD (ms)	General Verb Latency (ms)	SD (ms)
Simple	951	244	1012	251
Complex	1028	406	1202	442

Table 14: Mean eye-movement latencies to targets (in ms) across the Specific, General and Simple, Complex conditions, relative to the verb and noun onsets. [Experiment 4, full data set].

	Specific Simple (ms)	**General Simple (ms)**	***Difference between conditions***	**Specific Complex (ms)**	**General Complex (ms)**	***Difference between conditions***
Verb onset	951	1012	61	1028	1202	174
Noun onset	232	290	58	388	562	174

Note: The specific latencies are under 400ms which means that these eye-movements are mediated by the verb information not the noun (see Chapter 3 discussion).

Complex data set.

The data was further analysed to examine the effects of verb position and sentence type. The factors analysed in this model included verb type, verb position, and sentence type.

Random effects model.

The preferred random effects structure included items and participants as random effects on the intercept; (1|PptNum) + (1|Item).

Fixed effects model.

The model comparison preferred the full parameter model due to a significant three-way interaction between Verb Type, Sentence Type and Verb Position ($\chi^2 = 7.698$, $p < .01$). In light of this, I examined the two sentence types (other and embedded) separately to highlight where this three-way interaction comes from. Analysis of the embedded sentence structures revealed no interaction between verb position and verb type ($\chi^2 = 0.127$, $p = 0.72$), but a main effect of both (verb position: $\chi^2 = 22.664$, $p < .001$; and, verb type: $\chi^2 = 5.609$, $p < .05$). Analysis of the other sentence structures also revealed no interaction between verb position and verb type ($\chi^2 = 0.127$, $p = .72$) but a main effect of both (verb position: $\chi^2 =15.305$, $p < .001$; and, verb type: $\chi^2 = 10.355$, $p < .01$). Close inspection of the means reveals that this three way interaction appears due to a larger verb type effect (specific-general difference) in embedded sentences where the verbs appears late compared to early, and a larger effect of verb type in other sentences where the verbs appears early compared to late (see Table 15 below). Therefore, in this experiment it appears as though it is easier to use verb information when it is presented late on in an embedded sentence, however, conversely is easier to use when it is presented early on in other sentence structures.

Table 15. The mean latencies to the target across the two verb type conditions (specific, general) when presented within the two different sentence types (embedded, other), and the two different verb positions (early, late). [Experiment 4, complex data set].

	Embedded Sentence		**other Sentence**	
Verb Type	Early Verb Position	Late Verb Position	Early Verb Position	Late Verb Position
Specific (latency in ms)	1217	827	1139	841
General (latency in ms)	1317	960	1330	999
Prediction Effect: Specific-General Difference (ms)	100	***133***	***191***	158

Note: The bottom row shows the strength of the prediction effect across the sentence types and verb positions. The emboldened numbers highlight where three way interaction between verb type, sentence type and verb position stems from; with a stronger effect in late embedded compared to early embedded, and a stronger effect in early other compared to late other.

Does a concurrent production task affect online comprehension?

Effect of speech onsets on eye-movement latencies.

The next step in analysis involved examining whether the participants' speech production affected eye-movement latencies in the visual world. One possible outcome here could be that faster speech onsets are associated with slower eye-movement latencies; a trade-off between the two tasks. Such a finding could indicate that participants are using the comprehension task to plan their own utterance. To examine this possibility the speech onsets and eye-movements were compared across trials.

Random effects model.

The preferred random effects structure included a by-participants slope for speech onsets (with different estimations of covariance) (1 + SpeechOnset | PptNum), and by-item slope for complexity (1 | Item:SimpComplex).

Fixed effects model.

When examining the full data set the model comparison preferred a simple no interaction model with no three-way interaction between speech onset, verb type and complexity (χ^2 = 0.346, p = .56) and no two way interactions between speech onset and verb type (χ^2 = 1.984, p = .16), speech onset and complexity (χ^2 = 0.076, p = .78), or verb type and complexity (χ^2 = 0.241, p = .62). The model also *excluded* speech onset (χ^2 = 0.062, p = .80) demonstrating that speech onset did *not* significantly impact upon the model of eye-movement data[3].

Effect of recall on eye-movement latencies.

The next model to run in conjunction with the eye-movement data is one that examines the impact of 'recall'. Participants were asked to include three objects in their spoken utterance, the number of items that were recalled and used in speech have been calculated across all participants and all trials. This linear model examines whether the number of items recalled in speech impacts upon eye-movement latencies. For example, if recalling two items is easier than

[3] A comparison was also conducted with the complex data set. This analysis also demonstrated no effect of speech onsets. The output for this model can be found in Appendix 2D.

recalling three does this therefore mean that participants have more resources available to predict at the verb? The factors analysed here include recall, verb type and complexity.

Random effects model

When analysing the full data set the preferred random effects structure included participants as random effects on the intercept (1|PptNum), and a by-item slope for complexity (1|Item: SimpComplex).

Fixed effects model

A simple no interaction model was preferred with no three way interaction between recall, verb type and complexity (χ^2 = .018, p = .89) and no two way interactions between recall and complexity (χ^2 = 0.113, p = .74), verb type and complexity (χ^2 = 2.274, p = .13), or verb type and recall (χ^2 = 2.188, p = .14). Importantly, the model did *not* find a significant effect of recall (χ^2 = 0.067, p = .79) and thus recall was excluded from the model[4].

Cross-Experiment Comparison

Does a concurrent production task affect online prediction?

(Comprehension with Production vs. Comprehension Only)

In order to examine in more depth whether the production task has impacted upon the comprehension task, I must compare eye-movement latencies in this experiment to those documented in Experiment 3 where the comprehension task was performed alone. This will allow us to explore whether there is a delay in processing speed during the comprehension task here, and also whether the ability to predict (as measured by the difference between specific and

[4] A comparison looking at the complex data also demonstrated no effect of recall. The output for this model can be found in Appendix 2D.

general verbs) is altered by the presence of this concurrent production task. The factors analysed in this model include verb type, complexity and experiment.

Random effects model.

The preferred random effects structure included participants as random effects on the intercept (1|PptNum), and a by-item slope for experiment and complexity (with different estimations of covariance) (0 + Exp:SimpComplex | Item).

Fixed effects model.

Model comparison revealed no significant three-way interaction between verb type, complexity and experiment (χ^2 = 2.151, p = .14) and no two-way interaction between verb type and complexity (χ^2 = 1.043, p = .31), or experiment and complexity (χ^2 = 0.732, p = .39). The preferred model, however, did include an interaction between experiment and verb type (χ^2 = 9.758, p < .01). Further investigation of the means reveals that the Experiment:Verb type interaction appears due to a smaller difference in eye-movement latencies across specific and general verbs in Experiment 4; i.e., the effect of verb type decreases by 86ms in Experiment 4 relative to Experiment 3 (see Table 17). Therefore demonstrating that although there is still an advantage for specific verbs (and thus a significant prediction effect), this advantage decreases when a concurrent speech production task is present. In other words, online prediction *is* affected by the presence of a secondary task.

The model also illustrated a main effect of experiment ($\chi^2 = 0$, $p < .001$). As can be seen in Tables 16 and 17, the eye-movements in the comprehension task in Experiment 4 are sigificantly longer than those seen in Experiment 3 where there was no secondary task[5].

Table 16. The mean and standard deviations for eye-movement latencies in simple and complex sentences as shown across Experiments 3 (comprehension only) and Experiment 4 (production and comprehension). [Experiment 4 and Experiment 3, full data sets].

	Simple Sentences		Complex Sentences		Simple-Complex Difference (ms)
	Mean Latency (ms)	SD (ms)	Mean Latency (ms)	SD (ms)	
Experiment 3	957	260	1053	283	96
Experiment 4	977	348	1109	431	132

Note: The final column demonstrates how the difference between simple and complex conditions is larger in the dual-task experiment.

Table 17. The mean eye-movement latencies to the target across the specific and general verb conditions in Experiment 3 and Experiment 4. [Experiment 4 and Experiment 3, full data sets].

Experiment	Specific Verb Latency (ms)	General Verb Latency (ms)	Prediction Effect (General-Specific Difference)
3 (Single-Task)	917	1134	*217*
4 (Dual-task)	996	1129	*133*

Note: The final column demonstrates the change in strength of prediction in experiment 4 where a secondary task is present.

[5] A comparison was also conducted looking at the complex data set. This comparison demonstrated no interactions between experiment and verb position or experiment and sentence type. The output for this model can be found in Appendix 2D.

Production Task: Onset Latencies and Items Recalled

Does a concurrent comprehension task affect online production?

(Production with Comprehension vs. Production Only)

This model examines whether the comprehension task impacted upon speech production. Here, we compare the speed of production (via speech onsets) across production only (PO) trials and production with comprehension (PC) trials, and, in addition, examine differences in speech content (via recall of items) across these two trial types[6]. Possible outcomes here include speech onsets being increased in the presence of the comprehension task, or conversely participants recalling fewer items from the production display in order to maintain similar speeds across the two types of trial. The dependent variable in this model is speech onset and the factors included are recall (number of items recalled and included in the participants' utterance) and trial type (PO, PC).

Random effects model.

The preferred random effects model included a by-participant slope for trial type (0+ TrialType|PptNum).

Fixed effects model.

The model preferred a simple no interaction model with no interaction found between trial type and recall ($\chi^2 = 0.378$, $p = .54$). The model did however find a significant main effect of trial type ($\chi^2 = 13.738$, $p < .001$) indicating that speech onsets differed depending on which type of trial was encountered. Through comparing the means this difference was found to be due to faster

[6] The fluency of speech across the two trial types was also examined to ensure that speech in the production only trial was not significantly more fluent than speech in the dual-task. The number of disfluencies were calculated across trial types (PO and PC) in all of the dual-task experiments. These comparisons revealed very similar disfluency rates across the two types of trials demonstrating that participants are producing speech effectively in both the single and dual task conditions. The comparison is described in Appendix 2D.

speech onsets in the *dual-task* condition. This is an unexpected finding as logically the speech production should be harder when a second task is present. Therefore I went on to examine whether these faster onsets were mediated by recall; is it that speech is faster in the dual-task condition because participants include less items in their speech? Indeed, from looking at Table 18 below it would appear as though fewer items are recalled in total in the dual-task trials.

Table 18. The number of trials where participants recalled, 0, 1, 2, or 3 items, as shown across production only (PO) and production with comprehension (PC) trials. [Experiment 4, Speech production data].

Recall	PO Trials (N)	PC Trials (N)
0	15	138
1	37	59
2	386	413
3	1017	845
Total no. of items recalled:	3860	3420

Note: The bottom column shows the total number of items recalled in PO and PV trials.

The model revealed that there *was* a significant main effect of recall on speech onset latencies ($\chi^2 = 47.867$, $p < .001$). However, comparison of these means revealed that this effect was not due to a speed-accuracy trade off as described above: participants were in fact *quicker* to produce speech when more items were included in their utterance (see Table 19)[7].
This therefore suggests that participants' speech benefits from having a concurrent comprehension task present. Reasons for this unusual finding are explored in the discussion below. For a detailed list of the mixed effect model outputs for this experiment see Appendix 2D.

[7] The number of items recalled and speech onsets were analysed separately to assess whether they are differently affected by simple or complex sentences, or specific or general verbs. The models revealed, however, that there was no significant effect of verb type or complexity on either the speech onsets or number of items recalled. See Appendix 2D.

Table 19. The mean and standard deviations of speech onset latencies when 1, 2 or 3 items were recalled, as shown across production only (PO) and production with comprehension (PC) trials. [Experiment 4, Speech production data].

Recall	1		2		3	
	Mean Speech Onset Latency (ms)	SD (ms)	Mean Speech Onset Latency (ms)	SD (ms)	Mean Speech Onset Latency (ms)	SD (ms)
PO Trials	1967	758	1869	797	1671	911
PC Trials	2040	1238	1644	864	1464	734

Summary.

The results demonstrate that online prediction at the verb is maintained in the comprehension task, even when a concurrent production task is present. Similarly, the difference in processing simple and complex sentences is also repeated here with faster eye-movement latencies associated with simple syntactic structures. However, in contrast to Experiment 2 where the secondary task did not impact upon prediction, here we notice that a concurrent production task does affect online prediction processes. The ability to predict at the verb, as measured by the difference between the specific and general verb conditions, is reduced here relative to Experiment 3 where no secondary task was present.

Conversely, the results also illustrate that the effect of the production task on comprehension is not mediated by the speed of participants' speech onset or how many items they recall and include in their speech. Faster speech onsets and recalling more items does not affect participants' eye-movements in the visual world.

There was also a significant impact of the comprehension task on production. Surprisingly, however, the effect was due to faster speech onsets in the dual-task condition when a second task is present. That is, participants are actually faster to produce speech when they have to coordinate their production with comprehension. The findings demonstrate that these faster onsets are not due to recalling fewer items; in fact participants are *faster* to speak when more items are recalled and included in their utterance.

5.4.4 Participant Feedback.

Participants were asked to subjectively explain how they coordinated the two tasks and try to pinpoint where in the task they plan their speech. The majority of participants (21 out of 27) described how they planned their utterance prior to the comprehension task starting. Figure 1 illustrates that there are two displays in between where the three objects are presented in the production display and where the sentence plays and comprehension starts. These two displays allow for 3500ms where individuals can plan their speech ready for when the production cue appears. Although participant feedback here is a subjective measure of where planning occurs, nearly all participants claimed to use this specific timeframe to plan their speech so that speech production did not have to be coordinated with the speech production task. This therefore highlights a crucial flaw in our design which reduces the amount of coordination needed to complete both tasks successfully. The implications of this design flaw are examined in more detail in the discussion.

5.5 Discussion

The results demonstrate a sustained effect of verb type, whereby participants can process information at the verb online, and use this information to predict an upcoming theme or referent. This verb effect is maintained when the verb is presented within a complex sentence structure, thus supporting our previous work where online comprehension in simple and complex sentences was examined without the addition of a production task. I also demonstrate how this ability to predict online is preserved even in the presence of a secondary task. When comprehension must be coordinated with a capacity demanding speech production task participants are still able to use verb information online to predict an upcoming referent. This sustained verb type effect illustrates that comprehension processes proceed online in this dual-task scenario.

One new finding with this eye-movement data, however, is an interaction between verb type, sentence type and verb position. In embedded sentence structures, the ability to predict (as measured by the difference between specific and general verbs) is greater at late verb positions. When we consider the two different verb positions it is easy to see why this effect might arise, as the embedded clause always appears at the beginning of the sentence and, in the early verb condition (1a), the verb information is presented within the embedded clause which is arguably the most difficult part of the sentence to interpret. Whereas, in the late verb position (1b), the verb is presented outside of the embedded clause and, therefore, may require less processing effort.

(1a) The woman who *will thread/catch the needle* has very nimble fingers.

(1b) The boy who is the captain of the football team *is eating/moving the cake*.

Conversely, in other sentence structures, the ability to predict is greater at early verb positions. Once again, on examining the two different verb positions the reason for this advantage at early verb positions becomes apparent. These sentence structures often contain a number of content words and the interpretation of these structures is difficult due to having to remember the word order. Therefore, when the vital verb information is presented early on (2a), before all the content words start appearing, the processing demands are lower allowing prediction to proceed quickly. Conversely, in the late verb positions (2b), the verb information follows after the content words at a time where processing demands are inevitably higher.

(2a) The woman *will rip/win the bag* whilst working on the small tombola stall in the corner of the school fair.

(2b) The woman is sitting on the high brown stool with a wooden seat and metal legs whilst *frying/washing the mushrooms*.

A possible reason why this interaction only appears here rather than in Experiment 3 could be due to the additional processing demanded in this dual-task setting; as processing demands are increased the differences between sentence structures and verb positions become more apparent.

Crucially, in this experiment I demonstrate that coordinating speech comprehension with speech production leads to a reduction in the ability to predict online. That is, although the prediction effect is still present here, it is significantly reduced relative to when comprehension is performed alone in Experiment 3. This finding could be taken to suggest that production processes are in some way interfering with comprehension, and thus the two tasks must be overlapping in time. We also see that comprehension has some impact on production too, with a net decrease in the number of objects recalled in the dual-task condition compared to when production is performed alone (n= 3420 and 3860 respectively). Therefore, perhaps I have shown that planning and comprehension are tightly coordinated in the same window of time, and, as resources are shared between the two tasks, less capacity is available to process verbs and thus prediction processes are slowed.

However, the feedback from our participants would suggest that this is not the case. At the end of the experiment participants were asked at what point did they plan their sentence, and whether they thought the planning of their own sentence and comprehension of the auditory sentence overlapped in time. 21 out of 27 participants said they had a sentence ready prior to the onset of the auditory sentence and all 27 participants believed that their planning and comprehension did *not* overlap in time. When asked to pinpoint which part of the trial they used to plan, the majority (21 out of 27) claimed to have used the preview section of the visual world task. To recap, the preview stage allows participants to become familiar with the four visual world objects, they have 3000ms to look around the screen before the sentence begins (gaze-dependent on the face that appears in the centre, see Figure 6). This 3000ms interlude is sufficient time for participants to plan a sentence and store this idea in working memory ready for later. This idea that a sentence is already planned prior to comprehension could explain why there is no interaction between number of items recalled and eye-movement latencies. One would expect that if separate items are being stored and rehearsed then the number of items would have an effect on the secondary task, with one item causing less interference than three. If however the items have been combined into one sentence plan, the number of items included has less importance. Crucially, we note that eye-movements in the comprehension task are completely unaffected by participants' recall of items in speech.

In light of these responses, I suggest that *planning* processes are not interfering with comprehension, however storage (and rehearsal) of the sentence *plan* are. That is, holding the fully planned sentences in working memory consumes processing capacity, and therefore reduces the resources available to online prediction. This processing strategy could also explain why speech onset latencies are significantly faster in the dual-task condition; in contrast to production only trials where participants are cued to speak immediately, the dual-task condition allows an additional three seconds of planning time. Thus, when the production cue appears the sentence is fully prepared and ready to articulate. This idea is further corroborated by the absence of any effect of speech onset on eye-movement latencies. One may expect that if planning and comprehension overlap in time then faster speech latencies should be associated with longer listening times; that is, a sort of trade-off between whether planning or comprehension is given more attention. However, planning and comprehension appear to be separated in time with no planning necessary during comprehension processes, therefore, there is no impact on eye-movements *and* fast speech onsets.

This notion could also offer an explanation for the surprising findings that speech onsets are faster, and more items are recalled, in the dual-task condition. Participants' speech somehow benefits from having to coordinate speech production with a concurrent comprehension task. If the notion proposed above is true, and participants have more time to plan in the dual-task condition, it follows that speech can be produced more quickly and the speech content can be well-formed.

As touched upon in the results, the secondary task employed here seems to be more effective than the digit task presented in Experiment 2. This may be due to employing a conceptual task that requires semantic working memory alike to sentence comprehension (see discussion of Experiment 2), however, may also be attributable to greater task demands. That is, not only will the speech production task be more difficult that digit matching, the experiment now involves two tasks that are both linguistic in nature; therefore we could be observing some domain-specific interference. If production and comprehension use similar pathways as well as working memory capacities then having to coordinate two linguistic tasks will be harder than coordinating comprehension and a non-linguistic task such as digit-matching. Indeed, the aforementioned work by Hohlfeld et al. (2004) demonstrated that when a semantic processing task was performed alongside another linguistic task the interference between tasks 1 and 2 was greater compared to when the second task was non-linguistic.

Finally, one more important question is whether this separation also makes the coordination process harder in some respects. That is, participants had a gap in time in which they could create a full sentence plan and, potentially, formulate at least some of the sentence structure. This means that participants *could* be storing and rehearsing a partial or full sentence as they complete the comprehension task. They may have begun (or completed) formulating their speech. This is inevitably harder than holding on to an idea or concept. Therefore, participants in this experiment may be giving themselves more to do than necessary, and more than is done in dialogue.

