An ERP investigation of individual differences in the processing of *wh*-dependencies by native and non-native speakers

By

Lauren Covey

Submitted to the graduate degree program in Linguistics and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Co-Chair: Dr. Alison Gabriele

Co-Chair: Dr. Robert Fiorentino

Dr. Utako Minai

Dr. Annie Tremblay

Dr. Yan Li

Date Defended: June 12, 2018

The dissertation committee for Lauren Covey certifies that this is the
approved version of the following dissertation:

An ERP investigation of individual differences in the processing of *wh*-dependencies by native and non-native speakers

Co-Chair: Dr. Alison Gabriele

Co-Chair: Dr. Robert Fiorentino

Date Approved: June 12, 2018

Abstract

This study examines the processing of *wh*-dependencies by native English speakers and Mandarin Chinese-speaking learners of English. *Wh*-dependencies involve a long-distance relationship between a fronted *wh*-word (e.g., *who*) and the position in the sentence where it originated, called a gap site. The examination of *wh*-dependency resolution presents an interesting test case for whether or not grammatical knowledge is used online because, in languages such as English, *wh*-movement is constrained such that extraction is only possible from certain positions and is barred from other positions, called islands (Ross, 1967). In examining whether native speakers and second language (L2) learners are sensitive to island constraints online, this study tests predictions of prominent L2 processing theories which argue that adult learners are unable to utilize abstract grammatical information during processing (e.g., Clahsen & Felser, 2006). In the native literature, the processing of *wh*-dependencies has been proposed to involve at least two distinct processes, a predictive process in which the parser searches for a potential gap site, and an integrative process, when the dependency is successfully resolved at the gap site. The broader electrophysiological literature has linked these qualitatively different processes to distinct event-related potential (ERP) components: the N400 for prediction (e.g., Federmeier, 2007; Lau et al., 2008; 2013; Michel, 2014; Van Berkum et al., 2005), and the P600 for syntactic integration (e.g., Gouvea et al., 2010; Kaan et al., 2000; Phillips et al., 2005). Although previous ERP studies have examined these components independently, few studies have tracked the dynamics of *wh*-dependency resolution across the sentence, examining both prediction and integration to investigate whether these processes are indexed by unique components. The present study takes this approach, focusing on the processing of *wh*-dependencies at three critical regions across the sentence, two of which are associated with prediction, and one with integration. This study additionally investigates the extent to which the use of grammatical knowledge during online processing and the ability to engage in predictive processing is modulated by proficiency in L2 learners, and performance on a range of cognitive measures in native speakers. Results show that both native speakers and highly proficient learners engage in gap prediction during processing, although this is limited to certain contexts for learners. In examining processing inside of an island, a position from which extraction is prohibited, the current study shows that native speakers and highly proficient learners are guided by grammatical knowledge. Finally, both natives and learners show evidence of successful dependency resolution at the actual gap site, even in sentences with islands. Overall, the results present a complex picture of processing *wh*-dependencies by native English speakers and Mandarin-speaking learners of English, showing that while both native speakers and learners with higher proficiency are able to use grammatical information during online processing, the contexts in which L2 learners are able to predict differ from native speakers.

Acknowledgments

This dissertation could not have been completed without the generous time and support of many people. First and foremost, I would like to thank my co-advisors, Dr. Alison Gabriele and Dr. Robert Fiorentino, who have gone above and beyond to support, challenge, and guide me throughout my doctoral studies. In and outside of the classroom, they helped me approach and understand the complexities of SLA, psycholinguistics, and neurolinguistics, and showed me what it means to be an excellent mentor. I thank Dr. Gabriele for her thoughtful feedback, advice, and insight in all aspects of my graduate studies. I thank Dr. Fiorentino for teaching me the intricacies of EEG, and for allowing me to learn from him about experimental design, journal reviews, conference presentations, and much more.

I also wish to thank my committee members, Dr. Utako Minai, Dr. Annie Tremblay, and Dr. Yan Li. I thank Dr. Minai for her guidance teaching me about first language acquisition and for her generosity in allowing me to work with her on an eye-tracking study. I thank Dr. Tremblay for her insightful feedback on each stage of many projects, which greatly helped strengthen my research skills in all areas. I also wish to thank other professors in the department who taught me in their classes, and often gave me important feedback about my research: Dr. Joan Sereno, Dr. Allard Jongman, Dr. Jie Zhang. I thank Dr. Zhang in particular, who made the important observation about the role of L1 Mandarin in interpreting my results.

I received grant funding for my dissertation, which I also gratefully acknowledge. A Dissertation Grant from the journal Language Learning allowed me to fund participant payment for the Mandarin Chinese learners of English. I also received a Doctoral Dissertation Improvement Grant from the National Science Foundation, as well as the Dissertation Fellowship in the Cognitive, Clinical, and Neural Foundations of Language from the William

Orr Dingwall Foundation. This funding was crucial in allowing me to conduct a large-scale study in order to pay participants, purchase lab supplies, and hire undergraduate research assistants.

I thank all of the Mandarin learners and native English speakers at the University of Kansas who spent three hours in the lab to make this dissertation possible. My research assistants also made this dissertation feasible. In particular, I thank Catherine Pham for her dedication and helpfulness in data collection. Thanks to the entire RA team for their many hours in the lab: Tom Lieber, Emily Jacobs, Ran Lu, Justin Nguyen, Haley Schippers, and Tristy Goetz.

