PUPILLOMETRY AND FILLER GAP DEPENDENCIES

LEIGH BREAKELL FERNANDEZ

MPhil

2013

PUPILLOMETRY AND FILLER GAP DEPENDENCIES

LEIGH BREAKELL FERNANDEZ

A thesis submitted in partial fulfilment

of the requirements of the University

of Northumbria at Newcastle for the

degree of a Masters of Philosophy

Research undertaken in the School of

Life Sciences

November 2013

Abstract

In this thesis, pupillometry was used to investigate how the parser comprehends filler gap dependencies. It has been shown that the parser automatically searches for and forms dependencies as soon as a gap is encountered. The parser utilizes verb information when positing upcoming gaps; however, in languages in which the verb occurs late in the sentence (e.g. Japanese) it has been shown that pre-verbal information is used to predict upcoming gaps. In this thesis a series of studies was conducted investigating filler gap dependencies. These dependences are typically unbounded in length; however there are some exceptions known as islands, which do not allow a filler and a gap to be associated. There are also exceptions to island constraints known as parasitic gaps. The first set of studies investigates processing costs involved with parasitic gap-like constructions in order to adjudicate between competing accounts of island constraints. The data suggests that the parser actively searches for gaps and forms dependencies as soon as possible (regardless of semantic fit). The second set of studies in this thesis investigates pre-verbal gap prediction in a verb-medial language (i.e. English). The results suggest that the parser consults transitivity information when positing upcoming gaps. These data support theories of grammar in which the parser forms the simplest (syntactic) interpretation in line with good-enough theories, in order to minimize processing costs when forming filler gap dependencies.

List of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank first and foremost my supervisor Paul Engelhardt for his unwavering support and constructive input, and without whom this work would never have been conducted. I would like to also thank Andre Bester for his amazing Matlab skills, I could never have analyzed my data without his help. I would also like to thank Rebecca Jennings and Allison Brown for their help with recording stimuli and assistance with the research. Thank you to the Cognition and Communication lab at Northumbria, without their ideas, input, shoulders, and presence, I would never had stay sane long enough to finish this project. And of course a giant thank you to the psychology department at Northumbria for being so welcoming, warm, and supportive. Last, but not least, I would like to thank Paul Agnew for navigating this work through all the hoops.

Declaration

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others.

Any ethical clearance for the research presented in this thesis has been approved. Approval has been sought and granted by the School Ethics Committee in 2011.

Name:

Leigh Fernandez

Signature:

Date:

March 24, 2014

CHAPTER 1: FILLER GAP DEPENDENCIES - AN INTRODUCTION

This thesis focuses on how the parser processes linguistic constructions known as filler gap dependencies using a measure of cognitive processing cost called pupillometry. The first chapter provides an overview of the theories and research surrounding filler gap dependencies. Chapter 2 provides an in depth review of pupillometry highlighting what the pupil can reveal about language processing. Chapter 3 and 4 are empirical studies that used pupillometry to investigate the competing theories of filler gap dependencies and the constraints that govern them. The final chapter discusses the implications of the empirical findings within the context of previous research and provides novel theoretical arguments of filler gap dependencies.

This chapter begins by explaining constituent movement and why these constructions are inherently ambiguous. I then briefly discuss transformational and non-transformational accounts of filler gap dependences. The first section ends with an outline of key studies that have investigated filler gap dependences, and highlights the more prominent theories that explain them. The second section of the chapter focuses on hyper-active gap filling. The third section focuses on island constraints, which are the main set of constraints that govern filler gap dependencies. The fourth and final section focuses on a construction known as a parasitic gap, these constructions involve a filler being posited within an island.

1.1. Movement

English is a subject-verb-object language (see Example 1). Thus, the grammatical relationships in (1) are considered canonical. However, word order by itself is not enough to determine grammatical relationships.

1) The boy bought the toy.
2) The boy bought what?
3) What did the boy buy?

In Example (2), the object was replaced by *what* to form a question and the grammatical relationships are still linear. In (3) however, the wh-word has been moved and the word order becomes non-canonical. These types of constructions are known as filler gap dependencies or unbounded dependencies. Transformational grammar (Chomsky, 1965) assumes that sentences such as (3 and 4) are derived via movement. Movement occurs when a constituent, called the filler (e.g. $what_i$), is moved from its canonical position and it leaves a trace at the original location. This trace (or gap (t_i)) is phonologically silent but remains syntactically relevant because the moved constituent is assigned grammatical properties through the gap site. Therefore, the gap site assigns a thematic role to the filler thereby allowing the comprehender to interpret the sentence. For the remainder of this paper the filler will be identified using a subscript ($_i$) and the gap identified using (t_i). The subscripts indicate co-reference (see example 4).

4) What$_i$ did the boy buy t_i?

1.1.1. Ambiguity resolution

Much linguistic research has focused on the interpretation of filler gap dependencies and how the parser incrementally interprets sentences with these kinds of dependencies. The correct interpretation of (4) can only be formed after the gap site is located; however there is no overt phonological signal establishing the gap, which begs the question of how the parser knows when it has found the correct gap site? The gap site is also potentially ambiguous making it even harder for the parser to correctly determine where the filler was moved.

In example (5) there is a potential gap site following the verb *want*; however, this is incorrect because the correct gap actually occurs at the end of the sentence. Gap identification problems can also be exacerbated by particular types of verb. For example, transitive verbs typically require a direct object; however optionally transitive verbs such as *eat* can take a direct object (e.g. *I already ate a pickle*) but these verbs do not obligatorily require a direct object (e.g. *I already ate*).

5) What$_i$ did you want *t_i Bob to answer t_i?

Verb argument structure refers to the characteristics of a verb and the complements (or arguments) that the verb is allowed to take. The argument structure of a verb is lexically specified in terms of both syntax and semantics. Syntactic arguments refer to the subcategorization frame of the verb, and these are the arguments that a verb licenses. For example, the dative verb *donate* in (6) subcategorizes for a noun phrase (NP) *her clothes* as well as a prepositional phrase (PP) *to the shop*. The verb *donate* can take

both an NP and PP argument at the same time (6a), or individually (6b and 6c). The semantic characteristics of the verb, on the other hand, refer to the thematic roles dictated by the verb. The verb *donate* in (6) dictates an agent (*the girl*), theme (*clothing*), and/or a recipient (*the shop*).

6) A. The girl donated her clothes to the shop

B. The girl donated her clothes

C. The girl donated to the shop

It is clear that both subcategorization and semantics play a role in parsing; however it is unclear how and when they are utilized during parsing. In this thesis I focus primarily on the role of syntax during online parsing. Ambiguity resolution has been used to assess what subcategorization information is available during online parsing; traditionally experimental paradigms manipulate verb type with either a plausible or an implausible argument. Evidence concerning online parsing has been mixed. In the next section I will highlight some of the contrasting studies and theories within the context of subcategorization manipulations and within the context of filler gap dependencies.

1.1.2. Transformation based parsing accounts

Janet Fodor (1978) outlined and evaluated three theories concerning the parsing of moved constituents: the *Last Resort Model, the First Resort Model,* and the *Lexical Expectation Model.* The first model she evaluated was the Last Resort Model*,* which assumes that the parser does not fill gaps until it is forced to do so. By this theory the parser will not fill an ambiguous

gap (this is a potential gap site that may potentially turn out to be the incorrect gap site) until evaluating whether there are any other potential gap sites in the sentence. Sentences such as (5) have an ambiguous gap site, and by the Last Resort theory, are predicted to be more difficult to process compared to sentences such as (4). However, Fodor provided evidence refuting the idea that the parser evaluates the whole sentence before filling an ambiguous gap site and that sentences like (4) and (5) are processed differently, and thus, Fodor rejected the Last Resort Model as a viable explanation of how the parser processes filler gap dependencies.

The second theory that Fodor evaluated was the First Resort Model. This model is based on the idea that the parser fills gaps, ambiguous or unambiguous, immediately. This is the basis for the *Active Filler Hypothesis*, which will be discussed in more detail below (Frazier, 1987; Frazier & Flores d'Arcais, 1989). However, Fodor argued the parser would expend too much effort and resources parsing sentences, such as (7), because gaps would be posited at every potential site. Based on a processing resources perspective Fodor rejected the First Resort Model.

7) What$_i$ did you want *t_i Bob to persuade *t_i John to inform *t_i Mary that she forgot t_i?

The third model she evaluated was the Lexical Expectation Model. This model suggests that the location where gaps are posited relies on lexical items within in the sentence, particularly the verb. The parser is assumed to rank different syntactic structures based on frequency (for example) and the structure with the highest frequency will guide the initial

interpretation of a sentence. Therefore, the parser relies on subcategorization tendencies associated with particular lexical items to rank or evaluate alternative syntactic structures. Thus, gaps will be posited where they are most likely to occur given the words in the sentence. This theory circumvents the weaknesses of the previous two models.

Obligatorily transitive verbs require a direct object, while strictly intransitive verbs do not. By the Lexical Expectation Model, the parser should be expecting an NP after encountering a transitive verb and should posit a gap after the verb if no NP is present. Clifton, Frazier, and Connine (1984) and Tanenhaus, Stowe, and Carlson (1985) found support for the Lexical Expectation Model. However, Clifton and Frazier (1989) argued that the model has cross-linguistic problems. The Lexical Expectation Model assumes that in languages with different word orders or verb locations the assignment of dependencies may be delayed.

Using a frame-by-frame reading paradigm Frazier (1987) investigated syntactic processing in Dutch. Unlike English, verb phrases are head-final in Dutch, as can be seen in example (8). Lexical models assume that the head of a phrase is encountered and arguments are postulated based on the head preferences.

8) Ik heb Joost gezien

I have Joost seen

In Dutch the verb occurs at the end of the sentence and from a syntactic point of view, processing should be delayed given the location of the verb. However, Frazier found that Dutch participants were postulating gaps before

encountering the verb which is inconsistent with lexical based theories like the Lexical Expectation Model. This lead Frazier to propose the Active Filler Hypothesis, which assumes the parser posits a gap and forms a dependency at the first available gap site. She argued that the parser postulates an upcoming gap as soon as it encounters a filler, and the parser is driven by fillers, not gaps. Frazier and Flores d'Arcais (1989) also found evidence to suggest that the parser uses an Active Filler strategy; they found that participants were positing a gap at the first potential site and that the parser does not consult lexical information when assigning a gap.

1.1.3. Non-transformation based parsing accounts

Non-transformational theories of grammar assume that filler gap dependency mechanisms are unnecessary and only a direct association needs to be formed between the verb and the filler (Pickering & Barry, 1991). Pickering and Barry argue that the covert phonological trace is not necessary, and the data supporting gap filling can also be explained through direct association. They call this the *Direct Association Hypothesis*. Given that most examples of gap filling are observed in constructions where the verb and the gap are next to each other the findings can just as easily explained through Direct Association. They provided evidence that the parser does not wait until the gap to form the dependency when the verb and gap are not next to each other. Instead, the parser associates the verb and the filler directly. However, the authors do not provide experimental evidence to support their theories; rather they provide intuitive evidence (Traxler & Pickering, 1996). Pickering and Barry provided many examples

and constructions that support a gap-free grammar (9) and (10) highlight this.

9) In which box did you put the very large and beautifully decorated wedding cake bought from the expensive bakery?

10) Which box did you put in the very large and beautifully decorated wedding cake bought from the expensive bakery in?

Pickering and Barry argue that in both (9) and (10) the gap site is located at the end of the sentence, therefore both sentences would impose the same processing resources under a gap-based theory of grammar. However, (9) is easier to process than (10) and the authors argue this can only be explained via Direct Association. Example (9) is easier to process because the filler, *which box*, can be coindexed to subcategorizer *put* without processing the NP. Example (10), on the other hand, is more difficult to process because the filler, *which box,* must be held in memory until it reaches the subcategorizer *in* which does not occur until the parser has processed the entire sentence.

Gorrell (1993) and Gibson and Hickok (1993) both argued against the gap- free/Direct Association grammar proposed by Pickering and Barry. Gorrell argued that the examples provided for the Direct Association Hypothesis were far too narrow, and by investigating a wider range of examples there was little evidence to support the Direct Association Hypothesis. Gibson and Hickok argued that the conclusions drawn by Pickering and Barry concerning a gap-free grammar are far too strong. Instead, they argued that gap filling does not occur at the site of the gap,

rather the parser can posit a gap as soon as it is licensed by the grammar (i.e. at the verb). They proposed a predictive parsing algorithm that entails gaps and accounts for the examples given by Pickering and Barry.

Pickering (1993) provided additional examples and refuted the criticisms of Gorrell and Gibson and Hickok. He argued that a predictive account of gap filling is more complicated than a simpler grammar that does not assume gaps. While the work in this thesis focuses on filler gap dependencies, it was not designed to investigate the presence/absence of gaps. Instead, this study focuses on how the parser creates the link between a verb and its arguments, and in particular, in cases where the argument precedes the verb.

1.1.4. Empirical filler gap dependency research

Garnsey, Tanenhaus, & Chapman (1989) used evoked brain potentials to investigate First and Last Resort-type models. Their stimuli consisted of sentences with ambiguous gap sites in which the filler was either a plausible or implausible object. If the parser employed a First Resort strategy they anticipated a N400 response at the verb, which has been shown to reflect a semantic anomaly. If the parser employed a Last Resort strategy an N400 would not be evoked until later in the sentence when the parser realizes there are no other gap sites. They found an N400 response at the verb when the filler was implausible compared to when a plausible filler preceded the verb. These data support a First Resort strategy as it lends itself to the idea that the parser is assigning the filler to the first potential gap site. This data also supports the idea that the parser may not take plausibility into effect

when forming filler gap dependencies, as the dependency was formed despite the implausibility.

Boland, Tanenhaus, Garnsey, and Carlson (1995) tested several types of verbs to evaluate how subcategorization information was used during online parsing. Examples of the verb types (i.e. object control, dative, and typical transitive) are shown in 11 – 15 below. They advocate a *Constraint-Based Lexical Model* in which both syntactic and semantic characteristics of the verb are utilized to form filler gap dependencies, as soon as the parser encounters the verb. Constraint-Based Lexical Models assume that verb argument information is accessed in parallel, encountering the verb activates the set argument structures associated with that verb and ranked by frequency (MacDonald, Pearlmutter, & Seidenberg, 1994).

Object control verbs, such as *remind,* subcategorize for an object NP as well as an infinitive complement. In example (11), *remind* takes *Lana* as a direct object, and *to eat* as its complement. The NP *Lana* serves not only as the direct object of *remind* but it also serves as the subject of the verb *eat*. In (12A and 12B), there are two possible gap sites for the verb remind.

11)She reminded Lana to eat the cake

12)A. Which friend$_i$ did she remind t_i to eat the vanilla cake?

B. Which cake$_i$ did she remind you to eat t_i?

The gap in (12B) is assigned its thematic role via the infinitive verb *to eat*, while the gap in (12A) is assigned its thematic role from the object control verb *remind*.

When the parser encounters a dative verb like *donate* (see 13 and 14), the argument structure and syntactic constraints dictate that an NP gap

must be posited adjacent to the verb or after the PP. Semantic constraints of each of the potential gap sites give a framework into which thematic roles can be evaluated online. In examples (13) and (14), c*lothes* is a more likely theme of *donate* compared to *store* (and vice versa).

13)Which store$_i$ did the girl donate her clothes to t_i?

14)Which clothes$_i$ did the girl donate t_i to the store?

Dative verbs and object control verbs have multiple gap sites. However, unlike object control verbs, the theme of dative verbs (like transitive verbs) is not dictated by selectional restrictions. Transitive verbs (15) typically take one NP, and they offer only one gap site.

15)Which store$_i$ did the girl visit t_i before her trip?

Boland et al. (1995) conducted a series of five studies investigating the role of verb argument structure in filler gap dependencies using a stop making sense task. The authors manipulated plausibility to see whether the parser consults argument and subcategorization information before forming a filler gap dependency or whether the parser waits for the gap itself to form the dependency; they looked at the three types of verbs just reviewed (i.e. object control verbs, dative verbs, and transitive verbs).

They found that argument structure plays a role in filler gap dependencies, resulting in plausibility effects when the filler is an implausible direct object of the verb, similar to what was reported by Garnsey et al. However, this was only observed when the verb subcategorized for one direct object (i.e. with transitive verbs). With transitive verb types, a

plausibility effect was evident at the verb, which shows that argument structure does play a role in assigning a filler to a gap. The argument structure of object control and dative verbs provide two potential gap sites, and plausibility effects were not found at the verb when there was an implausible direct object of the verb. A plausibility effect was not evident until the second gap site, taken together this led Boland et al. to conclude that the parser consults the argument structure of the verb before positing a gap.

These findings therefore support the assumptions of the Constraint Based Lexical Model with respect to gap filling. The parser uses the argument structure of a verb to evaluate potential gap sites using both syntactic and thematic information. These findings provide evidence that subcategorization and thematic roles are assigned in parallel and are available immediately at the verb. This led the authors to suggest that when a verb has multiple arguments, semantic information affects ambiguity resolution.

Traxler and Pickering (1996) point out that the stop making sense task utilized by Boland et al. does not reflect normal reading, and may reflect strategies that are incompatible with gap filling theories. Traxler and Pickering used a plausibility mismatch paradigm to investigate gap filling and the constraints that dictate them. They found processing difficulty associated with the implausible condition at the verb compared to the plausible condition, suggesting that the parser immediately forms a filler gap dependency at the verb, and thus supporting the Direct Association Hypothesis.

In their second experiment they investigated filler gap dependencies with island constraints. Island constraints are linguistic constraints that block fillers from being associated with gaps (Ross, 1967). The specific details of island constraints are discussed in section 1.3. Traxler and Pickering found that there was no effect of plausibility when the verb was inside an island, suggesting that the parser is sensitive to island constraints. These results taken together suggest that the parser consults the argument structure of the verb and does not wait for the gap site to form filler gap dependencies.

The majority of the studies above have been based on obligatory syntactic assignment in real-time using stop making sense tasks, in the next few studies the authors focus on thematic role assignment using a Visual World Paradigm. The Visual World Paradigm involves participants hearing auditory stimuli while having their eyes tracked during viewing of an array of images depicting objects from the auditory stimuli. This task is more naturalistic and can provide online information about the processing of the entirety of the sentence, and reveals what the participant is anticipating during comprehension.

Altmann (1999) and Altmann and Kamide (1999) investigated if the same obligatory syntactic filler gap assignment is observed in optionally ambiguous dependencies using the Visual World Paradigm. They found that the parser was predicting upcoming thematic roles (as evidenced by anticipatory looks to the direct object) even in the absence of a filler, suggesting that the parser uses verb information to postulate upcoming thematic roles similar to the way it uses the verb to postulate upcoming gap

sites. These findings suggest that the parser is utilizing verbs to predict syntax and context (mediated by real world experience).

Sussman and Sedivy (2001) used a Visual World Paradigm to investigate the projection of upcoming gap sites. They were able to investigate the anticipatory nature of the parser, and whether the parser was actively searching for a gap after encountering a verb. They found that the parser, only in conditions with moved constituents, associated a moved constituent with the first verb it encountered. If this served to be the incorrect gap site participants later shifted their attention towards the correct object. In their second study the manipulated transitivity and found when a verb is transitive (and can take a direct object) the parser actively formed a dependency with the filler. This was not seen with the intransitive verb, as evidenced by similar proportion of looks to all objects in the visual array. These findings taken together show that the parser was actively searching for a gap site after encountering a filler, it also showed that the parser used verb information to restrict what is considered a filler. These findings support a Constraint Based approach to gap filling, given that the parser is utilizing verb information to posit gaps.

Ferreira and Patson (2007) note that approaches that assume several sources of information are being postulated in parallel during sentence comprehension would require a lot of cognitive resources. They argue that this is not a sustainable way for language comprehension to occur given the limited comprehension resources available to the parser at any moment. Rather the parser system seems to make shallow and incomplete representations during processing (Ferreira, Bailey, & Ferraro, 2002). For

example the anomaly in the sentence *The authorities had to decide where to bury the survivors* may be undetected by the comprehender (Barton & Sanford, 1993). Ferreira and her colleagues put forward the idea of a *Good Enough* approach to comprehension in which the parser forms an interpretation during comprehension to ensure continuity, and this interpretation may be incomplete and potentially incorrect in some instances; essentially the parser forms an interpretation that is "good enough".