To summarise here, participants created a strategy whereby they separated out the production and comprehension tasks. Participants first focused on production processing, preparing their utterance and storing this in working memory. They then switched tasks to comprehension, expending their available resources on listening and predicting online. This strategy allowed participants to perform well in the speech production task; as demonstrated by fast speech onsets in the dual-task condition. Further to this, regardless of a reduced strength in prediction, I still demonstrate online speech comprehension, as illustrated by a significant effect of verb type in the dual-task condition. What is not clear from this experiment however, is whether this interesting strategy of separating out production and comprehension is one that is used during normal, everyday, conversation, or whether it is a consequence of this experimental design. This study has highlighted a crucial design flaw; the preview stage in the visual world adds a gap in time where there wouldn't always be in normal conversation. Dialogue involves a very quick interchange of information, often we don't have long to plan what we want to say. The speed of interchange increases further with multiple interlocutors where you have to be quick to get your point in first. Therefore, to develop this design I need to create a way of removing this time gap, and forcing more coordination between production and comprehension processing.

CHAPTER 6

THE EFFECTS OF A CUED PREVIEW ON COORDINATION

6 Experiment 5

6.1 Goals and Motivation.

This study set about removing the design fault evident in Experiment 4. The purpose here was to remove the gap in time permitted by the preview that allowed participants to plan prior to any speech input. The preview of the visual world is an important stage of the comprehension task, it allows participants to identify the objects (and potential referents of the sentence), and makes participants aware of the object locations. This becomes important after sentence onset, whereby once participants have acquired verb information they can quickly target the noun, without having to waste time searching the screen for the relevant item. Without this preview, the pattern of eye-movements would be much less accurate and I could not compare subsequent results with those documented in Experiments 1-4. Therefore, I cannot remove this preview stage. I can, however, prevent participants from using this time to plan.

This experiment makes the preview stage of the visual world an active part of the experiment. Here, the names of the four object pictures were mentioned in a randomised order and participants must look to the objects immediately after they are heard. This task allows for all four objects to be recognised, both in terms of identity and location, and yet should prevent any planning from taking place. This stage of the visual world will now be referred to as the “cued preview”. The cued preview also created a situation in which participants must quickly coordinate planning processes with comprehension processes, first in the form of single word comprehension and then whole sentences.

Operation span task.

Another goal of this experiment was to examine whether storage and processing information in the dual-task here correlates with other measures of working memory. A common working memory task used to examine executive function is the operation span task; this requires individuals to store words in memory whilst processing mathematical equations. Tasks of this nature have been shown to highlight individual differences in working memory capacity (Engle, 2002; Unsworth, Heitz, Schrock, & Engle, 2005).

Previous research has revealed that working memory capacity is related to dual-task performance, with individual differences reported in the ability to perform dual-tasks and the strategies employed in dual-task situations. For example, Kemper, Schmalzried, Herman, Leedahl, & Mohankumar (2009) examined how individual differences in working memory, processing speed, and Stroop interference affect dual-task performance. The dual-task involved tracking a moving target whilst producing speech. Through comparing dual-task performance with baseline measures of these three variables, the researchers found no effect of Stroop interference on language production, but did find an effect of processing speed and working memory. Processing speed was found to affect speech rate and the grammatical complexity of utterances: faster processors produced longer utterances and spoke at a faster rate than slower processors. Working memory was also found to affect grammatical complexity: those individuals with a superior working memory produced more grammatically complex utterances.

These individual differences in working memory and processing speed also influenced how individuals were affected by dual-task demands. Faster processors who used longer sentences and spoke faster were also more vulnerable to changing their speech rate under dual-task demands. Whilst individuals with better working memories, who naturally use more complex sentences, suffered less decline in grammatical complexity under dual-task conditions relative to those with less enhanced working memories. Kemper et al. (2009) propose that an advantage in working memory can act as a buffer to dual-task demands.

The link between working memory capacity and dual-task performance lies on the assumption that individuals with larger working memory capacities can process information in dual-tasks more efficiently, and switch between tasks more successfully. In fact, in an investigation of the most important predictors of dual-task (or multi-task) performance, Konig, Buhner and Murling (2005) found that working memory was the most important factor, alongside attention and intelligence.

Further to this, research has also examined how working memory capacity affects prediction. Although little research exists in this area that examines spoken language, there are some studies that have been conducted using written language comprehension. For instance, the aforementioned study by Estevez and Calvo (2000), whose results indicated that high working-memory capacity speeds up the time course of prediction. And similarly, the study by Linderholm (2002), which speculated that only those with high working memory capacities are able to made predictive inferences.

Additionally, Otten and Van Berkum (2009) used an ERP technique to examine online prediction in Dutch and whether this was influenced by working memory capacity. They compared participants with high and low working memory capacities, as calculated by their reading span scores. Based on previous findings, such as those documented above, Otten and Van Berkum hypothesised that readers with lower working memory capacity will be less likely to predict upcoming information. This hypothesis was tested using predictive stories, where the story was constructed to allow for the prediction of a particular upcoming noun. ERP responses to predictive stories were compared to control stories where there was no predictable continuation. An anticipatory effect was measured through presenting matching or mismatching gender information prior to the predictable noun. That is, in Dutch the definite article varies depending on the gender of the noun; common gender nouns are preceded by the article “de”, whilst neuter gender nouns are preceded by “het”. This allowed the researchers to manipulate the relationship between the article and the noun. The prior discourse could be predictive (experimental) or non-predictive (control condition), and the relationship between the article and the noun could be consistent (de ketting (*necklace*)) or inconsistent (het ketting).

The ERP results demonstrated a negative deflection at 200-600ms after hearing predictive-inconsistent determiners; this early ERP effect was present for readers with both high and low working memory capacities, demonstrating that working memory capacity did not affect participants’ ability to predict upcoming nouns. However, the researchers did note a late ERP effect specific to readers with low working memory capacity; a late negative deflection at 900-1500ms was found in response to predictive-inconsistent stories. This indicates that although readers with low working memory capacity are anticipating language they are dealing with prediction in a different way; Otten and Van Berkum suggest that these participants need to carry out additional processing when they come across inconsistent information.

Although this research links prediction with written rather than spoken comprehension, it does provide some evidence that working memory capacity plays an important role in predicting upcoming themes.

I speculated here that coordination of production and comprehension is a skill that engages the central executive, and should correlate with sophisticated measures of working memory capacity. To this end, I presented participants with an operation span task and correlated the scores from this task with the ability to predict in the dual-task condition.

6.1 Hypotheses.

If planning is now prevented from occurring prior to comprehension processes then we should observe an increased overlap in the two tasks; this may present itself in delayed latencies to the target and perhaps a further reduction (or loss) of the prediction effect.

If production processes are affected by the changes to the comprehension task then we should see slower speech onset latencies in the production and comprehension trials relative to production only trials.

Finally, I expected to see that performance in this dual task correlates well with measures of working memory capacity. However, I also noted that this correlation would depend on whether we observe a large enough range in individual differences within the student population tested here.

6.2 Method.

Participants.

Twenty-four students from the University of Birmingham took part in the study in return for course credits (22 female, ages 19 to 31 years, mean age 20 years 6 months). All were native English speakers with normal to corrected vision and hearing.

Materials.

All materials were taken from Experiment 4. An additional 119 audio files were created that contained the names of the visual world objects. These audio files were recorded using a female English native speaker, with a neutral accent (mono, 64 kHz sampling rate, 16 bit sampling resolution).

Design & procedure.

The design and procedure are the same as Experiment 4, apart from the alternate version of the preview stage; the cued preview. The preview remains 3000ms in duration, during this time the audio files for the four object names are played. The audio files are played straight after one another; however each audio file has 100ms of silence at the end to ensure that all words can be easily segmented and identified.

The experiment takes approximately 30 minutes to complete without the operation span task (see below), or 50 minutes including the operation span task.

Operation Span Task.

Twenty of the participants in this experiment also took part in the operation span task (18 female, ages 19 to 31 years, mean age 20 years 9 months). Participants are presented with a word alongside a mathematical equation, for example:

2 x 6 + 7 = 19 Boat

The participant must process the equation and decide whether the answer given is true or false, responding via a button press on the keyboard (“z” if true, “m” if false). Additionally, participants had to remember the word presented alongside the equation and when cued (by the word “RECALL”) they had to recall and write down all the words that were presented during that block. Each block is made up of between 2 and 6 sets of words and equations. There are 2 practice blocks and 15 experimental blocks. Participants are asked to voice aloud the equation and the word on each trial. The task takes approximately 20 minutes to complete.

6.3 Results.

6.3.1 Eye-movement data.

Eye-movement variables:

All eye-movement variables are described in Experiment 3.

Exclusion criteria:

The exclusion criteria are also described in Experiment 1. 132 trials where participants did not fixate on the target were removed from analysis (63 trials were removed from the general verb condition, 69 from the specific condition). 16 trials were excluded due to fixating before the 200ms time-point (8 trials from the general condition, 8 trials from the specific condition). 38 trials were removed due to eye-movements that were over 2.5 standard deviations above the participant mean (25 from the general condition, 13 from the specific condition).

In total, participants fixated on the target in 89% of the trials. They fixated the target on 90% of trials in the general verb condition, and on 89% of the trials in the specific verb condition.

Table 20. The number and percentage of valid fixations across the specific and general verb conditions. [Experiment 5].

	Number of Valid Fixations	Total Number of Trials	% of Trials with Valid Fixations
Specific Verb Condition	555	624	89
General Verb Condition	561	624	90
Total	1116	1248	89

7.3.2 Linear mixed effects model.

Again, to make these analyses and comparisons clear this section is split into "Comprehension Task", "Cross-Experiment Comparison", and "Production Task".

Comprehension task.

Full data-set.

Random effects model.

When analysing the full data set the preferred random effects structure included participants as random effects on the intercept (1|PptNum), and a by-item slope for complexity (with different estimations of covariance) (0 + SimpComplex|Item).

Fixed effects model.

A no interaction model was preferred (χ^2 = 0.024, p = .88), with a main effect of verb type (χ^2 = 40.583, p < .001) but no main effect of complexity (χ^2 = 0.447, p = .50). The analysis demonstrates shorter latencies to the target after hearing a specific verb compared to a general verb (see Table 20). However, there is no longer a difference between latencies to simple and complex sentences, as seen in Table 20 below.

Complex data set.

The data was further analysed to examine the effects of verb position and sentence type.

Random effects model.

The best random effects structure modelled participants and items as random effects on the intercept, (1|PptNum) (1|Item), and a by-participant slope for verb type and verb position (1 | PptNum:GeneralSpecific:VerbPos).

Fixed effects model.

The model comparison preferred a simple no interaction model due to no significant three- or two-way interactions. There was no effect of sentence type ($\chi^2 = 2.950$, $p = .09$), however a highly significant main effect of both verb position ($\chi^2 = 22.274$, $p < .001$) and verb type ($\chi^2 = 16.519$, $p < .001$). The latency means demonstrate that participants are faster to look to the target after hearing specific compared to general verbs (see Table 21, see also Table 22 for a comparison of mean latencies in relation to the verb and noun onset times) and faster to respond to verbs that appear late on in the sentence relative to those that appear early (M= 1208ms, SD= 408ms and M= 1488ms, SD= 558ms respectively).

Table 21. The means and standard deviations of eye-movement latencies in the specific and general verb conditions and across simple and complex sentences. [Experiment 5, full data set].

Complexity	**Verb Type**			
	Specific Verb Latency (ms)	**SD (ms)**	**General Verb Latency (ms)**	**SD (ms)**
Simple	1244	460	1417	436
Complex	1260	510	1424	485

Table 22: Mean eye-movement latencies to targets (in ms) across the Specific, General and Simple, Complex conditions, relative to the verb and noun onsets. [Experiment 5, full data set].

	Specific Simple (ms)	**General Simple (ms)**	***Difference between conditions***	**Specific Complex (ms)**	**General Complex (ms)**	***Difference between conditions***
Verb onset	1244	1417	173	1260	1424	164
Noun onset	525	695	170	620	784	164

Does a concurrent production task affect online comprehension?

Effect of speech onsets on eye-movement latencies

The next step in analysis involves examining whether the speech production task affected eye-movement latencies in the visual world. That is, does the speed of speech onsets impact or interact with verb type or complexity.

Random effects model.

The preferred random effects structure included items and participants as random effects on the intercept; (1|PptNum) + (1|Item).

Fixed effects model.

When examining the full data set the model comparison excluded the three-way interaction between speech onset, verb type and complexity ($\chi^2 = 0.079$, $p = .78$), and the two-way interactions between speech onset and verb type ($\chi^2 = 2.075$, $p = 0.15$) and verb type and complexity ($\chi^2 = 0.72$, $p = .40$). However, the model revealed a significant two-way interaction between speech onset and complexity ($\chi^2 = 5.833$, $p < .05$). In light of this, I examined the two sentence structures (simple and complex) separately to highlight where this interaction comes from. Analysis of the simple sentence structures revealed that there was no interaction between verb type and speech onset ($\chi^2 = .49$, $p = .48$), but a main effect of both speech onset ($\chi^2 = 9.65$, $p < .01$) and verb type ($\chi^2 = 15.265$, $p < .001$). Similar analysis of the complex data demonstrated that there was only a significant main effect of verb type ($\chi^2 = 23.637$, $p < .001$), thus highlighting that the two-way interaction stems from a significant main effect of speech onset in simple sentences but no main effect in the complex sentence counterparts. This effect stems from slower speech onset latencies after listening to simple sentences.

Effect of recall on eye-movement latencies

The next model examines whether the number of items recalled in speech impacts upon eye-movement latencies.

Random effects model.

When analysing the full data set the preferred random effects structure included participants as random effects on the intercept (1|PptNum), and a by-item slope for complexity (1 | Item:SimpComplex).

Fixed effects model.

A simple no interaction model was preferred with no three- or two-way interactions. The model included only a significant effect of verb type ($\chi^2 = 32.884$, $p < .001$). Importantly, the model did *not* find a significant effect of recall ($\chi^2 = 1.390$, $p = .24$) and thus recall was excluded from the model[8].

Cross Experiment Comparison

Does a concurrent production task affect online prediction?

(Comprehension with Production vs. Comprehension Only)

This model compares eye-movement latencies in this experiment to those where no production task is present (in Experiment 3). This allows us to explore whether the ability to predict online is altered by the presence of a concurrent production task.

Random effects model.

The best random effects model included participants as random effects on the intercept, and a by-item slope for experiment and verb type (with different estimations of covariance) (0 + Exp:GeneralSpecific | Item).

Fixed effects model.

The model comparison excluded the three-way interaction between verb type, complexity and experiment ($\chi^2 = 0.093$, $p = 0.76$) and two-way interactions between verb type and

[8] A comparison looking at the complex data also demonstrated no effect of recall. The output for this model can be found in appendices 2E.

complexity ($\chi^2 = 0.001$, p = .97), and experiment and complexity ($\chi^2 = 0.748$, p = .39). However, the model revealed that the interaction between experiment and verb type was approaching significance ($\chi^2 = 2.813$, p = .09). Table 23 below illustrates how the difference between latencies in the specific and general condition decreases by 54ms in Experiment 5 relative to Experiment 3. Therefore demonstrating that although there is still an advantage for specific verbs (and thus a significant prediction effect), this advantage does appear to decrease slightly when a concurrent speech production task is present.

Table 23. The mean eye-movement latencies to the target across the specific and general verb conditions in Experiment 3 and Experiment 5. [Experiment 5 and Experiment 3, full data sets].

Experiment	Specific Verb Latency (ms)	General Verb Latency (ms)	Prediction Effect (Specific-General Difference)
3 (Single-Task)	917	1134	*217*
5 (Dual-task)	1261	1424	*163*

Note: The final column demonstrates the change in strength of prediction in experiment 5 where a secondary task is present.

Interestingly, there is no interaction between verb type and complexity even though the simple-complex difference disappears in this experiment but is significant in Experiment 3. As demonstrated in Table 24 below, this non-significant interaction may be down to the greater standard deviations in eye-movements in Experiment 5. The right hand column does demonstrate that the mean difference between simple and complex structures is drastically reduced in the dual task.

Table 24. The mean and standard deviations for eye-movement latencies in simple and complex sentences as shown across Experiments 3 (comprehension only) and Experiment 5 (production and comprehension with cued preview). [Experiment 5 and Experiment 3, full data sets].

	Simple Sentences		Complex Sentences		Simple-Complex Difference (ms)
Experiment	**Mean Latency (ms)**	**SD (ms)**	**Mean Latency (ms)**	**SD (ms)**	
3 (Single-Task)	957	282	1053	283	96
5 (Dual-Task)	1332	456	1340	504	8

Note: The final column demonstrates how the difference between simple and complex conditions is smaller in the dual-task experiment.

It should be noted here, however, that any differences in prediction between Experiments 3 and 5 could be due to the presence of the production task, or equally could be due to the effects of the cued preview on comprehension (which is not present in Experiment 3). The aim of this experiment was to force more coordination between comprehension and production by introducing a cued preview. Thus, it is important that I specifically examine the impact of the cued preview on prediction. To this end, I conducted a comparison of Experiments 4 and 5. Experiments 4 and 5 both involve speech production but only Experiment 5 contains a cued preview.

This comparison revealed similar differences to when Experiment 3 was examined; there was no interaction between experiment and complexity ($\chi^2 = 1.073$, p = .30), and an interaction between experiment and verb type that was approaching significance ($\chi^2 = 3.019$, p = .08) (see Table 25 and Appendix 2E). These findings suggest that the cued preview has had an effect on comprehension. The experiment x verb type interaction documented here is different to that documented in the cross-experiment comparison of Experiment 3 and 4. This suggests that the new dual-task condition is being dealt with differently than in Experiment 4, I explore possible reasons for this in the discussion.

Table 25. The mean and standard deviations for eye-movement latencies across the two verb conditions and simple and complex sentences in Experiments 4 (dual-task, no cued preview) and Experiment 5 (production and comprehension with cued preview). [Experiment 4 and Experiment 5, full data sets].

Experiment	**Simple Sentences**				**Complex Sentences**			
	General Verbs		**Specific Verbs**		**General Verbs**		**Specific Verbs**	
	Mean Latency (ms)	**SD (ms)**	**Mean Latency (ms)**	**SD (ms)**	**Mean Latency (ms)**	**SD (ms)**	**Mean Latency (ms)**	**SD (ms)**
4 (Dual-Task, No Cued preview)	1012	251	951	244	1202	442	1028	406
5 (Dual-Task, Cued preview)	1417	436	1244	460	1424	485	1260	510

Production Task: Speech Onset Latencies and Items Recalled

Does the concurrent comprehension task affect production?

This model assesses whether coordinating production with comprehension affected the speed (as measured by speech onsets) or accuracy (as measured by recall) of participants' speech production[9].

Random effects model

The preferred random effects model included trial type as a random effect on the slope of participants.

Fixed effects model

[9] Speech onsets that are below 300ms were removed from analysis. 1 trial was excluded from the analysis (the 1 PV trial was a complex sentence containing a specific verb) All trials where no items were recalled were also removed from analysis (35 PO trials, 273 PV trials; 142 from the complex condition, 131 from the simple condition; 140 from the specific condition, 133 from the general condition).

The model preferred the full parameter model due to a significant interaction between trial type and recall ($\chi^2 = 4.745$, $p < .05$). Alike to Experiment 4, this interaction was found to be due to *faster* speech onsets and *more* items being recalled in the dual-task condition (see Table 26)[10]. For a detailed output of the linear mixed models in this experiment please see Appendix 2E.

Table 26. The mean and standard deviations of speech onset latencies when 1, 2 or 3 items were recalled, as shown across production only (PO) and production with comprehension (PC) trials. [Experiment 5, Speech production data].

Recall	1		2		3	
	Mean Speech Onset Latency (ms)	SD (ms)	Mean Speech Onset Latency (ms)	SD (ms)	Mean Speech Onset Latency (ms)	SD (ms)
PO Trials	2781	1208	2168	1177	1868	940
PC Trials	2793	1270	1527	1163	1444	933

Summary

The results demonstrate that online prediction at the verb is maintained in the comprehension task, even when a concurrent production task is present. However, the difference in processing simple and complex sentences is no longer present here with similar eye-movement latencies across simple and complex structures.