I also wish to thank my wonderful friends for the constant support, advice, and listening skills they graciously gave me throughout the PhD. I thank Dr. Katrina Connell for her friendship and both academic and personal support every step of the way. Dr. Kate Coughlin provided me with statistics help regularly, and even more often, cat sitting help; I am lucky that one of my grad school role models is also my dear friend. I thank Lena Roesner and Longcan Huang for their friendship and support, especially during the semester we shared the Blake 113 office, one of the highlights of these five years. I am grateful to have found a group of kind and fun friends with whom I can also discuss research design and linguistics debates: thank you Nick Feroce, Xiao Yang, Seulgi Shin, and Charlie Redmon. Thanks also to the members of the Research in Acquisition and Processing (RAP) group, who provided me with excellent feedback and advice on my research and presentations.

I am also grateful to my family, who made every opportunity available to me to pursue graduate studies in Kansas. To my parents, I thank you for your unwavering support and love over the past five years, and for encouraging me to follow my dreams wherever they take me. Last, but certainly not least, I thank my husband, Thomas; I dedicate this dissertation to my partner and greatest supporter.

Table of Contents

Chapter 1: Introduction 1
Chapter 2: Processing of wh-dependencies by native English speakers 7
  Introduction 7
  Literature Review 8
    Psycholinguistic approaches to the processing of wh-dependencies 8
    Individual Differences in Predictive Processing 15
    Island sensitivity 18
    Neurolinguistic approaches to the processing of wh-dependencies 25
  Current Study 31
  Methods 34
    Participants 34
    EEG Experiment 34
    Offline Acceptability Judgment Task 37
    Cognitive Measures 38
  Data Analysis 39
    EEG Analysis. 40
    Individual Difference Analyses. 42
  Behavioral Results 43
  EEG Results 44
    Subject Position 45
    Object Position 48
    Actual Gap Site 52

Individual Difference Analyses 57

  Subject Position: 300-500 ms time-window analysis. 57

  Object Position: 300-500 ms time-window analysis. 58

  Object Position: 500-900 ms time-window analysis. 61

  Actual Gap Site: 500-900 ms time-window analysis... 63

Discussion 65

  Gap predictability and the N400 65

  Island sensitivity 70

  Dependency resolution and the P600... 71

  Late negativity at the object position 76

Conclusions.. 80

Chapter 3: Processing of wh-dependencies by L2 learners 81

Introduction.. 81

Literature Review 83

  Island sensitivity in L2 learners 83

  Predictive Processing in L2 Learners 95

  Neurolinguistic approaches examining L2 processing of filler-gap dependencies.............. 98

Current Study 104

Methods 108

  Participants... 108

Data Analysis 110

  EEG Analysis... 110

  Individual Difference Analysis. 110

  Behavioral Results 110
  EEG Results 111
    Subject Position 112
    Object Position 115
    Actual Gap Site 122
  Individual difference analysis examining offline island sensitivity 127
    Object Position: 500-900 ms time-window analysis. 128
  Discussion 128
    Subject Position 130
    Object Position 133
    Actual Gap Site 138
  Conclusions 142
Chapter 4: Overall Discussion 144
  Prediction in language processing 145
  Island constraints in the processing of wh-dependencies 149
  Wh-dependency resolution 151
  Future Research 153
  Conclusions 154
References 156
Appendix A: Statistical Tables, Native Speaker study 169
Appendix B: Statistical Tables, L2 learner study 178

Chapter 1: Introduction

A central debate in the language processing literature focuses on how grammatical knowledge is employed during sentence processing. A related issue in the field of second language (L2) acquisition regards whether non-native speakers process complex syntactic dependencies in a native-like way. The processing of *wh*-dependencies provides an ideal test case to investigate these research questions in the native and L2 sentence processing literatures, and is the focus of the present study.

In languages such as English, questions and relative clauses have been argued to involve *wh*-movement (e.g., Chomsky, 1981, 1986). The relationship between the fronted *wh*-word and the position in the sentence from which it originated is called a *wh*-dependency. For example, in the embedded question in (1b), the *wh*-element has been displaced from its original position in the syntactic structure, called the gap site, and moved to the beginning of the clause.

(1) a. Harry ate chocolate.
b. I wonder *what* Harry ate ___.

During real-time processing, it is not possible to integrate a fronted *wh*-phrase into the syntactic and semantic representation immediately upon encountering it; instead, the *wh*-expression must be structurally integrated and interpreted at the actual gap site. Thus, upon identifying a *wh*-item, the parser must search for and identify the actual gap. In the native literature, this process has been characterized as predictive, such that the parser predicts potential gap sites prior to encountering information in the bottom-up signal confirming the location of the actual gap (e.g., Frazier & Clifton, 1989). A second, distinct process takes place at the gap site, where the dependency is successfully resolved and the *wh*-item is integrated into the sentence.

The current study examines whether *wh*-dependency resolution proceeds similarly in L2 processing, testing whether learners predict upcoming gaps. One prominent account, the Shallow Structure Hypothesis (Clahsen & Felser, 2006), argues that learners are unable to rapidly utilize abstract syntactic information during sentence processing and instead must rely on semantic or pragmatic information to resolve syntactic dependencies (Felser & Roberts, 2007; Marinis, Roberts, Felser, & Clahsen, 2005). Under this hypothesis, L2 learners do not posit syntactic gaps during processing, but instead rely on semantic/pragmatic information from the subcategorizing verb to match thematic arguments. Another proposal states that in general, learners have a reduced ability to generate expectations as compared to natives, and thus predicts that L2 learners are unlikely to show effects of prediction during processing (Grüter, Lew-Williams, & Fernald, 2012; Grüter, Rohde, & Schafer, 2017). Others propose that predictive processing is possible for L2 learners, but the ability to predict may be modulated by a range of factors, including proficiency and individual differences in cognitive abilities (Hopp, 2013; Kaan, 2014). This study examines gap prediction in two contexts, pre-verbally in the subject position and post-verbally in the object position, providing insights on the conditions under which predictive processing in the L2 is possible, allowing us to better understand the possibilities and limitations of adult L2 acquisition.