1.1.5. Summary

Filler gap dependencies have been debated for many years, and there have been many theories put forward to explain how the parser comprehends these constructions. In this section I began by explaining filler gap dependencies, ambiguity, and ambiguity resolution that arises from these dependencies. I also explored the importance of verbs and the role they play in predictive processing. I then discussed transformation based accounts, highlighting the three more prominent accounts: Two-Stage accounts, Lexically-based accounts, and the Good Enough approach. I also discussed non-transformation based accounts of filler gap dependencies and Direct Association Hypothesis and for the remainder of this paper I will make no attempt to tease apart transformational and non-transformation accounts of filler gap dependencies. While there is not a decisive and comprehensive account of comprehension I have highlighted and discussed the strengths and weaknesses of the more influential accounts. In the following empirical chapters I attempt to provide data within the framework of the previous theories in an attempt to put forward a comprehensive account of filler gap dependencies.

1.2. Hyper-active gap filling

As the previous section have outlined, the verb is important in terms of filler gap dependences. English is considered a verb-medial language, meaning that the verb occurs earlier than later in the sentence. Japanese, on the other hand, is a verb-final language, and is characterized by verbs following their arguments. Aoshima, Phillips and Weinberg (2004) found that in Japanese the parser is consulting pre-verbal gap formation to posit upcoming gaps, suggesting that the verb does not need to be encountered to construct an upcoming gap site. The authors found that gap filling was not driven by a need to fill the gap as soon as possible; rather it is driven by the need to satisfy the thematic requirements of the filler as soon as possible. These findings suggest that active gap filling does occur in Japanese, but it is not verb driven, rather gap construction can be driven other grammatical categories.

There is a direct processing cost associated with holding a filler in memory and Omaki, Lau, White, and Phillips (2012) speculate that pre-verbal gap construction seen in Japanese may be a processing strategy to avoid holding a filler in memory until encountering the verb (at the end of the sentence). In English the verb is encountered relatively early in a sentence and while there is a processing cost involved with holding the filler until encountering the verb, the filler can be posited in the gap and released from working memory early in the sentence (e.g. Gibson, 1998). Therefore, waiting to encounter the verb to form a filler gap dependency in English is not as costly as it is in Japanese. This strategy in English allows the parser to consult the verb to form dependencies thereby the parser does not

prematurely construct a gap that may need reanalysis if the verb turns out to be incapable of taking an object (i.e. intransitive verbs). Chapter 4 will specifically investigate whether pre-verbal gap information plays a role in English when predicting upcoming gap sites.

1.3. Island constraints

Filler gap dependencies are considered unbounded in length because multiple clauses can occur between the moved constituent and the trace site. Research has shown that the parser actively searches for the gap site after encountering a filler (e.g. Frazier & Flores d'Arcais, 1989; Stowe, 1986; Sussman & Sedivy, 2003). However, there are constraints that block the association between a filler and potential trace sites. These constraints are collectively known as island constraints (Ross, 1967), and are of interest to both theories of syntax and language processing. Island constraints do not permit a filler to be associated with a gap inside of an island. Essentially, island clauses do not allow for movement in or out of the clause. There are many types of island constraints, and they allow us to investigate the role that grammatical knowledge plays when forming filler gap dependencies. An example of an island constraint can be seen in (16).

16)*Roberta couldn't find the magazine$_i$ that Jane said the editor who read t_i last week.

In this example you can see that the filler *the magazine* cannot be associated with the gap after *read*. This sentence is considered ungrammatical; interestingly, not because of the subcategorization or thematic violations. (16) is considered ungrammatical because of syntactic constraints.

As suggested by Sprouse, Wagers, and Phillips (2012) there are three broad groups of theories that focus on explaining the processing of island constraints: Grammatical, Reductionist, and Grounded theories. Grammatical theories assume that the parser can precisely apply grammatical knowledge of island constraints during online processing (Phillips, 2006). Thus, these theories advocate a representational basis for island constraints that assume they are inherent within the architecture of grammar.

Reductionist theories assume the graded acceptability of these sentences occur because of the demand they place on processing resources. Essentially the harder a sentence is to process, the less acceptable it will be (Deane, 1991; Pritchett, 1991). The complex syntactic construction that occurs when attempting to parse islands overloads processing capacity and ultimately leads to ungrammaticality. (Limitations in this case could involve several different executive functions and cognitive abilities, such as working memory.)

Grounded theories are a combination of Reductionist and Grammatical theories. Like Grammatical theories, Grounded theories argue that the immediate cause of island constraints is inherent in the grammar. Like Reductionist accounts, Grounded theories assume island constraints are ultimately motivated by the parser. Grounded theories make similar predictions to Grammatical theories, and will not be a focus in this body of work.

There have been many studies investigating a variety of island constraints, using a variety of different research methods; several of these

will be outlined below. Results have been mixed. Some studies have shown the parser does not posit a gap in an island (e.g. McElree & Griffith, 1998; Stowe, 1986; Traxler & Pickering, 1996), other studies have shown that the parser will in some cases violate island constraints during online processing (e.g. Freedman & Forster, 1985), and one study has shown both (Pickering, Barton, & Shillcock, 1994).

Stowe (1986) was the first to report that the parser does not attempt to associate a filler with a potential gap in an island. Stowe used a self-paced reading task to investigate active gap filling in sentences that contained island clauses. If the parser was not sensitive to island constraints it would attempt to associate a filler with gap in an island, which would ultimately be ungrammatical. She found that when a gap was embedded in an island no active gap filling was observed; this suggest that the parser is sensitive to island constraints and will not attempt to form a dependency when it is not licensed by the syntax.

As mentioned previously, Traxler and Pickering (1996) used a plausibility mismatch paradigm to investigate filler gap dependencies with island constraints. They found that the parser did not associate a filler with a verb that was embedded within an island thus leading the authors to argue the parser consults the argument structure of a verb (unless it is embedded within an island). While Stowe and Traxler and Pickering found an immediate application of island constraints, it remains unclear how the parser applies these constraints in real-time.

McElree and Griffith (1998) investigated the real-time processing of island constraints in an attempt to better explain filler gap dependencies.

The authors used reaction time and speed accuracy trade off tasks to explore the lexical constraints employed during the association of a filler and a gap. The task looks at the reaction time in detecting whether a sentence is grammatical or not, the authors argue that the time it take to detect various grammatical and ungrammatical strings will reflect when the time in which the violation information is available. McElree and Griffith found island (e.g. **It was the essay that the writer scolded the editor who admitted)* conditions were detected faster than other sentence types. This supports the idea that island constraints block the parser form positing a gap inside an island, as found by Stowe (1986). They also found that detection of subcategorization violations (e.g. **It was the essay that the writer knew the editor had gloated)* was quicker than detection of thematic role violations (e.g. **It was the essay that the writer knew the editor had amazed*). They argue that subcategorization information is employed before thematic information during parsing. Therefore the parser is sensitive to a rich set of syntactic constraints that are employed immediately.

Not all research has shown this sensitivity to island constraints, Freedman and Forster (1985) used a same-different matching task to explore island constraints. This task presents two sentences that match or do not match, and participants are timed in how long it takes to judge if the sentences are the same. This in turn can give insight into overgeneration: if a sentence is overgenerated it will be constructed and a filter will later rule the sentence as ungrammatical. For example sentences like (17) and (18) were presented and participants were timed in how long it took to make a judgement whether they matched.

17)*Who did the duchess sell Turner's portrait of?

18) Who did the duchess sell a portrait of?

(17) is ungrammatical because it violates island constraints, while (18) does not. In this type of matching task ungrammatical overgenerated sentences will be quickly matched with grammatical controls because they are originally parsed as a grammatical sentence (and it is not until later that a filter deems it ungrammatical). The authors found the participants were quickly matching ungrammatical island violation constructions and grammatical controls thus leading them to argue that these sentences are overgenerated by the parser. Therefore the sentences violation island constraints, like (17), are parsed (and thus matched quickly with grammatical control sentences) and are later filtered (and deemed ungrammatical) which lead the authors to argue the parser is not immediately sensitive to island constraints.

Pickering, Barton, & Shillcock (1994) conducted a series of studies employing self-paced reading and eye-tracking to investigate the time course of island constraints. In their first study they found an increased processing cost associated with an embedded verb regardless whether it was in an island construction or not. However, in experiment 2, the authors found that the parser was sensitive to island constraints. While these results are conflicting the authors argued that the parser immediately forms filler gap dependencies despite island constructions. Therefore, they claim their data supports the idea that island constructions are not part of grammar; rather they are filtered out after the first parse.

Three studies have used event-related potentials (ERPs) to investigate island constraints, all of which found a sensitivity to the boundary

of an island (Kluender & Kutas, 1993; McKinnon & Osterhout, 1996; Neville et al., 1991). McKinnon and Osterhout, for example, found that in sentences with subjacency violations there was a P600 effect, this effect has been shown to reflect difficulty with building semantic representations (i.e. Osterhout & Holcomb, 1992; 1993). Neville et al. (1991) and Kluender and Kutas (1993) also found a similar sensitivity to island constraints; these findings suggest that movement constraints are employed very rapidly. However, ERPs only tell us that a syntactic anomaly was detected, it does not tell us what is happening after detection. The parser seems to be sensitive to island boundaries but it is unclear if the parser attempts to construct a dependency within an island.

While reports have been conflicting, it is generally believed that the parser is in fact sensitive to island constraints. However, how these sentences are incrementally parsed and interpreted online are highly debated. In their recent paper Sprouse et al. (2012) attempt to adjudicate between Reductionist and Grammatical theories; they focus on the most prominent Reductionist approach, which is the capacity-based theory (Kluender and Kutas, 1993). This theory argues that island constraints arise due to the resources associated with forming both a long distance dependency and building an island structure. The simultaneous processing of these two things taxes the limited processing pool available and ultimately lead to island constraints. Grammatical theories, on the other hand, assume island constraints arise through grammar. If island constraints arise due to processing capacity, the authors logically argue, individuals with greater processing capacity should have the resource capacity to form these

constructions and thus they should deem more of these types of constructions acceptable than those with smaller processing capacities. If island constraints are grammatical in nature than the processing capacity of a participant should play no role in the acceptability of island constructions.

In their study Sprouse et al. used a serial recall task and an n-back task as a measure of processing capacity (working memory) for each participant. They used a differences-in-differences measure to determine the strength of island constructions and two acceptability rating tasks. They found no relationship between working memory capacity and acceptability of island constructions, leading the authors to argue island constraints are a result of grammar, not a result of processing resources.

1.2.1. Summary

In this section I discussed a set of constraints that block filler gap dependencies from being formed, known as island constraints. I provided empirical evidence that the parser is sensitive to island constraints, as well as evidence that the parser is not sensitive to island constraints. I then discussed some ERP research; while there is evidence that the parser is sensitive to the boundaries of island constructions it is unclear what is occurring after the detection of the boundary. I then provided the three prominent accounts of island constraints: Reductionist, Grammatical, and Grounded accounts, and provided some evidence that working memory plays no role in the acceptability of island constraints, suggesting that island constraints my be grammatically dictated.

1.3. Parasitic gaps

Interestingly, there are some exceptions to island constraints, called parasitic gaps. Parasitic gaps are a constructions in which a gap occurs with an in island and an additional gap occurs that is not within an island, and both of these gaps must be associated with the same filler (as illustrated by examples 19-21). Example (19) is ungrammatical because a gap occurs inside an island. However, when the illegal gap in (19) is combined with the legal gap in (20) a grammatical construction is formed (i.e., example 21). It is the combination of two gaps that creates the exception and "rescues" what should be and ungrammatical or illegal structural association.

19) *What did the attempt to fix _ ultimately damage the car?

20) What did the attempt to fix the car ultimately damage _?

21) What did the attempt to fix _ ultimately damage _?

Parasitic gaps pose an interesting set of problems for comprehension. First, they show that gaps can occur within an island, but only when combined with another gap. Second, the illegal gap typically occurs prior to the legal gap and the comprehension system must keep track of multiple gap sites in order to determine the correct association. If island constraints are Reductionist in nature we assume that the parser will not attempt to associate the filler with the first gap in (21), given that this construction is cognitively too hard to parse. Grammatical theories, on the other hand, assume that island constraints are represented within grammar and violations of island constraints are possible if represented by the grammatical construction (i.e. parasitic gaps).

Parasitic gaps are therefore ideal constructions to test Grammatical and Reductionist theories. As stated above Reductionist based theories assume that island constraints are the result of processing limitations and therefore predict that under no circumstance should the island constraint be violated, given that they are cognitively too complex to construct. Pritchett (1991) for example assumes a head-driven parsing architecture which does not allow for a filler and a gap to be associated within a subject noun-phrase. He argues that island violations are not ungrammatical per say, rather they lead to a processing overload that ultimately leads to ungrammaticality. He suggests that island violations are a type of Garden Path that results from local parsing decisions that are inconsistent with the global grammatical representation; these parsing decisions are ultimately wrong and unrecoverable. In that light, the parser should not be able to construct a gap within an island, even if allowed by a parasitic gap. As Phillips (2006) points out, if the parser is able to posit a gap within an island, it is hard to argue island constraints are a result of processing limitations.

Grammatical theories, on the other hand, assume that the parasitic gap constructions are represented within grammar, and when allowed for (i.e. parasitic gap constructions), the parser will posit an gap within an island. However, if parasitic gaps are inherent within grammar, there is still the question of how these constructions are parsed. The “illegal” gap within an island is only made legal by a gap that occurs later in the sentence, and this presents a problem for real-time and incremental language comprehension. The parser may posit a gap within an island only after encountering the gap that would form a parasitic gap construction. Or the parser may posit a

parasitic gap within and island as soon as it encounters a verb that is capable of hosting this type construction. If the latter, the parser would be employing a very rich set of syntactic rules and would be in direct opposition of the Reductionist based theories.

Phillips (2006) conducted a series of studies to investigate how parasitic gaps were constructed. The first study was an acceptability study paradigm to investigate different filler gap dependences, including “good gap”, “bad gap” and parasitic gap-like constructions (with both types of gaps) using two verb types. He looked at infinitival verbs which can host a parasitic gap construction (examples 22-24) and finite verbs which cannot host a parasitic gap construction (examples 25-27).

22) The outspoken environmentalist worked to investigate what the local campaign to preserve the important habitat had harmed____ (good gap, infinitive).

23) The outspoken environmentalist worked to investigate what the local campaign to preserve ____ had harmed the migration (bad gap, infinitive).

24) The outspoken environmentalist worked to investigate what the local campaign to preserve ____ had harmed____ (both gaps, infinitive).

25) The outspoken environmentalist worked to investigate what the local campaign that preserved the important habitat had harmed____ (good gap, finite).

26) The outspoken environmentalist worked to investigate what the local campaign that preserved ____ had harmed the migration (bad gap, finite).

27) The outspoken environmentalist worked to investigate what the local campaign that preserved ____ had harmed____ (both gaps, finite).

He found that participants were correctly judging the items in the good gap condition as acceptable, and the items in the bad gap condition as unacceptable. He also found that the infinitive both gap condition (which can host a parasitic gap construction) showed a similar acceptability rating as the good gap infinitive condition; this pattern suggests that parasitic gaps are acceptable constructions and that parasitic gap are only acceptable when they occur within a certain subclass of islands (i.e. with infinitive verbs)

In his second study Phillips used a self-paced word-by-word reading paradigm with parasitic-gap like constructions (examples can be seen in Table 1), he found that the parser did not posit a gap in an island unless it was warranted by the verb (i.e. parasitic gap construction). The results are in direct contradiction to Reductionist theories that assume island constraints are the result of processing costs. It appears that the parser is incrementally building these constructions despite the fact that the parasitic gap occurs before the gap that licenses it. These findings support a Grammatical based theory, in which the parser is incrementally employing a very rich set of syntactic constraints in real-time. Parasitic gaps will be explored in more detail and is the focus or the first empirical chapter, chapter 3.

1.3.1. Summary

In this section I discussed an exception to island constraints known as parasitic gaps. Parasitic gaps are constructed in a way that a gap is posited in an island as a result of a legal gap that occurs in later in the sentence, these constructions allow us to test two competing processing theories: Reductionist and Grammatical theories. In terms of parasitic gaps Reductionist accounts assume positing a gap in an island (even when in a parasitic gap construction) will be cognitively too difficult, while Grammatical theories assume that the parser can posit a gap in and island (when warranted by a parasitic gap construction). Phillips (2006) provided evidence that the parser is capable of posting a gap in an island construction only when warranted by a parasitic gap construction; these findings give evidence for a Grammatical based theory of island constraints.

1.4. Conclusions

There are several theories that have been proposed to explain how the parser constructs and comprehends filler gap dependencies and the constraints that govern them. Above, I have provided evidence supporting several theories; the most prominent of these are the Two-Stage models, Lexically-based models, and the Good Enough account. Two-Stage models (e.g. Frazier, 1987; Frazier & Flores, 1989; Stowe, 1986) assume that the parser forms a gap at the first potential gap site without consulting thematic and semantic information from the verb. Lexically-based models (e.g. Boland et al., 1995; Fodor, 1978; Sussman and Sedivy, 2001; Traxler & Pickering, 1996) assume that the parser takes into account thematic and semantic verb information when parsing and will posit a gap based on this

information. I also discussed the Good Enough strategy (e.g. Ferreira, Bailey, & Ferraro, 2002) of comprehension, in which the parser merely forms a “good enough” linguistic representation. While there is not comprehensive account of filler-gap dependencies the empirical chapters in this thesis attempt to adjudicate between these competing theories of filler gap dependency processing.

I have also discussed the role that verbs play in postulating upcoming gaps in English, while touching upon pre-verbal gap creation as observed in verb-final languages. Most research has shown a filler-driven strategy for gap filling in verb-medial languages (e.g. Stowe, 1986) but it has been found that preverbal information can be used for verb-final languages (e.g. Aoshima, Phillips and Weinberg, 2004) which may represent a strategy shift associated with the late position of the verb. Chapter 4 uses pupillometry in an attempt to investigate whether pre-verbal gap creation is evident in English or whether it is a processing strategy apparent only in verb-final languages.

I have also presented research and theories accounting for island constraints and exceptions to these constraints known as parasitic gaps. Island constraints and parasitic gaps provide an interesting problem for theories of parsing as well as providing testing grounds to investigate competing theories. Two theories were put forward to explain the graded acceptability of island constraints: the Reductionist theory (that assumes island constructions are ungrammatical because they are cognitively too difficult to parse) and the Grammatical theory (which assumed that island constraints arise from grammar). Parasitic gap constructions are formed by

positing a gap in an island and are only made grammatical by a proceeding gap. Phillips (2006) provide evidence that the parser will posit a gap in an island but only when warranted by a parasitic gap construction; these results are difficult to explain via a Reductionist approach (given that a gap in an island should be too difficult to construct). In Chapter 3, pupillometry is used to investigate parasitic gap constructions in an attempt to tease apart on Grammatical and Reductionist approaches to island constraints and parasitic gap constructions.

CHAPTER 2: PUPILLOMETRY

The purpose of this chapter is to provide a thorough review of pupillometry; in particular, I focus on what pupillometry can reveal about language comprehension. Pupillometry is essentially the study of pupil size changes. The constriction and dilation of the pupil is believed to directly reflect processing associated with mental work, and can be measured using an eye tracker. This chapter starts with an overview of the classic studies of pupillometry focusing on the seminal work of Hess and Kahneman. This section focuses on Kahneman's criteria regarding physiological measures of cognitive processing load (i.e. within-task variation, between-task variation, and between-individual variation). This review demonstrates the validity and usefulness of pupillometry. The second section of the chapter what the pupil reveals about language. Finally, the third section focuses on methodological issues and the way in which the pupil data, in this thesis, was analysed.