In contrast to Experiment 4 where the secondary production task impacted upon prediction, here we notice an interaction between experiment and verb type that is only approaching significance. However, when examining the means the difference between the

[10] The number of items recalled and speech onsets were analysed separately to assess whether they are differently affected by simple or complex sentences, or specific or general verbs. For example, if complex sentences add an additional memory load then perhaps the speed of speech onsets are increased in these trials. Similarly, complex-sentence trials may be associated with fewer items being recalled in speech. The models revealed, however, that there was no significant effect of verb type or complexity on either the speech onsets or number of items recalled. See Appendix 2E.

specific and general verb conditions, is reduced here relative to Experiment 3 where no secondary task was present.

Conversely, there was also a significant impact of the comprehension task on production. Alike to Experiment 4, the effect was due to faster speech onsets in the dual-task condition when a second task is present. That is, participants are actually faster to produce speech when they have to coordinate their production with comprehension. The findings demonstrate that these faster onsets are not due to recalling fewer items; in fact participants are faster to speak when more items are recalled and included in their utterance.

6.4.4 Participant Feedback.

Once again, I asked participants to describe how they coordinated the two tasks. A large majority of participants (21 out of 24) claimed to have planned their speech prior to comprehension starting, even with the alteration in design to restrict the amount of gaps available for planning. They professed that the 1500ms during which the production display is present for was sufficient to create a sentence plan.

6.4.5 Operation Span Measure.

Finally, I report the results from the operation span task. Here, I was interested whether the scores from the operation span task would correlate with the ability to predict online in a dual-task setting. Both tasks necessitate the central executive and require the processing and storage of information. Participants' performance in the mathematical equations was analysed for accuracy; participants who did not calculate the equations correctly more than 70% of the time were removed from analysis. Two participants were removed from analysis for this reason, with scores of 53 and 61 percent accuracy. The number of words recalled in the operation span task was calculated across participants; this is the storage aspect of the task and these scores are used in the correlation. Tests of normality, linearity and homoscedasticity on these scores revealed that two further participants should be removed from analysis; one for an exceptionally high score of 56 out of 58, and one for an exceptionally low score of 15 out of 58. The remaining scores formed one of the variables in the bivariate correlation.

The second variable consisted of individuals' mean prediction effect. That is, the mean difference between eye-movement latencies in the specific condition and latencies in the general condition. These differences in latencies capture participants' ability to predict online at the verb (as I am specifically interested in how the operation span scores correlate with the ability to use information to predict in dual-task setting). The relationship between performance on an operation span task and the ability to use specific verbs online was investigated using Pearson's correlation coefficient. There was a medium, negative, correlation between the two variables, $r = -.432$, $n = 18$, $p = .07$, with the variables sharing 19% of their variance. However, this correlation was only approaching significance. For a detailed look at this correlation and the means across participants please see Appendix 1I.

6.5 Discussion

Again, I demonstrate online prediction; participants' latencies to targets were significantly faster following specific verbs compared to general verbs. This therefore indicates that participants are still able to process and use verb information online to predict themes, even in this new dual-task condition. This illustrates that online prediction is an extremely resilient process which is employed even in difficult, resource demanding settings.

A noteworthy finding here is that there is no longer a significant main effect of complexity; simple and complex sentence structures appear to be dealt with in the same way. This is an unexpected finding as the effect of complexity has been highly significant in Experiments 3 and 4. One possible explanation for this finding is that now processing demands are so high participants are just listening for the vital information in the sentence- the verb- and ignoring the overall structure of the sentence. This kind of strategy can be implemented here as participants do not need an overall interpretation of the sentence to perform well in the comprehension task; all they need is the relevant verb phrase. If I were to implement a task where participants must make a judgement about the whole sentence we would probably see this difference between simple and complex sentences reappear. It is clear from this finding that as the task is becoming more complex, more resources must be recruited and, as a result, participants create more inventive ways of dealing with the two tasks.

One finding that has remained unaltered, regardless of the new design, is significantly faster speech only latencies in the dual-task condition compared to when production is performed alone. In Experiment 4, I discussed how this result indicated how more planning can be achieved during the dual-task condition due to the 3000ms preview time. Why, therefore, is this effect maintained when I have made this preview an active part of the experiment? Once again, the effect seems to be due to the extra processing time available for production within the dual-task. In speech production only trials, participants have 1500ms to take in the three object pictures for production and then are immediately presented with a cue to speak. Therefore, in this situation participants have 1500ms only to plan their speech. In the speech production and comprehension trials, participants *could* use additional time on top of this, i.e., the time that should be taken to complete the comprehension task.

During the comprehension task, it is assumed that participants are focusing on comprehension processes, however, it cannot be ruled out that participants use some of this time to perfect or practice their utterance. Therefore, the increase in speech onsets may be due to participants getting further in processing; whilst in the production only trials they may have achieved up to the conceptual (planning) stages of speech, in the production and comprehension trials they may have also began formulating their response. What is of importance here is how far participants get with the production process. If, in both types of trial, participants have completed the most demanding stage of production- conceptualisation- during the initial production display then the degree of coordination between the two tasks is severely reduced. In other words, if 1500ms is sufficient to conceptually plan utterances then participants will not have to coordinate the capacity demanding aspects of planning with comprehension processes during the visual world, the conceptual plan would already be in place.

Therefore, in order to examine the time frame of planning I now refer to participants' feedback. When asked which part of the task they used to plan their utterance, the vast majority (18 out of 24 participants, or 75%) claimed to plan immediately, using the 1500ms during which the production items were presented. The remaining six participants stored the objects separately and rehearsed them (n= 3), used the preview time (n= 2), or used a combination of techniques (n= 1). This feedback suggests that participants' planning is often achieved immediately and the 1500ms production display offers sufficient time to allow for conceptualisation stages to be completed. As a result, planning and comprehension processes can be kept separate. Indeed, when asked whether participants believed the two tasks overlapped or were dealt with separately, 91.7% (22 out of 24 participants) claimed to deal with the two tasks separately.

Here, it is important to assess the plausibility of participants' claims; the questionnaire administered at the end of the experiment is a subjective measure and the accuracy of participants' judgements is uncertain. Therefore, I now look to the psycholinguistic literature to examine whether planning could be completed in this 1500ms timeframe.

Other research has suggested that conceptual processing (of single objects) can be achieved as quickly as 200ms after picture onset. For example, Indefrey and Levelt (2004) performed a meta-analysis on brain imaging studies that examined word production processes. Through reviewing research on the different levels of processing including conceptual processing, lexical selection, and encoding the phonological and phonetic form, they formed estimations of the time course of these different processes. For instance, conceptual processing has previously been studied using electrophysiology (ERP) and Indefrey & Levelt noted that the time frame for accessing conceptual information is between 150 and 200ms. They therefore took a median estimate for conceptual processing of 175ms. Similarly, they offer an estimate of lexical selection to take around 75ms, beginning at around 150-200ms and being completed by 150-350ms, and the various word-form processes (including retrieval of the phonological code, syllabification processes, and phonetic encoding) to be completed in the subsequent 350ms. Taken together these estimations suggest that word production processes are completed and a word is ready to articulate by 600ms after a picture is presented (see Cheng, Shafer, & Akyurek, 2010; Costa, Strijkers, Martin, & Thierry, 2009; Indefrey, 2011; Laganaro, Morand, Schwitter, Zimmermann, Camen, & Schnider, 2009; and Rahman & Sommer, 2008 for similar claims).

Consequently, if participants only need a conceptual plan for their sentence before going on to comprehension, and can process each object by around 175ms, there could be plenty of time available to combine these objects into one full concept/ sentence plan. Participants may have even gotten as far as accessing the lemmas or phonology. The speed in which the three objects are combined could be further increased if multiple planning processes can be achieved in parallel (Bock, 1986; Ferreira & Swets, 2002). Participants may be phonetically encoding the first object whilst conceptually processing the second, and so on. Furthermore, as participants progress in the experiment, creating sentences will become easier due to previously used (and heard) sentence structures becoming active, thereby allowing processes such as interactive alignment to be employed.

An interesting change here appears when comparing eye-movement latencies in this experiment to those found in Experiment 3 where comprehension was performed alone. Here we see that the interaction between experiment and verb type is only approaching significance, which highlights better prediction in this experiment relative to Experiment 4 (where the difference in prediction was significantly worse than that in Experiment 3). This indicates that participants' performance is in fact enhanced by the new cued preview design. I believe that there could be three possible reasons for this; 1) participants have less time to plan in this experiment (due to the cued preview) and therefore are only conceptualising their utterance rather than formulating it, decreasing the amount of information that must be stored during comprehension, 2) participants are better able to use the verb information as they are ignoring the overall structure of the sentence (as described above), and/or 3) the cued preview allows items in the visual world to be better recognised, thereby allowing faster identification of the referent on hearing the specific verb.

As mentioned in Experiment 4, the separation of planning and comprehension means that the sentence plan must be stored as comprehension is completed, I further argued that during the 3000ms gap participants may have also started or completed the formulation stage of speech production. This would therefore mean that the associated words and structures necessary for producing their utterance would also have to be stored during comprehension. It could be the case here, however, that because of the ‘cued preview’ participants have less time to plan and thus cannot complete as many stages of speech production prior to comprehension. This in turn means that less information has be stored in working memory as comprehension processes completed, leaving more capacity available for prediction processes. Further to this, I suggest that the resources required during comprehension are further reduced due to simple and complex sentences being dealt with in the same fashion; if participants are only listening out for the verb phrase then more capacity is available for processing it.

Alternatively, the increased ability to predict at the verb may be due to the cued preview. The cued preview focuses participants attention on each of the visual world objects in turn. Although participants are asked to process these objects in Experiments 3 and 4, it may be the case that explicitly hearing each object being mentioned creates a better “mental map” of where the items are located. This effect may be even stronger when the items have several pseudonyms; hearing the word “sofa” in the cued preview clarifies that this how the item will be referred to in the auditory sentence (rather than “couch” or “settee”, for example). The cued preview as a facilitatory effect is an interesting notion; instead of the additional processing making prediction harder it may instead make using the specific verb easier.

Unfortunately, there is no clear comparison of a comprehension only and dual-task scenario, as Experiment 3 did not include a cued preview. However, following from the argument above I think it unlikely that any subsequent decrease in prediction in the dual-task condition (e.g., in Experiment 6) can be mistaken for effects of the cued preview. If the cued preview actually aids comprehension, as it appears here, then we can reasonably assume that any detrimental effects on prediction (from forcing more coordination between tasks) are due to changes in the production task and how it is dealt with online[11].

Finally, I noted a correlation between our eye-movement measure and the operation span scores that was only approaching significance. Beforehand I had expected that the ability to store and process information within our dual-task condition would correlate well with other working memory measures requiring executive functioning. However, this lack of correlation is not unsurprising when we consider the sample population used; all participants were undergraduate students with relatively similar (high) levels of intelligence. Therefore, the working memory span of this population did not vary dramatically and did not allow for a successful examination of individual differences. Indeed, research into working memory capacity (WMC) often splits participants into a high and low WMC groups and compares each group on a variety of measures, for instance, this was done in Otten and Van Berkum's study into how working memory capacity affects prediction in written comprehension. It is likely that if this experiment was conducted on several groups of participants with individual differences in working memory capacity then we would then see a correlation between eye-movement latencies and operation span scores. Furthermore, correlations of this type largely rely on sample size (Pallant, 2010). This correlation

[11] Since submitting this thesis a control experiment has been conducted. This control experiment introduced a cued preview into the comprehension only task (i.e., an Experiment 3 control). The results demonstrate that there is no significant experiment*verbtype interaction, thereby indicating that the cued preview does not affect the processing of verb information, There was however a significant main effect of experiment, with faster latencies in the cued preview experiment. This suggests that the cued preview does speed up latencies to targets (e.g., by clarifying the name of the referent and/or its location), but this advantage is equivalent across the two verb types. See Appendix 2G.

was based a low number of participants and it could be the case that a larger sample would have yielded a significant correlation.

In conclusion here, this experiment indicates that participants once again separate out planning and comprehension processes, creating a strategy whereby limited coordination is required. I believe that this separation of tasks is made possible because of the 1500ms designated to the production display. This 1500ms is sufficient to create, at the very least, a conceptual plan for an utterance. Again, participants are given the opportunity to deal with coordination in a manner that may not be possible during normal conversation; as interlocutors we don't typically get a time window in which we can plan utterances without the presence of incoming speech. A typical pattern of conversation will involve listening to speech and using this input to create one's own utterance, therefore creating a situation where conceptualisation must be tightly coordinated with listening rather than being completed even before speech is heard. Therefore, this experiment has illustrated where our design must be further tweaked; in order to force more coordination between speaking and listening I must prevent planning from being completed during the production display time. In order to do this, I now shorten this production display time to 700ms and examine what effect this has on eye-movement and speech onset latencies.

CHAPTER 7

THE COORDINATION OF SPEAKING AND LISTENING: PREDICTION AS A CAPACITY SAVING PROCESS

7 Experiment 6

7.1 Goals and Motivation.

This experiment aimed to force more coordination between production and comprehension processes. To create a situation in which planning *must* overlap with comprehension I reduced the time available for planning in the production display; instead of 1500ms, the duration of the production display was only 700ms. 700ms should give participants enough time to look at and recognise the three object pictures, but not enough time to have completed their sentence plan (Indefrey & Levelt, 2004).

7.2 Hypotheses.

With an incomplete sentence plan participants should have to use time during the comprehension task to plan. If this is the case, we should see a disruption or loss in the strength of the prediction effect. Conversely, participants might have to wait until comprehension processes are complete before commencing speech production; in this case we should see delayed speech onset latencies in comparison to production only trials.

Now that the coordination task becomes even harder I expected that participants will continue to create strategies to deal with the two tasks, for example, dealing with simple and complex sentences in a similar fashion by listening out for only vital information.

7.3 Method.

Participants.

27 native English speaking participants were recruited from the University of Birmingham (24 female, ages 18-27, mean age 20 years 2 months). All participants had normal to corrected vision and hearing and took part in return for course credits.

Materials.

All materials were taken from Experiment 5.

Design & procedure.

The design and procedure were the same as in Experiment 5 apart from the alternation in production display time from 1500 to 700ms. The stop signal previously presented between the production display and the cued preview was also removed to avoid including any points where planning could take place prior to comprehension.

7.4 Results.

7.4.1 Eye-movement data.

Eye-movement variables.

All eye-movement variables described in Experiment 3 are similarly used here.

Exclusion criteria.

The exclusion criteria are also described in Experiment 1. 87 trials where participants did not fixate on the target were removed from analysis (39 trials were removed from the general verb condition, 48 from the specific condition). 5 trials were excluded due to fixating before the 200ms time-point (3 trials were excluded from the general condition, 2 trials from the specific condition). Any eye-movements that were over 2.5 standard deviations above the participant mean were also removed from analysis. 45 trial were removed in total; 25 from the general condition and 20 from the specific condition.

In total, participants fixated on the target in 94% of the trials. They fixated the target on 94% of trials in the general verb condition, and on 93% of the trials in the specific verb condition.

Table 27. The number and percentage of valid fixations across the specific and general verb conditions. [Experiment 6].

	Number of Valid Fixations	Total Number of Trials	% of Trials with Valid Fixations
Specific Verb Condition	654	702	93
General Verb Condition	663	702	94
Total	1317	1404	94

7.4.2 Linear mixed effects model.

Comprehension task.

Full data-set.

Random effects model.

When analysing the full data set the preferred random effects structure included participants as random effects on the intercept (1 | PptNum), and a by-item slope for verb type and complexity (with different estimations of covariance) (0 + GeneralSpecific: SimpComplex | Item).

Fixed effects model.

The model comparison preferred a simple no interaction model (χ^2= 0.7018, p = .40), and further excluded complexity (χ^2= 0.076, p = .78). The model, however, revealed a significant main effect of verb type (χ^2= 6.506, p < .05), with faster latencies in the specific verb condition compared to the general condition (see Table 28, see also Table 29 for a comparison of mean latencies in relation to the verb and noun onset times).

Table 28. The means and standard deviations of eye-movement latencies in the specific and general verb conditions and across simple and complex sentences. [Experiment 6, full data set].

Complexity	**Verb Type**			
	Specific Verb Latency (ms)	**SD (ms)**	**General Verb Latency (ms)**	**SD (ms)**
Simple	1291	450	1358	348
Complex	1255	489	1375	433

Table 29: Mean eye-movement latencies to targets (in ms) across the Specific, General and Simple, Complex conditions, relative to the verb and noun onsets. [Experiment 6, full data set].

	Specific Simple (ms)	**General Simple (ms)**	***Difference between conditions***	**Specific Complex (ms)**	**General Complex (ms)**	***Difference between conditions***
Verb onset	1291	1358	67	1255	1375	120
Noun onset	571	636	58	615	735	120

Complex data.

(The factors analysed in this model included verb type, verb position, and sentence type).

Random effects model.

The best random effects structure modelled items and participants as random effects on the intercept (1 | PptNum) (1 | Item); a by-participant slope for verb position (1 | PptNum: VerbPos); and a by-item slope for verb type (1 | Item:GeneralSpecific).

Fixed effects model.

The model comparison preferred a no interaction model with no three- or two-way interactions. Sentence type was also excluded from the model (χ^2= 0.218, p = .64). The model did reveal a significant main effect of verb position (χ^2= 20.697, $p < .001$) (see Table30) and verb type (χ^2= 5.543, $p < .05$).

Table 30. The means and standard deviations of eye-movement latencies to early and late verb positions in Experiment 6. [Experiment 6, complex data set].

Early Verb Position		***Late Verb Position***	
Mean Latency (ms)	***SD***	***Mean Latency (ms)***	***SD***
1408	466	1234	383

Does a concurrent production task affect online comprehension?

Effect of speech onsets on eye-movement latencies

The next step in analysis involves examining whether the speech production task affected eye-movement latencies in the visual world. The factors analysed in the full model here included verb type, complexity and speech onset.

Random effects model

The preferred random effects structure included participants as random effects on the intercept (1 | PptNum), and a by-item slope for complexity (1 | Item:SimpComplex).

Fixed effects model

When examining the full data set the model comparison preferred the full parameter model due to a significant three-way interaction between speech onset, verb type and complexity (χ^2 = 5.290, $p < .05$). Further analysis revealed that this interaction stems from a significant interaction between speech onset and verb type in the complex data set (χ^2 = 6.771, $p < .01$), with faster speech onset latencies after hearing specific verbs compared to general verbs (Specific, Mean= 1267ms, SD = 700ms; General, Mean = 1422ms, SD = 770ms). This interaction is not significant in the simple data set (χ^2 = 0.041, p = .84). (See Appendix 2F for the full output from this analysis).

Effect of recall on eye-movement latencies

The next model to run in conjunction with the eye-movement data is one that examines the impact of recall; whether the number of items recalled in speech impacts upon eye-movement latencies. The factors analysed here include recall, verb type and complexity.

Random effects model

The preferred random effects structure included participants as random effects on the intercept (1 | PptNum), and a by-item slope for complexity (1 | Item:SimpComplex).

Fixed effects model

When analysing the full data set the preferred random effects structure included items and participants as random effects on the intercept. The model comparison found no three way interaction between recall, verb type and complexity (χ^2 = 0.052, p = .82), and no two way interactions between recall and complexity (χ^2 = .0004, p = .99), or verb type and complexity (χ^2 = 1.354, p = .24). However, the model highlighted a significant interaction between recall and verb type (χ^2 = 4.312, $p < .05$). Further examination revealed that more items were recalled during speech production after hearing a specific verb compared to a general verb (M = 1.50, SD = 1.12 ; M = 1.48, SD = 1.11 respectively, see Table 31). Possible reasons for this are explored in the discussion.

The model also highlighted a significant main effect of verb type ($\chi^2 = 13.447$, $p < .001$), however found no significant main effect of complexity ($\chi^2 = 0.088$, $p = .77$) or recall ($\chi^2 = 0.069$, $p = .79$) [12].