A second line of native and L2 research examines to what extent grammatical information is used to predict potential gap sites, investigating whether the parser avoids predicting gaps within islands. *Wh*-movement in English is constrained, such that extraction is only possible from certain positions and is barred from other positions, which have been called ‘islands’ (Ross, 1967). For example, (2a) contains a relative clause, one type of island domain, and (2b) demonstrates that extracting a *wh*-item from the relative clause renders the sentence ungrammatical.

(2) a. Meghan likes the store [that sells dark chocolate].
b. *Which chocolate does Meghan like the store [that sells __]?

While it has been observed that native speakers respect island constraints during the processing of *wh*-dependencies (e.g., Phillips, 2006; Stowe, 1986; Traxler & Pickering, 1996), a debate remains regarding whether non-native speakers are similarly guided by grammatical knowledge in the online resolution of *wh*-dependencies (e.g., Aldwayan, Fiorentino, & Gabriele, 2010; Canales, 2012; Felser, Cunnings, Batterham, & Clahsen, 2012; Johnson, Fiorentino, & Gabriele, 2016; Kim, Baek, & Tremblay, 2015; Omaki & Schulz, 2011). The Shallow Structure Hypothesis, for example, argues that because learners cannot utilize abstract syntactic information in the course of processing *wh*-dependencies, they should not show sensitivity to syntactic islands online. In other words, this account predicts that learners will attempt to resolve a *wh*-dependency (via thematic argument linking) inside of an island domain. However, several studies have reported native-like processing of islands, showing that like natives, learners do not attempt to posit gaps inside islands (e.g., Johnson et al., 2016; Kim et al., 2015; Omaki & Schulz, 2011). The current study addresses the island sensitivity debate, and further investigate the extent to which the use of grammatical knowledge during online processing is modulated by L2 proficiency.

By examining the processing of islands, this dissertation also addresses a theoretical debate in the native literature regarding the nature of island constraints (e.g., Hofmeister, Casanto, & Sag, 2012a/b, 2013; Hofmeister & Sag, 2010; Kluender, 2004; Sprouse, Wagers, &

Phillips, 2012). Grammatical accounts propose that gap-filling in islands is avoided due to knowledge of the syntactic constraints which govern *wh*-movement (Sprouse et al., 2012), while processing accounts argue that the parser avoids positing gaps within islands because they present processing bottlenecks (e.g., Hofmeister & Sag, 2010; Kluender, 2004; Kluender & Kutas, 1993b). Under this second approach, the parser avoids positing gaps within islands due to increased difficulty resolving *wh*-dependencies in complex structures. It has been proposed that one way to tease apart these two kinds of accounts is to examine the role of individual differences in the processing of *wh*-dependencies within islands. Sprouse et al. (2012) argued that processing accounts predict that positing gaps in islands should be possible for individuals with increased processing resources. The current study addresses this debate by testing native speakers on a battery of cognitive measures, allowing us to test for a relationship between processing abilities and the positing of gaps inside island structures.

This dissertation investigates the processing of *wh*-dependencies by native speakers and L2 learners utilizing electroencephalography (EEG), a technique that can shed light on both the time-course of processing *wh*-dependencies and can provide information about whether native speakers and L2 learners process these dependencies using qualitatively similar mechanisms. The broader electrophysiological literature has shown two distinct event-related potential (ERP) components to be involved in the processing of *wh*-dependencies. The N400 component has been shown to be modulated by prediction (e.g., Federmeier, 2007; Lau, Holcomb, & Kuperberg, 2013; Van Berkum, Brown, Zwisterlood, Kooijman, & Hagoort, 2005; for a review see Lau, Phillips, & Poeppel, 2008), and a recent study by Michel (2014) has suggested that the N400 may also reflect prediction of gap sites during *wh*-dependency resolution. The second component, P600, has been elicited in studies examining dependency resolution, and is suggested to reflect syntactic integration processes resulting from integrating the *wh*-word at the gap site (Felser, Clahsen, & Münte, 2003; Gouvea, Phillips, Kazanina, & Poeppel, 2010; Kaan, Harris, Gibson, & Holcomb, 2000; Phillips, Kazanina, & Abada, 2005). The ERP literature has thus far generally examined these components independently, and few studies have tracked the dynamics of *wh*-dependency resolution across the sentence to examine both prediction and integration. The present study utilizes this approach to examine the processing of *wh*-dependencies at three critical regions across the sentence, two of which are associated with prediction, and one with integration, to investigate whether these processes are indexed by unique components in both native speakers and L2 learners. The study is one of the first ERP studies to examine sensitivity to grammatical constraints in the processing of *wh*-dependencies using grammatical sentences, making it one of the first L2 ERP studies to not rely on a violation paradigm and allowing us to examine more natural language comprehension.

The structure of this dissertation is as follows. Chapter 2 includes the native speaker study, first reviewing evidence from psycholinguistic and neurolinguistic studies examining gap prediction and dependency resolution during the processing of *wh*-dependencies. This chapter outlines the linguistic manipulation used in the study (filled-gap paradigm) and describes all experimental methods. Results for the native speakers are then described and interpreted in light of relevant literature. Chapter 3 focuses on the L2 learner population, and describes the two major theoretical SLA debates addressed by the study. A review of studies which investigate L2 learners’ processing of *wh*-dependencies and island constraints will follow. The learner population and the role of the native language (L1) is outlined, followed by the results and discussion for the study. Chapter 4 includes an overall discussion of the study and a detailed

comparison of the native and non-native results. Future directions for this line of research are also discussed.