2.1. Pupillometry

2.1.1. Classic pupillometry studies

The pupil is the hole in the iris, and two sets of efferent ganglion control the smooth (dilator and sphincter) muscles within the eye. Essentially, pupil size is determined by a balance of tonus between these two types of muscles. The size of the pupil typically ranges from 1.5 – 8 mm. Constriction-related ganglia are parasympathetic fibres from the oculomotor nerve, and dilation-related ganglia are sympathetic fibres from a cervical ganglia. Pupillometry is

essentially the measurement of pupil diameter change in response to a stimulus. Pupil size can be thought of as the tachometer of the brain, essentially reflecting processing load (Hess & Polt, 1960; Kahneman & Beatty, 1966). Hess and Kahneman are the seminal researchers in the study of pupillometry, and Hess (1975) coined the term *pupillometrics* to describe the evaluation of pupil diameter as a direct measure of processing cost. Their work, beginning in the 1960s, revealed the usefulness of the pupil to measure cognitive processing costs associated with a range of behavioural tasks.

2.1.2. Kahneman's criteria

Kahneman (1973) suggested that for physiological measures to be valid measures of cognitive processing they need to meet certain criteria. First, they should show variation within a task depending on level of task difficulty. Second, they should show that different tasks that employ different cognitive operations reflect different processing loads. Third, they should reflect individual differences in the allocation of effort to different tasks. I will begin with a brief review of the classic literature supporting Kahneman's criteria; I will then review pupillometry in language comprehension, methodological issues in pupillometry research, and finally, relate this to research with filler gap dependencies.

Many studies have shown within-task variation in a range of cognitive tasks. I will briefly mention some of the classic studies that explored short-term memory (Kahneman & Beatty, 1966), mental arithmetic (Hess & Polt, 1964), and pitch discrimination (Kahneman & Beatty, 1967). In a short-term memory study, Kahneman and Beatty (1966) aurally presented a string of

digits (from 3-7) and found that pupil diameter increased with each digit presented, and during recall, the pupil diameter decreased with each digit recalled. They also found a greater pupil slope (pupil slope is the is the slope of the pupil diameter change over time) associated with longer strings of digits (i.e., a 7 digit presentation had larger pupil slope than did a 3 digit presentation). These results support within-task variation, as both the number of digits presented and difficulty were reflected in pupil diameter change.

In terms of mental arithmetic, Hess and Polt (1964) were the first to show the pupil is sensitive to mental multiplication. In their study, the pupil increased in relation to the difficultly of the multiplication problem. They presented their results as percentage of dilation, and found an ordered effect of problem difficulty reflected in pupil dilation percentage, thus reflecting sensitivity to within-task variation.

Kahneman and Beatty (1967) conducted a pitch discrimination study, in which a standard tone was presented to participants, and after a short interval, additional tones were individually presented at varying frequencies. The participants had to judge if the additional tones were higher or lower than the standard tone. They found that pupil size was smaller for easy discrimination and larger for hard discrimination. Thus, difficulty of the task was inversely related to the pitch of the second tone, which also supports within-task variation.

The second criterion is between-task variation; an ordered effect of processing effort across different tasks is expected. Beatty (1982) highlights

the pupil variance elicited by different tasks. Hard multiplication elicits the largest peak response, followed by memory tasks, language tasks, and perception tasks evoke the lowest peak response. This provides evidence to support Kahneman's second criterion of between-task variation.

The third criterion is between-individual variation. This was shown in the work of Ahern and Beatty (1979; 1981). They used scores on a U.S. aptitude test, known as the SAT, to group participants into high and low intelligence. They then used a variety of tasks (i.e., vocabulary, multiplication, digit recall, and sentence comprehension) with both hard and easy levels of each to investigate the relationship between pupil size and intelligence. They found that more difficult tasks evoked larger pupil responses, and that in three of the four tasks, the high intelligence participants showed smaller pupil responses than those in the low intelligence group. These studies provide evidence to support Kahneman's third criterion (i.e., the measure should be sensitive to individual differences in cognitive performance).

This brief review of classic pupillometry research provides evidence that pupillometry is an effective measure of cognitive processing effort, and that pupillometry adheres to the criteria proposed by Kahneman. There have been many technological advances since the 1960's; however, Hyönä, Tommola, and Alaja (1995) point out that while pupillometry research flourished in the 1960's and 1970's there was little work in the 1980's and 1990's. This is surprising given the usefulness of the measure and the widespread use and affordability of eye trackers, which makes recording and

analysing pupillometry data much easier. Since Hyönä et al.'s paper, there has been a steady resurgence of pupillometry studies in the literature.

2.2. Pupillometry and language review

In this section I turn to language research, and I highlight what pupil diameter changes can reveal about language comprehension from both theoretical and methodological perspectives. I begin with research at the letter level, and then move on to word level, and sentence recall. In the last section, I focus on syntactic processing and translation in both comprehension and production. The final section considers pupillometry in combination with measures of event-related potentials.

2.2.1. Letter encoding

At the orthographic level, Beatty and Wagoner (1978) investigated pupil dilation during visual letter encoding. Letters were either orthographically similar (DD or dd) or orthographically dissimilar (Dd), and participants were asked to judge whether the letters were the same or different. Results showed larger pupil size when the letters were orthographically dissimilar compared to orthographically similar. In a second experiment, participants had to judge orthographic similarity, and whether the letters matched in letter category (i.e. vowels and constants); Beatty and Wagoner found that pupil dilation was largest in the category mismatch condition. The orthographically dissimilar condition showed less dilation, and finally the smallest pupil response was in the matching condition. These findings reveal different levels of processing load involved with the encoding of letters both in terms of orthography and letter category.

2.2.2. Letter and word recall

Early language research using pupillometry investigated short-term memory in sentence comprehension, including both full and partial sentence recall. Studies also looked at word recall, including both full and partial list recall. Wright and Kahneman (1971) investigated whether pupil diameter reflected syntactic phrase boundaries; they hypothesized that sensitivity to phrase boundaries could benefit psycholinguistic research, and allow researchers to gain information about the allocation of mental effort during online sentence comprehension. In their experiments participants heard a sentence or list of words, and were asked to recall some or all of the stimuli. In the sentence comprehension experiment, participants were also asked a question about the sentence. The stimuli were presented with a pause of either 3 or 7 seconds, and pupil diameter was measured for the duration of the trial.

Results were similar for both studies: pupil dilation was larger in the recall conditions compared to the question condition for both pause lengths. However, the materials were different between the two studies: one employed sentence recall and the other word recall. These findings were expected for word recall because of rehearsal strategies, but were somewhat unexpected for sentence recall. The findings for sentence recall were believed to result from the effort that was reflected in retention and rehearsal for repetition, as opposed to the comprehension question condition. Unexpectedly, Wright and Kahneman found no evidence of sensitivity to syntactic phrase boundaries. However, they argued that pupillometry is a less contaminated dependent variable compared to behavioural responses, such as comprehension questions. While there were

no syntactic results in this particular study, the next section will focus on research that has shown pupil sensitivity to syntactic manipulations.

2.2.3. Syntax

Schluroff (1982) conducted a study that investigated pupil diameter change in relation to syntactic complexity. He was interested in the reliability between participant ratings of comprehensibility and what was reflected in pupil diameter. Schluroff found that the more complex the sentence, the larger the pupil dilation. Additionally, he found that comprehensibility ratings were not reliably related to pupil size, suggesting that pupil diameter is a better indicator of grammatical complexity than participants' subjective ratings. In another study Schluroff et al. (1986) explored syntactic complexity and processing effort. In this study, participants heard ambiguous sentences with verb-oriented and subject-oriented readings, and were required to transform the sentences into passives. For example, a subject-oriented passive voice sentence is *The man on the motor bike was chased by Peter,* and an example of an object-oriented passive sentence is *The man was chased by Peter on the motor bike*. Thus, the task elicits either a verb-oriented reading (which is grammatically more complex) or an object-oriented reading (which is grammatically less complex). They found that the verb-oriented reading evoked larger pupil diameter compared to the less complex object-oriented sentences. This shows the usefulness of pupillometry as a measure of cognitive effort associated with syntactic processing, and provides evidence that pupillometry is a better indicator of grammatical difficulty than offline ratings.

Ben-Nun (1986) investigated pupil diameter changes in response to lexical ambiguity resolution. In particular, he focused on homophone ambiguity which is the ambiguity that arises from words that are spelled the same but have different meanings (e.g. *The man threw the **straw** on the floor* vs. the *The farmer threw the **straw** on the floor*). Ben-Nun found that when homophones were presented in an ambiguous context they evoked larger pupil sizes compared to when the homophones were presented in an unambiguous context. He also found a spike in pupil diameter at the point of the homophone. Additionally, he varied the task to investigate the depth of processing involved with homophone ambiguity. The tasks were to define the homophone, to recall it, or to choose-the-best-meaning (in either an ambiguous or unambiguous context). He found that in the choose-the-best-meaning condition there was a post-sentential pupil increase, which he attributed to integration processes and a greater depth of processing.

Just and Carpenter (1993) conducted a series of studies measuring the intensity in which mental resources are consumed using task evoked pupil responses during sentence comprehension. In their first experiment, they looked at object relatives (e.g. *The reporter that the senator attacked admitted the error.*) and subject relatives (e.g. *The reporter that attacked the senator admitted the error.*). Object relatives are more difficult to process for several reasons; they have non-canonical word order, an embedded clause, and atypical thematic roles (as mentioned in the previous chapter thematic roles refer to the semantic characteristics of the verb). The subject relatives on the other hand are easier to process because of canonical word order and more typical thematic role assignment.

Just and Carpenter found that object relative sentences elicited larger pupil size compared to subject relative sentences. The peak in pupil diameter occurred approximately 1000 ms after the main clause verb, the main clause verb is the point of greatest processing demand in both conditions. When the authors adjusted for first pass reading time, the point at which pupil diameter was largest was between 1250-1475 ms, which is similar to the findings of Ahern and Beatty (1979). Difficulty with object relatives was also reflected in gaze duration and error rates. The larger pupillary response to object relatives compared to subject relatives indicates that more resources are being consumed to process these sentence types. Therefore, Just and Carpenter concluded that pupil diameter changes are a viable and sensitive measure of the processing costs associated with sentence complexity.

In their second study, Just and Carpenter investigated unbounded dependencies involving a moved wh-phrase, and a plausibility-mismatch paradigm (see examples 28-31). To my knowledge this is the only peer-reviewed paper investigating unbounded dependencies using pupillometry and it has set the stage for the studies in this thesis.

28) The confused police didn't know which leader$_i$ the rioters followed t_i noisily down the street after the meeting (plausible, movement).

29) The confused police didn't know whether the rioters followed the leader noisily down the street after the meeting (plausible, no movement)

30) The confused police didn't know which blanket$_i$ the rioters followed t_i noisily down the street after the meeting (implausible, movement)

31) The confused police didn't know whether the rioters followed the blanket noisily down the street after the meeting (implausible, no movement)

Just and Carpenter hypothesized that holding the moved constituent in memory would evoke greater pupil dilation as would a plausibility mismatch.

The mean pupil diameter was measured beginning at the first fixation to the verb in the subordinate clause and ending 3 seconds later or when the end of the sentence was encountered. Just and Carpenter found there was a larger pupil change in the condition with a moved constituent (i.e. a wh-phrase) when compared to the condition that did not involve movement. They also found that the implausible conditions evoked larger pupil dilations as compared to plausible conditions. Their results demonstrated that pupil diameter changes are sensitive to both difficulty of a moved constituent and sensitive to plausibility manipulations. These findings have set the stage for the next two empirical chapters as they provide evidence that pupil diameter change increases when holding a filler in memory.

Engelhardt, Ferreira, and Patsenko (2010) investigated pupil change in a syntactic ambiguity resolution task that manipulated prosody and visual context. The goal of the study was to investigate whether prosody or visual context has a larger effect on processing effort, and whether the two information sources interact during online comprehension. In their first

experiment Engelhardt et al. varied the prosody within a Garden Path sentence, prosody had three levels: cooperative with 200 ms break, cooperative with 400 ms break, and conflicting prosody with no prosodic break. Participants were asked a comprehension question following the sentence. For example participants would hear *While the woman cleaned* (#) *the dog that was big and brown stood in the yard*, and the (#) represents the location of the manipulated prosodic break. The question was either a "yes" question (e.g. *Did the dog stand in the yard?*) or a "no" question (e.g. *Did the woman clean the dog?*). They found that pupil diameter increased with conflicting prosody as well as eliciting more incorrect final interpretations, while cooperative prosody showed a flat or slightly negative pupil slope that led to more correct final interpretations.

In their second experiment the authors used the same auditory stimuli with three levels of prosody (no prosodic break, a 200 ms break, or a 400 ms break) but they added visual context, which consisted of images that either conflicted or were consistent with the correct interpretation of the sentence. This allowed them to investigate how processing load and final interpretations were affected by both prosody and visual context. In terms of final interpretations, there were more incorrect interpretations with conflicting prosody, and in terms of pupil slope there was an interaction between prosody and visual context, which was driven by the inconsistent visual context. The cooperative prosody with the shorter prosodic break and the conflicting prosody condition did not differ. The cooperative prosody with the longer break did differ from the conflicting condition, but this was not reflected in comprehension accuracy. This highlights the difference between

online and offline performance, thus showing that pupillometry is sensitive to syntactic ambiguity resolution, as well as showing the influence of visual context during online sentence processing.

In a more recent study investigating subject relative, object relatives and digit recall, Piquado and colleagues looked at pupil responses in both younger and older adults (Piquado, Isaacowitz, & Wingfield, 2010). Age is of particular interest because in general, older adults have smaller pupil sizes; a smaller pupil size may indicate less responsiveness and suggest that pupil diameter is not a useful tool for measuring cognitive processing costs in older adults (i.e. it may underestimate their cognitive effort). In the digit recall task participants were asked to retain an aurally presented list of digits to recall after a 3 second retention period. Similar to previous research the authors found, for both age groups, an increase in pupil size with each digit heard, and pupil size decreased with each digit recalled. Because Piquado et al. adjusted pupil sizes based on a pre-test, they found that older adults actually exhibited larger pupil sizes during encoding and retention which suggests that older adults invested more cognitive resources compared to the younger adults. However, when they analyzed the data without adjusting for baseline pupil size, they found that younger adults exhibited larger pupil sizes than the older adults, which leads to the incorrect interpretation that younger adults invested more cognitive resources compared to older adults.

In their second experiment, Piquado et al. (2010) investigated pupil change in both older and younger adults, using the same sentence types as Just and Carpenter (i.e. subject and object relatives). However in this study they added additional modifiers to their stimuli to increase memory demands

during sentence recall. The participants heard object relative and subject relative sentences that were either 9 or 12 words long, and participants were asked to recall the sentences as completely and accurately as possible after a 2 second retention interval. After normalizing pupil size for both age groups, the authors found pupil size was larger during retention for younger adults in the object relative condition, with longer sentences evoking larger pupil diameter. The older adults only showed a change in pupil diameter when comparing the different sentence lengths. The lack of difference in pupil diameter between the two sentence types may reflect the linguistic experience that comes with age.

What these two experiments tell us about pupil change is that increasing memory load (be it digits or words) is directly measurable across age group by pupil change. The sentence data provide evidence that older adults do not invest the same processing resources compared to younger adults. The study also highlights the importance of establishing a baseline pupil diameter, particularly when comparing participants of different ages. On a higher or more global-level, the next section focuses on translation with simultaneous comprehension, interpretation, and production.

2.2.4. Translation

In terms of language interpretation and translation, Hyönä, et al. (1995) conducted a series of studies investigating pupil change during interpretation and translation between English and Finnish. Their first study involved either listening to a passage with no other task, listening to a passage and repeating it back, or simultaneous translation of English into Finnish. (In the studies all participants were native speakers of Finnish). Hyönä et al. found

that the translation condition evoked the largest pupil sizes, followed by the repeating condition, and then the listening condition. The second experiment consisted of single words instead of passages of text which allowed the authors to investigate moment-to-moment processing as opposed to overall processing from the whole text (as in Experiment 1). In the reading condition, participants responded with a “yes” after word presentation. In the second condition, they had to read the word aloud. In the third condition, they had to give the meaning of the word presented (stimuli were presented in either English or Finnish to be translated). Results were similar to the first study, with the largest pupil size in the translation condition, followed by repetition, and the smallest pupil size was observed in the recognition condition. They also found words that were harder to translate evoked a larger pupil size than easy to translate words, and there was a smaller pupil change associated with repeating Finnish words compared to repeating English words. These results highlight the sensitivity of pupillometry as a measure of cognitive processing when simultaneously interpreting and translating.

2.2.5. Pupillometry and ERPs

Pupillometry has also been used in conjunction with other online measurements, such as event-related potentials (ERPs). This work provides information about brain responses and processing costs associated with cognitive processing. Gutiérrez and Shapiro (2010) investigated pupil changes in response to subject-verb violations, filler gap dependencies, and subcategory violations (e.g. **The dog that runs fast asleep the cat around the yard*). They found that subject-verb violations and filler gap

dependencies yielded pupil changes that corresponded to changes in ERP waveforms. Essentially, the time in which the pupil dilates after the disambiguating word corresponds to the Left Anterior Negativity (for subject-verb violations) and Early Left Anterior Negativity (for filler gap dependencies). These results show that pupil diameter change is sensitive to the information being processed and can be time locked to particular ERP components. The subcategory violations refer to violations to the arguments licensed by the verb, on the other hand, showed an immediate pupil diameter decrease, which may indicate processing overload.

In line with this research, Kuipers and Thierry (2011) conducted a study where they presented picture-word pairs, some matching and some mismatching, and recorded pupil change and ERPs simultaneously. The authors anticipated an N400, which reflects semantic anomaly detection with picture-word mismatch (Lau, Phillips, Popple, 2008). However, Kuipers and Thierry were unsure whether mismatches would be reflected in pupil change. As predicted, they found an N400 in the mismatch condition, and found increased pupil size in the mismatch condition. The pupil size reflected the processing cost associated with a mismatching word-picture combination. They also found a significant negative correlation between N400 amplitude and pupil size (an N400 response is a negative deflection). These results can be interpreted as a direct connection between ERP amplitude and pupil diameter size, suggesting that pupil diameter is correlated with brain responses during language processing. The above sections have shown the usefulness of pupillometry, particularly in terms of

language research. In the next section, I focus on the methodology and the analyses associated with pupil slope.

2.3. Methodology

The pupil is inherently variable, and can change in response to light, emotion, and cognitive variables. These factors can make it difficult to differentiate pupil changes in response to a task or to some other confounding variable. Infrared cameras have been the most popular way to measure the pupil. Infrared light illuminates the pupil and provides a measure of the number of pixels the pupil encompasses (Klingner, 2010).

2.3.1. Pupil slope

Traditional statistical techniques have investigated the averaged pupil response curve (Kahneman & Wright, 1971; Wright & Kahneman, 1971), which is often time locked to a critical point in a stimulus (Beatty, 1982); this can then be compared across conditions. In terms of language comprehension, Just and Carpenter (1993) looked at the change from baseline to peak pupil size after the critical word in the sentence. In their study the time to peak was approximately 1200 ms, and this has been taken as the standard time required for peak amplitude following a point of difficulty in a sentence (Beatty, 1982; Just & Carpenter, 1993). Pupil measures are often coupled with gaze duration or some other (overt) behavioural task (e.g., comprehension question responses, mathematical computation, etc.).

More recently, Engelhardt et al. (2010) analyzed pupil slope by first filtering out blinks using linear interpolation, and then pupil diameter was analyzed in a 1.2 second time window from the critical point in the sentence

(Just & Carpenter, 1993). Individual trials were averaged together and the resulting vectors for each participant were submitted to a simple regression, in which time was the independent variable and pupil diameter was the dependent variable. The unstandardized regression coefficient (i.e. the linear slope of the pupil diameter change over time) served as the dependent variable. While there are many ways to analyze pupil data, pupil slope has been shown to be an effective technique and will be used in this thesis.