Table 31. The average eye-movement latencies by the number of items recalled across the two verb types. [Experiment 6, full data set with speech production data].

	Average Eye-movement Latency (ms) by Number of Items recalled				Average Number of Items recalled (n)
Verb Type	0	1	2	3	
Specific	1305	1335	1297	1224	1.50
General	1327	1438	1357	1414	1.48

Note: The right-hand column show the average number of items recalled across the specific and general verb conditions.

Cross-Experiment Comparison

Does the new production task affect online prediction?

(Comprehension with Production vs. Comprehension Only)

I now wanted to examine in more depth whether the production task has impacted upon the comprehension task. Previously, I have compared eye-movement latencies in this experiment to those documented in Experiment 3 where the comprehension task was performed alone. However, in this case there were too many elements that had changed from Experiment 3 to 6 to allow for an effective comparison. Therefore, in this case I compared comprehension in Experiment 5 with comprehension in Experiment 6. This allowed me to specifically explore what effect the shortened production display has on prediction.

[12] A comparison was also conducted with the complex data set. This analysis also demonstrated no effect of recall. The output for this model can be found in Appendix 2F.

Random effects model.

The preferred random effects structure included participants and items as random effects on the intercept; (1 | PptNum) + (1 | Item).

Fixed effects model.

Model comparison revealed no interaction between experiment and complexity ($\chi^2 = .547$, $p = .46$), however, did reveal a significant interaction between experiment and verb type ($\chi^2 = 4.628$, $p < .05$) Alike to Experiment 4, when examining the experiment x verb type interaction, the means revealed that the difference between latencies in the specific and general condition decreases by 63ms in Experiment 6 relative to Experiment 5 (and 125ms relative to Experiment 3, see Table 32). Therefore, once again demonstrating that although there is still an advantage for specific verbs (and thus a significant prediction effect), this advantage decreases when a concurrent speech production task is present. Interestingly, this decreased prediction effect documented here is more prominent than that found in Experiments 4 and 5 where coordinating the two tasks was less demanding (see Table 32).

Table 32. The mean eye-movement latencies to the target across the specific and general verb conditions in Experiments 3-6. [Experiments 3, 4, 5 and 6, full data sets].

Experiment	**Specific Verb Latency (ms)**	**General Verb Latency (ms)**	**Prediction Effect (Specific-General Difference, in ms)**
3 (Single-task)	917	1134	*217*
4 (Dual-task 1)	996	1129	*133*
5 (Dual-task 2)	1262	1417	*155*
6 (Dual-task 3)	1274	1366	*92*

Note: The final column demonstrates the change in strength of prediction across experiments, from a single task condition (Experiment 3) to when the most demanding secondary task is present (Experiment 6).

Production Task: Speech Onset Latencies and Items Recalled

Does a concurrent comprehension task affect production?

(Production with Comprehension vs. Production Only)

This model examines whether the comprehension task impacted upon speech production[13]. Here, we compare the speed of production (via speech onsets) across production only (PO) trials and production with comprehension (PC) trials, and, in addition, examine differences in speech content (via recall of items) across these two trial types[14]. The dependent variable in this model is speech onset and the factors included are recall (number of items recalled and included in the participants' utterance) and trial type (PO, PC).

Random effects model.

The preferred random effects model included participants as a random effect on the intercept (1| PptNum).

Fixed effects model.

The model preferred a simple no interaction model with no interaction found between trial type and recall ($\chi^2 = 3.684$, $p = .06$). The model did however find a significant main effect of trial type ($\chi^2 = 6.945$, $p < .01$) indicating that speech onsets differed depending on which type of trial was encountered. Through comparing the means this difference was found to be due to faster speech onsets in the *dual-task* condition (see PC trials in Table 32). As done previously, I went on to examine whether these faster onsets are mediated by recall. The analysis revealed

[13] Speech onsets that are below 300ms were removed from analysis. 2 trials were excluded from the analysis (2 PV; 1 complex, 1 simple, 1 general, 1 specific). All trials where no items were recalled were also removed from analysis (63 PO trials, 402 PV trials; 202 from the complex condition, 200 from the simple condition; 204 from the specific condition, 198 from the general condition).

[14] The fluency of speech across the two trial types was also examined to ensure that speech in the production only trial was not significantly more fluent than speech in the dual-task. The comparison is described in Appendix 1J.

that there was a significant main effect of recall ($\chi^2 = 10.591$, $p < .01$), however, alike to Experiments 4 and 5, the effect stems from faster speech onsets when more items are recalled in speech (again, see Table 33). Once again the results suggest that speech production benefits from a concurrent comprehension task. (For a detailed list of the mixed effect model outputs for this experiment please see Appendix 2E).

Table 33. The mean and standard deviations of speech onset latencies when 1, 2 or 3 items were recalled, as shown across production only (PO) and production with comprehension (PC) trials. [Experiment 6, Speech production data].

Recall	1		2		3	
	Mean Speech Onset Latency (ms)	SD (ms)	Mean Speech Onset Latency (ms)	SD (ms)	Mean Speech Onset Latency (ms)	SD (ms)
PO Trials	2470	1968	1690	1013	1686	785
PC Trials	1720	1274	1525	1276	1236	731

Summary

The results demonstrate that online prediction at the verb is maintained in the comprehension task, even when a concurrent production task is present. However, the difference in processing simple and complex sentences is no longer present here with similar eye-movement latencies across simple and complex structures.

Whilst in Experiment 5 we noted that the interaction between experiment and verb type only approached significance, here we see that the interaction is significant again (alike to Experiment 4). The means show that the difference between the specific and general verb conditions, is reduced here relative to Experiment 3 where no secondary task was present (also see Appendix 2F for a cross-experiment comparison of Experiments 3 and 6 which shows a significant experiment x verb type interaction).

There was also a significant impact of the comprehension task on production. Alike to Experiments 4 and 5, the effect was due to faster speech onsets in the dual-task condition when a second task is present. That is, participants are actually faster to produce speech when they have to coordinate their production with comprehension. Once more, the findings demonstrate that these faster onsets are not due to recalling fewer items; participants are in fact faster to speak when more items are recalled and included in their utterance.

7. 5 Discussion.

The results once again demonstrate the sustained effect of verb type; through using information gleaned at the verb participants are able to predict which of the four objects will be later referred to in the auditory sentence. The results also demonstrate that the overall processing demands experienced here are increased in relation to Experiment 5. The latencies to targets are longer in this dual-task setting, indicating that shorter production display time (now 700ms instead of 1500ms) has successfully increased the level of task difficulty. Regardless of task difficulty, however, the ability to predict has again remained significant.

Interestingly, we once again observe no main effect of complexity; with similar eye-movement latencies across simple and complex sentences. This sustained finding therefore supports our claim that with these higher processing demands participants save on resources by listening in for the vital information only. This is an interesting, and surprising, strategy that no doubt helps with coordinating two difficult tasks and allows more attention to be allocated to speech production.

Further to this, when studying participants' eye-movement latencies we see a significant interaction between speech onsets and verb type in our complex data set. The means indicate that when listening to complex sentences participants are faster to speak after hearing specific verbs compared to general verbs. This thereby suggests that there is something about the specific condition (with complex sentences) which allows for more planning of speech so that speech can be produced more quickly. It could be that predicting online at the verb may serve to preserve capacity and allow resources to be expended elsewhere; in this case on speech production processes. That is, predicting allows for a speedier interpretation of the verb phrase and enables participants to quickly choose the correct referent; once this referent has been targeted participants can then allocate all of their attention on the competing production task, resulting in faster speech onsets. I believe that this effect was borne out here (rather than in previous experiments) because the task demands and overlap between tasks are considerably higher. The design of this experiment successfully forced coordination of speaking and listening. In order to deal with the two tasks effectively capacity saving processes are essential.

We also observe a significant interaction between recall and verb type, with more objects recalled in the specific verb condition compared to the general verb condition. This could further indicate that prediction is a capacity saving process whereby by prediction allows participants to concentrate on planning their own utterance and, with more time available to plan, can incorporate more objects into their speech.

These latter two findings suggest that planning is no longer completed prior to comprehension starting; if this were the case we should see no interaction between the speed of speech, or the amount of objects mentioned, and the type of verb presented.

This notion that planning is no longer separated in time from comprehension is also corroborated by the experiment by verb type interaction. In Experiment 5 I noted only a marginally significant interaction between these factors, with the prediction effect actually improving in strength from Experiment 4. Here, however, I demonstrate that this interaction is now highly significant again with the strength of the prediction effect decreasing to only 92ms (see Table 6). This therefore demonstrates that when coordinating the speech production and comprehension tasks cognitive resources must be shared between the two and this impacts upon comprehension processes, both in the way sentences are interpreted (dealing with simple and complex sentences in the same way) and how information from the sentence is used (reduced prediction).

Finally, I note that participants tend to prioritise their own speech production, as evident in the significantly faster speech onset latencies in the dual-task condition in Experiments 4 through 6. Moreover, the capacity saving strategies of predicting online and listening only for vital information allows for more resources to be expended on perfecting one's own speech. This strategy use is further evaluated in the general discussion (Chapter 8), alongside considerations of how this might relate to dialogue and natural conversation, and applications of this new dual-task paradigm.

CHAPTER 8

GENERAL DISCUSSION AND IMPLICATIONS

8.1 Summary of experiments and results

The overall topic of this thesis is the coordination of speech production and speech comprehension. In particular, I have been interested in how interlocutors manage to coordinate the planning of their own speech with the comprehension of their speech partners' utterances. In addition to this, I have examined a particular comprehension process: prediction, and how this is affected by various capacity demands and task manipulations.

In Experiment 1, I replicated the finding that verb information can be processed online to predict upcoming referents. This effect was originally documented by Altmann and Kamide in 1999, and demonstrated the incremental nature of language processing. Using this visual world paradigm here offered a method of examining speech comprehension that is time-sensitive, does not interrupt speech input, and can be employed alongside various task manipulations.

In Experiment 2, I examined comprehension, via online prediction at the verb, in a dual-task setting where participants had to also complete a concurrent digit-matching task. Results demonstrated that, in this case, prediction remained unaffected by a secondary task. When examining why there was no reduction in prediction strength from single- to dual-task condition, I found that recent research has highlighted a distinction between phonological working memory (which digit tasks rely on) and semantic working memory (which sentence comprehension relies on) (Hanten & Martin, 2000; Martin, 2005; Martin & He, 2004; Martin et al., 1994). This distinction indicated that a secondary task that employs semantic working memory *would* lead to a disruption in prediction (this hypothesis was later confirmed in Experiments 4 & 6).

In Experiment 3, I examined whether syntactic complexity affects online prediction processes; previous research suggested that syntactically complex sentences place greater demand on listeners and that certain sentence structures require offline processing (Fallon et al., 2006; Gordon et al., 2002; Tun et al., 2010). Here, however, I confirm the online nature of prediction and demonstrate that verb information can be processed locally and used online to pick out future references. I do however, support the idea that syntactically complex sentences require more processing effort than their simple counterparts, as demonstrated by longer latencies overall to complex sentence structures. This demonstration of online verb processing in complex sentences allowed me to extend my stimuli set in future experiments; all subsequent studies could now examine comprehension using a wide range of different spoken sentences.

In Experiment 4, I started to focus in on my overall research goal of examining the coordination of speech production and comprehension. Here, I reverted back to a dual-task scenario and examined prediction in comprehension alongside a speech production task. I demonstrated that capacity was shared between the two tasks with a disruption to prediction relative to performance in Experiment 3. I also noted however, that participants created a strategy in which they separated out production and comprehension processes; planning their speech prior to comprehension starting. This strategy thereby highlighted a design flaw in our experiment, whereby participants had a gap in processing time (the visual world preview) to plan their speech rather than having to coordinate the two tasks at once. Therefore, to rectify this problem I made the preview an active part of the experiment to prevent any planning at this stage (Experiment 5).

In Experiment 5, I noted that participants began to deal with simple and complex sentences in the same way; that is, they now listened only for the vital verb information and ignored the overall structure of the sentence. I believe this to be a way of saving capacity in a resource-demanding situation. I also noted that participants still separated out the production and comprehension tasks- they were able to use the 1500ms production display time to conceptually plan their utterance and then, possibly, start to formulate their utterance (which is a less demanding stage of production) during comprehension. Again, participants were given the opportunity to deal with coordination in a manner that may not be possible during normal conversation. Therefore, in my final attempt I forced coordination of the two tasks by preventing conceptualisation from being completed prior to comprehension beginning; this was achieved by reducing the production display time from 1500ms to 700ms.

In my final experiment, I did observe coordination between production and comprehension tasks. The shorter production display time prevented participants from forming a full conceptual plan prior to speech comprehension beginning. This was evident in the increased disruption to prediction, with a much smaller difference between specific and general verbs noted here than in Experiments 1-5. I also noted that prediction, although reduced, is still present with participants being able to use specific verb information online to anticipate the correct referent. I proposed that prediction is a capacity saving process that allows resources to be expended elsewhere; in this case on speech production. This is evident in the increased onsets and number of objects mentioned in participants' speech after hearing a specific verb.

In the following sections I discuss the results and examine the importance of predicting in language, I explore how the results are and are not representative of normal language, and assess the future applications of this new dual-task paradigm.

Table 34: Overview of main results

Experiment	Verb Type Effect	Complexity Effect	Speech onsets on fixations to target	Recall on fixations to target	Dual-task vs. Single-task	Production Only vs. Production & Comp
Exp 1 Visual World (VW) only	Faster latencies to specific verbs, $p < .001$	-	-	-	-	-
Exp 2 VW and Digit Task (Phonological WM)	Faster latencies to specific verbs, $p < .001$	-	-	-	No effect of digit task $p = .74$: Separation of phonological and semantic WM	-
Exp 3 Simple/ Complex sentences	Faster latencies to specific verbs, $p < .001$	Faster latencies to simple sentences, $p < .001$	-	-	-	-
Exp 4 Comp with Production (1)	Faster latencies to specific verbs, $p < .001$	Faster latencies to simple sentences, $p < .001$	No effect of speech onsets	No effect of recall	Prediction effect significantly reduced in dual-task $p < .001$	Faster speech onset latencies in PC trials $p < .001$
Exp 5 Comp with Production (2: Cued preview)	Faster latencies to specific verbs, $p < .001$	Similar latencies across simple and complex sentences, $p = .50$	No effect of speech onsets	No effect of recall	Prediction effect is reduced in dual-task relative to single-task, and the difference is approaching significance $p = .09$	Faster speech onset latencies in PC trials $p < .001$, and more items recalled in PC trials
Exp 6 Comp with Production (3: 700ms production display and cued preview)	Faster latencies to specific verbs, $p < .001$	Similar latencies across simple and complex sentences, $p = .78$	Faster onsets associated with specific verbs (in complex sentences)	Higher recall of items associated with specific verbs	Prediction effect significantly reduced in dual-task $p < .01$	Faster speech onset latencies in PC trials $p < .001$

8.2 Discussion of the results.

8.2.1 Online prediction as a capacity saving process.

The ability to anticipate language is a vital part of language processing. It is evident from our own experiences with language that we are able to predict what a person might say next; for example, the feeling that you able to finish another's sentences or the ability to provide them with the appropriate word when they are struggling (Kamide, 2008). In support of Altmann and Kamide (1999), I have demonstrated the ability to anticipate language. Through using information gleaned at the verb, participants can predict upcoming referents based on thematic relationships. That is, in the context of cake, ball, glasses and handbag, participants can use the verb "eat" to predict the future noun "cake". Here, the ability to predict during comprehension allowed for faster processing speeds.

In Experiments 2-6 I illustrate the resilience of online prediction in capacity demanding situations. That is, when having to share capacity between speech comprehension and a secondary task (or increasing processing demands via syntactic complexity) the prediction effect always remains significant, with participants using specific verbs online to predict upcoming referents. These findings suggest that prediction is an important part of language processing that interlocutors engage in, even when processing demands are high.

Within this thesis I begin to illustrate why prediction is so important. In Experiment 6 I have demonstrated the value of prediction as a capacity saving mechanism. That is, when coordinating speech production and speech comprehension, participants have engaged in prediction processes which in turn have allowed for better coordination between tasks. Inthis experiment, I observed that speech onsets and recall are greater after hearing specific verbs (in complex sentences). These findings indicate that predicting online allows for a speedier interpretation of the verb phrase, enables the future referent to be chosen, and now that comprehension processes are complete resources can be expended elsewhere; in this case on speech production processes. Therefore, participants are benefiting from the ability to predict online. This capacity saving nature of anticipating language could explain why the prediction effect remains so strong across all experiments, even when demands are extremely high. Participants are in fact in a better position to coordinate tasks if they engage in prediction processes. Alike to processes such as interactive alignment, prediction appears to be a mechanism that helps to both speed up comprehension and to preserve cognitive resources (Kamide, 2008).

This capacity saving nature of prediction is not necessarily restricted to verbs. For instance, in the introduction I discussed how tense information can be used to predict online. Altmann & Kamide (2007) demonstrated how participants could not only process verbs online but also base their eye-movements on the statement’s tense: “will drink” generated looks to a full drinking vessel, whilst “has drunk” generated looks to an empty glass. Kamide, Altmann and Haywood (2003) demonstrated that world knowledge regarding the grammatical agent of a sentence can drive predictions: when choosing between the referents “motorbike”, “carousel” and “sweet jar”, hearing “the girl will ride” generates fixations to the carousel, whilst hearing “the man will ride” instigates looks to the motorbike. These types of prediction could similarly be implemented to help simplify the comprehension of sentences and save on cognitive resources (Kamide, 2008).

Further predictive clues that could help to make comprehension faster and easier are disfluencies. For example, a study by Arnold and colleagues (2004) examined how the incidence of disfluencies affects processing. They demonstrated how disfluencies are processed incrementally and used to aid comprehension processes; for example when hearing a disfluency participants were biased towards looking to an object that had not yet been mentioned, thereby speeding up sentence comprehension. Such discreet clues in speech, when used correctly, allow participants to save time and resources in comprehension.

I believe that the experiments documented here demonstrate that anticipating language is a beneficial process that allows for effective and efficient language comprehension. When performed successfully, making predictions about future themes and referents is a capacity-saving process that allows resources to be conserved and used elsewhere. It appears (from these experiments) that prediction is a robust effect. It is not an optional extra that one engages in only if time or resources allow.

Of course, prediction does not always lead to success. Sometimes one can predict wrong and have to expend more resources back-tracking and correcting errors (Kamide, 2008). However, considering the overall capacity demands of speech production and comprehension, you may gain more than you lose by predicting upcoming information.

8.2.2 Processing demands and strategy use.

Another interesting finding from this experiment is the strategies employed by participants as processing demands increase. For instance, the separation of production and comprehension processes in Experiments 4 and 5 and the comparable processing of simple and complex sentence structures in Experiments 5 and 6. The development of such strategies not only demonstrates the ingenuity of participants but also shows how the processing demands associated with each experiment continually increases (this is also evident in overall lengthening of eye-movement latencies through from Experiment 1 to 6).

Although it is not possible to say whether such strategies are used in normal conversation or just employed in this highly demanding experimental set-up, there are obvious advantages to the strategies employed here. For example, dealing with simple and complex sentences in an equivalent manner by listening only for the vital verb phrase is a remarkable way of reducing processing demands in complex sentence structures. It also has real world applications; when comprehending a complex speech or attempting to multitask whilst listening we often seek only the vital information necessary to piece together a general interpretation of what is being said (Treisman, 1969). Therefore, this strategy *could* be typically employed in highly demanding situations.

Moreover, I also observe how participants use anticipatory techniques to help improve production performance (as evidence in faster speech onset latencies after hearing specific verbs in complex sentences in Experiment 6, and faster onsets in the production and comprehension trials compared to production only trials). This again could be a strategy employed in normal, everyday conversation. Coordinating production and comprehension is a convoluted process that requires organising multiple discrete processing stages, therefore engaging in actions such as prediction may be essential to conserve enough capacity to complete both tasks successfully. However, normal conversation benefits from several additional elements that help to make dialogue easier (see section 8.2.3) therefore where prediction may be vital in this experimental setting it may be less essential in normal dialogue.