Chapter 2: Processing of *wh*-dependencies by native English speakers

Introduction

Early psycholinguistic studies revealed that during real-time language comprehension, native speakers do not wait to construct a syntactic structure and interpret the phrase or sentence for meaning, but rather do so incrementally (e.g., Altmann & Steedman, 1988; Boland, Tanenhaus, Garnsey, & Carlson, 1995; Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Kutas & Hillyard, 1984; Marslen-Wilson, 1975; Marslen-Wilson & Tyler, 1980; Pickering, 1994; Steedman, 1989; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Traxler, Bybee, & Pickering, 1997; Tyler & Marslen-Wilson, 1977). Furthermore, recent psycholinguistic and neurolinguistic research has suggested that native speakers not only process sentences incrementally, but actively form predictions about what might come next (e.g., Altmann & Kamide, 1999; Altmann & Mirkovic, 2009; DeLong, Urbach, & Kutas, 2005; Jaeger & Snider, 2013; Kamide, Altmann, & Haywood, 2003; Levy, 2008; MacDonald, 2013; Pickering & Garrod, 2011, 2013; Van Berkum et al., 2005; Van Petten & Luka, 2012; for a review see Kutas et al., 2011). One major focus of this literature is on lexical-semantic prediction occurring in semantically-constraining sentence contexts (e.g., the word *kite* in "*The day was breezy so the boy went outside to fly…*", DeLong et al., 2005). This dissertation focuses on syntactic prediction in particular, examining whether native speakers of English anticipate upcoming syntactic structure during the processing of *wh*-dependencies. As reviewed below, it has been proposed in the psycholinguistic literature that *wh*-dependency formation involves active prediction of syntactic structure; in what follows, we discuss findings and open questions from the existing psycholinguistic and neurolinguistic literature which examine the processing of *wh*-

dependencies, as well as examining how the resolution of *wh*-dependencies may be modulated by individual differences in cognitive abilities.

Literature Review

Psycholinguistic approaches to the processing of wh-dependencies

As discussed in Chapter 1, during real-time processing it is not possible to integrate a fronted *wh*-element into the syntactic and semantic representation immediately upon encountering it. Instead, the *wh*-expression must be structurally integrated and interpreted at the actual gap site. In the psycholinguistics literature, researchers have investigated the process by which the parser posits a gap within the structure in order to link the *wh*-element with its licensor, examining *where* in the structure gaps are predicted, as well as the time-course of this process, asking *when* gap prediction occurs. The formation of *wh*-dependencies has been characterized as proactive, such that the parser attempts to resolve a *wh*-dependency (i.e., posit a gap) in a top-down manner before confirming evidence in the bottom-up input that confirms the position of an actual gap site (i.e., a missing constituent) (e.g., Crain & Fodor, 1985; de Vincenzi, 1991; Frazier, 1987; Fraizer & Clifton, 1989; Nicol, 1993; Nicol & Swinney, 1989; Pickering & Traxler, 2001, 2003). One prominent hypothesis that makes specific claims about *wh*-dependency resolution is the *Active Filler Strategy*. This hypothesis states that the parser prioritizes the resolution of a *wh*-dependency during processing, ranking the assignment of the gap site above all else. This means that the search for a potential gap site begins immediately upon encountering a filler and continues until the *wh*-dependency is successfully resolved. The parser will predict potential gap positions at each grammatically possible position in the sentence until the dependency is completed, rapidly constructing a possible syntactic structure during the prediction of possible gap sites (e.g., Clifton & Frazier, 1989; Frazier, 1987; Frazier & Clifton, 1989; Frazier & Flores D'Arcais, 1989).

Evidence for the active search of gap positions has primarily come from two types of behavioral studies examining the online processing of filler-gap dependencies, one of which utilizes a plausibility mismatch manipulation. For example, in their influential eye-tracking study, Traxler and Pickering (1996) manipulated whether a filler semantically fit with a verb which was a potential gap licensor. For example, in (3a) below, the filler *book* thematically matches the verb *wrote*, although the filler *city* in (3b) is not a plausible object of the verb. Note that (3b) is ultimately grammatical, as the actual gap site in both sentences actually follows the preposition *about* rather than the optionally transitive intermediate verb.

(3) a. We like the *book* that the author **wrote** unceasingly and with great dedication about __ while waiting for a contract.

b. We like the *city* that the author **wrote** unceasingly and with great dedication about __ while waiting for a contract.

Traxler and Pickering (1996) reported that native English speakers showed longer first-pass reading times at the critical verb *wrote* for sentences such as (3b), in which the filler was a semantically implausible object, as compared to when the filler was plausible (3a). This difference in reading times is referred to as a plausibility mismatch effect. This effect is taken to suggest that the parser relies on an immediate association strategy, forming a link between the filler and the verb as soon as the verb is encountered, even when the filler was not a semantically plausible argument for the verb (e.g., Pickering, 1993). Plausibility mismatch effects have been reported in other eye-tracking and self-paced reading studies (e.g., Marinis et al., 2005; Omaki et al., 2015; Stowe, Tanenhaus, & Carlson, 1991; Tanenhaus, Boland, Garnsey, & Carlson, 1989;

Williams, 2006), as well as EEG studies (e.g., Dallas, DeDe, & Nicol, 2013; Garnsey, Tanenhaus, & Chapman, 1989; Jessen & Felser, 2018).