2.4. Conclusion

The studies above have shown the usefulness of pupillometry to measure processing costs across a wide range of cognitive tasks. The latter part of the chapter has demonstrated the usefulness of pupillometry to investigate different facets of language processing. Not only has it been shown to be sensitive to syntactic difficulty, syntactic ambiguity resolution, and prosodic manipulations, but it has also been shown to correspond to underlying brain activity.

CHAPTER 3: PARASTIC GAPS

3.1. Background

This chapter focuses on the parsing and comprehension of parasitic gap constructions. As mentioned above parasitic gap constructions are exceptions to island violations; these constructions are formed when an illicit gap in an island clause (32) is combined with a later occurring acceptable gap (33) to form an acceptable parasitic gap construction (34).

32) *What did the attempt to fix _ ultimately damage the car?

33) What did the attempt to fix the car ultimately damage _?

34) What did the attempt to fix _ ultimately damage _?

These dependencies are interesting because they are evidence that violating and island constraint is not always illegal. They also pose a problem from theories of incremental processing given that the gap that makes this construction legal occurs after the illegal gap; this presents a look-ahead problem for the parser.

Two main accounts have been put forward to explain parasitic gap constructions: Grammatical accounts and Reductionist accounts. Grammatical accounts assume that theses constructions are inherent to the grammar of the language while Reductionist accounts assume these constructions are simply too difficult to parse. If island constraints are more processing based then we assume that the parser will not attempt to associate the filler (i.e. *what*) with the first gap in (34). Grammatical theories,

on the other hand, assume that island constraints are represented within grammar and violations of island constraints are possible if represented by the grammatical construction (i.e. parasitic gaps). If the comprehension system attempts to associate the wh-filler "*what"* with the gap in an island (32) – (34), then there should be a noticeable processing cost compared to non-parasitic constructions. In the parasitic gap construction (34), if the system does associate *what* in the first gap then there should be measurable processing costs and it could be argued that the parser is predicting a parasitic gap construction; thus supporting Grammar based theories. If there is no evidence of positing the filler in the first gap of (34) then it can be argued that the parser does not violate island constraints without confirming the parasitic gap construction; thus supporting processing-based accounts.

In a self-paced reading study that investigated parasitic gap constructions, Phillips (2006) argued that participants spent more time reading a verb in an implausible condition compared to a plausible condition (see Table 1) because they were immediately creating a dependency between the wh-phrase and the verb in the subject NP. The plausibility mismatch led to a slowdown in reading time. The other conditions produced more similar mean reading times which would suggest that in these conditions participants did not create the dependency and thus favour a grammatical representation of parasitic gaps. I conducted a study which measured people's pupil diameter as they processed sentences similar to the ones used by Phillips (2006) in order to further investigate Reductionist and Grammatical accounts of parasitic gap.

3.1.1. Current study

In this study the critical windows occurred following (1) the onset of the verb in the first embedded clause and (2) the word following the verb, which is referred to as N + 1. The first time window is important because it allows us to test whether the parser is forming a dependency within and island condition (we would only expect a dependency to be formed only if the parser is anticipating a later occurring parasitic gap). The second time window allows us to investigate any spill over effects that may occur after the parser encounters the gap (again we only anticipate spill over effects if the parser forms a dependency between the filler and the gap occurring within an island). Grammatical accounts assume that the parser will form a dependency with the critical verb only when warranted by syntax (i.e. with and infinitive verb), and we would expect to see a decrease in pupil size at the two critical windows as the filler would be released from working memory. Reductionist accounts assume that the parser will not be able to posit a gap within an island so we would expect to see in an increase in pupil size between the two time windows in all conditions given that the parser must maintain the filler in working memory. I will discuss this in a bit more detail below.

In the critical item the wh-phrase (i.e. "*which*") was either an implausible or plausible direct object of the embedded verb in a subject island clause (refer to Table 1). The embedded verb within the island was either infinitive (e.g. *expand*) or finite (e.g. *expanded*); infinitive islands permit parasitic gap constructions, while finite subject islands do not. Given the results from Phillips (2006) I assume a Grammatical account, and

therefore predict the parser is sensitive to parasitic gap constructions (and thus will posit a gap in an island construction) and there will be an increase in processing load (and thus an increase in pupil size) in the implausible (infinitive) subject islands because the embedded verb is an implausible object of the wh-filler. However, once the gap has been identified there should be a decrease in processing load (and thus a decrease in pupil size) as the parser no longer has to hold the filler in working memory. There should not be an effect of plausibility (with finite verbs) as the embedded clause in the subject noun phrase does not support a parasitic gap. However, there should be an overall higher processing cost compared to the infinitive conditions given that the parser has to keep the filler in working memory. However, if the parser does not permit parasitic gap structures but instead relies on island constraints (as predicted by Reductionist accounts) then we expect to see no differences between the finite and the infinitival conditions as the verbs are embedded within an island.

Table 1. Example sentences and predicted processing for study 1.

	Example sentence	Predicted processing effect
Infinitive Plausible	The superintendent learned **which schools** the proposal to **expand** upon the curriculum would overburden ____ during the following semester	There will be a decreased processing cost/pupil size: infinitival verbs allow for parasitic gap constructions and the wh-filler can be associated with the verb (*expand*) and therefore no longer maintained in working memory. Plausibility will play no role.
Infinitive Implausible	The superintendent learned **which high school students** the proposal to **expand** upon the curriculum would motivate ____ during the following semester	There will be a decreased processing cost/pupil size: infinitival verbs allow for parasitic gap constructions and the wh-filler can be associated with the verb (*expand*) and therefore no longer maintained in working memory. There will be a plausibility mismatch.
Finite Plausible	The superintendent learned **which schools** the proposal that **expanded** upon the curriculum would overburden - ____ during the following semester	There will be an increased processing cost/pupil size: finite verbs do not allow for a parasitic gap construction and the wh-filler must be maintained in working memory. Plausibility will play no role.
Finite Implausible	The superintendent learned **which high school students** the proposal that **expanded** upon the curriculum would motivate ____ during the following semester	There will be an increased processing cost/pupil size: this construction does not allow for a parasitic gap construction and the wh-filler must be maintained in working memory. Plausibility will play no role.

3.2. Study 1 : A pupillometry study

In this study changes in pupil diameter were measured to investigate the mechanism that allows people to comprehend long distance dependencies

(i.e. parasitic gap constructions). Pupil diameter has been shown to reflect syntactic complexity and the processing effort required during language comprehension (e.g. Engelhardt et al., 2010; Just & Carpenter, 1993). For example Just and Carpenter found that pupil diameter increased in conditions with a moved constituent compared to one without, and they argue that the pupil reflects the additional cognitive resources allocated to hold a filler in working memory. Parasitic gaps pose an interesting question in terms of island constraints, given that island constraints may not necessarily prevent the parser from searching for gaps, at least in some islands. Phillips (2006) found in a self-paced reading study that the parser only posits gaps in islands in which a parasitic gap dependency is allowed. We used similar materials to investigate parasitic gaps using changes in pupil diameter over time as an index of processing load. The critical items were taken from Phillips (2006) and were simplified for auditory comprehension.

Some of the theoretical debate surrounding filler gap dependencies and island constraints focus on whether comprehension operates via grammatical knowledge (e.g. Phillips, 2006), or whether comprehension is limited by processing resources (e.g. Deane, 1991; Pritchett, 1991). To investigate processing load we measured pupil diameter as participants heard parasitic gap constructions. We hypothesized that the parser will not posit gaps inside islands unless it is in an acceptable (parasitic) environment. As mentioned previously, encountering the illegal gap should increase processing load (in the finite conditions), and we expect this to be reflected in an increase in pupil size. To date there have only been a handful

of studies that used online techniques to assess how comprehenders process sentences with island constraints (e.g. McKinnon & Osterhout, 1996; Neville, 1991). Therefore, we seek to add to the literature by adjudicating between Reductionist and Grammatical accounts of island phenomena.

3.2.1. Method

3.2.1.1. Participants

Sixteen undergraduate psychology students from Northumbria University participated as a requirement for their undergraduate psychology course. All participants were native speakers of British English with normal or corrected-to-normal vision.

3.2.1.2. Materials

There were four sentences types and Table 1 shows example sentences and includes summaries of predicted processing effects. There were a total of 20 critical trials with five critical sentences in each condition, and participants began with eight practice trials. The sentences were recorded by a female native speaker of British English at normal speaking rate. The files were saved as .wav, and normalized for volume. The recordings were placed into four lists and rotated in a Latin square design. There were 52 filler sentences (35 were ungrammatical and 17 were grammatical). Thus, each participant completed eight practice trials, 20 experimental trials, and 52 fillers. The critical items contained a verb embedded in either an infinitival subject island or a finite relative clause island. Infinitival subject islands allow parasitic gaps, while finite relative clauses do not allow for parasitic gap

constructions. It is important to note that there were no parasitic gap dependencies in any of the items we tested thus ensuring the results were not primed by the experimental materials. The critical items could all plausibly have been a parasitic gap construction, but they were not (see also Phillips, 2006).

3.2.1.3. Apparatus

Stimulus presentation was programmed with SR Research Experiment Builder Software. Participant's pupil diameter was recorded with an Eyelink 1000, sampling at 1000 Hz. Viewing was binocular, but only the right eye was tracked. The eye tracker and a 46 cm (18") CRT colour display monitor were connected to a 3-GHz Pentium 4 PC, which controlled the experiment and monitored the position of the eye throughout the experiment. Participants were seated in a chair at a comfortable height approximately 100 cm from the display. Head movements were minimized using a chin rest, and eyes were calibrated and validated using a 9-point sequence.

3.2.1.4. Design and procedure

The design was 2 × 2 (plausibility × verb type). Plausibility refers to whether the filler was a plausible direct object of the embedded verb (i.e. – it is a plausible to expand a school, but it is implausible to expand a student). Verb type was either infinitive (*to expand*) or finite (*expanded*). The infinitive condition allowed for a potential parasitic gap construction, while the finite condition did not. Both variables were manipulated within subject. Critical items were counterbalanced across participants, and the order of the critical and filler trials was randomly assigned for each participant.

Participants were informed that a fixation dot would appear for 2000 ms (this was to allow the pupil to adjust to the luminance of the screen) and this would be followed by a fixation cross. The sentence recording was then presented and the fixation cross remained on the screen; the fixation cross stayed 2000 ms post utterance offset. Participants were required to look at the cross while the sentence was playing and were asked not to blink. After the sentence was over an acceptability scale appeared and participants had to rate the acceptability of the sentence on a scale from 1 – 7. (One was an ungrammatical/unacceptable sentence and seven was grammatical/perfectly acceptable.) The acceptability rating was followed by a second (1 – 7) rating scale that asked participants to rate their confidence in the grammaticality judgment. The acceptability and confidence scale ratings were input by pressing the corresponding key on the keyboard.

3.2.1.5. Pupil data processing

The few blinks that did occur were filtered out and the missing values were replaced using linear interpolation. We analyzed the data from only those trials in which the participants' acceptability and confidence ratings were four or above. Pupil diameter was measured at (1) the onset of the verb in the first embedded clause and (2) the word following the verb, which we refer to as N + 1. We analyzed a 1.2 second time window as this has been previously reported to be the length of time that it takes for the pupil to reach maximum diameter following a point of difficulty (Just & Carpenter, 1993). Individual trials were averaged together and the resulting four vectors for each participant were submitted to a simple regression in which time was the independent variable and pupil diameter was the dependent variable. The

unstandardized regression coefficient (i.e. the slope of the pupil diameter change over time) served as the main dependent variable in the study. In addition to the within subject comparisons we also conducted one-sample t-tests to determine whether the slope of the pupil diameter was significantly differently from zero (i.e. flat) or no change. All data were screened prior to the inferential analysis; any data point greater than 2.5 standard deviations or more away from the mean in any condition was replaced with the mean for that condition. This affected approximately 5% of the data.

3.2.2. Results

Refer to table 1 for a summary of predictions.

3.2.2.1. Acceptability and confidence judgments

Participants rated sentences on an acceptability scale of 1 – 7, and also judged the confidence in their response. Figure 1 shows the mean grammaticality ratings, and Figure 2 shows the mean confidence ratings; in all graphs error bars represent the standard error of the mean.

In terms of grammaticality ratings a 2 × 2 (plausibility × verb type) repeated measures ANOVA revealed a significant main effect of plausibility $F(1,15) = 6.86$, $p < .02$ (see Figure 1). The plausible conditions were rated higher than the implausible conditions. There was no effect of verb type, and the interaction was also not significant. A 2 × 2 (plausibility × verb type) repeated measures ANOVA conducted on the confidence ratings showed no significant effects.

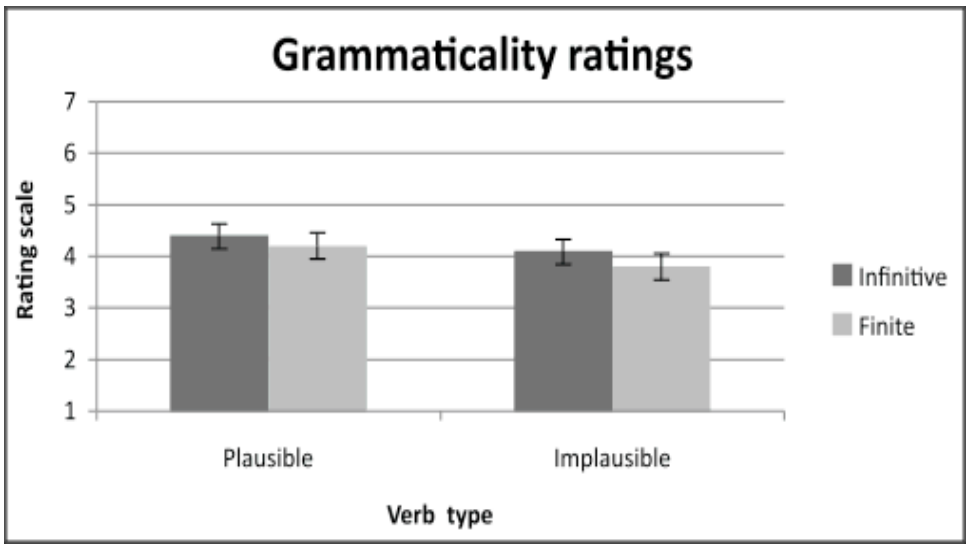

Figure 1. Grammaticality ratings for study 1.

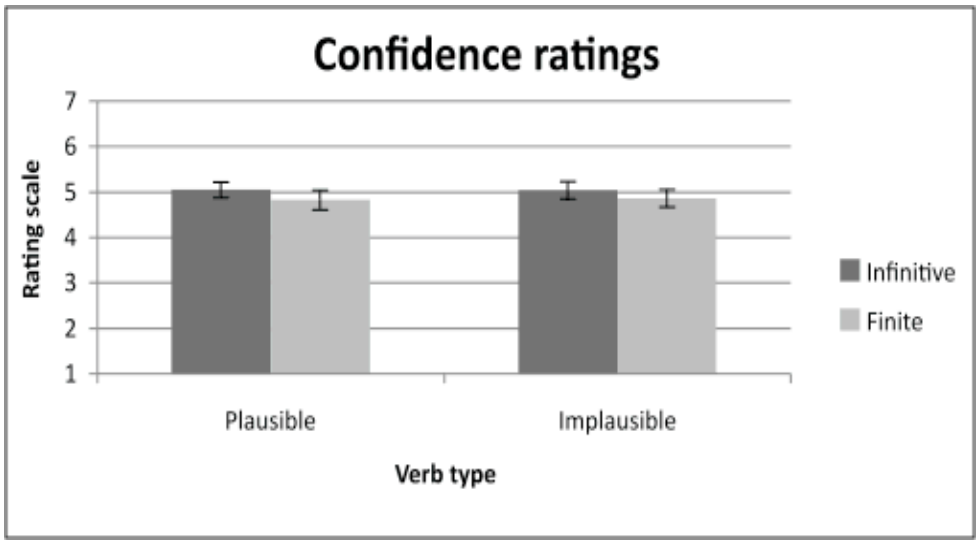

Figure 2. Confidence ratings for study 1.

3.2.2.2. Pupil diameter at the verb

This set of analyses looked at trials with an acceptability rating of four or greater, **and** a confidence rating of four or greater at the verb; Table 2 shows the number of trials meeting these criteria. At the verb there was a main effect of verb type $F(1,15) = 4.71$, $p < .05$ (see Figure 3). The finite implausible condition evoked a larger decrease in pupil size compared to the

finite plausible condition, which goes against our predictions. We therefore compared the infinitive plausible and infinitive implausible conditions despite the lack of interaction and found no difference between the two conditions $t(15) = .43$, $p = .67$. This analysis is important because it shows that the pupil slope in the implausible condition does not significantly differ from the pupil slope in the plausible condition.

Table 2. Number and percentage of trials rated four or greater for study 1.

	Plausible	**Implausible**
Infinitive	41(51.3%)	54(67.5%)
Finite	30(37.5%)	38(47.5%)

Note. There were 80 trials in each condition.

One-sample *t*-tests revealed that the pupil slope was significantly greater than zero in both infinitive conditions: plausible $t(15) = -2.51$, $p < .05$ and implausible $t(15) = -2.46$, $p < .05$. In the finite conditions, pupil size was not different from zero in either condition: implausible $t(15) = -1.59$, $p = .88$ and plausible $t(15) = -1.03$, $p = .32$. These findings are consistent with our prediction as there should be a decrease in pupil diameter in the infinitive conditions at the verb because the wh-filler can be associated with the verb, and thus reduces the demand on cognitive resources.

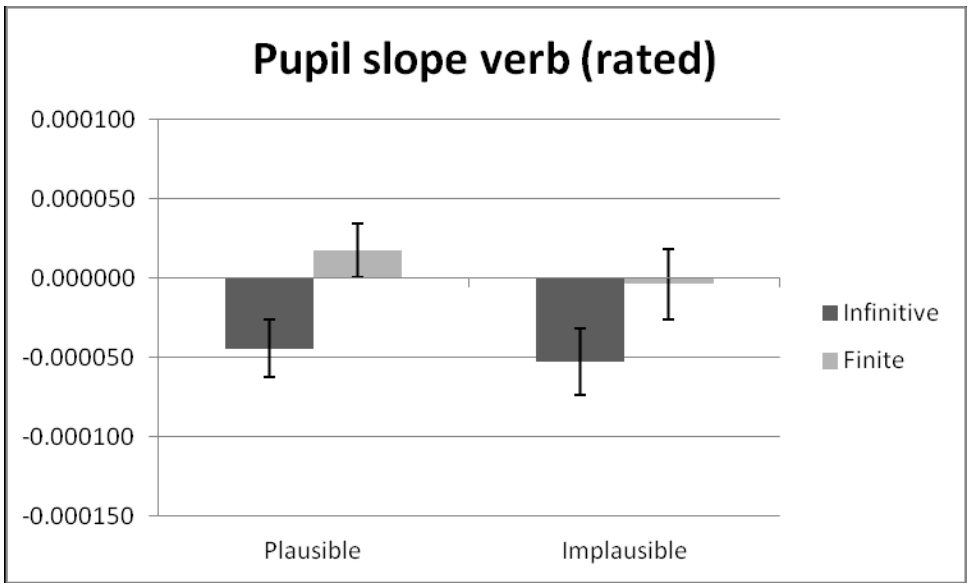

Figure 3. Mean pupil slope at the verb (rated trials) for study 1.

3.2.2.3. Pupil diameter at N + 1

We then looked at the trials that were given an acceptability rating of four or greater, and a confidence rating of four or greater a N+1(see Table 2). N +1 trials showed a main effect of plausibility $F(1,15) = 5.39$, $p < .05$. There was also a significant interaction between plausibility and verb type $F(1,15) = 8.17$, $p < .02$ (see Figure 4). Paired comparisons showed a significantly larger pupil slope in the infinitive-implausible condition compared to the finite-implausible condition $t(15) = -2.39$, $p < .05$, and in the finite-plausible condition compared to the finite-implausible condition $t(15) = 2.91$, $p < .05$. There were no significant differences when comparing the infinitive-plausible condition and the infinitive-implausible condition $t(15) = .46$, $p = .65$ and when comparing the infinitive-plausible condition and the finite-plausible condition $t(15) = 1.71$, $p = .11$.