One factor that I cannot evaluate here is what stage of speech production participants have reached before they are cued to speak. That is, have they only completed the conceptualisation stage, or have they started (or even completed) the formulation stages of speech production? As discussed in the introduction of Chapter 5 (section 5.1.2) not all of an utterance needs to be planned prior to articulation, although participants need an overall concept of what they want to say, they can build some of the sentence as they are speaking. Here, it is not possible to pinpoint how much of their speech has been formulated prior to production; participants could have planned only the first word, as suggested by Griffin and Bock (2000), or perhaps the first clausal unit, as found in Allum & Wheeldon (2007). It is also possible that participants have planned even further, using the gaps in processing time during comprehension to formulate a more advanced plan.

One final strategy employed here is the prioritisation of speech production processes. That is, participants actively engage in planning processes during the comprehension task rather than wait until comprehension processes are complete. This is clearly evident when comparing speed of speech onset across the single- and dual-task conditions. This could be a *voluntary* strategy choice in order to ensure that production is always performed successfully. A similar claim has been offered in the dual-task literature where participants emphasise one task over the other (Roelofs, 2008; Ruthruff, Pashler, & Hazeltine, 2003; Ruthruff, Pashler & Klaassen, 2001; Tombu & Joliceour, 2004). Prioritisation may occur here due to the nature of the task itself; unsuccessful planning processes is immediately evident to participants as they struggle to think of a sentence, whilst unsuccessful comprehension is less explicit. In normal dialogue successful speech production depends on successful comprehension of others' speech. We can only produce an appropriate response if we have correctly understood what has come before. Whilst in this experiment, however, comprehension and production are not linked in this fashion. This therefore makes comprehension processes less significant, and could compel participants to prioritise their own speech production. I come back to this idea in section 8.2.3.

8.2.3 How does this differ from natural dialogue?

An important topic that needs raising here is how coordination in our paradigm could differ from that in normal conversation. In other words, what are the limitations of using the experimental paradigm documented here?

The most striking difference between our paradigm here and dialogue is, of course, that there is only one speaker. Participants are not directly interacting with another interlocutor, they are instead conversing alongside a recorded voice. This lack of direct interaction highlights some other major disparities between our experiments and dialogue. For example, the two strands of speech (one created by the participant, one presented via a recording) are not interlinked; there is no continuation of a theme between the two tasks. This is obviously a striking difference from normal dialogue.

Conversation is thought to be governed by certain rules which have been labelled Grice's maxims, to recap here, the maxims include quality, quantity, manner, and of most importance here, relevance (Grice, 1975). The maxim of relevance dictates that you must make your contribution relevant and appropriate to what has gone before, a bad conversationalist would be one that started talking about something completely different (Harley, 2008). In this experiment, however, I am asking participants to do just this- they are given object pictures that are in no way related to the sentence they hear and must create a sentence that bears no relevance to what has gone before. This kind of structure also means that there is no common ground or shared knowledge between participants. Shared knowledge allows interlocutors to model one another's 'internal state' (Pickering & Garrod, 2004), that is, understanding what the other person does and does not know. This then helps speakers to gear their speech accordingly and include all the necessary amount of information for the listener to understand. Although the demands placed on the listener may be reduced as no record of speech is necessary, the processes that help to aid dialogue are also missing.

Another crucial process that is limited within our design is interactive alignment (Branigan et al., 2000; Pickering & Garrod, 2004). As broached upon in the introduction (section 1.5.1), interactive alignment is argued to be an automatic mechanism that helps to simplify conversational exchange. Representations in speech are thought to be shared between interlocutors, for example interlocutors might share lexical representations by referring to objects using the same word, or share structural representations by employing the same syntactic frame (Branigan et al., 2000; Brennan & Clark, 1996; Brennan, Pickering & Cleland, 2000; Garrod & Anderson, 1987). According to Pickering and Garrod (2004), these shared representations subsequently become activated and therefore interlocutors are more likely to continue using these particular words and syntactic structures. They also suggest that this type of alignment is thought to take place automatically and help to reduce the demands enforced by dialogue. It is further thought that these alignment processes are not possible in monologue as the speaker has no one to align his representations with (Pickering & Garrod, 2004). In this design, although there could be some activation of specific object names and syntactic structures, representations cannot be as easily shared across speakers. Therefore, the benefits associated with aligning one's speech with another's is reduced.

Alignment *between* production and comprehension processes is also absent in this paradigm, or at the very least extremely diminished. As described in the introduction, Pickering and Garrod's model of interactive alignment incorporates a bi-directional channel of alignment, with alignment passing from comprehension to production, and from production to comprehension. They argue that this alignment occurs due to priming and the channels to alignment are direct and automatic; activation of specific representations in one individual automatically leads to activation of the same representations in the other individual. This premise goes along way to explaining how the demand for cognitive resources can be reduced in dialogue; when structures and representations have already been accessed this subsequently make them more accessible for future use. The lack of this type of alignment in my paradigm entails that more cognitive resources are required here in comparison to normal dialogue.

Another difference between our design and naturally occurring speech relates to turn-taking. Turn-taking refers to the flow of conversation from one speaker to the next, for example, person A speaks whilst person B listens, then person B takes their turn to speak as A listens, and the exchange continues (Sacks, 1974). A study by Ervin-Tripp in 1979 demonstrated that people are typically able to start talking within a few tenths of a second of when their speech partner finishes. However, we notice in the studies documented here a much longer latency between turns (>1000ms). I believe the reason for this once again lies in the style of interaction. In normal conversation we have many cues as when our speech partner will finish talking, including intonation patterns, eye-contact and gestures (Cutler & Pearson, 1986; Duncan, 1972; Kendon, 1967). All these cues are missing when interacting with a recorded voice, and thus participants are left in the dark as to when it will be their turn. Further to this, I have actively tried to prevent participants from being able to guess the upcoming structures they hear by including different levels of syntactic complexity, different types of sentences and different verb positions. Therefore, the cues that are available for anticipating turn-taking are severely reduced in this paradigm. This is not to mention, that the paradigm is non-interactive, therefore, the participants may not have been motivated to tightly coordinate their speech onset with the end of the auditory sentence.

This study clearly differs from 'real' dialogue in several ways; most noticeably there is only one speaker, but other differences include a lack of shared knowledge and common ground, an artificial flow of information, and an inability to engage in interactive alignment

Additionally, Pickering and Garrod (2007) have suggested that language production can be used to make predictions in comprehension. They suggest that prediction uses the language production system in order to help with comprehension processes. Evidence offered to support this claim includes the activation of the articulatory system in comprehension, whereby muscles in the tongue and lips have been found to become active during comprehension tasks. Prediction together with imitation (a part of interactive alignment) simplifies communication processes, relieving the pressure on the comprehension system. However, there are some problems with this theory. For example, it has been well documented that during language development the ability to comprehend language develops faster than production. Therefore, it is not clear how during this time children would use production processes to aid comprehension. In regards to our study, comprehension does not appear to benefit from production processes, conversely, when production processes must be employed in a secondary task there is a delay to comprehension and a reduction in prediction strength. However, here, production and comprehension are not linked; what is comprehended is unrelated to the objects that must be spoken about. It is possible that we would observe a different result during dialogue. This could be an interesting expansion of this dual-task research.

Gordon et al. (2002) suggest that language has special characteristics which allow for information to be retrieved efficiently from memory and for mental representations to be generated rapidly. They also claim that these special characteristics reduce the amount of interference that is experienced within memory. Therefore, in naturally occurring conversation, when these special characteristics are in operation, there may be a huge reduction in interference between production and comprehension.

However, regardless of these additional processes that aid coordination, participants still perform exceptionally well in these dual-task studies. Participants' speech remains fast and fluent and comprehension processes are still achieved online. Indeed, coordination of these two types of tasks is a highly practiced ability, and, therefore, even when some of the capacity saving components of dialogue are missing (e.g., shared knowledge and interactive alignment) participants can perform effectively and efficiently in both tasks. Whilst processing demands may be higher here than in naturally occurring conversation, I have demonstrated how flexible the coordination process can be and how adaptable participants are to coordinating speech tasks. Individuals are able to make use a wide variety of different cues and strategies to produce and comprehend speech with speed and accuracy, even within a rather un-natural experimental setting.

There is one final discrepancy between this design and real-world dialogue which I would like to highlight here might go some way to explaining why participants are performing so well under such highly demanding circumstances. This discrepancy lies in the paradigm itself. The foundation of the visual world paradigm is the link between eye-movements and language: the verbal information presented affects where participants allocate their attention. It has thought that participants direct their attention to visual objects that overlap with the auditory input not just because they wish to identify the objects but because looking at these objects helps them to process information about that particular object (Huettig et al., 2011). Therefore, there are obvious advantages to processing language alongside a visual display. However, it is clearly not the case that an interlocutor will *always* talk about objects that are nearby or items that have been mentioned in prior utterances. In these situations, the type of prediction we have demonstrated here will not be as strong; when hearing “The boy will eat”, it is not immediately possible to know that the upcoming referent is a cake. And so, whilst the domain of references will be restricted to edible items, the target itself may not be as easy to hit. Furthermore, the cued preview section of this experiment will have also impacted on their processing, whereby objects were highlighted and located prior to sentence comprehension. This forms another distinction from natural dialogue which will have also aided their language processing. It would seem, therefore, that the visual world paradigm in these experiments may go someway towards making comprehension easier than it might otherwise be.

8.2.4 How does this relate to natural dialogue?

This section examines how our tasks relate to everyday dialogue and how I have tried to mirror the demands involved in coordinating production and comprehension. Firstly, speech production must be tightly coordinated with comprehension processes; during conversation we rarely see gaps of silence where interlocutors are thinking of what to say (Ervin-Tripp, 1979). Therefore, I needed to create a situation where planning and comprehension overlap in time, allowing participants to speak almost immediately after they are cued. In Experiment 6 (where a couple of design flaws had been ironed out) I believe I have managed to create such a situation. My results suggest that planning can no longer be performed prior to comprehension beginning, and participants subsequently make use of strategies, such as online prediction, to save capacity and help plan speech. The set of results I have documented in our final experiment describes an online coordination pattern that has not been seen before. When studying participants’ eye-movement latencies we see a significant interaction between speech onsets and verb type in our complex data set. This indicates that there is something about the specific verb condition which allows for more planning of speech so that utterances can be produced more quickly. I have proposed that predicting online at the verb may serve to preserve capacity and allow resources to be expended elsewhere; in this case on speech production processes. There was also a significant interaction between recall and verb type, with more objects recalled in the specific verb condition compared to the general verb condition. This could further indicate that prediction is a capacity saving process whereby prediction allows participants to concentrate on planning their own utterance and, with more time available to plan, they can incorporate more objects into their speech.

I have also demonstrated how the storing and processing of information in working memory impacts upon comprehension. Within conversation it is vital that we keep a record of what has gone before, who has said what, and hold onto our own ideas whilst we listen to others. In Experiments 4 and 5, although we see less overlap between the two tasks, what we do see is the successful storage of ideas as comprehension processes are complete. Further to this, we see that a high working memory load does slow down comprehension (as evident in longer eye-movement latencies and a disruption to prediction processes), however, comprehension processes are still occurring online as the sentence unfolds. This demonstrates how resilient comprehension processes are and how skilful we are as listeners; indeed, it is a rare situation where we can't quickly understand our native language.

Another advantage to this work is the link between language and the visual world. The visual world paradigm allows us to examine how the world around us may aid comprehension processes (Huettig et al., 2011). As reviewed in the introduction (sections 1.2 and 1.3) multiple sources of linguistic information can be used in conjunction with the visual world to help speed up comprehension processes. For example, we can compare incoming phonology to pinpoint what object is being referred to (Allopenna et al., 1998), or use semantic information to activate relevant associations in language (Huettig & Altmann, 2005). Here, I further demonstrate how the visual world along with world knowledge can be used to preserve cognitive resources and help individuals deal with multiple tasks at once. This relationship between language and eye-movements demonstrates how multiple streams of information interact to guide our behaviour (Huettig et al., 2011). The visual world paradigm is a wonderful technique to examine this, and as I have illustrated here, the ways in which it can be used are endless.

Whilst the generalisability of the results I have documented here are in some ways limited (as highlighted in section 8.2.3 above), they do speak to natural dialogue in several ways. For example, there are situations in which planning of one's speech is coordinated with comprehension of another's, however the two strands are unrelated in terms of content. Imagine, for instance, a case where you are exchanging anecdotes about your most embarrassing moment with a friend. In this situation, you will be comprehending your speech partner's story, yet planning how to convey your own anecdote immediately after they have finished; these two anecdotes are likely to involve very different themes. Similarly, when listening to a speech partner you may be eager to change the topic of conversation; again the planning of your subsequent speech could be very different and not necessarily related to the sentence you are currently comprehending.

Another finding which may mirror (some) cases of normal dialogue is listening in for vital information only. The results of Experiments 5 and 6 demonstrate how listeners can predict within simple and complex sentences comparably; to do this they listen out for the vital part of the sentence (in this case the verb phrase). Cases where this strategy is used can also be seen in normal dialogue. Take, for example, the case of listening to questions at the end of a presentation. When you are presented with a long, complicated question it can be the case that you focus in on one part (the part that you are most happy to answer). Similarly, when listening to speech to gain one specific fact comprehension may wane after this piece of knowledge is acquired. Or perhaps you are listening to set of instructions, it is likely that you will focus in on the instruction which describes the next step you need to take.

Experiments 5 and 6 illustrate how this comparable processing of simple and complex structure occurs when processing demands in the dual-task are particularly high. This too may resonate with normal language processing. For instance, when attempting to coordinate two tasks which interfere with one another (e.g., typing and listening; reading and listening; listening whilst completing a complex car manoeuvre, and so on) cognitive resources can be reduced by only partially comprehending spoken utterances. In such cases, you may listen enough to gain a general ‘gist’ of what is being spoken about, but do not comprehend enough to repeat the sentence verbatim.

One final similarity between our findings and normal conversation is the prioritisation of speech production. I mentioned above that our particular design may encourage the prioritisation of production processes; however, this prioritisation may also arise in natural dialogue. Speaking intuitively, as interlocutors we often endeavour to make our voices heard. When holding onto an important statement or detail it is not uncommon to pay less attention to others’ speech. Therefore, although prioritisation in this manner may not *always* be present in conversation, it may be a strategy that does appear naturally in speech.

8.2.5 The paradigm and potential applications.

This research has highlighted a new way of implementing the visual world paradigm, and has similarly demonstrated that it is possible to examine both speech production and comprehension together in an experimental setting. In this final section I discuss the future use of this paradigm and the potential applications of this research.

First of all, measuring prediction has proved to be a valuable way of tracking comprehension processes; in order to predict upcoming themes and referents participants must be able to process and use the verb information online. The visual world paradigm does not necessitate any disruption to speech input and has provided a time-sensitive method of examining comprehension processes. More specifically, however, examining prediction has allowed me to monitor online comprehension through numerous types of task manipulation and varying levels of task demand. Studying the strength of predictions made has enabled us to continuously examine how comprehension is being affected by the presence of a secondary task and has thereby created an extremely effective dual-task technique. Such a technique could equally be used to examine predictions using tense information, semantic associations, prediction in other languages, or prediction of non-native speakers. The combination of tasks could equally be extended to examine comprehension alongside logical or spatial reasoning tasks, various working memory spans... the list is endless.

A further way in which this paradigm might be extended could be to examine speech production and comprehension in real-life dialogue. That is, with a careful and intricate design it could be possible to track comprehension through eye-tracking as participants converse with a speech partner. This would allow researchers to examine the coordination process in a more realistic setting where there is shared knowledge, turn-taking, and an opportunity for interactive alignment.

Further to this, the paradigm could be adapted for use alongside imaging techniques. For example, previous research such as that by Rommers, Meyer, Praamstra and Huettig (2012) have examined the role of shape representations in prediction through measuring the N400 effect using an ERP technique. The N400 is elicited in response to semantically anomalous information, for example, when encountering a word that is inconsistent with the preceding context (e.g., "I take coffee with cream and dog", Kutas & Hillyard, 1980). This component was employed by Rommers et al. to examine the strength of N400 activation when participants encountered different sentences where online predictions could be made. They discovered that after a predictable word ERPs remained relatively flat, however, when encountering semantically unrelated words an N400 effect was elicited. This kind of technique could be similarly used to examine whether prediction and planning can be processed in parallel; by measuring the strength of the N400 effect in the presence and absence of a secondary planning task it would be possible to observe how planning impacts upon prediction and a illustrate a more precise time-course of coordination.

Additionally, the long term applications of this research could have broader implications. For example, it could be possible to examine what groups of individuals are good at coordinating production and comprehension and which groups find it more challenging. With an in depth individual differences project you could compare performance from a wide range of individuals with varying abilities. Once more is known about the coordination process it may be possible to develop strategies that help minimize the costs involved, and specifically target those who struggle.

One final issue that I would like to raise here concerns the *type* of prediction I have studied. My question is whether the results I have demonstrated here might be generalizable to other types of prediction, or whether there is something special about predicting at the verb? Indeed, prediction at the verb (as documented throughout this thesis) is a very specific type of prediction which occurs due to knowledge of thematic relationships; i.e., when acquiring the English language you learn that the verb 'eat' (typically) relates to an edible object. It is this knowledge that drives eye-movements to the cake. However, would the same eye-pattern that was documented here also be driven by predictions about world knowledge? For example, in Kamide, Altmann & Haywood (2003) the researchers illustrated that eye-movements can be driven not only by verb information but by the grammatical agent. For instance, participants were presented with a visual scene containing two agents; a man and a girl, and multiple objects including a motorbike, a carousel, a sweet jar and a beer. Alongside this scene participants could hear (2a), (2b), (2c), or (2d).

(2a) The man will ride the motorbike.

(2b) The girl will ride the carousel.

(2c) The man will taste the beer.

(2d) The girl will taste the sweets.

Kamide et al. proposed that it is world knowledge about the grammatical agent and the verb that is used to predict the upcoming noun (with increased looks to motorbike after hearing the first part of (2a), and increased saccades to carousel when hearing (2b)).

These types of prediction require more knowledge of the world than simple verb predictions. Would this therefore mean that more working memory capacity is required in order to make these predictions? If so, we would surely see a different pattern of eye-movements to those documented in Experiments 2-6. It could be the case that this form of prediction does not act as a capacity saving mechanism; it may be that in these cases we in fact lose more than we gain by engaging in prediction processes. This would be a very interesting avenue of research to pursue in the future.

To conclude, my research demonstrates that speech production and speech comprehension can be effectively coordinated in an experimental setting. I have further demonstrated that cognitive resources must be shared between the two tasks when they are presented within the same time-frame, however, participants actively engage in strategies that help to reduce processing demands including separation of production and comprehension where necessary and using online prediction to preserve capacity. This is the first research of its kind to specifically examine the coordination process and I have demonstrated here that it is not only possible to assess production and comprehension processes together but that it is also a fruitful venture. There is still a long way to go until we fully understand how this type of coordination works.

References

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*(3), 299–319.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419–439.

Allum, P. H., & Wheeldon, L. R. (2007). Planning scope in spoken sentence production: The role of grammatical units. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 791–810.

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

Altmann, G. T. M. & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57*, 502–518.

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new disfluency and reference resolution. *Psychological Science*, *15*(9), 578–582.

Baars, B. J. (2002). The conscious access hypothesis: origins and recent evidence. *Trends in cognitive sciences*, *6*(1), 47–52.

Baayen, R. H., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language, *59,* 390–412.

Baddeley, A. D. (1978). The trouble with levels: A reexamination of Craik and Lockhart's framework for memory research. *Psychological Review*, *85*(3), 139–252.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*(10), 829–839.

Baddeley, A., Eldridge, M., & Lewis, V. (1981). The role of subvocalisation in reading. *The Quarterly Journal of Experimental Psychology*, *33*(4), 439–454.