Another psycholinguistic approach that researchers have utilized to examine the processing of *wh*-dependencies is a ‘filled-gap’ manipulation. Filled-gap effects occur when the parser predicts a gap in a position that is already filled, resulting in a reading time slowdown (e.g., Boland et al., 1995; Bourdages, 1992; Crain & Fodor, 1985; Frazier & Clifton, 1989; Stowe, 1986). For example, in Stowe’s (1986) seminal study, participants read sentences such as (4) which contained three potential gap positions: the subject position (e.g., *Ruth*), the direct object position (e.g., *us*), and the prepositional object position (e.g., following *to*), which is the actual gap position in sentence (4a) below.

(4) a. My brother wanted to know *who* **Ruth** will bring **us** home to __ at Christmas.
b. My brother wanted to know *if* **Ruth** will bring **us** home to Mom at Christmas.

Upon encountering the *wh*-filler (e.g., *who*), the parser is hypothesized to predict a gap site at the immediately adjacent subject region. As this region is filled with lexical material (*Ruth*), it is predicted to yield a reading time slowdown in the *wh*-extraction sentence (4a) as compared to the declarative sentence without extraction (4b). A reading time slowdown at *Ruth* in (4a) vs. (4b) is referred to as a subject filled-gap effect. The parser is expected to continue the search for an actual gap site, and next predict an object gap. Since the object position is also filled (with *us*), a filled-gap effect is also predicted at the object position, with longer reading times for *us* in (4a) as compared to (4b). The dependency is ultimately completed upon encountering the actual gap site, in the prepositional object position. Stowe (1986) compared reading times at the subject and object positions in sentences with *wh*-extraction (4a) to declarative sentences without extraction

(4b). The results revealed an object filled-gap effect, which Stowe interprets as the parser actively predicting an abstract gap position before confirming that the direct object position was available.

Stowe (1986) did not observe reading time slowdowns in the subject position, and subsequent studies have provided mixed results with respect to the presence or absence of subject filled-gap effects (e.g., Aldwayan et al., 2010; Canales 2012; Johnson, 2015; Johnson et al., 2016; Lee, 2004), leaving open the question of to what extent gaps are predicted in the subject position. Answering this question is critical, since, as Pickering and Barry (1991) point out, object filled-gap effects may reflect either gap prediction or *direct association* of the filler with the licensing verb. Under the Direct Association Hypothesis, object filled-gap effects are a consequence of a thematic argument relationship in which the *wh*-filler is directly associated as an argument of the licensing verb or preposition (e.g., Pickering & Barry, 1991). In other words, instead of predicting syntactic gaps during processing, this account proposes that the parser may simply be searching for an open argument position. Object filled-gap effects emerge under this hypothesis due to the establishment of a thematic argument relationship in which the parser directly associates the filler as an argument of the licensing verb. The Direct Association Hypothesis is also compatible with Stowe's (1986) finding that no subject filled-gap effects surfaced; at the subject position within the sentence, the parser has not yet encountered a subcategorizer, and thus subject filled-gap effects are not expected to surface under a direct association view. Investigating whether subject filled-gap effects emerge is crucial for teasing apart the direct association and syntactic gap prediction hypotheses, as these pre-verbal effects would provide direct evidence that gap positions are anticipated using syntactic knowledge.[1]

Another possibility regarding why subject filled-gap effects may not surface is due to the parser having limited time to generate (or commit to) a prediction for a subject gap (Clifton & De Vincenzi, 1990; Clifton & Frazier, 1989; De Vincenzi, 1991; Gibson, Hickock, & Schütze, 1994). Because Stowe (1986) employed sentences in which the subject position (e.g., *Ruth*) immediately followed the filler (e.g., *who*), it is possible that readers were unable to generate structural predictions rapidly enough for subject filled-gap effects to emerge. To test whether subject filled gaps would emerge when the distance between the filler and subject position was increased, Lee (2004) manipulated the distance between the filler and the subject position. A prepositional phrase (e.g., *on two different occasions*) was inserted between the filler and subject, as shown in (5).

(5) a. That is the laboratory *which* (on two different occasions) **Irene** used a courier to deliver the samples to ___.

b. That is the laboratory *to which* (on two different occasions) **Irene** used a courier to deliver the samples.

The intervening phrase is shown in parentheses because a short distance condition was also included, in which the filler and potential gap were immediately adjacent, as they were in Stowe (1986). The filled gap manipulation used a preposition stranding construction in (5a), in which

---

[1] Another way to address this debate is to examine head-final languages in which gap positions occur prior to the verb, such as Japanese (e.g., Aoshima, Phillips, & Weinberg, 2004; Nakano, Felser, & Clahsen, 2002). Aoshima et al. (2004), for example, reported a Japanese counterpart of a subject filled-gap effect (Experiment 2), and argue in favor of an active filler view.

the *wh*-filler *which* is extracted from the prepositional phrase at the end of the sentence where the actual gap site is located. The subject region *Irene* is a subject filled gap site in (5a). In contrast, the declarative version in (5b) involves “pied piping” in which the entire prepositional phrase is fronted such that no *wh*-extraction is involved.

The results indicated a significant overall subject filled-gap effect across conditions, with reading time slowdowns emerging for the *wh*-extraction condition (5a) as compared to the baseline condition (5b). Although there was not a significant interaction between the factors Extraction (*wh*-extraction, no extraction) and Distance (short, long), Lee conducted follow-up tests to examine whether this effect was significant across both long and short conditions. He found a marginal subject filled-gap effect for the long condition and no effect in the short condition. These results reveal that native English speakers are able to generate a prediction for a gap in subject position, and Lee suggests the parser may need increased distance to be able to generate a prediction or to commit to a subject gap prediction. As Johnson (2015) points out, however, there are several concerns with Lee’s analysis. The primary issue relates to the follow-up analysis examining the long and short conditions separately, given that there was not statistical motivation (i.e., a significant interaction) to do so. Examining only the significant findings, the results suggest that there is an overall effect of extraction, with subject filled-gap effects across both long and short conditions.