A one-sample *t*-test, with a test value of zero revealed that the pupil slope was significantly different than zero in two conditions: the infinitive-

implausible condition $t(15) = 2.37$, $p < .05$ and the finite-plausible condition $t(15) = 2.61$, $p < .05$. Pupil slope did not differ from zero in the infinitive-plausible condition $t(15) = -1.04$, $p = .32$ and the finite-implausible condition $t(15) = 1.68$, $p = .11$.

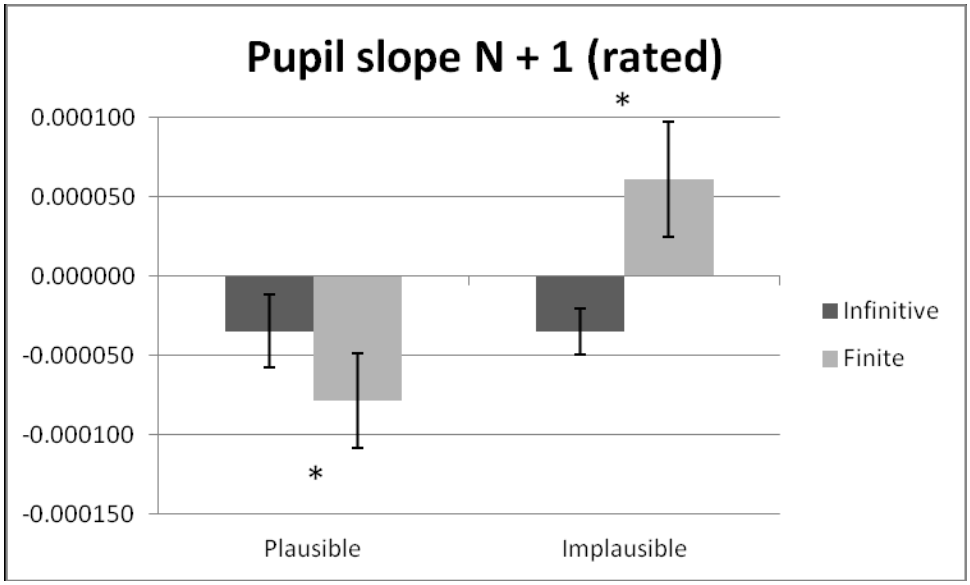

Figure 4. Mean pupil slope at N+1 (rated trials) for study 1.

3.2.2.4 Summary

The findings taken together present an interesting story. In the infinitive conditions, we found a decrease in pupil size which suggests that the parser is associating the filler with the potential gap site when it is in a parasitic gap-like environment. This is consistent with previous ERP work that shows the comprehension system can almost immediately detect the island boundary (Kluender & Kutas, 1993; McKinnon & Oosterhout, 1996; Pickering et al.,1994). However the decrease in this case was not significant but it is consistently negative across the time windows; this could reflect a decrease in memory load given that the filler is no longer actively held in memory (Baddeley, 1986). The finite conditions are not consistent with predictions

and a bit harder to interpret; they may reflect processing further downstream from the critical island clause. We return to this issue in the General Discussion.

3.3. Study 2 : Misinterpretations

In this study plausibility was tested to determine whether participants associated the direct object with the embedded verb. It has been found that temporary thematic role assignments linger and lead to misinterpretation despite the downstream ambiguity resolution (Christianson et al., 2001). We anticipated that participants would associate the plausible direct object with the embedded verb as a way to decrease processing costs during auditory language comprehension.

3.3.1. Method

3.3.1.1. Participants

Sixteen undergraduate psychology students from Northumbria University participated as a requirement for their undergraduate psychology course. All participants were native speakers of British English with normal or corrected-to-normal vision.

3.3.1.2. Materials

Materials were the same as those used in study 1.A. However, instead of a grammaticality and confidence scale, a yes/no comprehension question was asked to probe misinterpretations (see Table 3). Comprehension questions probed whether the participant associated the filler with verb embedded within the island clause. In the plausible conditions the filler was a plausible

direct object of the embedded verb (i.e. it is possible to expand a school), and ultimately the incorrect interpretation of the sentence. In the implausible conditions the filler was an implausible direct object of the embedded verb (i.e. it is not possible to expand a student), and ultimately the incorrect interpretation of the sentence. Therefore, the critical comprehension questions were all correctly answered with "no". The fillers were the same as in the previous experiment and a comprehension question was created for each sentence. There was an equal amount of correctly answered "yes" and "no" questions over the whole experiment.

Table 3. Example sentences and comprehension questions for study 2.

	Example sentence	**Comprehension question**
Infinitive Plausible	The superintendent learned **which schools** the proposal to **expand** upon the curriculum would overburden ____ during the following semester	Was the proposal to expand the schools?
Infinitive Implausible	The superintendent learned **which high school students** the proposal to **expand** upon the curriculum would motivate ____ during the following semester	Was the proposal to expand the students?
Finite Plausible	The superintendent learned **which schools** the proposal that **expanded** upon the curriculum would overburden ____ during the following semester	Was the proposal to expand the schools?
Finite Implausible	The superintendent learned **which high school students** the proposal that **expanded** upon the curriculum would motivate ____ during the following semester	Was the proposal to expand the students?

3.3.1.3. Apparatus

Stimulus presentation was programmed with SR Research Experiment Builder Software. The experiment was presented on a 46 cm (18”) CRT colour display monitor connected to a 3-GHz Pentium 4 PC, which recorded comprehension responses. Participants sat at a comfortable distance from the monitor and used a standard keyboard to input comprehension answers.

3.3.1.4. Design and procedure

The design was the same as in study 1. The procedure was the same as above, however the grammaticality and confidence scales were replaced with a comprehension question. Participants answered the comprehension questions by pressing the corresponding key on the keyboard (“y” for “yes” and “n” for “no”).

3.3.2. Results

Participants answered a comprehension question, probing interpretation of the stimuli and Figure 5 shows the comprehension accuracy in terms of percent correct; error bars represent the standard error of the mean.

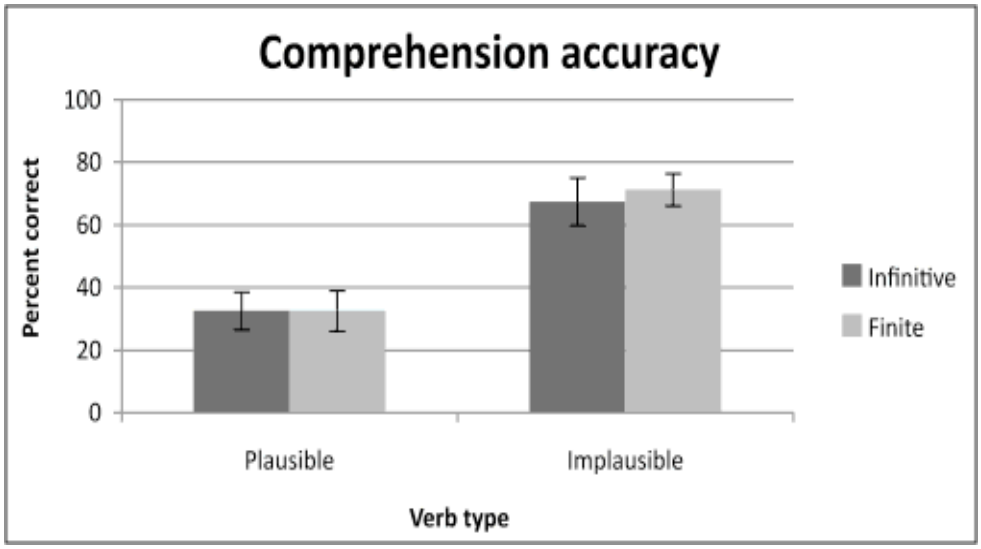

Figure 5. Comprehension accuracy for study 2.

A 2 × 2 (plausibility × verb type) repeated measures ANOVA showed a significant main effect of plausibility $F(1,15) = 29.29$, $p < .001$. The plausible conditions showed lower comprehension accuracy than the implausible condition. There was no effect of verb type and the interaction was not significant.

3.3.2.1. Summary

In a post hoc comprehension study the interpretation of the critical stimuli was probed. Comprehension accuracy showed that participants were incorrectly interpreting sentences when the filler was a plausible object; this suggests that the parser associates the filler in the first semantically acceptable position.

3.3. General discussion

The aim of this study was to investigate whether the parser is sensitive to island constraints in parasitic gap-like constructions. The results focused on two time windows: the critical verb (the first potential gap site) and N+1 (the

word following the verb). Our results focused on those trials rated above a certain grammatical threshold with a high confidence rating. We also investigated comprehension of the stimuli.

The infinitive conditions led to significantly negative pupil slopes in the two time windows suggesting less processing effort associated with the constructions that allow for the filler to be associated with the gap. A negative pupil slope indicates that the parser associates the filler with embedded verb which reduces or alleviates some load on memory resources. There should also be a semantic mismatch between the filler and verb. In previous research this incompatibility results in slowdowns in reading times (Phillips, 2006). The lack of an implausibility effect is interesting and this is consistent with models of sentence comprehension that assume that semantic information does not affect initial structural parsing decisions.

In the finite conditions there were fairly large effects of plausibility which is inconsistent with predictions. The relative clause is overtly marked and it is impossible to move out of which should prevent the filler from being associated with the embedded verb. At the verb, we observed that both plausible and implausible conditions were flat (i.e. not different from zero). In the N + 1 time window the two conditions diverge; the plausible condition is significantly negative, and the implausible condition is significantly positive.

Pupil diameter showed a main effect of verb type at the verb. The infinitive sentences resulted in a decrease in pupil size and finite sentences showed essentially no change in pupil size. Our data did not show an effect of plausibility with the infinitive sentences. If anything the pattern was the

reverse of what was found by Phillips (2006); the implausible sentences resulted in a slightly greater decrease in pupil diameter compared to plausible sentences which would suggest that in terms of processing load they are easier to process. Thus the pupil data might suggest that the dependency is actually created in both infinitive conditions. However, because we only observed a main effect of verb type and no effect of plausibility, one possibility is that pupil diameter is simply reflecting overall processing differences between the two types of sentences (Pickering et al., 1994).

The next time window (N +1) did show a significant interaction between variables. (Phillips only showed a main effect of verb type at N+1 and N+2). There was no difference between plausible and implausible sentences with infinitive verbs, and both were negative (i.e. pupil diameter decreased) which is the same as in the previous window. There were large differences with finite verbs: in the plausible condition, pupil diameter significantly decreased compared to the implausible condition. The implausible condition showed a significant increase in pupil diameter. Thus, there is an effect of semantics on pupil diameter but it occurs downstream from the verb. The pattern at N + 1 indicates that the parser creates the dependency when the verb is finite and wh-phrase is a plausible object.

Behavioural measures showed a large effect of semantics: grammaticality and comprehension showed a main effect of plausibility. The plausible sentences were rated more acceptable compared to implausible sentences, and the plausible sentences led to worse comprehension performance compared to implausible sentences (.31 vs. .69). The higher

comprehension accuracy with implausible sentences seems straightforward: people do not make mistakes when the proposition is not possible given real-world knowledge. The low rate for plausible sentences is more interesting because it indicates that participants are creating a dependency between the wh-phrase and the verb in the subject NP, even when the verb in the subject NP is finite. Research has shown that thematic assignments made during online processing often linger and result in systematic errors in offline comprehension performance (Christianson et al., 2001).

Theories of island constraints can be divided into two main camps. The most prominent being the Reductionist accounts, in which island constraints are a result of a processing overload of processing and the Grammatical accounts, in which island constraints are a result of grammatical representations within the brain. The current data support Reductionist accounts of island constraints; the parser is forming a dependency in the infinitive conditions, regardless of plausibility. In the finite condition the parser forms a dependency only when the filler is a plausible object of the embedded verb. This study employed auditory input and I argue that when the input rate of the stimuli is not controlled by the listener the parser may posit a filler in an ultimately incorrect gap site to alleviate processing costs. These findings may also be the result from a good enough processing strategy (Ferreira et al., 2002).

3.3.3. Conclusions

Semantics has less of an immediate effect when participants don't control the rate of input, while online processing load showed a different pattern of results. One interpretation of these differing results is that participants create

the dependency when the verb is infinitive, regardless of (im)plausibility, but also when the wh-filler is plausibly the direct of object of a finite verb. The latter of these conclusions is strongly supported by the offline comprehension measures. These results are consistent with theories that place an emphasis on processing resources, however they cannot completely explain the results; if the construction was too taxing to parse then there should be no evidence of a dependency formation at the site of the island clause. One reason why the parser may have an active gap filling strategy, even in cases where it is not possible, is because it leads to decreases in processing load. Alternatively, it could represent a strategy shift in which participants tend to adopt good enough representations (Ferreira et al., 2002).

CHAPTER 4: HYPER-ACTIVE GAP FILLING

4.1. Background

This aim of this chapter is to investigate the point at which the parser starts forming filler gap dependencies and postulating upcoming gap sites. As discussed in chapter 2 English is a verb-medial language, and there has been much debate on how the parser constructs filler gap dependencies in verb-medial languages. There have been many accounts put forward that differ in many aspects, but are all similar in that they assume that verbs play an important role in parsing filler gap dependencies. However it has been found that in verb-final languages, like Japanese, pre-verbal information is used to construct upcoming gaps and the parser does not wait for the verb to construct a gap (Aoshima, Phillips, & Weinberg, 2004). These findings suggest that in Japanese the parser is employing an active gap filling strategy; however this strategy does not consult verb information when postulating gap sites. Rather, gaps are postulated at the first possible position in which a thematic role can be assigned.

As mentioned previously holding a filler in memory taxes processing resources, and Omaki et al. question whether active gap filling in English is driven by the early release of the filler at the verb site, given that the verb is encountered relatively early in the sentence. By holding the filler in memory it allows the parser to consult verb information and potentially avoid positing an incorrect gap. The potential processing costs involved with holding a filler

in memory is less taxing (given the typically early placement of the verb in English) than potentially positing a gap and forcing the parser to later reanalyze. Japanese, on the other hand, is a verb-final language so processing costs involved with holding a filler in memory are much higher, as the filler must be held for a longer period of time. Therefore, in Japanese, the parser has been shown to employ a strategy that uses pre-verbal information to construct gaps.

Omaki et al. refer to the verb-medial strategy of gap filling as *Conservative gap filling* because construction of the gap does not occur until verb information is consulted. *Hyper-active gap filling*, on the other hand, is characterized by constructing upcoming gaps before encountering a verb. English may not employ hyper-active gap filling as this type of gap filling is more risky and can potentially lead to structural reanalysis if the later occurring verb cannot host an object. Omaki et al. questioned whether hyper-active gap filling is a language specific adaptation to verb-final languages, or whether it is inherent to all parsing architectures regardless of language. Both types of gap filling assume that there is a structural integration between the moved constituent and the gap but the strategies differ in terms of what triggers integration and when it happens.

Staub (2007) conducted an eye-tracking study using a transitivity mismatch paradigm to investigate whether the parser consults verb information when forming filler gap dependencies. The transitivity mismatch paradigm tests whether the transitivity properties of the verb influence parsing; intransitive verbs do not take an object and therefore no dependency can be formed with this verb type. Transitive verbs do take an

object and are capable of hosting filler gap dependencies. Staub found that the parser did not construct a gap until it consulted verb transitivity information, which supports a conservative gap filling hypothesis.

Omaki et al. employed a self-paced reading and transitivity mismatch paradigm similar to the one used by Staub. However, Omaki and colleagues point out that the stimuli in Staub's study were not consistent across conditions (making direct comparisons difficult) and thus Staub's findings may have been confounded by the different structures tested. Omaki et al. constructed a baseline condition in which they embedded the critical verb within an island, this has been shown to prohibit filler gap dependencies from occurring with the verb in the island (see Table 4). By doing this they were able to keep the stimuli consistent across conditions and avoid the potential confound in Staub's study. Hyper-active gap filling predicts that the non-island conditions will produce slower reading times than the island conditions. In terms of the intransitive (non-island) condition this will be due to the mismatch between the anticipated subcategorization properties of the transitive verb (which turns out to be intransitive and thus incapable of hosting a filler gap dependency). A slowdown in the transitive (non-island) condition would be the result of a plausibility mismatch between the filler and the verb. The conservative gap filling hypothesis predicts that there will be a difference in reading times between the transitive non-island and transitive island conditions resulting from the plausibility mismatch, but no reading time difference is anticipated between the two intransitive conditions (the parser should consult transitivity before positing a gap).

Omaki et al. found a slowdown in reading in the transitive non-island condition as a result of the plausibility mismatch compared to baseline, as well as a slowdown in the non-island intransitive condition compared to baseline. These results suggest that the parser is positing gaps before consulting the subcategorization properties of the verb and are in direct contrast to those of Staub (2007) thus favouring a hyper-active gap filling explanation.

Omaki and colleagues note, however, their stimuli were comprised of both unergative and unaccusative intransitive verbs. Unergative intransitive verbs are capable of hosting a semantic agent as an argument, so in special cases they can host a direct object noun phrase, unlike unaccusative intransitive verbs which cannot. In their second experiment Omaki et al. revised their stimuli to include only unaccusative intransitive verbs, and employed an eye-tracking methodology (see Table 5 for example stimuli). They found longer first fixation durations in the intransitive non-island condition compared to baseline, and no such effect was seen in the transitive conditions. This suggests that the parser is positing the upcoming gap before consulting the verb, leading to a reading disruption and supporting a hyper-active gap filling hypothesis. However, they found that regression paths were shorter in the transitive non-island condition when compared to the other three conditions and they speculated that an additional word in the island conditions may have been a factor in their results. These findings taken together support the hypothesis that verb-medial languages, like English, employ hyper-active gap filling, suggesting that the parser posits gaps before consulting verb information.

4.1.1. Current studies

In this study pupil diameter change was measured to investigate processing load and the point in which the parser forms filler gap dependencies. Pupil diameter has been shown to reflect syntactic complexity and processing effort required during language comprehension (e.g. Engelhardt et al., 2010; Just & Carpenter, 1993). Research has been mixed regarding the point at which the parser uses subcategorization information to form filler gap dependencies, and at what point the parser predicts an upcoming gap. Recently it has been found that the parser may not consult the verb before positing a gap and may actually employ hyper-active gap filling (Omaki et al.). In the experiments below, the materials from Omaki et al. were used to adjudicate between hyper-active gap filling and conservative gap filling accounts using changes in pupil diameter over time as an index of processing load. Pupil diameter change at the critical verb and the word following the verb (which was always an adverb) were investigated. The verb site provides us with information about pre-verbal postulation of gaps and the adverb site provides post-verbal information (i.e. whether subcategorization and island constraints play a role in gap positing).

4.2. Study 3

4.2.1. Method

4.2.1.1. Participants

Sixteen undergraduate psychology students from Northumbria University participated as a requirement for their undergraduate psychology course. All

participants were native speakers of British English with normal or corrected-to-normal vision.

4.2.1.2. Materials

There were four sentences types and Table 4 shows examples. There were a total of twenty-eight critical trials, with six or eight of each sentence type in each condition, and participants began with eight practice trials. The sentences were recorded by a female native speaker of British English at normal speaking rate. The files were saved as .wav, and normalized for volume. The recordings were placed into four lists, rotated in a Latin square design. There were 52 filler sentences (35 were ungrammatical and 17 were grammatical). Thus, each participant completed eight practice trials, 28 experimental trials, and 52 fillers.