Baddeley, A. D, & Hitch, G. J. (1993). Working memory. In G.A. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–89). New York: Academic Press.

Baddeley, A. D., Vallar, G. & Wilson, B. A. (1987). Sentence comprehension and phonological memory: Some neuropsychological evidence. In M. Coltheart (Ed.), Attention and Performance XII: The psychology of reading. Hove, UK: Lawrence Erlbaum Associates Ltd., pp. 509–529.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency and control in social cognition. In R. S. Wyer & T. K. Srull (Eds.) Handbook of social cognition (Vol 1, pp. 1–40). Hillside, NJ: Lawrence Erlbaum Associates Inc.

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

Bates, D. M. (2005). Fitting linear mixed models in R. R News, 5, 27–30.

Bates, D. M., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0.

Baum, S. R. (1989). On-line sensitivity to local and long-distance syntactic dependencies in Broca's aphasia. *Brain and Language, 37,* 327–338.

Bavelas, J. B., Black, A., Lemery, C. R., & Mullett, J. (1986). I show how you feel": Motor mimicry as a communicative act. *Journal of Personality and Social Psychology*, *50*(2), 322–329.

Bell, L., Gustafson, J., & Heldner, M. (2003, August). Prosodic adaptation in human-computer interaction. In *Proceedings of ICPhS* (Vol. 3, pp. 833–836).

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive psychology*, *18*(3), 355–387.

Brennan, S. E. (1991). Conversation with and through computers. *User Modeling and User-Adapted Interaction*, *1*(1), 67–86.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482–1493.

Branigan, H. P., Catchpole, C. M., & Pickering, M. J. (2011). What makes dialogues easy to understand? *Language and Cognitive Processes*, *26*(10), 1667–1686.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, *75*(2), B13–B25.

Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon Press.

Broadbent, D. E. (1982) Task combination and selective intake of information. *Acta Psychologica, 50*, 253–290.

Brysbaert, M. (2007). The language-as-fixed-effect-fallacy: some simple SPSS solutions to a complex problem. *London: Royal Holloway, University of London*.

Bundesen, C. (1990). A theory of visual attention. *Psychological review*, *97*(4), 523–547.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multi-model inference: a practical information-theoretic approach*. Springer Verlag.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference understanding AIC and BIC in model selection. *Sociological methods & research*, *33*(2), 261–304.

Butterworth, B. (1992). Disorders of phonological encoding. *Cognition, 42*, 261–286.

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

Caramazza, A., & Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: evidence from the tip-of-the tongue phenomenon. *Cognition*, *64*(3), 309–343.

Carpenter, P. A., Miyake, A., & Just, M. A. (1994). Working memory constraints in comprehension: Evidence from individual differences, aphasia, and aging. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 1075–1122). New York: Academic.

Chambers, C. G., & Juan, V. S. (2008). Perception and presupposition in real-time language comprehension: Insights from anticipatory processing. *Cognition*, *108*(1), 26–50.

Chartrand, T. L. & Bargh, J. A. (1999) The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology,* 76, 893–910.

Cheng, X., Schafer, G., & Akyürek, E. G. (2010). Name agreement in picture naming: an ERP study. *International Journal of Psychophysiology*, *76*(3), 130–141.

Clark, H. H. (1996). *Using language* (Vol. 4). Cambridge: Cambridge University Press.

Cleland, A. A., Gaskell, M. G., Quinlan, P. T., & Tamminen, J. (2006). Frequency effects in spoken and visual word recognition: Evidence from dual-task methodologies. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 104–119.

Cook, A. E., & Meyer, A. S. (2008). Capacity demands of phoneme selection in word production: New evidence from dual-task experiments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 886−899.

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

Corrigan, R. (1986). The internal structure of English transitive sentences. *Memory & cognition*, *14*(5), 420–431.

Costa, A., & Caramazza, A. (2002). The production of noun phrases in English and Spanish: Implications for the scope of phonological encoding in speech production. *Journal of Memory and Language*, *46*(1), 178–198.

Costa, A., Strijkers, K., Martin, C., & Thierry, G. (2009). The time course of word retrieval revealed by event-related brain potentials during overt speech. *Proceedings of the National Academy of Sciences*, *106*(50), 21442–21446.

Cutler, A., & Pearson, M. (1986). On the analysis of prosodic turn-taking cues. *Intonation in discourse*, 139–156.

Damian, M. F., & Dumay, N. (2007). Time pressure and phonological advance planning in spoken production. *Journal of Memory and Language*, *57*(2), 195–209.

Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450–466.

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of memory and language*, *27*(2), 124–142.

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of personality and social psychology*, *23*(2), 283–292.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current directions in psychological science*, *11*(1), 19–23.

Ervin-Tripp, S. (1979). Children's verbal turn-taking. *Developmental pragmatics*, 391–414.

Estevez, A. & Calvo, M. G. (2000). Working Memory Capacity and Time Course of Predictive Inferences. *Memory, 8,* 51–61.

Fallon, M., Peele, J. E. & Wingfield, A. (2006). Spoken sentence processing in young and older adults modulated by task demands: Evidence from self-paced listening. *Journal of Gerontology, 61B*, 000–000.

Fernald, A., & Hurtado, N. (2006). Names in frames: Infants interpret words in sentence frames faster than words in isolation. *Developmental science*, *9*(3), F33–F40.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, *30*(2), 210–233.

Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 1187−1199.

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*(1), 57–84.

Fischer, R., Miller, J. & Schubert, T. (2007). Evidence for parallel semantic memory retrieval in dual tasks. *Memory & Cognition, 35,* 1685–1699.

Frauenfelder, U. H., & Peeters, G. (1998). Simulating the time course of spoken word recognition: An analysis of lexical competition in TRACE. *Localist connectionist approaches to human cognition*, 101–146.

Frauenfelder, U. H., & Tyler, L. K. (1987). The process of spoken word recognition: An introduction. *Cognition*, *25*(1), 1–20.

Frazier, L. (1985). Syntactic complexity. In D. Dowty, L. Karttunen, & A. Zwicky (Eds), *Natural language processing: Psychological, computational and theoretical perspectives* (pp. 129–189). Cambridge, UK: Cambridge University Press.

Frazier, L. (1987). Syntactic processing: evidence from Dutch. *Natural Language & Linguistic Theory*, *5*(4), 519–559.

Friedrich, F. J., Martin, R., & Kemper, S. J. (1985). Consequences of a phonological coding deficit on sentence processing. *Cognitive Neuropsychology*, *2*(4), 385–412.

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language Production. Vol. 1: Speech and Talk* (pp. 177–220). London: Academic Press.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181–218.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in cognitive sciences*, *8*(1), 8–11.

Garrod, S., & Pickering, M. J. (2007). Automaticity in language production in monologue and dialogue. In A. S. Meyer, L. R Wheeldon, & A. Krott (Eds.) (2007), *Automaticity and Control in Language Processing* (pp. 1–21). New York: Psychology Press.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Lawrence Erlbaum Associates, Inc.

Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, *14*(3), 225–248.

Grice, H. P. (1975). Logic and conversation. 41-58.

Goldrick, M., & Rapp, B. (2002). A restricted interaction account (RIA) of spoken word production: The best of both worlds. *Aphasiology*, *16*(1-2), 20−55.

Goldman-Eisler, F. (1967). Sequential temporal patterns and cognitive processes in speech. *Language and Speech*, *10*(2), 122–132.

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science*, *13*(5), 425–430.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological science*, *11*(4), 274–279.

Haarman, H. J., & Kolk, H. H. J. (1994). On-line sensitivity to subject-verb agreement violations in Broca's aphasics: The role of syntactic complexity and time. *Brain and Language, 46,* 493–516.

Hanten, G., & Martin, R. C. (2000). Contributions of phonological and semantic short-term memory to sentence processing: Evidence from two cases of closed head injury in children. *Journal of Memory and Language*, *43*(2), 335–361.

Harley, T. A. (1993). Phonological activation of semantic competitors during lexical access n speech production. *Language and Cognitive Processes, 8,* 291–309.

Harley, T. A. (2008). The psychology of language. *Psychology Press, Hove.*

Henderson, A., Goldman-Eisler, F., & Skarbek, A. (1966). Sequential temporal patterns in spontaneous speech. *Language and Speech*, *9*(4), 207–216.

Hitch, G. J. (1980). Developing the concept of working memory. In G. Claxton (Ed.), *Cognitive psychology: New directions* (pp. 154–196). London: Routledge & Kegan Paul.

Hohlfeld, A., Sangaks, J. & Sommer, W. (2004). Effects of Additional Tasks on Language Perception: An Event-Related Brain Potential Investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1012–1025.

Holmes, V. M. & Forster, K. I. (1970). Detection of extraneous signals during sentence recognition. *Perception & Psychophysics, 7,* 297–301.

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm. *Cognition, 96,* B23-B32.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta psychologica*, *137*(2), 151–171.

Humphreys, G. W., Riddoch, M. J., & Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology, 5,* 67–104.

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Frontiers in psychology*, *2*.

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

Ivanova, I., Pickering, M. J., Branigan, H. P., McLean, J. F., & Costa, A. (2012). The comprehension of anomalous sentences: Evidence from structural priming. *Cognition*, *122*(2), 193–209.

Jescheniak, J., & Schriefers, H. (1998). Discrete serial versus cascaded processing in lexical access in speech production: Further evidence from the coactivation of near-synonyms. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24,* 1256–1274.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological review*, *99*, 122–149.

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass*, *2*(4), 647–670.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*, 133–59.

Kemper, S., Schmalzried, R., Herman, R., Leedahl, S., & Mohankumar, D. (2009). The effects of aging and dual task demands on language production. *Aging, Neuropsychology, and Cognition*, *16*(3), 241–259.

Kendon, A. (1967). Some functions of gaze direction in social interaction. Acta Psychologica, 26, 1–47.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior research methods*, *42*(3), 627–633.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*(5), 580–602.

Klapp, S. T, Marshburn, E. A., & Lester, P. T. (1983). Short-term memory does not involve the "working memory" of information processing: The demise of a common assumption. *Journal of Experimental Psychology: General, 112*, 240–264.

Konig, C. J., Buhner, M., & Murling, G. (2005). Working memory, fluid intelligence, and attention are predictors of multitasking performance, but polychronicity and extraversion are not. *Human Performance*, *18*(3), 243–266.

Kubose, T. T., Bock, K., Dell, G. S., Garnsey, S. M., Kramer, A. F., & Mayhugh, J. (2005). The effects of speech production and speech comprehension on simulated driving performance. *Applied Cognitive Psychology*, *20*(1), 43–63.

Kuhlen, A. K., Allefeld, C., & Haynes, J. D. (2012). Content-specific coordination of listeners' to speakers' EEG during communication. *Frontiers in human neuroscience*, *6,* 1–15.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*(4), 389–433.

Laganaro, M., Morand, S., Schwitter, V., Zimmermann, C., Camen, C., & Schnider, A. (2009). Electrophysiological correlates of different anomic patterns in comparison with normal word production. *Cortex*, *45*(6), 697–707.

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

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech, 6*, 172–187.

Linderholm, T. (2002). Predictive Inference Generation as a Function of Working Memory Capacity and Causal Text Constraints. *Discourse Processes, 34,* 259–280.

Love, T., Haist, F., Nicol, J., & Swinney, D. (2006). A functional neuroimaging investigation of the roles of structural complexity and task-demand during auditory sentence processing. *Cortex*, *42*(4), 577–590.

Luce, P.A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics, 39,* 155–8.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive psychology*, *24*(1), 56–98.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological review*, *101*(4), 676–703.

MacKay, D. G. (1987). *The organization of perception and action: A theory for language and other cognitive skills.* New York: Springer-Verlag.

Marslen-Wlison, W. D. (1984). Spoken Word Recognition: A tutorial review. In H. Bouma & D.G. Bouwins (Eds.) *Attention and Performance X: Control of Language Processes* (pp. 125–150). Hove, UK: Lawrence Erlbaum Associates Ltd.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1), 71–102.

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

Marslen-Wilson, W.D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*, 29–63.

Martin, R. C. (1990). Neuropsychological evidence on the role of short-term memory in sentence processing. *Neuropsychological impairments of short-term memory*, 390.

Martin, R. C. (2005). Components of Short-Term Memory and Their Relation to Language Processing Evidence From Neuropsychology and Neuroimaging. *Current Directions in Psychological Science*, *14*(4), 204–208.

Martin, R.C. and He, T. (2004). Semantic short-term memory and its role in sentence processing: A replication. *Brain and Language, 89,* 76–82.

Martin, R. C., Miller, M., & Vu, H. (2004). Lexical-semantic retention and speech production: further evidence from normal and brain-damaged participants for a phrasal scope of planning. *Cognitive Neuropsychology*, *21*(6), 625–644.

Martin, R.C., Shelton, J.R. and Yaffe, L.S. (1994). Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language, 33,* 83–111.

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Attention, Perception, & Psychophysics*, *53*(4), 372–380.

Mattys, S. L. & Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language, 65,* 145–160.

McCann, R. S., Remington, R. W., & Van Selst, M. (2000). Automaticity in visual word processing. *Journal of Experimental Psychology: Human Perception & Performance*, *26*, 1352–1370.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology*, *18*(1), 1–86.

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, *10*(8), 363–369.

McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review, 15,* 1064–1071.

McQueen, J. M., Cutler, A., Briscoe, T., & Norris, D. (1995). Models of continuous speech recognition and the contents of the vocabulary. *Language and Cognitive Processes*, *10*(3-4), 309–331.

McQueen, J. M. (2005). Speech perception. *The handbook of cognition*, 255–275.

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

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture–word interference experiments. *Journal of Memory and Language*, *35*, 477–496.

Meyer, A.S., Wheeldon, L.R. & Krott, A. (2007). *Automaticity and Control in Language Processing.* Hove & New York: Psychology Press.

Meyer, A. S., & Van Der Meulen, F. F. (2000). Phonological priming effects on speech onset latencies and viewing times in object naming. *Psychonomic Bulletin & Review*, *7*(2), 314–319.

Miller, G. A. (1974). *The psychology of communication*. Voice of America.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2012). Adaptation in gesture: Converging hands or converging minds? *Journal of Memory and Language,* 66, 249–26.

Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology, Learning, Memory, and Cognition, 28(3)*, 555–563.

Navarrete, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. *Journal of Memory and Language*, *53*(3), 359−377.

Nicol, J., Swinney, D., Love, T. & Hald, L. (2006). The On-line Study of Sentence Comprehension: An Examination of Dual Task Paradigms. Journal of Psycholinguistic Research, 35, 215–231.

Norris, D., Cutler, A., McQueen, J. M., & Butterfield, S. (2006). Phonological and conceptual activation in speech comprehension. *Cognitive Psychology*, *53*(2), 146–193.

Otten, M. & Van Berkum, J. J. A (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Research, 1291*, 92–101.

Pallant, J. (2010). *SPSS survival manual: A step by step guide to data analysis using SPSS*. Open University Press.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116,* 220–244.

Parrill, F., & Kimbara, I. (2006). Seeing and hearing double: The influence of mimicry in speech and gesture on observers. Journal of Nonverbal Behavior, 30, 157–166.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–189.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension?. *Trends in Cognitive Sciences*, *11*(3), 105–110.

R Development Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Raaijmakers, J. G. W. (2003). A further look at the “language-as-fixed-effect fallacy’. *Canadian Journal of Experimental Psychology, 57*, 141–151.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with “the language-as-fixed-effect fallacy”: Common misconceptions and alternative solutions. *Journal of Memory and Language, 41*, 416–426.

Rabovsky, M., Alveraz, C. J., Hohlfeld, A., & Sommer, W. (2008). Is lexical access autonomous? Evidence form combining overlapping tasks with recording event-related potentials. *Brain Research, 1222,* 156–165.

Rahman, R. A., & Sommer, W. (2008). Seeing what we know and understand: how knowledge shapes perception. *Psychonomic bulletin & review*, *15*(6), 1055–1063.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review; Psychological Review*, *107*(3), 460−499.

Rizzolatti, G. & Arbib, M. A. (1998) Language within our grasp. *Trends in Neurosciences,* 21, 188–94.

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

Roelofs, A. (2008) Attention, gaze shifting and dual-task interference from phonological encoding in spoken word planning. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1580–1598.

Rommers, J., Meyer, A. S., Praamstra, P., & Huettig, F. (2013). The contents of predictions in sentence comprehension: Activation of the shape of objects before they are referred to. *Neuropsychologia*, *51*(3), 437–447.

Ruthruff, E., Pashler, H. E., & Hazeltine, E. (2003). Dual-Task Interference with equal Task Emphasis: Graded Capacity-Sharing or Central Postponement? *Perception and Psychophysics*, 65, 801–816.

Ruthruff, E., Pashler, H., & Klaassen, A. (2001). Processing bottlenecks in dual-task performance: Structural limitation or strategic postponement? *Psychonomic Bulletin & Review, 8,* 73–80.

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

Saffran, E. M., & Marin, O. S. (1975). Immediate memory for word lists and sentences in a patient with deficient auditory short-term memory. *Brain and Language*, *2*, 420–433.

Schriefers, H. (1992). Lexical access in the production of noun phrases. *Cognition*, *45*(1), 33–54.

Sedivy, J. C., Tanenhaus, M., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147.

Shapiro, L. P., Zurif, E., & Grimshaw, J. (1987). Sentence processing and the mental representation of verbs. *Cognition*, *27*(3), 219–246.

Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition*, *73*(3), 205–246.

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: Human Learning and Memory, 6,* 174–215.

Strauss, T. J., Harris, H. D., & Magnuson, J. S. (2007). A reimplementation and extension of the TRACE model of speech perception and spoken word recognition. *Behavior Research Methods, 39*, 19–30.

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., Lu, C.C., Pechmann, T., Pleh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., IIyer, G., Kohnert, K., Mehotcheva, T., Orozco-Figueroa, A., Tzeng, A., Tzeng, O., Arevalo, A. Vargha, A., Butler, A.C., Buffington, A., & Bates, E. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory & Language, 51,* 347–250.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, *29*(6), 557–580.

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.

Tombu, M., & Jolicoeur, P. (2004). Virtually no evidence for virtually perfect time-sharing. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 795–810.

Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological review*, *76*(3), 282.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33,* 285–318.

Tun, P. A., Benichov, J. & Wingfield, A. (2010). Response Latencies in Auditory Sentence Comprehension: Effects of Linguistic Versus Perceptual Challenge. *Psychology and Aging, 25,* 730–735.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505.

Vallar, G., & Baddeley, A. D. (1984). Phonological short-term store, phonological processing and sentence comprehension: A neuropsychological case study. *Cognitive Neuropsychology*, *1*(2), 121–141.

Vallar, G., & Baddeley, A. (1987). Phonological short-term store and sentence processing. *Cognitive Neuropsychology*, *4*(4), 417–438.

Van den Broek, P. (1994). Comprehension and memory of narrative texts: Inferences and coherence. In D.A. Balota, G.B. Flores d'Arcais, & K. Rayner (Eds.), Comprehension processes in reading (pp. 423–445). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Van den Broek, P., Fletcher, C. R., & Risden, K. (1993). Investigations of inferential processes in reading: A theoretical and methodological integration. *Discourse Processes*, 16, 169–180.

Vasishth, S., Suckow, K., Lewis, R. L., & Kern, S. (2010). Short-term forgetting in sentence comprehension: Crosslinguistic evidence from verb-final structures. *Language and Cognitive Processes*, *25*(4), 533–567.

Waters, G. S., & Caplan, D. (2001). Age, Working Memory, and On-Line Syntactic Processing in sentence comprehension. *Psychology and Aging, 16,* 128–164.

Waters, G. S., Caplan, D., & Yampolsky, S. (2003). On-line syntactic processing under concurrent memory load. *Psychonomic Bulletin & Review, 10,* 88–95.

Winter, B. (2013). Linear Mixed Effects Models in R. (Tutorial 2) from http://bodowinter.com/resources.html

Yudes, C., Macizo, P., & Bajo, M. T. (2012). Coordinating comprehension and production in simultaneous interpreters: Evidence from the articulatory suppression effect. *Bilingualism: Language and Cognition*, *15*(02), 329–339.