Following Lee (2004), Johnson (2015) crossed the factors distance (short vs. long) and extraction (*wh*-extraction vs. no extraction) in a self-paced reading experiment, testing a large sample of native English speakers (n=110). (6) below provides an example set of stimuli, with the prepositional phrase on parentheses demonstrating the short and long-distance conditions.

(6) a. The principal questioned *who* (during the difficult test) **Diana** will put the girl near __ for the exam.

b. The principal questioned if (during the difficult test) **Diana** will put the girl near for the exam.

At the critical filled subject region (*Diana*), a significant interaction emerged between the factors distance and extraction. Follow-up analyses revealed a subject filled-gap effect for the short distance condition only. This finding indicates that native English speakers predicted a gap immediately after encountering the filler, providing some of the strongest evidence that native speakers of English engage in pre-verbal gap prediction.[2]

In a separate analysis, Johnson (2015) examined whether individual differences in processing capabilities modulated the ability to predict gaps and process subject filled-gaps. Participants were tested on measures targeting cognitive abilities such as working memory and attentional control, and results showed that attentional control was shown to significantly affected processing the subject filled-gap. Specifically, individuals with higher attentional control resources (as measured by a number Stroop task) were more likely to show a reading time slowdown, or a subject filled-gap effect. This may suggest that during the processing of *wh*-dependencies, individuals with greater attentional control may be able to focus attention on the early subject position, yielding a larger reading time slowdown upon finding this potential gap

---

[2] In explaining why the long-distance subject filled gap effect did not emerge as in Lee (2004), Johnson (2015) points out several key changes that were implemented in her study, including changing the design of the target stimuli so that the infrequent *to which* construction was avoided, increasing the number of filler sentences, as well as testing a large sample of participants (n=110 vs. n=24). These improvements make it unlikely that the long-distance subject filled gap did not emerge due to the experimental approach. In sum, Johnson points out that there is little evidence, including Lee's (2004) analysis, which was statistically unmotivated, that the parser requires additional distance to predict a subject gap.

position filled. This finding is in line with studies which have revealed effects of attentional control in language processing, and in the following section this body of literature is briefly reviewed.

Individual Differences in Predictive Processing

Attentional control has been shown to play a role in cognitive processes such as maintaining attention in the presence of distractors, retrieving correct information during interference, and inhibiting habitual responses (e.g., Kane & Engle, 2002). These capabilities are arguably involved in prediction as well, an effortful process that entails generating and maintaining a predicted element while simultaneously processing the bottom-up input during comprehension. Indeed, brain imaging studies have suggested that the prefrontal cortex is engaged both in tasks that target attentional control as well as during expectation-driven processing in general (e.g., Bar, 2009; Kane & Engle, 2002).

Relatively few studies, however, have directly examined the relationship between attentional control and prediction during language processing. Studies in the psycho- and neurolinguistic literatures which have included measures targeting attentional control have primarily examined lexical-semantic processing (e.g., Boudewyn et al., 2012, 2013; Hutchison, 2007). For example, Hutchison (2007) probed for individual differences in attentional control by assessing native English speakers on three attentional control measures (operation span, anti-saccade, and color Stroop tasks). In two word-pair semantic priming experiments, all participants showed an overall priming effect of related prime-target pairs. Crucially though, increased attentional control was associated with sensitivity to a context cue, with greater priming effects emerging when the context cue was predictive. Hutchison proposed that sufficient attentional control is required to engage in the process of generating predictions when it is most beneficial to do so.

A second, closely related cognitive ability examined in the language processing literature is working memory. Tasks which measure working memory require participants to keep information in short-term storage (storage component) while simultaneously processing interfering information (processing component) (e.g., Conway et al., 2005). Individual differences in working memory capacity have been argued to reflect differences in computational resources recruited during processing, or in the amount of information that can be actively maintained during ongoing comprehension (e.g., Daneman & Carpenter, 1980; Just & Carpenter, 1992; Just et al., 1996). It should be noted that some proposals have argued that the abilities underlying working memory may not be separable from the construct attentional control. Engle (2002), for example, notes that a variety of studies have reported overlap between performance on working memory and executive attention tasks. The link between these cognitive abilities may be related to the nature of working memory tasks, specifically in the processing component (Kane & Engle, 2002). That is, successful performance on both attentional control and working memory tasks require that participants be able to inhibit irrelevant information and divert attention to the task at hand. Following Johnson (2015) and Boudewyn et al. (2012), this study includes separate measures of working memory and attentional control in order to examine the relationship between the two capabilities as well as their independent contribution to the processing of *wh*-dependencies.

Several researchers have noted the importance of working memory in the processing of filler-gap dependencies, although relatively few studies have included measures of working memory to directly test this proposal (e.g., Gibson, 1998, 2000; King & Just, 1991). Gibson

(1998), for example, proposed that the processing of long-distance dependencies places a burden on memory resources, such that there is a processing cost associated with maintaining an unresolved dependency across the span of the sentence. Evidence for this proposal comes from studies which have shown that longer dependencies produce *locality effects* at the actual gap site: a reading time slowdown at the gap site as compared to in shorter distance dependencies (e.g., Bartek, Lewis, Vasishth, & Smith, 2011; Gibson, 2000; Grodner & Gibson, 2005; Vasishth & Drenhaus, 2011). One recent study by Nicenboim and colleagues (2015) directly probed individual differences in working memory in the processing of filler-gap dependencies, examining locality effects in Spanish. Target sentences contained either a short- or long-distance filler gap dependency. Results from an eye-tracking experiment showed that working memory capacity, as measured by the operation span task, modulated processing at the head (gap licensor) of the dependency. At the critical region, increased working memory capacity was associated with *decreased* locality effects (lower probability of increased regression durations). In other words, individuals with increased working memory capacity are less affected by locality effects, and Nicenboim et al. (2015) suggest that these individuals may be better able to maintain the filler in memory for a longer time.