The critical items were taken from Omaki et al., there were four sentences types: transitive non-island, transitive island, intransitive non-island, and intransitive island. The transitive conditions contained an optionally transitive verb, while the intransitive conditions contained either an unergative or unaccusative intransitive verb. The verbs were all semantically implausible fillers for the first gap site, but the sentences were all globally plausible. We do not focus much on (im)plausibility because our interests lie mainly in the processing costs associated with maintaining a filler in memory while crossing island boundaries. In the island conditions the verb was embedded within a relative clause island thereby disallowing any filler gap dependencies to be formed with the embedded verb. In the island conditions, implausibility and transitivity should have no effect on processing, as the verb is inaccessibly located in an island.

If a conservative gap filling approach is correct then the parser will consult verb properties before positing a gap (see Table 4 for example stimuli). Therefore we would anticipate no difference in the two intransitive conditions, as no gap should be anticipated. Also a larger pupil size is anticipated in the transitive island condition compared to the transitive non-island condition given that the parser is unable to form a dependency and will continue to hold the filler in memory. If a hyper-active gap filling approach is correct then the parser will not consult verb properties before positing a gap. Therefore, a larger pupil size should be evident in the intransitive island condition because the parser will have anticipated a gap site, but the intransitive verb will not allow for one (and thus the filler will continue to be maintained in working memory). Pupil size should also be larger in the transitive island condition compared to the transitive non-island condition given that the parser is unable to form a dependency and will continue to hold the filler in memory. Both hypotheses predict the same pupil increase in the transitive island condition (compared to transitive non-island condition) resulting from maintaining a filler in memory. We are therefore primarily interested in what is occurring in the intransitive conditions.

Table 4. Example sentences for study 3.

	Example sentence
Transitive Non-Island	The city$_i$ that the author **wrote** regularly about ____ was named for an explorer
Transitive Island	The city$_i$ that the author who **wrote** regularly saw ____ was named for an explorer
Intransitive Non-Island	The city$_i$ that the author **chatted** regularly about ____ was named for an explorer
Intransitive Island	The city$_i$ that the author who **chatted** regularly saw ____ was name for an explorer

4.2.1.3. Apparatus

The apparatus is the same as in Chapter 3.

4.2.1.4. Design and procedure

The design was 2 × 2 (island-hood × transitivity). Island-hood refers to whether the critical sentence contained an island clause or not. Transitivity refers to verb-type, which was either transitive or intransitive. Critical items were counterbalanced across participants, and the order of the critical and filler trials was randomly assigned for each participant.

Participants were informed that a fixation dot would appear for 2000 ms (this was to allow the pupil to adjust to the luminance of the screen) and this was followed by a fixation cross. The sentence recording was then presented and the fixation cross remained on the screen for 2000 ms post utterance offset. Participants were required to look at the cross while the sentence was playing and were asked not to blink. After the sentence was over an acceptability scale appeared and participants had to rate the acceptability of the sentence on a scale from 1 – 7. (One was an ungrammatical/unacceptable sentence and seven was grammatical/perfectly

acceptable.) The acceptability scale ratings were input by pressing the corresponding key on a keyboard.

4.2.1.5. Pupil data processing

The few blinks that did occur were filtered out and the missing values were replaced using linear interpolation. Pupil diameter was analyzed in two ways: first, we examined the data from all trials, and second, we examined only those trials in which the participants' acceptability ratings were four or above. Given the high amount of trials that were rated ungrammatical we look at both sets of data independently. Pupil diameter was measured at (1) the onset of the verb and (2) the word following the verb (which was always an adverb). We analyzed a 1.2 second time window as this has been previously reported to be the length of time that it takes for the pupil to reach maximum diameter following a point of difficulty (Just & Carpenter, 1993). Individual trials were averaged together, and the resulting the four vectors for each participant were submitted to a simple regression, in which time was the independent variable and pupil diameter was the dependent variable. The unstandardized regression coefficient (i.e. the slope of the pupil diameter change over time) served as the main dependent variable in the study. In addition to the within subject comparisons, we also conducted one-sample *t*-tests to determine whether the slope of the pupil diameter was significantly differently from zero. All data were screened prior to the inferential analysis, any data point greater than 2.5 standard deviations or more away from the mean in any condition was replaced with the mean for that condition. This affected approximately 5% of the data.

4.2.2. Results

4.2.2.1. Acceptability judgments

Participants rated sentences on an acceptability scale of 1 – 7. Figure 6 shows the mean grammaticality ratings. In all graphs error bars represent the standard error of the mean.

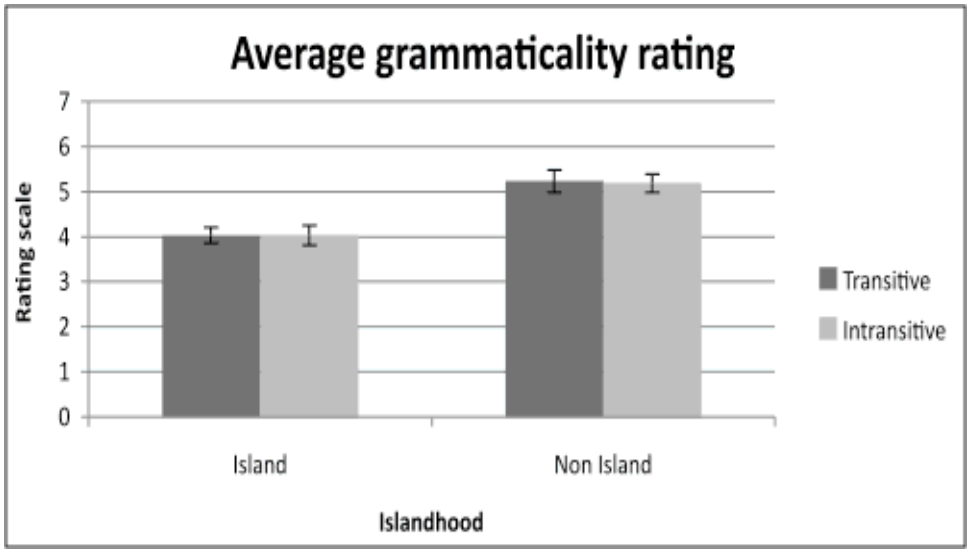

Figure 6. Average grammaticality rating for study 3.

A 2 × 2 (island-hood × transitivity) repeated measures ANOVA showed a significant main effect of island-hood $F(1,15) = 49.96$, $p < .001$, with the island condition having lower grammaticality ratings than the non-island condition. There was no effect of transitivity and the interaction was not significant (all *p's* > .05).

4.3.2.2. Pupil diameter at the verb

The first set of analyses looked at all the trials regardless of acceptability rating. There were no significant main effects or interaction (see Figure 7). One-sample *t*-tests revealed that the pupil slope was not significantly

different than zero in any condition: transitive island $t(15) = 1.55$, $p = .14$, transitive non-island $t(15) = -.58$, $p = .57$, intransitive island $t(15) = .23$, $p = .82$ and intransitive non-island $t(15) = .82$, $p = .43$.

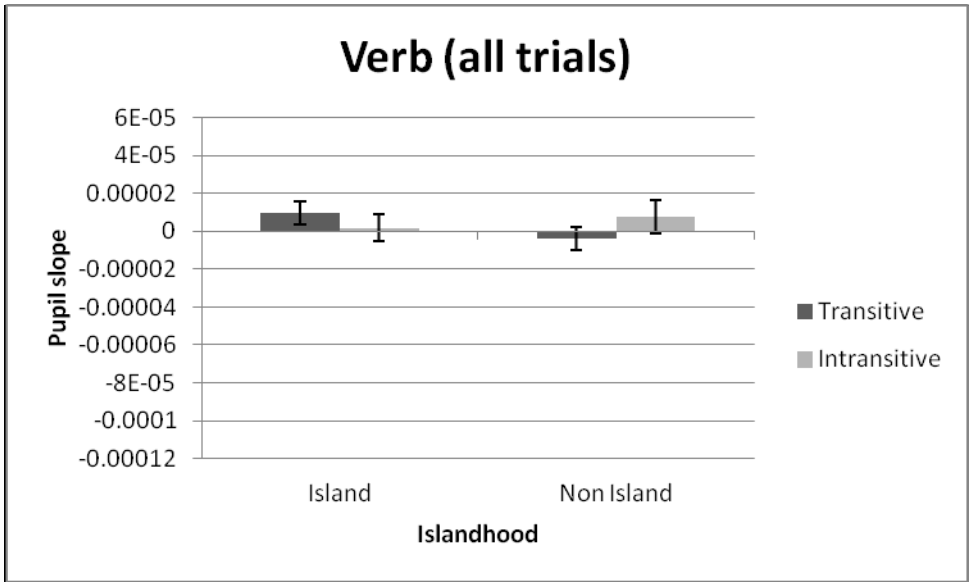

Figure 7. Pupil slope at the verb (all trials) for study 3.

In the second set of analyses we looked at trials with an acceptability rating of four and higher. Table 5 shows the number of trials meeting these criteria.

Table 5. Number and percentage of trials rated four or greater for study 3.

	Transitive	**Intransitive**
Island	57(50.8%)	68(60.7%)
Non Island	99(88.3%)	95(84.8%)

Note. There were 112 trials in each condition.

For the trials rate at 4 or above, at the verb, there were no significant main effects or interaction (see Figure 8). One-sample *t*-tests revealed that the pupil slope was not significantly different than zero in any condition: transitive island $t(15) = -1.17$, $p = .26$, transitive non-island $t(15) = .66$, $p =$

.52, intransitive island $t(15) = .001$, $p = .99$ and intransitive non-island $t(15) = .67$, $p = .16$.

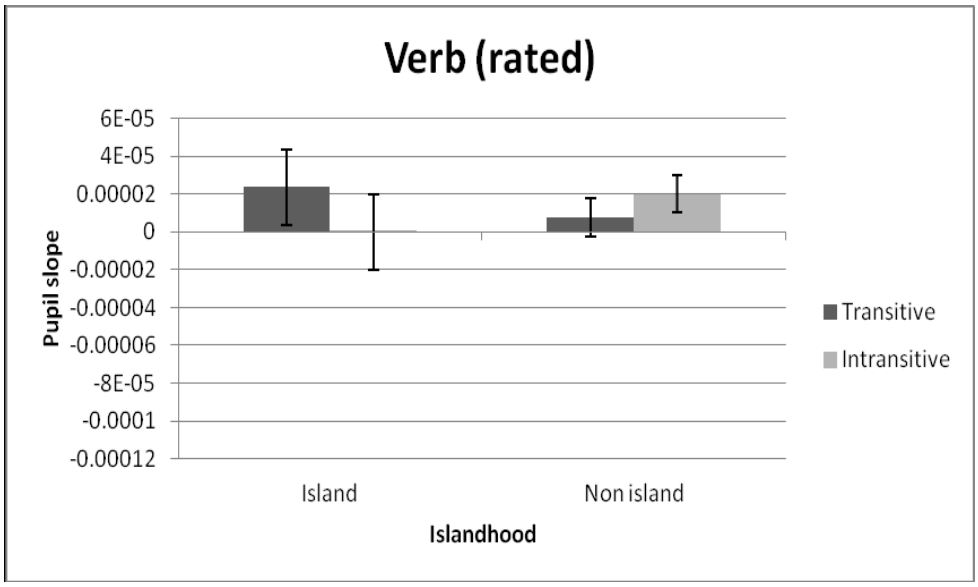

Figure 8. Pupil slope at the verb (rated 4 or above trials) study 3.

4.3.2.3. Pupil diameter at the adverb

The first set of analyses looked at all the trials regardless of acceptability rating. There was no main effect, but the interaction was approaching significance $F(1,15) = 3.04$, $p = .08$ (see Figure 9). One-sample *t*-tests revealed that the pupil slope was significantly different than zero in all conditions: transitive island $t(15) = 3.82$, $p < .01$, transitive non island $t(15) = -4.48$, $p < .01$, intransitive island $t(15) = -3.05$, $p < .01$ and intransitive non island $t(15) = -4.03$, $p < .01$.

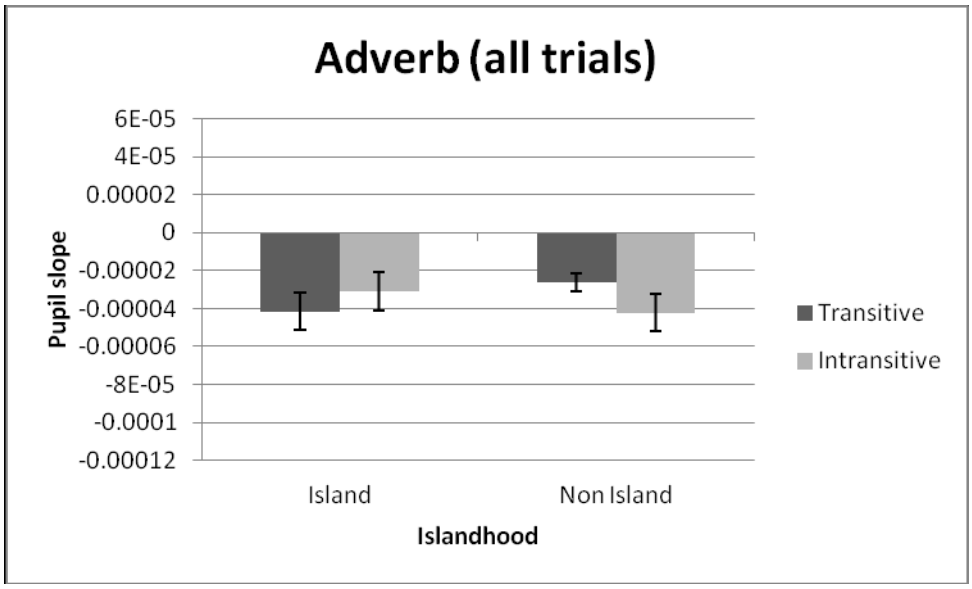

Figure 9. Pupil slope at the adverb (all trials) study 3.

The second set of analyses looked at trials with an acceptability rating of four or greater. Table 5 shows the number meeting these criteria. Looking at the rated trials at the adverb there were no significant main effects or interaction (see Figure 10). One-sample *t*-tests revealed that the pupil slope was significantly different than zero in all conditions: transitive island $t(15) = -3.68$, $p < .01$, transitive non island $t(15) = -4.41$, $p < .01$, intransitive island $t(15) = -3.08$, $p < .01$ and intransitive non island $t(15) = -4.12$, $p < .01$.

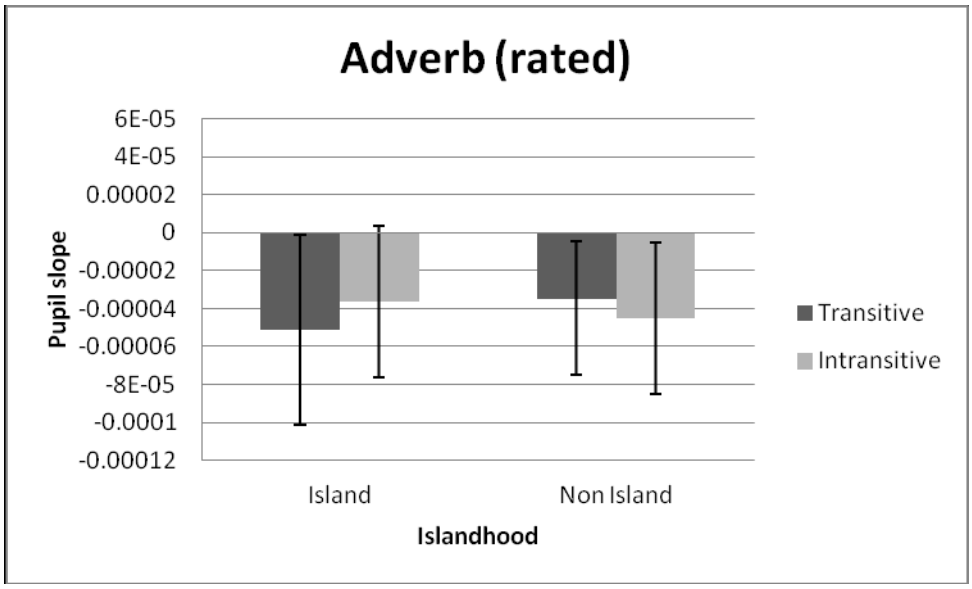

Figure 10. Pupil slope at the adverb (rated 4 or above trials) study 3.

4.3.2.4. Summary

At the verb (for all the trials and for the rated trials) there was a similar pattern; there were no significant main effects or interactions and pupil slope did not significantly differ from zero. While these findings were not predicted by either hypothesis and are based on non-significance, I believe they are more consistent with a conservative gap filling strategy. Given that there were no apparent differences in processing costs between the intransitive non-island and the intransitive island conditions it appears that the parser consulted the verb properties and did not attempt to from a dependency. In terms of the transitive conditions there were also no differences, which is surprising. Both hypotheses assume the parser will posit a gap after the transitive verb, and therefore an increase in pupil size should be evident in the island condition as the filler will have to be held in working memory.

At the adverb the verb has already been encountered so this allowed us to investigate what happens after the verb has been consulted. When

looking at both sets of analyses (for all trials and for the rated trials) a similar pattern emerged, and again there were no significant effects. We do see, when looking at all the trials, an interaction that is approaching significance, but when we take out the unacceptable sentences we no longer see this effect. Overall, regardless of verb type or island-hood, there was a decrease in pupil slope in all conditions; this suggests that the parser has posited the gap immediately following the verb regardless of syntax. This may be a result of the unergative intransitive verbs in the stimuli (which can take a direct object noun phrase in some cases). Or this sort of gap creation may be in line with a Garden Path Model in which the parser is positing a gap and forming a dependency at the first potential gap site. This is an interesting finding as it suggests that the parser may not be sensitive to island constraints, rather it is trying to build the simplest interpretation, thus alleviating processing costs.

4.3. Study 4

As Omaki et al. pointed out, they used unergative intransitive verbs as well as unaccusative verbs in their study. Unergative intransitive verbs are sometimes capable of hosting a direct object noun phrase (e.g. sneezed a big sneeze); this complicates the generalizability of the findings because the parser may be anticipating a potential direct object with the intransitive conditions. Unaccusative intransitive verbs, on the other hand, do not take a direct object noun phrase. In this experiment the same paradigm from Experiment 3 was employed, but the verbs in the stimuli were strictly unaccusative intransitives. And again, participants were asked to rate the sentences on a 1 – 7 grammaticality scale; additionally they were asked to

rate their confidence in their grammaticality rating. The predictions are the same as in Study 3.

4.3.1. Method

4.3.1.1. Participants

Sixteen undergraduate psychology students from Northumbria University participated for course credit. All participants were native speakers of British English with normal or corrected-to-normal vision. None participated in Experiment 3.

4.3.1.2. Materials

There were four sentences types and Table 6 shows example sentences. There were a total of 23 critical trials, with five or six of each sentence type in each condition. The sentences were recorded by a female native speaker of British English at normal speaking rate. The files were saved as .wav, and normalized for volume. The recordings were placed into four lists, rotated in a Latin square design. There were 52 filler sentences (35 were ungrammatical and 17 were grammatical). Thus, each participant completed eight practice trials, 23 experimental trials, and 52 fillers.

The critical items were taken from Omaki et al., there were four sentences types: transitive non-island, transitive island, intransitive non-island, and intransitive island. The transitive conditions contained an optionally transitive verb, while the intransitive conditions contained an unaccusative verb. Unlike study 3, the fillers were all semantically plausible objects of the first gap site, and again, we do not focus much on plausibility. In the island conditions the verb was embedded within a relative clause

island constraint, thereby disallowing any filler gap dependencies to be formed with the embedded verb. Plausibility and transitivity should have no effect in the island conditions since the verb is inaccessibly located within an island.

Table 6. .Example sentences for study 4.

	Example sentence
Transitive Non-Island	The airport$_i$ that the ambassador **left** rapidly for ____ during the unrest was closed to most traffic.
Transitive Island	The airport$_i$ that the ambassador who **left** rapidly had visited ____ during the unrest was closed to most traffic.
Intransitive Non-Island	The airport$_i$ that the ambassador **departed** rapidly from ____ during the unrest was closed to most traffic.
Intransitive Island	The airport$_i$ that the ambassador who **departed** rapidly had visited ____ during the unrest was closed to most traffic.