Appendix A: Sentence stimuli for Experiments 1 & 2.

The right hand column shows the target picture and three distractor pictures that appear in the visual display. (*)Marks which sentences were taken from Altmann and Kamide (1999).

Label	Sentence (Specific Verb/General Verb)	Target (and Distractors)
It1_A	The boy will eat/move the cake.*	Cake (toy car, ball, toy bear)
It2_A	The woman will drink/try the wine.*	Wine (cheese, lipstick, chair)
It3_A	The woman will bathe/touch the baby.*	Baby (crown, plant, stool)
It4_A	The boy will bounce/throw the ball.*	Ball (acorn, paper plane, stool)
It5_A	The man will climb/draw the mountain.	Mountain (deer, moon, cactus)
It6_A	The woman will fry/wash the mushrooms.*	Mushrooms (knife, jug, weighing scale)
It7_A	The woman will inject/check the child.*	Book (television, microscope, child)
It8_A	The woman will play/dust the piano.*	Piano (table, television, telephone)
It9_B	The woman will read/shut the book.*	Book (door, bag, window)
It10_B	The man will repair/see the washing* machine.	Washing machine (tiger, cloud, dog)
It11_B	The girl will ring/kick the bell.*	Bell (bricks, drum, duck)
It12_B	The man will sail/watch the boat.*	Boat (bird, sun, car)
It13_B	The man will smoke/collect the cigarette.*	Cigarette (diary, glasses, briefcase)
It14-B	The boy will grow/feed the plant.	Plant (dog, clown, hen)
It15_B	The man will wear/forget the hat.*	Hat (wallet, bread, magnifying glass)
It16_B	The boy will ride/stroke the horse.	Horse (hair, hand, scarf)
It17_C	The woman will cook/sample the chicken.	Chicken (lipstick, perfume, ice-cream).
It18_C	The man will drive/admire the car.	Car (television, watch, church)
It19_C	The woman will boil/clean the kettle.	Kettle (plate, table, ring)
It20_C	The boy will smash/pass the plate.	Plate (balloon, cushion, slipper)
It21_C	The man will chew/buy the sweet.	Sweet (robot, kite, car)
It22_C	The man will light/fetch the match.	Match (remote, pepper, milk)
It23_C	The woman will kiss/ignore the man.	Man (television, radio, cooker)
It24_C	The girl will lick/steal the lolly.	Lolly (ball, book, toybear)
It25_D	The woman will sharpen/break the pencil.	Pencil (bottle, computer, telephone)
It26_D	The woman will cut/polish the nail.	Nail (car, table, glass)
It27_D	The man will plant/smell the flower.	Flower (perfume, cheese, candle)
It28_D	The boy will taste/avoid the celery.	Celery (teacher, snake, spider)
It29_D	The man will fasten/carry the coat.	Coat (guitar, apple, rabbit)
It30_D	The girl will write/receive the letter.	Letter (parcel, trophy, medal)
It31_D	The boy will chop/spot the carrot.	Carrot (gorilla, motorbike, owl)
It32_D	The girl will wipe/sketch the desk.	Desk (tree, bra, fire).

Appendix B: Script and output from Experiment 1's Linear Mixed Effect Model

The text below describes how the linear mixed model was run and documents the raw outputs from the models.

All the text in black relates to the programming (or syntax) of the models, and blue text is the output from running the script. The main stages of analysis are highlighted with the lines of hashmarks (#)

After opening the program, we start by loading the necessary libraries:

```
###########################################################################
#Loading the appropriate libraries that contain the functions we need to use in the analysis
###########################################################################
library(plyr)
library(MASS)
library(ggplot2)
library(reshape)
library(nlme)
library(lme4)
setwd("C:/Users/Blaire/Desktop/R analyses_new")
###########################################################################
# Read in functions that we'll use -- IA_Analysis_functions.R
###########################################################################
source("IA_Analysis_functions.R")
###########################################################################
#Open the file we are working from (the raw data) and code which variables are factors in the analysis
###########################################################################
InputFilename<-"VW1.csv"
IADataRaw<-read.csv(InputFilename)
TargFixAfterV<-as.numeric(as.character(IADataRaw$Latency_verb))
PptNum<-as.factor(IADataRaw$Subject)
Item<-as.factor(IADataRaw$Item)
GeneralSpecific<-as.factor(IADataRaw$VerbType)
IAData<-
data.frame(PptNum=PptNum,Item=Item,GeneralSpecific=GeneralSpecific,TargFixAfterV=TargFixAfterV)
> # get rid of blank data
IAData<-IAData[IAData$PptNum != "                    ",]
########################################
# Examine the Means and Standard Deviations of the raw data
########################################
PptMeanSD<-NULL
PptMeanSD<-ddply(IAData,.(PptNum),function(df)mean(df$TargFixAfterV,na.rm=TRUE))
Temp<-ddply(IAData,.(PptNum),function(df)sd(df$TargFixAfterV,na.rm=TRUE))
PptMeanSD<-cbind(PptMeanSD,Temp$V1)
colnames(PptMeanSD)<-c("PptNum","MeanRT","SDRT")
##############################################
# Look at whether the data is normally distributed (in this case with a histogram)
##############################################
 truehist(IAData$TargFixAfterV)
##############################################
# Shape the data so that it reads correctly; i.e., what signifies different participants, different items and different
conditions
##############################################
```

```
EData<data.frame(PptNum=IAData$PptNum,Item=IAData$Item,GeneralSpecific=IAData$GeneralSpecific,TargFix
AfterV=IAData$TargFixAfterV)
EData$PptNum<-as.factor(EData$PptNum)
EData$Item<-as.factor(EData$Item)
EData$GeneralSpecific<-as.factor(EData$GeneralSpecific)
###########################################################################
#Remove outliers and latencies that are less than 200ms
###########################################################################
SDOutlierCutoff<-2.5
OutlierCutoff<-PptMeanSD$MeanRT + (PptMeanSD$SDRT * SDOutlierCutoff)
PptMeanSD<-cbind(PptMeanSD,OutlierCutoff)
# add ppt means/sd/outlier cutoff
IAData<-merge(IAData,PptMeanSD)
# remove values that are missing and outliers
IAData<-IAData[(!is.na(IAData$TargFixAfterV)) & (IAData$TargFixAfterV <= IAData$OutlierCutoff),]

# take out RT < 200 after V
IAData<-IAData[IAData$TargFixAfterV>200,]
MinRT<-min(IAData$TargFixAfterV,na.rm=TRUE)
MaxRT<-max(IAData$TargFixAfterV,na.rm=TRUE)
#####################################################
# Examine Standard Deviations across participants and conditions
#####################################################
PptCondSD<-cast(MData,PptNum ~ GeneralSpecific,sd)
PptCondSD
#Remove participants with large SDs here
######################################################
# Examine the different random effects structures for modelling the variance
######################################################
NumObs<-nrow(IAData)
TempGroup<-rep(NumObs,1)
IAData$TempGroup<-TempGroup

Model1<-lmer(TargFixAfterV ~ factor(GeneralSpecific) + (1 + GeneralSpecific|PptNum) +
(1|Item),data=IAData,REML=FALSE)
#This random model is modelling the slope of participants on verb type (that is whether the different participants
show different variances in the fixed effect) and the effect of items on the intercept (if participants deal with different
items differently)

Model2<-lmer(TargFixAfterV ~ factor(GeneralSpecific) + (1|PptNum) + (1|Item),data=IAData,REML=FALSE)
#Modelling the effects of participants and items on the intercept (differences in mean eye-movement responses
across both items and participants)

Model3<-lmer(TargFixAfterV ~ factor(GeneralSpecific) + (1|PptNum),data=IAData,REML=FALSE)
#Modelling just the effect of participants on the intercept

Model4<-lmer(TargFixAfterV ~ factor(GeneralSpecific) + (1|Item),data=IAData,REML=FALSE)
#Modelling just the effect of items on the intercept

Model5<-lmer(TargFixAfterV ~ factor(GeneralSpecific) + (1|PptNum) +
(1+GeneralSpecific|Item),data=IAData,REML=FALSE)
#Model that includes the effect of participants on the intercept and the slope of verb type on items (whether the
difference between general and specific verbs depends on the items used)
```

```
#############################################
# Now test which model is best
#############################################
anova(Model1,Model2,Model3,Model4,Model5)
Data: IAData
       Df   AIC   BIC  logLik   Chisq Chi Df Pr(>Chisq)
Model3  4 16921 16941 -8456.4
Model4  4 16970 16991 -8481.2  0.0000      0     1.0000
Model2  5 16894 16919 -8442.1 78.1895      1     <2e-16 ***
Model1  7 16898 16933 -8441.9  0.3418      2     0.8429
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#################################################################################
# The best model from those above is the one with the lowest AIC, in this case it is Model 2
# This means that the random effects model should include participants and items on the intercept: this random model is then used as a basis to look at the fixed effect (verb type).
# The CurModel means the Current Model (including random effects model 2); this is the full model that includes all fixed effects. We now compare this full model that includes our fixed effect of verb type to a model that does not include the fixed effect (an intercept only model).
# If removing the fixed effect results in a significant p value and the AIC value increasing, then the fixed effect (verb type) adds significant weight to the model (i.e., the effect of verb type is significant)
#################################################################################
CurModel<-Model2
Model1<-update(CurModel, ~ . - GeneralSpecific)
#Comparison of model with and without the fixed effect of verb type included

Data: IAData
Models:
Model1: TargFixAfterV ~ (1 | PptNum) + (1 | Item)
CurModel: TargFixAfterV ~ factor(GeneralSpecific) + (1 | PptNum) + (1 |
CurModel:     Item)
         Df   AIC   BIC  logLik  Chisq Chi Df Pr(>Chisq)
Model1    4 16970 16990 -8481.0
CurModel  5 16894 16919 -8442.1 77.938      1  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# The effect of verb type is significant at the level of 0.001

#################################################################################
# For comparison purposes I show below what the same result in an ANOVA would be
#################################################################################
library(ez)
# standard within ppts ANOVA for comparison
ezANOVA(IAData,dv=.(TargFixAfterV),wid=.(PptNum),within=.(GeneralSpecific))
$ANOVA

           Effect DFn DFd        F            p p<.05       ges
2 GeneralSpecific   1  33 95.89187 2.730222e-11     * 0.2417644
# The effect of verb type is significant (p = 2.73...e-11)
```

Appendix C: The verb onsets, noun onsets and verb-noun distances for all auditory sentences in Experiment 2 (after being corrected from Experiment 1).

Specific Item	Verb Onset	Noun Onset	Verb-Noun	General Item	Verb Onset	Noun Onset	Verb-Noun	Specific-General Verb
It1_A_S	0.998	1.605	0.607	It1_A_G	1.154	1.757	0.603	3.927
It2_A_S	1.035	1.689	0.654	It2_A_G	1.108	1.767	0.659	-4.716
It3_A_S	1.162	1.791	0.629	It3_A_G	1.145	1.781	0.636	-6.674
It4_A_S	0.978	1.711	0.733	It4_A_G	0.958	1.689	0.731	1.494
It5_A_S	0.979	1.661	0.682	It5_A_G	1.076	1.760	0.684	-1.395
It6_A_S	1.027	1.730	0.704	It6_A_G	1.093	1.790	0.697	6.482
It7_A_S	0.986	1.688	0.702	It7_A_G	1.111	1.824	0.712	-10.156
It8_A_S	0.974	1.756	0.781	It8_A_G	0.905	1.678	0.773	8.345
It9_B_S	1.048	1.859	0.811	It9_B_G	1.009	1.823	0.815	-3.613
It10_B_S	0.911	1.814	0.903	It10_B_G	1.119	2.022	0.903	0.000
It11_B_S	1.120	1.867	0.747	It11_B_G	1.112	1.860	0.748	-1.000
It12_B_S	0.950	1.748	0.798	It12_B_G	0.981	1.768	0.787	10.693
It13_B_S	0.951	1.739	0.788	It13_B_G	1.007	1.807	0.799	-11.169
It14_B_S	1.018	1.702	0.684	It14_B_G	1.013	1.707	0.695	-10.413
It15_B_S	1.075	1.783	0.708	It15_B_G	0.974	1.688	0.714	-6.521
It16_B_S	1.039	1.778	0.739	It16_B_G	1.071	1.799	0.728	10.633
It17_C_S	1.033	1.780	0.747	It17_C_G	1.207	1.946	0.739	7.217
It18_C_S	1.020	1.751	0.731	It18_C_G	1.055	1.784	0.729	2.644
It19_C_S	1.153	1.784	0.631	It19_C_G	1.114	1.751	0.637	-6.716
It20_C_S	1.002	1.873	0.871	It20_C_G	1.255	2.129	0.874	-2.905
It21_C_S	1.066	1.746	0.680	It21_C_G	1.182	1.869	0.687	-6.837
It22_C_S	0.976	1.658	0.682	It22_C_G	1.081	1.769	0.688	-5.855
It23_C_S	1.002	1.721	0.720	It23_C_G	1.077	1.793	0.716	3.708
It24_C_S	1.062	1.615	0.553	It24_C_G	1.092	1.647	0.554	-1.460
It25_D_S	0.943	1.754	0.811	It25_D_G	1.206	2.018	0.812	-1.396
It26_D_S	0.972	1.546	0.574	It26_D_G	1.198	1.770	0.572	2.120
It27_D_S	1.025	1.638	0.613	It27_D_G	1.055	1.676	0.621	-7.984
It28_D_S	1.012	1.667	0.655	It28_D_G	1.086	1.742	0.657	-1.839
It29_D_S	0.913	1.724	0.811	It29_D_G	1.161	1.979	0.818	-6.423
It30_D_S	0.963	1.651	0.688	It30_D_G	1.085	1.768	0.683	5.530
It31_D_S	1.049	1.662	0.614	It31_D_G	1.111	1.732	0.621	-7.004
It32_D_S	0.955	1.770	0.815	It32_D_G	1.139	1.945	0.806	8.801

Appendix D: Stimuli for the lexical decision task completed to ensure equivalent word recognition across specific and general verbs. The verb pairs are presented in adjacent rows below, with the specific verb followed by the corresponding general verb (e.g., sweeping (specific), selling (general). The adjacent column shows the matched pseudo-word created via 'Wuggy'.

Word	**Pseudo-Word**	**Word**	**Pseudo-Word**	**Word**	**Pseudo-Word**
sweeping	sweaming	ring	rint	chewing	choxing
selling	sinting	kick	keck	buying	bezing
cooking	ceaming	grate	grake	writing	kroting
sample	mimple	share	chare	receiving	remeaming
cutting	bitting	scare	prare	spend	spind
polishing	sadishing	poke	pove	pack	pank
fastening	hossening	hang	hant	call	rall
carrying	calcuing	fix	fif	dodge	donch
wiping	waming	eat	eet	thank	thunk
sketching	smutching	move	mive	dislike	dislyle
riding	pibing	smash	smank	light	lighs
stroking	scroding	pass	paff	lose	lolt
driving	droding	sharpen	sholfen	spot	spet
admiring	atcaring	break	cleak	smoke	smole
lick	bick	planting	spinting	collect	cottect
steal	steaf	smelling	glulling	bathe	barls
thread	throad	baking	bexing	follow	follap
catch	cates	fetching	lutching	rip	mip
iron	udon	unlocking	unbincing	win	hin
refund	refint	pushing	munding	melt	mech
milk	misk	folding	firting	hate	hant
feed	feem	choosing	cheebing	watering	nivering
tear	mear	reading	rouring	describing	desspazing
burn	burd	shutting	bretting	frying	spuing
kiss	kise	sailing	rairing	washing	ganding
ignore	idnord	watching	ganching	ring	rint
train	trawn	tasting	fosting	playing	spaging
replace	replert	saving	miving	dusting	runking
sipping	dapping	comb	cocs		
hiding	fibing	damage	dowage		
chasing	traming	wear	wour		
delivering	memitoring	forget	falget		
knitting	krotting	fly	fle		
borrowing	bompoming	drop	blop		
building	bourting	hurt	huys		
avoiding	alulging	examine	exexits		
drinking	drulling	repair	reboir		
trying	chuing	notice	notave		
peeling	poosing	boiling	burting		
weighing	weersing	cleaning	freening		
bouncing	beansing	climbing	plombing		
throwing	thraming	drawing	spaxing		

Appendix E: Output from paired samples t-test on lexical decision task data.
The first table shows the similarity of means across the general and specific verbs. The third table reveals that there is no significant difference between processing specific and general verbs.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	General_RT	623.47	17	84.378	20.465
	Specific_RT	627.94	17	77.209	18.726

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	General_RT & Specific_RT	17	.903	.000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	General_RT - Specific_RT	-4.471	36.341	8.814	-23.155	14.214	-.507	16	.619

Appendix F: Experiment 2 cross-over effects

The table shows how latencies in the comprehension (visual world) only task and dual-task become faster in the second experimental half where participants have already been exposed to the visual world stimuli once before.

	1st Experimental Half				**2nd Experimental Half**			
	Specific Verb Condition		General Verb Condition		Specific Verb Condition		General Verb Condition	
	Mean Latency (ms)	SD (ms)	Mean Latency (ms)	SD (ms)	Mean Latency (ms)	SD (ms)	Mean Latency (ms)	SD (ms)
Comprehension Only	962	232	1178	215	840	276	1004	315
Dual-Task	989	241	1157	255	921	252	1106	240

Appendix G: Sentence stimuli for Experiments 3-6.

Please note that the sentences marked by a (*) appeared in Experiment 3 but were removed before Experiment 4 as participant rarely fixated on the target in any of the conditions.

Labels that contain an S refer to simple sentence items, whilst C refers to complex sentence items.