Johnson and colleagues (2016) also found that working memory modulated the processing of *wh*-dependencies, reporting that native English speakers with higher working memory yielded increased slowdowns at a filled subject position. In other words, Johnson et al. found that higher working memory was associated with larger subject filled-gap effects, suggesting in line with Hutchison (2007) that individuals with increased processing resources may be better able to generate a prediction for an upcoming gap. However, not all studies have reported similar effects of working memory. For example, Johnson (2015), reviewed above, reported that while attentional control modulated processing of the subject filled gap, working memory capacity was not a significant predictor in the analysis. Johnson discusses several reasons why this relationship may not have emerged in her study, including differences in the scoring method for working memory tasks across studies (Conway et al., 2005) and the contribution of attentional control in accounting for variation in reading times.

To summarize, there is a growing body of literature investigating the role of individual differences in cognitive capabilities in native language processing. This dissertation examines the relationship between the cognitive abilities working memory and attentional control and prediction during *wh*-dependency resolution. Evidence for prediction in *wh*-dependency formation remains mixed in the crucial pre-verbal subject position, which can shed light on whether the parser generates a syntactic prediction prior to encountering the subcategorizing verb (e.g., Aldwayan et al. 2010, Canales, 2012; Johnson, 2015; Johnson et al., 2016; Lee, 2004; Stowe, 1986). Relatively few studies show a significant role of individual differences in the resolution of *wh*-dependencies in the behavioral literature (e.g., Johnson, 2015; Johnson et al., 2016), and no study thus far has utilized EEG to systematically examine individual differences in the processing of *wh*-dependencies. The current study takes this approach, and additionally examines whether native speakers are sensitive to grammatical constraints in the course of *wh*-dependency formation, an issue examined in the following section.

Island sensitivity

Many of the studies reviewed thus far have suggested that the search for potential gaps during *wh*-dependency resolution is active, involving prediction. As Wagers and Phillips (2009) point out, filled-gap effects also demonstrate the parser is “willing to make some mistakes,” and posit gaps before waiting to confirm the presence of a gap in the input (p. 399). It is reasonable to ask, then, precisely how active the search for gaps is, examining whether the parser ever posits gaps in grammatically unlicensed positions. Islands provide a perfect test case to examine whether grammatical information is utilized in the course of *wh*-dependency formation.

In languages like English, *wh*-movement is constrained such that *wh*-extraction is only possible from certain positions and is barred from other positions, termed islands (Ross, 1967). (7), repeated from (2) above, demonstrates that syntactic movement from an island domain is not allowed in English. (7a) contains a relative clause, one type of island domain, and (7b) demonstrates that extracting a *wh*-item from the relative clause renders the sentence ungrammatical.

(7) a. Meghan likes the store [that sells dark chocolate].
b. *Which chocolate does Meghan like the store [that sells __]?

For example, in (7b) above, it is semantically plausible that the filler *which chocolate* would be linked with the verb *sells*; it is only the syntactic constraint concerning extraction from within islands that rules out this analysis.

Using self-paced reading, several studies have shown reading time slowdowns at *licit* filled object positions, but no such slowdown an *illicit* filled object positions located within an island (e.g., Canales, 2012; Johnson et al., 2016; Stowe, 1986; see also Phillips, 2006; Traxler & Pickering, 1996). For example, in Experiment 2, Stowe (1986) presented participants with sentences containing complex noun phrase islands, as in (8) below.

(8) a. The teacher asked *what* the silly story [about **Greg's** older brother] was supposed to mean ___.

b. The teacher asked *if* the silly story [about **Greg's** older brother] was supposed to mean anything.

In (8a), *about* is a potential gap licensor for *what*, although crucially it is located within an island. If the parser avoids positing gaps in islands, there should be no evidence of a reading time slowdown at *Greg's* in (8a) as compared to *Greg's* in the declarative sentence (8b). As expected, Stowe (1986) did not find evidence of a filled-gap effect at the object position within the island, suggesting that during processing, native speakers do not attempt to posit gaps inside islands and the search for potential gap sites is grammatically constrained.

Note that is somewhat difficult to make a claim regarding the use of grammatical knowledge during dependency formation considering that island sensitivity surfaces as an *absence* of an effect (reading time slowdowns). In other words, the null effect is subject to alternative interpretations regarding why gap-filling in islands is avoided. One possibility is that gap positing within islands is avoided is due to grammatical knowledge of syntactic constraints which govern *wh*-movement (e.g., Phillips, 2006, 2013; Sprouse et al., 2012; Wagers & Phillips, 2009). However, a second account has argued that because island structures present processing bottlenecks, the parser avoids positing gaps within islands due to increased difficulty resolving *wh*-dependencies in such complex structures (Hofmeister & Sag, 2010; Hofmeister, Casasanto, & Sag, 2012a/b, 2013; Kluender, 2004; Kluender & Kutas, 1993b). Stowe's (1986) results from Experiment 2, for example, are consistent with both grammatical and processing accounts of islands. Under a grammatical view, the lack of filled-gap effect within the island is explained as the parser respecting grammatical syntactic island constraints. Proponents of the processing view may interpret this finding differently, as it could also be the case that no effects surface within the island because the parser cannot overcome the processing burden posed by the complex island structure to attempt to posit a gap within the structure.