4.3.1.3. Apparatus

Apparatus was identical to Study 3.

4.3.1.4. Design and procedure

Design and procedure were identical to the previous study with the exception of a confidence scale. After the participants rated the grammatical acceptability of the sentence they were presented a second scale immediately afterwards asking them to rate the confidence of their grammaticality rating, on a scale of 1-7. (One was an ungrammatical/unacceptable sentence and seven was grammatical/perfectly

acceptable). The confidence scale ratings were input by pressing the corresponding key on the keyboard.

4.3.1.5. Pupil data processing

Pupil data processing was identical to Study 3.

4.3.3. Results

4.3.3.1. Acceptability Judgments

Participants rated sentences on an acceptability scale of 1 – 7, and also judged the confidence in their response. Figure 11 shows the mean grammaticality ratings, and Figure 12 shows the mean confidence ratings. In all graphs error bars represent the standard error of the mean.

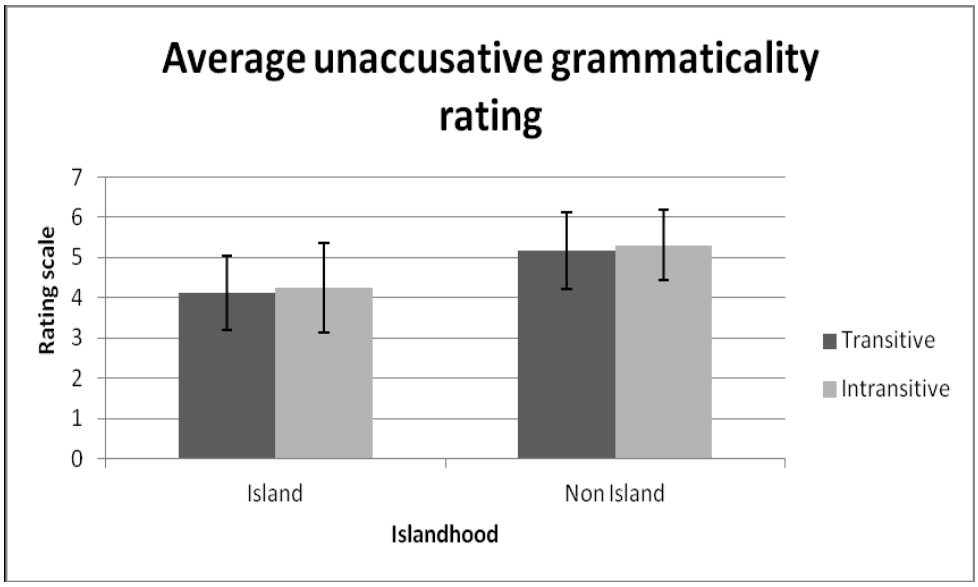

Figure 11. Average grammaticality ratings for study 4.

A 2 × 2 (island-hood × transitivity) repeated measures ANOVA showed a main effect of island-hood $F(1,15) = 24.45$, $p < .001$, with the island condition having lower grammaticality ratings than the non-island condition. There was

no effect of transitivity, and the interaction was also not significant (all *p's* > .05).

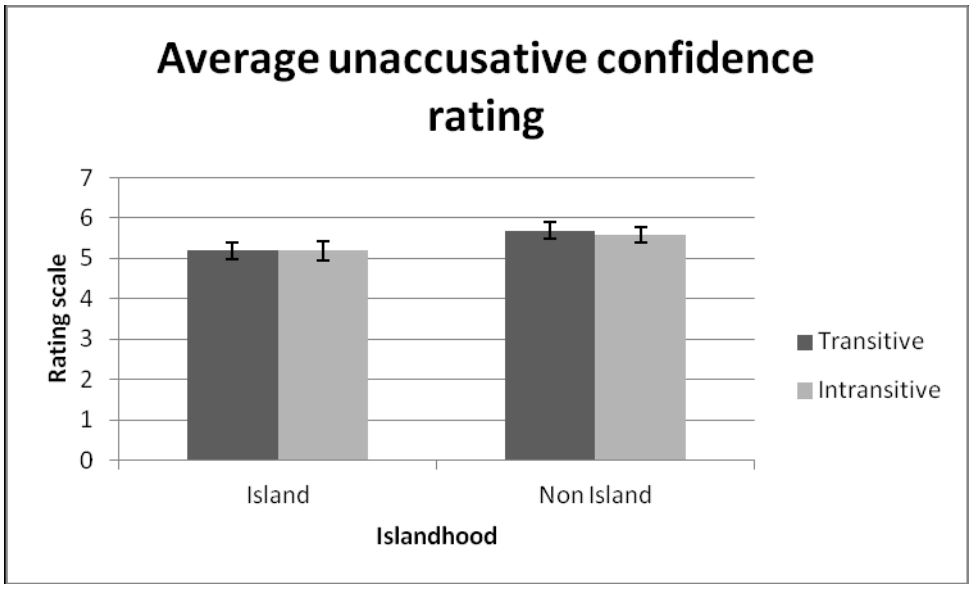

Figure 12. Average confidence ratings for study 4.

Confidence rating showed a main effect of island-hood $F(1,15) = 10.03$, $p < .01$, with the island condition having lower confidence ratings than the non-island condition. There was no effect of transitivity, and the interaction was also not significant (all *p's* > .05).

4.3.3.2. Pupil diameter at the verb

Our first set of analyses looked at all the trials regardless of acceptability and confidence ratings. There were no main effects or interaction (see Figure 13). One-sample *t*-tests revealed that the pupil slope was not significantly different than zero in any condition: transitive island $t(15) = -1.3$, $p = .69$, transitive non-island $t(15) = -.002$, $p = .99$, intransitive island $t(15) = -.86$, $p = .40$, and intransitive non-island $t(15) = -1.7$, $p = .11$.

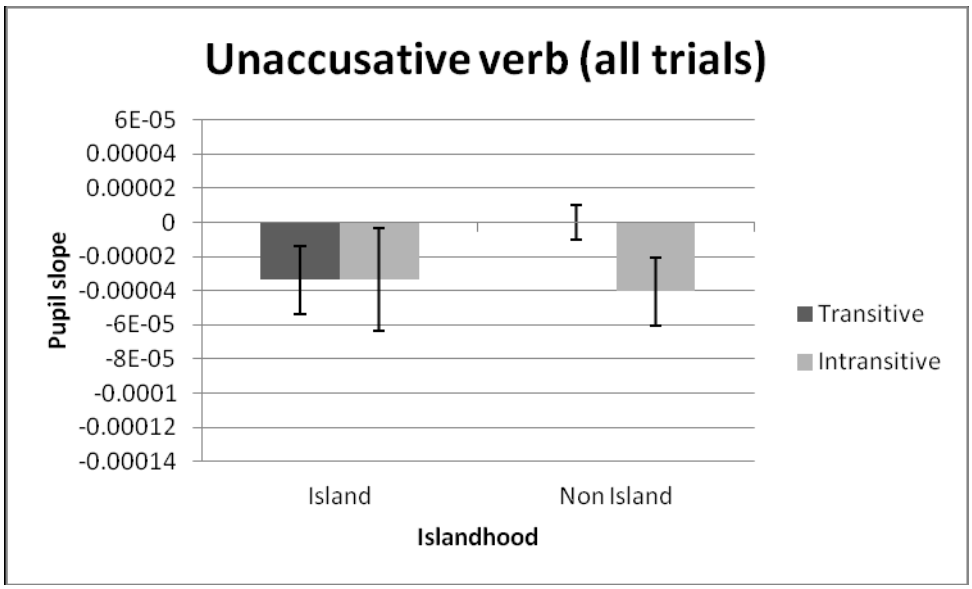

Figure 13. Pupil slope at the verb (all trials) for study 4.

The second set of analyses looked at trials with an acceptability **and** confidence rating of four or greater. Table 7 shows the number of trials meeting these criteria.

Table 7. Number and percentage of trials rated four or greater for study 4.

	Transitive	**Intransitive**
Island	51(55.4%)	53(57.6%)
Non Island	72(78.6%)	73(82.6%)

Note. There were 92 trials in each condition.

For the rated trials at the verb there were no main effects or interaction (see Figure 14). One-sample *t*-tests revealed that the pupil slope was not significantly different than zero in any condition: transitive island $t(15) = -.40$, $p = .69$, transitive non-island $t(15) = -.72$, $p = .49$, intransitive island $t(15) = -1.2$, $p = .23$ and intransitive non-island $t(15) = -1.3$, $p = .21$.

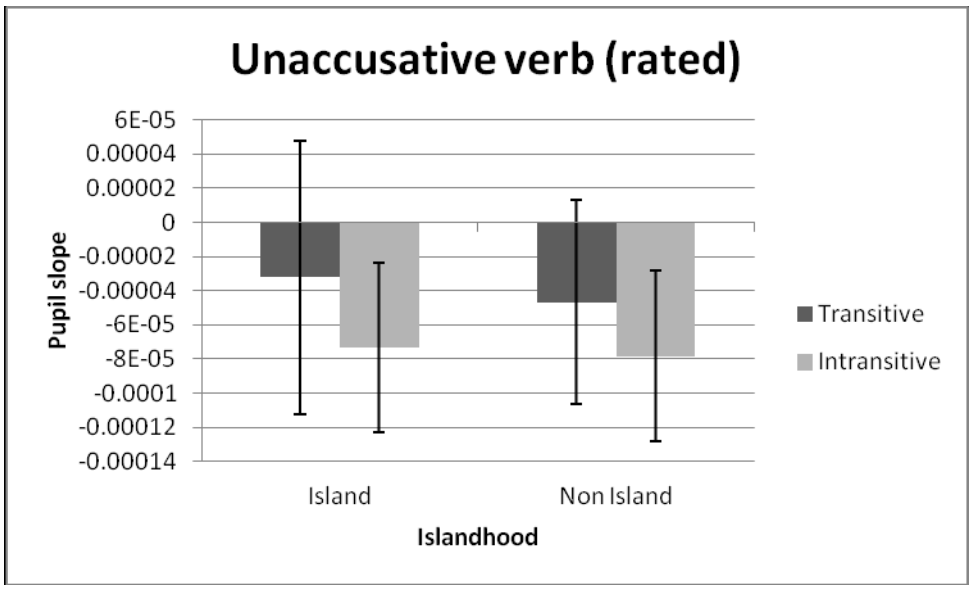

Figure 14. Pupil slope at the verb (rated 4 or above trials) study 4.

4.3.3.3. Pupil diameter at the adverb

The first set of analyses looked at all the trials regardless of acceptability or confidence ratings, there was no main effect or interaction (see Figure 15). One-sample *t*-tests revealed that the pupil slope was not significantly different than zero in any condition: transitive island $t(15) = .75$ $p = .47$, transitive non-island $t(15) = -.004$, $p = .99$, intransitive island $t(15) = .22$, $p = .83$ and intransitive non-island $t(15) = .47$, $p = .65$.

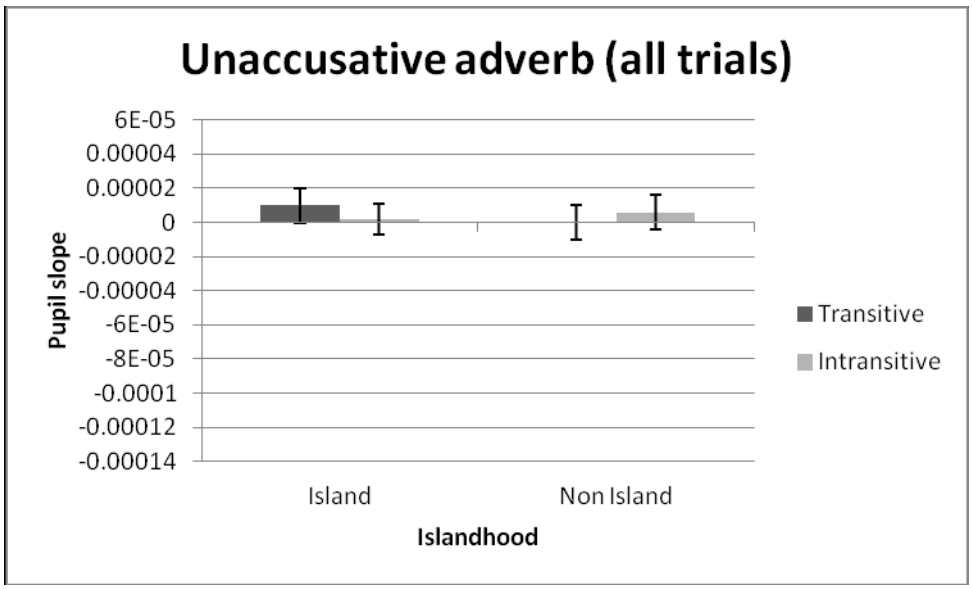

Figure 15. Pupil slope at the adverb (all trials) for study 4.

The second set of analyses looked at trials with an acceptability **and** confidence rating of four or greater. Table 7 shows the number of trials meeting these criteria. Looking at the rated trials at adverb there were no significant main effects or interaction (see Figure 16). One-sample *t*-tests revealed that the pupil slope was not significantly different than zero in any condition: transitive island $t(15) = 1.00$ $p = .33$, transitive non-island $t(15) = -.39$, $p = .71$, intransitive island $t(15) = -.79$, $p = .44$ and intransitive non-island $t(15) = -.06$, $p = .95$.

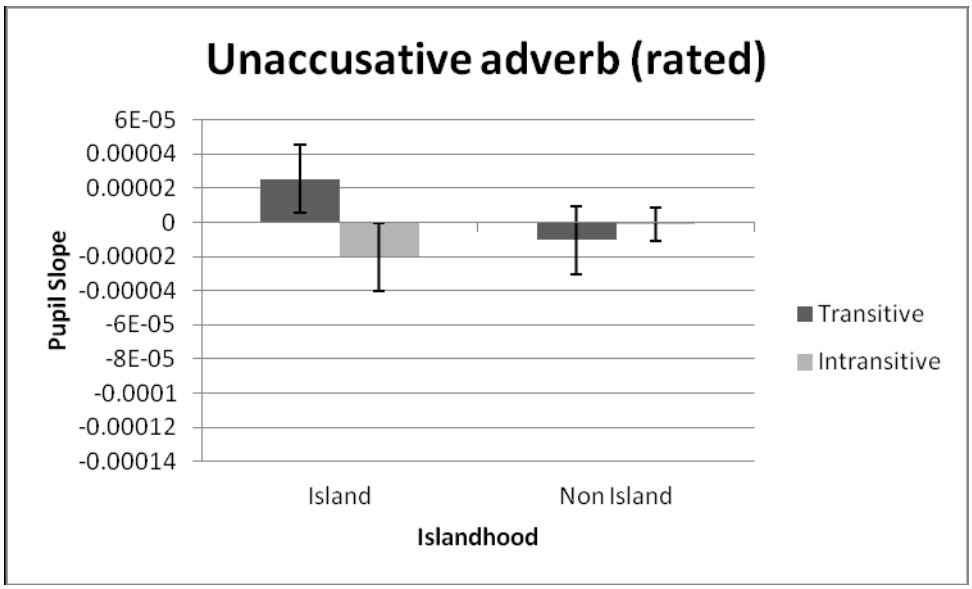

Figure 16. Pupil slope at the adverb (rated 4 or above trials) study 4.

4.3.3.4. Summary

Similar to experiment 3 we see no significant main effects of interactions (both sets of analyses). The pupil slopes in all of the conditions did not differ from zero (unlike Experiment 3). Again there were no differences between the two intransitive conditions and I believe that these findings are more consistent with predictions of a conservative gap filling hypothesis; the parser seems to be consulting verb information when postulating upcoming gaps. There were also no differences between the two transitive conditions, and given the lack of significance, I hesitate to make any strong claims favouring either hypothesis.

At the adverb, regardless of rating, there were no main effects or interactions and the pupil slope did not differ from zero in any of the conditions. Unlike Experiment 3 there was not a consistent negative slope in pupil diameter, and therefore it is unclear if the parser is positing the filler in a gap site.

4.4. Discussion

The current study used a pupillometry paradigm to investigate whether the parser utilizes pre-verbal information when constructing filler gap dependencies. This is a sensitive measure that allowed us to investigate the processing costs involved with forming filler gap dependencies. Research in verb-medial languages, like English, have shown that the parser actively searches for a gap (Fodor, 1978), and starts to postulate a gap as soon as it encounters the verb. It has been shown that in verb-final languages, such as Japanese, the parser is using pre-verbal information to posit gaps (Aoshima, Phillips, & Weinberg, 2004). Omaki et al. investigated whether this was a language specific adaptation of verb-final languages, or if this strategy is inherent across languages: they called the former strategy a conservative gap filling and the latter a hyper-active gap filling. They provided evidence that the parser forms gaps before encountering the verb in English, thus supporting a hyper-active gap filling strategy in English. The current studies specifically set out to investigate whether the parser utilizes pre-verbal information during gap construction using the stimuli from the Omaki et al. study with a pupillometry paradigm.

If we assume hyper-active gap filling we would anticipate an increase in pupil size in the intransitive island condition compared to the intransitive non-island condition resulting from the unanticipated transitivity mismatch. The conservative gap filling hypothesis, on the other hand, predicts no difference between the intransitive island and intransitive non-island condition (as the parser will have consulted transitivity information before positing a gap). Both hypotheses predict larger pupil size associated with

the transitive island condition compared to the transitive non-island condition given the parser is unable to release the filler from working memory.

In Study 3, there were no significant main effects or interactions at the verb or the adverb. At the verb the pupil size was not significantly different than zero in any condition, and the two transitive conditions did not differ (both hypotheses predicted a change). While based on non-significant findings I believe this supports a more conservative gap filling hypothesis, as the parser seems to be waiting to form dependencies until consulting the verb. This strategy suggests holding a filler in memory in verb-medial languages (like English) is more efficient than postulating upcoming gaps that may later force reanalysis.

When looking at the adverb analyses we saw that in all conditions the pupil size was significantly different than zero: this may reflect the parser associating the filler with the gap site, thus decreasing processing costs. This is interesting as it seems that verb type, implausibility, and island-hood play little role when the parser is positing a gap. The parser seems to be positing a filler in the first gap it encounters, this is consistent with an active gap filling strategy and with a good enough strategy in which the parser is forming the simplest syntactic structure (even if it is ultimately incorrect).

In the fourth study, again we saw no main effects or interactions at the verb and adverb; this data does not support a hyper-active gap filling strategy. There were no measurable differences across any of the conditions at the verb, suggesting that parser may be consulting transitivity information when forming filler gap dependencies. At the adverb we did not have the

same across the board negative pupil slopes as seen in the third experiment.

These two studies taken together do not support a hyper-active gap filing strategy as we found no effect of intransitivity. The lack of difference between the intransitive conditions may reflect the parser consulting transitivity information before positing a gap. This contrasts the work done by Omaki et al. who found evidence for hyper-active gap filling in English. However we did not see an effect in the transitive conditions either, which may show an overall insensitivity of pupillometry to the stimuli. I do not think that is the case given the across the board negative pupil slopes at the adverb in Study 3 suggest that there is some sensitivity of the pupil to filler gap dependency formation with these stimuli.

The non-significance of these studies makes strong conclusions difficult; I argue the data reflect a good enough strategy in which the simplest syntactic construction is formed by the parser. However there are many confounds that need to be addressed in future studies. When looking at the analyses of rated trials the power is undermined with the high trial removal rate, perhaps simpler sentences may be used in future research. Like studies one and two, an auditory paradigm was employed which may reflect different parsing strategies compared to reading paradigms. It may be of interest to manipulate the speed of the auditory input to see if speech rate plays a role in forming filler gap dependencies. The two time windows for analyses overlap thereby making comparisons more difficult. Pupillometry itself is not sensitive to moment-to-moment processing which may be

reflected in the lack of significant change in all conditions and across time windows.