Label	Sentence (Specific verb/General verb)	Target (and Distractors)	Production Display Items (Experiment 4-6)
It1_S	The man is sweeping/selling the rug.	Rug (horse, cake, robot)	Woman, cup, piano
It1_C	The man who is sweeping/selling the rug is moving to London on Thursday.	Rug (horse, cake, robot)	Woman, cup, piano
It2_S	The woman is cooking/sampling the chicken.	Chicken (lipstick, perfume, tissues)	Hair, mirror, scissors
It2_C	The woman who is cooking/sampling the chicken is having a dinner party this evening.	Chicken (lipstick, perfume, tissues)	Hair, mirror, scissors
It3_S	The woman is cutting/polishing the nail.	Nail (car, table, glass)	Farmer, cow bucket
It3_C	The woman who is cutting/polishing the nail is going on holiday to France.	Nail (car, table, glass)	Farmer, cow bucket
It4_S	The man is fastening/carrying the coat.	Coat (guitar, bell, rabbit)	Dog, ball, cactus
It4_C	The man who is fastening/carrying the coat is a maths teacher.	Coat (guitar, bell, rabbit)	Dog, ball, cactus
It5_S	The girl is wiping/sketching the desk.	Desk (policeman, cloud, fire)	Rabbit, carrot, slippers
It5_C	The girl who is wiping/sketching the desk will go to the theatre later.	Desk (policeman, cloud, fire)	Rabbit, carrot, slippers
It6_S	The boy is riding/stroking the horse.	Horse (hand, scarf, hair)	Table, paper, glass
It6_C	The boy who is riding/stroking the horse has to do his geography, chemistry and maths homework tonight.	Horse (hand, scarf, hair)	Table, paper, glass
It7_S	The man is driving/admiring the car.	Car (rose, watch, church)	Peg, clothes, sun
It7_C	The man who is driving/admiring the car has an awful cold.	Car (rose, watch, church)	Peg, clothes, sun
It8_S	The girl will lick/steal the lolly.	Lolly (ring, book, toybear)	Baby, highchair, egg
It8_C	The girl who will lick/steal the lolly is always getting in trouble.	Lolly (ring, book, toybear)	Baby, highchair, egg

It9_S	The woman will thread/catch the needle.	Needle (plant, egg, cup)	Horse, apple, fence
It9_C	The woman who will thread/catch the needle has very nimble fingers.	Needle (plant, egg, cup)	Horse, apple, fence
It10_S	The woman will iron/refund the clothes.	Clothes (piano, cooker, present)	Woman, toaster, watch
It10_C	The woman who will iron/refund the clothes hates her job.	Clothes (piano, cooker, present)	Woman, toaster, watch
It11_S	The girl will milk/feed the cow.	Cow (plant, doll, clown)	Bottle, stool, coat
It11_C	The girl who will milk/feed the cow lives in the big old tudor house with the statue on the lawn.	Cow (plant, doll, clown)	Bottle, stool, coat
It12_S	The girl will tear/burn the paper.	Paper (match, pan, candle)	Deer, acorn, hand
It12_C	The girl who will tear/burn the paper is very unhappy.	Paper (match, pan, candle)	Deer, acorn, hand
It13_S	The woman will kiss/ignore the man.	Man (telephone, door, kettle)	Girl, balloon, chair
It13_C	The woman who will kiss/ignore the man has just dropped out of medical school.	Man (telephone, door, kettle)	Girl, balloon, chair
It14_S	The girl will train/replace the cat.	Cat (desk, nail, butter)	Owl, wall, barn
It14_C	The girl who will train/replace the cat likes to show off to all her friends.	Cat (desk, nail, butter)	Owl, wall, barn
It15_S	The woman is sipping/hiding the wine.	Wine (letter, trophy, briefcase)	Man, bell, church
It15_C	The woman who is turning fifty next month is sipping/hiding the wine.	Wine (letter, trophy, briefcase)	Man, bell, church
It16_S	The man is chasing/delivering the rabbit.	Rabbit (sofa, newspaper, television)	Boat, cloud, tree
It16_C	The man who works every weekend is chasing/delivering the rabbit.	Rabbit (sofa, newspaper, television)	Boat, cloud, tree
It17_S	The woman is peeling/weighing the carrot.	Carrot (scarf, shuttlecock, pan)	Sofa, door, cat
It17_C	The woman who has a degree in Sociology is peeling/weighing the carrot	Carrot (scarf, shuttlecock, pan)	Sofa, door, cat
It18_S	The man is building/avoiding the church.	Church (beer, mirror, woman)	Teacher, cigarette, radio
It18_C	The man who has recently been fired is building/avoiding the church.	Church (beer, mirror, woman)	Teacher, cigarette, radio
It19_S	The woman is drinking/trying the wine.	Wine (cheese, lipstick,	Child, sock,

		pipe)	drawers
It19_C	The woman, who hates any confrontation, is drinking/trying the wine.	Wine (cheese, lipstick, pipe)	Child, sock, drawers
It20_S	The woman is peeling/weighing the carrot.*	Carrot (rucksack, baby, rabbit)	-
It20_C	The woman who has a degree in sociology is peeling/weighing the carrot.*	Carrot (rucksack, baby, rabbit)	-
It21_S	The boy is bouncing/throwing the ball.	Ball (paper plane, rope, bracelet)	Lipstick, scarf, chest
It21_C	The boy who took his dog to the vets this morning is bouncing/throwing the ball.	Ball (paper plane, rope, bracelet)	Lipstick, scarf, chest
It22_S	The girl will ring/kick the bell.*	Bell (cow, bricks, computer)	-
It22_C	The girl who reads crime novels in her spare time will ring/kick the bell.*	Bell (cow, bricks, computer)	-
It23_S	The girl will grate/share the cheese.	Cheese (perfume, paper, trophy)	Computer, lamp, desk
It23_C	The girl who is singing corny Christmas songs and decorating the tree with baubles will grate/share the cheese.	Cheese (perfume, paper, trophy)	Computer, lamp, desk
It24_S	The boy will scare/poke the clown.	Clown (needle, ice-cream, parcel)	Wheelchair, hill, sheep
It24_C	The naughty boy who always gets detentions will scare/poke the clown.	Clown (needle, ice-cream, parcel)	Wheelchair, hill, sheep
It25_S	The man will hang/fix the picture.	Picture (chair, computer, sewing machine)	Teapot, sugar cube, tray
It25_C	The man who is paid to do small jobs around the house will hang/fix the picture.	Picture (chair, computer, sewing machine)	Teapot, sugar cube, tray
It26_S	The boy will eat/move the cake.	Cake (bag, mirror, glasses)	Motorbike, crutches, flowers
It26_C	The boy who is the captain of the football team at school will eat/move the cake.	Cake (bag, mirror, glasses)	Motorbike, crutches, flowers
It27_S	The boy will smash/pass the plate.	Plate (balloon, mirror, slippers)	Boy, coins, sweets
It27_C	The boy who got a baseball for his birthday will smash/pass the plate.	Plate (balloon, mirror, slippers)	Boy, coins, sweets
It28_S	The woman will sharpen/break the	Pencil (bottle, computer,	Rucksack, toy

	pencil.	piano)	bear, crisps
It28_C	The woman, who is embarrassed about her terrible clumsiness, will sharpen/break the pencil.	Pencil (bottle, computer, piano)	Rucksack, toy bear, crisps
It29_S	The man is planting/smelling the flower.	Flower (perfume, cheese, candle)	Clown, plate, leopard
It29_C	The man is planting/smelling the flower on the newly laid lawn at the side of his quaint white cottage.	Flower (perfume, cheese, candle)	Clown, plate, leopard
It30_S	The woman is baking/fetching the bread.	Bread (money, waiter, knife)	Spider, bath tub, hairbrush
It30_C	The woman is baking/fetching the bread because she wants sandwiches for lunch.	Bread (money, waiter, knife)	Spider, bath tub, hairbrush
It31_S	The woman is unlocking/pushing the door.*	Door (skateboard, stool, wheelbarrow)	-
It31_C	The woman is unlocking/pushing the door which is old and creaky and made of oak.*	Door (skateboard, stool, wheelbarrow)	-
It32_S	The woman is folding/choosing the trousers.	Trousers (ring, apple, lipstick)	Old man, bench, duck
It32_C	The woman is quickly folding/choosing the trousers because she is in a rush to leave the house.	Trousers (ring, apple, lipstick)	Old man, bench, duck
It33_S	The girl is reading/shutting the book.	Book (door, bag, window)	Jumper, wool, barbed wire
It33_C	The little girl is reading/shutting the book whilst sitting in the pink and white bedroom with the large lamp.	Book (door, bag, window)	Jumper, wool, barbed wire
It34_S	The man is sailing/watching the boat.	Boat (bird, kite, sun)	Bag, letter, perfume
It34_C	The man is sailing/watching the boat because it is a lovely clear day outside.	Boat (bird, kite, sun)	Bag, letter, perfume
It35_S	The boy is tasting/saving the broccoli.	Broccoli (teacher, spider, kite)	Beer, trousers, beetle
It35_C	The boy is tasting/saving the broccoli after cleaning out the fridge.	Broccoli (teacher, spider, kite)	Beer, trousers, beetle
It36_S	The woman will comb/damage the hair.	Hair (watch, ball, medal)	Cooker, fire, sausages
It36_C	The woman will comb/damage the hair whilst having an argument on the phone.	Hair (watch, ball, medal)	Cooker, fire, sausages
It37_S	The boy will wear/forget the coat.	Coat (wallet, bread,	Wine, table, mop

		tape)	
It37_C	The boy will wear/forget the coat that he bought earlier that morning.	Coat (wallet, bread, tape)	Wine, table, mop
It38_S	The boy will fly/drop the kite.	Kite (sweet, kite, scarf, cigarette)	Hen, feather, nest
It38_C	The boy will fly/drop the kite because is it very windy outside.	Kite (sweet, kite, scarf, cigarette)	Hen, feather, nest
It39_S	The boy will hurt/examine the bird.	Bird (coat, globe, match)	Policeman, kite, glasses
It39_C	The boy will hurt/examine the bird that lives in the big oak tree at the end of his garden.	Bird (coat, globe, match)	Policeman, kite, glasses
It40_S	The man will repair/see the washing machine.	Washing machine (perfume, dog, chicken)	Wallet, belt, locker
It40_C	The man will repair/notice the washing machine and laugh about the bright red suitcase sat next to it.	Washing machine (perfume, dog, chicken)	Wallet, belt, locker
It41_S	The woman will rip/win the bag.	Bag (lolly, radio, teapot)	Monkey, picture, pencil
It41_C	The woman will rip/win the bag whilst helping on a small tombola stand in the corner of the school fair.	Bag (lolly, radio, teapot)	Monkey, picture, pencil
It42_S	The man will melt/hate the butter.*	Butter (deer, weighing scales, book)	-
It42_C	The man will melt/hate the butter that he bought that morning and will have to change his recipe.*	Butter (deer, weighing scales, book)	-
It43_S	The man is watering/describing the rose.	Rose (mountain, wine, church)	Bride, bike, arm
It43_C	The man with the gardening book on his desk is watering/describing the rose.	Rose (mountain, wine, church)	Bride, bike, arm
It44_S	The woman is frying/washing the mushrooms.	Mushrooms (knife, boat, window)	Jacket, ironing board, wardrobe
It44_C	The woman is sitting on the high brown stool with a wooden seat and metal legs whilst frying/washing the mushrooms.	Mushrooms (knife, boat, window)	Jacket, ironing board, wardrobe
It45_S	The woman is boiling/cleaning the kettle.	Kettle (plate, table, ring)	Paper plane, doll, hat
It45_C	The woman is listening to music on the radio and dancing wildly whilst	Kettle (plate, table, ring)	Paper plane, doll, hat

	boiling/cleaning the kettle.		
It46_S	The man is climbing/drawing the mountain.	Mountain (deer, moon, mushrooms)	Lock, gate, bat
It46_C	The man is talking about Eastenders whilst climbing/drawing the mountain.	Mountain (deer, moon, mushrooms)	Lock, gate, bat
It47_S	The woman is dusting/playing the piano.	Piano (picture, clock, telephone)	Donkey, chocolate, carousel
It47_C	The woman remembers she needs to go shopping whilst she is dusting/playing the piano.	Piano (picture, clock, telephone)	Donkey, chocolate, carousel
It48_S	The boy is chewing/buying the sweet.	Sweet (robot, kite, cat)	Girl, dress, bracelet
It48_C	The boy is in trouble because he is chewing/buying the sweet.	Sweet (robot, kite, cat)	Girl, dress, bracelet
It49_S	The girl is writing/receiving the letter.	Letter (parcel, trophy, money)	Snake, cage, branch
It49_C	The girl begins eating her Sunday lunch after writing/receiving the letter.	Letter (parcel, trophy, money)	Snake, cage, branch
It50_S	The man will spend/pack the money.	Money (clothes, glasses, newspaper)	Dentist, telephone, pizza
It50_C	The man is going home in the early hours of tomorrow morning and will spend/pack the money.	Money (clothes, glasses, newspaper)	Dentist, telephone, pizza
It51_S	The girl will call/dodge the policeman.	Policeman (table, wheelchair, cat)	Lorry, rocking chair, vase
It51_C	The girl in the blue, purple and pink stripy dress and shiny black stilettos will call/dodge the policeman.	Policeman (table, wheelchair, cat)	Lorry, rocking chair, vase
It52_S	The boy will thank/dislike the teacher.	Teacher (paper, computer, desk)	Snowman, stick, buttons
It52_C	The boy is moving classes because he is too bright and will thank/dislike his teacher.	Teacher (paper, computer, desk)	Snowman, stick, buttons
It53_S	The man will light/lose the match.	Match (pencil, tape, money)	Nurse, clipboard, bed
It53_C	The man in the silver car parked on the long pebbled drive will light/lose the match.	Match (pencil, tape, money)	Nurse, clipboard, bed
It54_S	The boy will chop/spot the carrot.	Carrot (man, fox, washing machine)	Queen, helicopter, tea

It54_C	The boy is stood next to the counter where he will chop/spot the carrot.	Carrot (man, fox, washing machine)	Queen, helicopter, tea
It55_S	The man will smoke/collect the cigarette.	Cigarette (dairy, broccoli, briefcase)	Ghost, house, rifle
It55_C	The man with the silver stud in his ear and the green dragon tattoo on his arm will smoke/collect the cigarette.	Cigarette (dairy, broccoli, briefcase)	Ghost, house, rifle
It56_S	The woman will bathe/follow the baby.	Baby (car, map, ball)	Briefcase, ruler banana
It56_C	The woman, with very long blonde hair tied in a neat bun will bathe/follow the baby.	Baby (car, map, ball)	Briefcase, ruler banana

Appendix H: The verb onsets, noun onsets and verb-noun distances of all sentences in Experiments 3-6.

Item	Simple Specific		Simple General			Complex Specific		Complex General		
	Verb Onset	Noun Onset	Verb Onset	Noun Onset	Specific-General Difference (in Verb-Noun Distance)	Verb Onset	Noun Onset	Verb Onset	Noun Onset	Specific-General Difference (in Verb-Noun Distance)
1	887	1788	1046	1957	-10	909	1658	944	1676	17
2	1019	1760	891	1643	-11	1082	1696	1041	1641	14
3	1118	1824	1089	1803	-8	1104	1840	1234	1980	-10
4	910	1740	1061	1883	8	1012	1669	992	1639	10
5	838	1577	879	1622	-4	1053	1875	1095	1910	7
6	1118	1975	889	1745	1	1104	1753	920	1574	-5
7	921	1806	942	1847	-20	990	1765	1117	1908	-16
8	974	1533	844	1409	-6	1162	1774	1056	1677	-9
9	838	1545	924	1629	2	1109	1855	1201	1957	-10
10	1182	1905	1065	1796	-8	1182	1872	1232	1924	-2
11	816	1506	864	1571	-17	973	1616	1066	1697	12
12	956	1638	1000	1674	8	947	1390	1066	1499	10
13	1028	1693	855	1529	-9	1176	1868	1146	1845	-7
14	832	1593	1129	1908	-18	1005	1622	993	1626	-16
15	762	1511	952	1686	15	3058	3864	3028	3814	20
16	977	1784	952	1764	-5	2670	3488	2685	3514	-11
17	989	1654	988	1668	-15	3606	4246	3275	3925	-10
18	1001	1731	873	1622	-19	2494	3155	2354	3030	-15
19	1005	1736	881	1606	6	3193	3847	3362	4017	-1
20	828	1428	994	1608	-14	5633	6389	5564	6310	10
21	873	1713	957	1809	-12	3306	4141	3194	4027	2
22	852	1472	922	1549	-7	3228	3759	3607	4126	12
23	876	1490	909	1516	7	5374	5982	5295	5900	3
24	885	1474	875	1453	11	3238	3779	3195	3741	-5

25	745	1388	736	1389	-10	3637	4174	4126	4660	3
26	956	1591	900	1536	-1	3499	4066	3526	4095	-2
27	952	1770	916	1732	2	3003	3675	2980	3650	2
28	816	1532	958	1659	15	3985	4548	3771	4321	13
29	813	1521	808	1520	-4	830	1491	713	1363	11
30	1023	1729	888	1603	-9	921	1698	885	1642	20
31	956	1725	979	1745	3	890	1590	1004	1708	-4
32	1012	1812	908	1719	-11	1229	1830	1347	1950	-2
33	747	1348	1119	1705	15	1124	1712	1102	1700	-10
34	919	1672	843	1590	6	920	1578	806	1453	11
35	770	1492	898	1640	-20	926	1569	838	1463	18
36	863	1436	1002	1565	10	961	1511	888	1421	17
37	834	1514	887	1568	-1	985	1460	862	1344	-7
38	949	1706	953	1703	7	666	1188	766	1269	19
39	889	1659	885	1675	-20	850	1495	702	1367	-20
40	885	1685	811	1600	11	998	1692	852	1542	4
41	936	1562	863	1503	-14	929	1485	870	1429	-3
42	817	1560	873	1607	9	956	1459	982	1491	-6
43	923	1724	959	1765	-5	2869	3624	2983	3737	1
44	1037	1840	987	1797	-7	5528	6315	5630	6437	-20
45	923	1734	1042	1856	-3	4536	5120	4565	5157	-8
46	878	1602	941	1657	8	2938	3554	2732	3361	-13
47	1146	1981	951	1771	15	3724	4450	3463	4176	13
48	1007	1736	1068	1814	-17	2141	2788	2160	2810	-3
49	851	1548	760	1465	-8	3227	3934	2898	3610	-5
50	863	1445	947	1522	7	4062	4678	4403	5000	19
51	819	1413	989	1578	5	5593	5970	5610	6007	-20
52	945	1776	918	1737	12	3507	4196	3875	4568	-4
53	984	1593	897	1525	-19	4028	4520	4080	4580	-8
54	922	1561	939	1590	-12	2945	3465	2788	3315	-7
55	851	1574	929	1635	17	4992	5582	5083	5682	-9
56	948	1631	876	1561	-2	3938	4445	3920	4439	-12

Appendix I: Means and correlation from the Operation Span task (Experiment 5)

-The first table shows the operation span scores and size of the prediction effect for each participant.
-The second table shows the means and standard deviations of the two variables.
-The final table shows the output from the bivariate correlation which examines how the two variables correlate with one another. The correlation is approaching significance, a larger sample size would likely have yielded a significant correlation.

Participant	**Prediction Effect (ms)** (Difference between average latency to Specific verb and average latency to General verb)	**Operation Span Score** (Number of words recalled)
1	111	36
2	211	43
3	255	36
4	-29	47
5	394	38
6	450	39
7	-64	40
8	-256	49
9	559	47
10	19	42
11	20	42
12	208	26
13	448	48
14	99	48
15	22	54
16	330	37
17	528	38
18	190	44

Descriptive Statistics

	Mean	Std. Deviation	N
PredictionEffect	194.1667	225.27663	18
Ospan_RT	6594.1667	1341.72230	18

Correlations

		PredictionEffect	Ospan_RT
PredictionEffect	Pearson Correlation	1	-.432
	Sig. (2-tailed)		.074
	N	18	18
Ospan_RT	Pearson Correlation	-.432	1
	Sig. (2-tailed)	.074	
	N	18	18

Appendix J: Number of disfluencies across production only and production with comprehension trial types in Experiments 4-6.

The number of disfluencies made during speech production were counted across experiments and trial types. There were 5 types of disfluency examined: grammatical errors, hesitations (e.g., err and erms), disfluencies mid-sentence (e.g "the err man…"), restarts, and semantic errors.

The number of disfluencies were very similar across the two types of trial.

Type of disfluency	Experiment 4		Experiment 5		Experiment 6	
	PO trials	PV trials	PO trials	PV trials	PO trials	PV trials
Grammatical Error	0	0	1	1	21	14
Hesitation	147	118	54	58	160	129
Disfluency (mid-sentence)	0	0	15	5	115	150
Restart	28	16	22	15	10	3
Semantic Error	0	0	0	0	5	2
Total	**175**	**134**	**92**	**79**	**311**	**198**

Appendix K: The AIC values for each random effects model in the LMEMs

The random model structure is chosen based on its AIC. Importantly, the model that is chosen must always have a lower AIC value than the fullest model tested (or at least are not significantly different from the fullest model). Below the random model AICs are shown for all models in Experiments 1-6. The best model AIC refers to the AIC value for the model I have chosen and modelled in the data, the full model AIC refers to the AIC value of the fullest model tested. The best model AIC in each case is lower than the full model AIC, or not significantly different.

Exp	Full data set		Complex data set		Speech onsets on eye-movements		Recall on eye-movements		Cross-experiment comparison		Production task (PO vs PV)	
	Best model AIC	Full model AIC	Best model AIC	Full model AIC	Best model AIC	Full model AIC	Best model AIC	Full model AIC	Best model AIC	Full model AIC	Best model AIC	Full model AIC
1	16894	16898										
2	39427	39429	(Data set with only 1st Exp half) 18221	18221								
3	19138 *not significantly different	19134	10617	10575								
4	11205	11231	6781	6813	8938	8964	10372	10414	30372	30400	44313	44332
5	15577	15600	7777	7822	12366	12417	16049	16105	35544	35586	38175	38175 #fullest model is best model
6	19860	19872	9534	9569	12688	12738	18816	18883	35854	35909	38379	38279 #fullest model is best model