4.5. Conclusions

In conclusion, this data suggests that in English the parser may be consulting transitivity before positing a gap, which would support a conservative gap filling account. It appears that hyper-active gap filling may reflect a processing strategy adopted in verb-final languages, and it is not a cross-linguistic strategy. Filler gap dependencies, in English, seem to be constructed in line with a more conservative approach in which the parser waits to encounter the verb information before forming a dependency. In terms of positing a filler in a potential gap site, we find the parser is adopting a good enough strategy and forming the simplest syntactic structure and positing the filler at the first potential gap site.

CHAPTER 5: GENERAL CONCLUSIONS

The research in this thesis focused on how the parser comprehends filler gap dependencies. These dependencies are characterized by a moved constituent, which by some accounts leaves a phonologically silent but syntactically relevant trace of itself at the extraction site. It is via this gap site that the parser can correctly interpret the meaning of the sentence. Much research was reviewed in the second chapter and many theories were provided explaining ways in which the parser builds and interprets these types of sentences. The most prominent theories provided were Active Filler Strategies, Lexically Based Strategies, and the Good Enough Strategy. The Active filler strategy assumes that as soon as the parser encounters a filler it is actively searching for a gap and will from a dependency as soon as a gap site is encountered. The Lexically Based Strategy assumes that lexical information plays a role when forming filler gap dependencies. The parser essentially ranks possible structures (typically based on the verb) in parallel and the highest-ranking structure dictates how the sentence is interpreted. While the Good Enough Strategy assumes the parser will produce the simplest syntactic structure. The verb was also shown to be important when parsing filler gap dependencies, but in verb-final languages (like Japanese) evidence suggests the parser adopts a hyper-active gap filling strategy, in which pre-verbal information is utilized to predict upcoming gap sites.

The previous chapters also outlined exceptions to filler gap dependencies: island constraints and parasitic gaps. Island constraints were shown to block a filler and a gap from being associated, while parasitic gaps were exceptions to island constraints in which a gap following an island clause can “rescue” the illegal gap and make the sentence grammatical. I outlined the competing theories that have been put forward to explain island constraints (and subsequently the grammatical acceptance of parasitic gap structures). These were Reductionist based accounts and Grammar based accounts. Reductionist based accounts assume that island constraints are too taxing on the parser and lead to ungrammaticality. Grammar based accounts assume that island constraints are inherent to the grammar of the language, and the ungrammaticality of these construction is motivated by linguistic knowledge.

The set of studies in this thesis focused on adjudicating between the differing accounts of filler gap dependencies using pupillometry. As explained in Chapter 2, pupillometry involves measuring the change in pupil diameter over time, and reflects processing costs. Pupillometry was shown to be sensitive to a variety of linguistic tasks and was shown to be a reliable method of measuring processing load. The data from the empirical chapters allowed for adjudication between the different comprehension accounts, and has laid the groundwork for theoretical advancement.

Chapter 3 looked at parasitic gap dependencies. Incrementally these constructions have an island clause followed by a legal gap at the end of the sentence. It is not until the second gap is encountered that the construction becomes a legal construction. Grammar based accounts, as mentioned

previously, theorize that island constraints and parasitic gaps are inherent in the grammar of the language, while Reductionist based accounts on the other hand theorize that these island clauses are too taxing on the parser and therefore are too complex to parse. The studies in chapter 3 investigated parasitic gaps to see whether the parser is incrementally sensitive to parasitic gaps or whether the parser posits an “illegal” gap in an island when warranted by the preceding legal gap. If the parser is positing a gap in an island (when warranted), then this would work against the assumptions of the Reductionist based theories, because they claim that under no circumstance should the parser cognitively be able to posit a gap in an island.

When looking at the site of the embedded verb in the parasitic gap-like constructions there was a main effect of verb type, with the infinitive implausible condition showing a decrease in pupil slope (the finite sentences pupil slope did not differ than zero). This is essentially the reverse of what was found by Phillips (2006). In the current study, there was a slightly larger decrease in the implausible conditions compared to the plausible, suggesting that the implausible conditions were easier to process than the plausible conditions. This decrease may be the result of a dependency being formed across the conditions, with no effect of plausibility. However only a main effect of verb type was found and this may be the result of the processing involved with the two different sentence types (Pickering, et al., 1994).

At the N+1 time window a significant interaction was found between plausibility and verb type. There was no effect of plausibility in the infinitive

conditions, though the pupil diameter decreased (which suggests a dependency is being formed despite plausibility). The finite verb showed a large effect of plausibility, with the implausible condition showing an increase in pupil size and the plausible showing a decrease in pupil size. This is unexpected because the parser should not be forming a dependency. Finite verbs do not allow for a parasitic gap dependency, so under no circumstance should the embedded verb in the island be capable of hosting a dependency. The plausibility of the embedded verb should therefore play no role in parsing, as it is inaccessible to the parser, but our data suggests that the parser is forming a dependency despite island clauses. The (im)plausibility of the filler and gap is not evident until after the verb was encountered. Therefore, the parser is forming a dependency when encountering a finite verb only when it turns out to be a plausible object of the filler.

In terms of verb type, there was a decrease in pupil diameter in all of the infinitive conditions (at the verb and at N+1); this suggests that the parser posited the filler in the first potential gap site. The decrease in pupil diameter observed at both the verb and N+1 time window represent a decrease in processing and can be argued to represent the parser removing the filler from memory. Interestingly there was no effect of plausibility; the parser does not seem to be taking semantic information into account when positing the filler in the gap. However, the parser does seem to take semantic information into account downstream from the verb, at the finite verb, and will only form the dependency if the filler is a plausible object of the verb. The parser seems to be positing a gap in a good enough fashion

(Ferreira et al., 2002), in which the simplest syntactic structure is parsed without semantic information.

Behaviourally, semantics had a larger role compared to verb type (i.e. there were significant main effects of plausibility). The plausible sentences were given higher acceptability than the implausible, and the plausible sentences were also inaccurately answered more than the implausible sentences. This comprehension performance suggests that in the plausible conditions the parser is associating the filler and the gap even when not syntactically acceptable. In the implausible conditions, as expected, the parser is not forming a dependency based on the semantic impossibility of the situations in real life. This data is in line with other research suggesting that original incorrect online interpretations persevere during offline performance (Christianson et al., 2001).

The data from the online pupillometry study with the offline comprehension data suggest that the parser is not always sensitive to island constraints. The parser will form a dependency with an embedded verb, this goes against the Reductionist based theories, however the data does not support a Grammar based theory either. The parser is capable of positing a gap in an island, but dependency formation does not seem to be an inherent grammatical rule. Rather the parser seems to be using an Active Filler Strategy to posit gaps in a good enough fashion.

In chapter 4, pupillometry was used to investigate whether the parser is utilizing pre-verbal information when forming filler gap dependencies, also known as hyper-active gap filling. On the other hand the parser could be employing what the authors call conservative gap filling, in which the parser

waits to consult verb information before positing a gap. Aoshima et al. (2004 found evidence suggesting that in Japanese (a verb-final language) the parser predicts upcoming gap sites before encountering the verb, suggesting that gap filling is an active process that may not be verb driven. Given the early placement of the verb in English, if the parser employs a hyper-active strategy it would be at the risk of a later reanalyses (if the verb is incapable of hosting a filler). A conservative gap filling strategy, on the other hand, would be at the risk of taxing the cognitive resources available to the parser, as the parser has to maintain the filler in working memory. Omaki et al. (2012) found evidence for hyper-active gap filling in English.

The data provided in Chapter 4 did not support a hyper-active gap filling account. Across both experiments, in all conditions and at all locations, no interactions or main effects were found. We did not see an effect of intransitivity at the verb in either study, suggesting that the parser consulted transitivity information before positing a gap. At the adverb in study 3 the parser seems to be positing a gap directly after the verb regardless of island constraints, verb type, or plausibility. In Study 4, the pupil slopes did not differ from zero in any conditions. The lack of difference between the two transitive conditions is surprising, indicating a lack of sensitivity to island constraints.

The findings from the two studies in Chapter 4 suggest that the parser is employing an Active Filler Strategy, and forming the simplest possible construction regardless of other linguistic information. We found no evidence to support a lexically based account of gap filler dependencies, rather, the data suggests a more conservative gap filling strategy in which the parser

waits to consult the verb before positing a gap. The data is interesting because even after consulting the verb the parser seems to posit a gap regardless if it is warranted by syntax or is semantically viable. The parser seems to be adopting a good enough strategy and is parsing the easiest possible construction; not taking any semantic or syntactic evidence into account and positing a gap immediately after the verb.

The two empirical chapters in this study support a parser that is actively forming dependencies with a good enough strategy, in which the simplest construction is parsed and produced (whether it is syntactically appropriate or not) and the interpretation lingers even in offline performance. The parser seems to be adopting an Active Filler Strategy in which a dependency is formed to alleviate processing costs as soon as possible. The data also suggests that the parser is not using pre-verbal information; rather it waits to consult verb properties before positing a gap. The parser does not predict the upcoming syntactic structure before reaching the verb given the relatively early position of the verb in English. The parser does not need a pre-verbal predictive strategy given that holding the filler in memory is not as costly as the potential of positing an incorrect gap and forcing reanalysis. Pre-verbal gap formation seems to be a parsing strategy for verb-final languages and not for verb-medial languages.

5.2. Limitations and future studies

As mentioned previously the lack of significance (particularly in Chapter 4) makes strong conclusions difficult and highlights the need for more research in the future. There are several confounds that should be addressed in future studies. Pupillometry reveals information about processing costs, but is not

sensitive to moment-to-moment processing, and given our overlapping critical time windows this may have driven some of this non-significance. Given the limitations of pupillometry and the limitations of ERPs (to reveal whether a gap has been posited after crossing an island boundary) I suggest a study that employs both ERP and pupillometry. Thus giving insight into the moment-to-moment processing of island boundaries, and the processing that occurs post-boundary crossing.

Additionally the speech rate of the auditory input may be causing the parser to adopt a good enough strategy when forming dependencies. Future research could investigate the role that auditory input plays in forming filler gap dependences. The stimuli themselves were also quite difficult, future research should address this with simpler island and parasitic gap constructions.

5.3. Conclusions

Overall this data suggests when forming filler gap dependencies the parser adopts the most efficient strategy and is actively filling gaps. Also, the parser does not take pre-verbal information into account, rather it waits to consult the verb and then form a dependency. This dependency occurs regardless of verb type or plausibility, suggesting the parser wants to reduce the demand on working memory as soon as possible. The parser will consult plausibility information with certain verb types (finite verbs that do not allow for parasitic gap dependencies) but this does not occur until after the verb has been encountered and the gap posited. Offline measures confirm that the original (often inaccurate) interpretation of the dependency lingers.

The data do not support the current filler gap dependency theories; the data cannot be accounted for with Reductionist or Grammar based accounts. I argue that the parser adopts an active strategy of gap filling and parses the simplest structure that is good enough. This may reflect the auditory nature of the stimuli; in real-time conversations the parser must construct and comprehend structures very rapidly to ensure the continuation of communication, and will therefore adopts a strategy in which structures are good enough.

REFERENCES

Ahern, S. K., & Beatty, J. (1979). Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science, 205 (4412),* 1289-1292.

Ahern, S.K., & Beatty, J. (1981). Physiological evidence that demand for processing capacity varies with intelligence. In M.P. Friedman, J.P. Das, & N. O'Conner (Eds.), *Intelligence and learning* (pp. 201-216). New York: Plenum Press.

Altmann, G.T.M.(1999). Thematic role assignment in context. *Journal of Memory and Language, 41,* 124-145.

Altmann, G.T.M, & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73(3),* 247–264.

Aoshima, S., Phillips, C., & Weinberg, A. (2004). Processing filler gap dependencies in a head-final language. *Journal of Memory & Language, 51,* 23-54.

Baddeley, A. D. (1986). *Working memory*. New York: Oxford Univ. Press.

Barton, S.B., & Sanford, A.J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition, 21,* 477-487.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91*, 276-292.

Beatty, J. & Wagoner, & B.L. (1978). Pupillometric signs of brain activation vary with level of cognitive processing. *Science, 199*, 1216-1218.

Ben-Nun, Y. (1986). The use of pupillometry in the study of on-line verbal processing: Evidence for depths of processing. *Brain and Language, 28*, 1-11.

Boland, J.E., Tanenhaus, M.K., Garnsey, S.M., & Carlson, G.N. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language, 34,* 774-806.

Bourdages, J.S. (1992). Parsing complex NPs in French. In H. Goodluck, & M.S. Rochemont (Eds.), *Island constraints: Theory, acquisition and processing* (pp. 61–87). Dordrecht: Kluwer.

Christianson, K., Hollingworth, A., Halliwell, J., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*, 368-407.

Chomsky, N. (1965). *Aspects of the Theory of Syntax.* Cambridge, MA: MIT Press.

Clifton, C., & Frazier, L. (1989). Comprehending sentences with long-distance dependencies. In M. Tanenhaus & G. Carlson (eds.) *Linguistic Structure in Language Processing* (pp. 273-317). Dordrecht: Kluwer.

Clifton, C., Frazier, L., & Connine, C. (1985). Lexical expectations in sentence comprehension. *Journal of Verbal Learning and Verbal Behavior, 23,* 696 – 708.

Deane, P. (1991). Limits to attention: A cognitive theory of island constraints. *Cognitive Linguistics, 2*, 1–63.

Dooly, R. P., & Lehr, D. J. (1967) Critique of a pupillary response experiment. *Perceptual and Motor Skills, 25*, 603-604.

Engelhardt, P.E., Ferreira, F., & Patsenko, E.G. (2010). Pupillometry reveals processing load during spoken language comprehension. *The Quarterly Journal of Experimental Psychology, 63*, 639-645.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25,* 348-368.

Ferreira, F., Bailey, K.G.D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*, 11-15.

Ferriera, F., & Patson, N.D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistic Compass,1,* 71-83.

Fordor, J.D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry, 9,* 427-474.

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

Frazier, L.,& Flores D'Arcais, G.B. (1989). Filler driven parsing: a study of gap filling in Dutch. *Journal of Memory and Language, 28(3)*, 331–344.

Freedman, S.E., & Forster, K.I. (1985). The psychological status of overgenerated sentences. C*ognition* 19, 101–31.

Garnsey, S.M., Pearlmutter, N.J., Myers, E., & Lotocky, M.A. (1997). The

contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language, 37*, 58-93.

Garnsey, S.M., Tanenhaus, M.K., & Chapman, R.M. (1989). Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research: Special Issue on Sentence Processing, 18*, 51-60.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*, 1-76.

Gibson, E., & Hickok, G. (1993). Sentence processing with empty categories. *Language and Cognitive Processes, 8,* 147-161.

Gorrel, P. (1993). Evaluation the direct association hypothesis: A reply to Pickering and Barry (1991). *Language and Cognitive Processes, 8,* 129-146.

Gutiérrez, R.S., & Shapiro, L.P. (2010) *Measuring the Time-course of Sentence Processing with Pupillometry*. Poster presented at the 23rd annual meeting of the CUNY conference on human sentence processing. New York, New York.

Hess, E.H. (1975). *The Tell-Tale Eye: How Your Eyes Reveal Hidden Thoughts and Emotions*. New York: van Nostrand Reinhold

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science, 132,* 349-350.

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science, 143*, 1190-1192.

Hyönä, J., Tommola, J., & Alaja, A-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology, 48A,* (3), 598-612.

Janisse, M.P. (1973). Pupil size and affect: a critical review of the literature since 1960. *The Canadian Psychologist, 14*(4), 311-329.

Just, M.A., & Carpenter, P.A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology, 47*(2), 310-339.

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, N.J.: Prentice-Hall.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154*, 1583-1585.

Kahneman, D., & Beatty, J. (1967). Pupillary response in a pitch discrimination task. *Perception & Psychophysics, 2*, 101-105.

Kahneman, D., & Wright, P. (1971). Changes of pupil size and rehearsal strategies in a short-term memory task. *Quarterly Journal of Experimental Psychology, 23*, 187-196.

Klingner, J. (2010). Measure cognitive load during visual tasks by combining pupillometry and eye tracking. (Doctoral dissertation, Stanford University, 2010).

Kluender, R., & Kutas, M. (1993). 1993. Subjacency as a processing phenomenon. *Language and Cognitive Processes, 8*, 573–633.

Kuipers, J.R., & Thierry, G. (2011). N400 amplitude reduction correlates with an increase in pupil size. *Frontiers in Human Neuroscience, 5*(61), 1-5.

Lau, E., Phillips, C., & Poepell, D. (2008). A cortical network for semantics: (de)constructing, the N400. *Nature Reviews Neuroscience, 9,* 920-933.

MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101,* 676-703.

McElree, B., & Griffith, T. (1998). Structural and lexical constraints on filling gaps during sentence comprehension: A time-course analysis. *Journal of Experimental Psychology, 24,* 432-460.

McKinnon, R., & Osterhout, L. (1996). Constraints on movement phenomena in sentence processing: Evidence from event-related potentials. *Language and Cognitive Processes, 11*, 495–523.

Neville, H., Nicol, J., Barss, A., Forster, K., & Garrett, M. (1991). Syntactically-based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience, 3,* 151-165.

Omaki, A, Lau, E., Davidson White, I., & Phillips, C. (2012). *Hyper-active gap filling: Pre-verbal object gap creation in English filler gap dependency processing.* Unpublished manuscript.

Osterhout, L., & Holcomb, P.J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31,* 785-806.

Osterhout, L., & Holcomb, P.J. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes, 8,* 413-438.

Phillips, C. (2006). The real-time status of island phenomena. *Language, 82*, 795-832.

Pickering, M.J. (1993). Direct association and sentence processing: A reply to Gorrell and to Gibson and Hickok. *Language and Cognitive Processes, 8,* 163-196.

Pickering, M.J., & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes, 6,* 229-259.

Pickering, M.J., Barton, S., & Shillcock, R. (1994). Unbounded dependencies, island constraints and processing complexity. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on Sentence Processing* (pp. 199-224). London: Lawrence Erlbaum.

Pickering, M.J., & Traxler, M. J. (2003). Evidence against the use of subcategorisation information in the processing of unbounded dependencies. *Language and Cognitive Processes, 18,* 469-503.

Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology, 47*, 560-569.

Pritchett, B.L. (1991). Subjacency in a principle-based parser. In, R.C. Berwick (Ed.), P*rinciple-Based Parsing: Computation and Psycholinguistics* (pp. 301-345). Dordrecht: Kluwer.

Ross, J.R. (1967). *Constraints on variables in syntax*. Cambridge, MA: MIT dissertation.

Scott, T. R., Wells, W. H., Wood, D. Z. & Morgan, D. I. (1967). Pupil response and sexual interest re-examined. *Journal of Clinical Psychology, 23*, 433-438.

Schluroff, M. (1982). Pupil responses to grammatical complexity of sentences. *Brain and Language, 17*, 133-145.

Schluroff, M., Zimmermann, R.B., Freeman, Jr., Hofmeister, K., Lorscheid, T., & Weber, A. (1986). Pupillary responses to syntactic ambiguity of sentences. *Brain and Language, 27*, 322-344.

Sprouse, J., Wagers, M., & Phillips, C. (2012). A test of relation between working memory and syntactic island effects. *Language, 88 (1),* 82-123.

Staub, A. (2007). The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 550-569.

Stowe, L.A. (1986). Evidence for on-line gap-location. *Language and Cognitive Processes, 1*, 227–245.

Sussman, R.S., & Sedivy, J.C.(2003). The time course of processing syntactic dependencies: evidence from eye movements. *Language and Cognitive Processes, 18,* 143–161.

Traxler, M., & Pickering, M J. (1996). Plausibility and the processing of unbounded dependencies: an eye-tracking study. *Journal of Memory and Language, 35*, 454–475.

Tanenhaus, M.K., Stowe, L.A. & Carlson, G.N. (1985). *The interaction of lexical expectation and pragmatics in parsing filler gap constructions*. Proceedings of the Seventh Annual Cognitive Science Society Meetings. Irvine, California.

Wright, P., & Kahneman, D. (1971). Evidence for alternative strategies of sentence retention. *Quarterly Journal of Experimental Psychology, 23*, 197-213.