One approach that has been used to tease apart predictions of the grammatical account and processing account of islands is to examine whether gaps are posited inside an island structure when that structure could ultimately be part of a parasitic gap construction, 'rescuing' the gap inside the island and making it licit (e.g., Phillips, 2006; Wagers & Phillips, 2009). For example, Phillips (2006) examined sentences in which a *wh*-item was extracted from a subject-island, which are ungrammatical (9a) unless they are rescued from a subsequent gap, as in (9b).

(9) a. *What did [the attempt to repair __] ultimately damage the car?
b. What did [the attempt to repair __] ultimately damage __?

Note that a parasitic gap construction requires that the verb within the island be non-finite (e.g., *to repair*) for the sentence to be considered grammatical. Under a processing-based account of islands, Phillips argues that the parser should not attempt to posit a gap inside a subject island, even if the verb inside the island is non-finite, because processing the island structure while holding the *wh*-phrase in working memory is too burdensome.

In a self-paced reading experiment, Phillips (2006) found that native English speakers posited a gap within a subject island when the verb was non-finite, where the gap inside the island is potentially rescuable by a subsequent gap. There was no evidence that readers posited a gap in sentences in which the verb in the island was finite. These results are taken to be in line with the grammatical account of islands, showing that the parser can posit a gap inside an island when it is (potentially) grammatically licensed to do so. Note that this interpretation has been challenged by processing-based accounts. Hofmeister and colleagues (2013) clarify that under the processing account, gap-positing within islands depends on how difficult the island context is, and that gap-positing inside islands “not impossible or prohibited” (p. 15). This means that factors which decrease processing difficulty, such as verb finiteness, may lead to an increased ability to posit a gap within an island.

A second approach to addressing the islands debate has been to examine the role of individual differences in processing capabilities. Sprouse et al. (2012) argue that processing accounts predict that individuals with higher processing abilities (e.g., greater working memory) should be more likely to posit gaps within islands as they have greater resources available to attempt to resolve the *wh*-dependency inside complex island structures. Grammatical accounts, on the other hand, do not predict such a relationship between individual differences in processing resources and island sensitivity, given that all native speakers of a language should be similarly sensitive to a grammatical constraint regardless of their processing abilities. To address the debate between the grammatical and processing account of islands, Sprouse et al. (2012) examined the relationship between working memory and island sensitivity in two large-scale acceptability judgment experiments. Sprouse et al. argue that under a processing account, individuals with higher processing resources (e.g., working memory) would be better able to overcome the processing burden of an island structure in order to posit a gap inside the island, and therefore be more likely to provide higher acceptability judgements. Individuals with lower processing resources, on the other hand, would have greater difficulty positing a gap within the island due to the increased processing burden, and would therefore reject these sentences to a greater degree.

Sprouse et al. (2012) tested 142 native English speakers on sentences containing one of four island types: whether islands, complex noun phrase islands, subject islands, and adjunct islands. The design crossed the factors dependency length (short vs. long) and island structure (non-island vs. island). For example, (10) below shows a set of stimuli from the Adjunct Island condition. This design allowed Sprouse and colleagues to examine acceptability ratings of sentences containing a long-distance dependency (10b), an island structure (10c), each of which is expected to be a source of processing difficulty, and in the crucial condition, both a long-distance *wh*-dependency and island structure (10d), which is ungrammatical.

(10) NON-ISLAND / MATRIX
a. Who __ suspects that the boss left her keys in the car?

NON-ISLAND / EMBEDDED
b. What do you suspect that the boss left __ in the car?

ISLAND / MATRIX
c. Who __ worries [if the boss leaves her keys in the car]?

ISLAND / EMBEDDED
d. *What do you worry [if the boss leaves __ in the car]?

Both the processing and grammatical accounts predict that the ungrammatical island violation sentence (10d) will receive the lowest acceptability ratings; however, the two theories differ regarding the source of the ratings. According to the grammatical account, low acceptability ratings for this condition result from the syntactic violation of an island constraint. Under the processing account, low acceptability judgments are a consequence of a processing burden posed by the combined effect of processing a long-distance dependency and processing an island structure (i.e., a superadditive effect of the two factors). As Sprouse et al. outline, such an account would then predict that individuals with increased processing resources would be able to overcome the processing burden and posit a gap inside an island. In other words, for those individuals with greater processing resources, the superadditive effect would be reduced.

Sprouse et al.’s results showed that no relationship emerged between scores on two working memory tasks (serial recall, n-back) and ratings in the judgment task, which the authors take to suggest that island sensitivity is grammatically guided rather than related to limited processing resources. However, Hofmeister et al. (2012a,b) raised several criticisms of Sprouse et al.’s interpretation, challenging their statistical analysis, which relied on R^2 values rather than *p*-values for hypothesis testing, and the tasks used. Because the stimuli relied on decontextualized questions with bare *wh*-fillers, Hofmeister et al. argue the target sentences may have been too hard to process, which would not allow variability in acceptability ratings to emerge.

One recent study by Johnson, Fiorentino, & Gabriele (2016) examined the relationship between working memory and island sensitivity in the course of online processing using self-paced reading. Using a filled-gap paradigm, Johnson et al. compared filled-gap effects in licit positions and within island domains. Reading time results from 54 native English speakers showed slowdowns at a licit object filled gap, suggesting that gaps were posited in grammatically licensed positions; no filled-gap effects emerged within islands. Participants were additionally assessed on two measures of working memory, and crucially, no relationship emerged between working memory and filled-gap effects within islands. This finding stands in contrast to Sprouse et al.’s (2012) predictions for processing-based accounts, in which increased processing resources should be associated with a greater ability to establish *wh*-dependencies in islands. Johnson et al.’s findings are better accounted for by the grammatical account.

The psycholinguistic literature summarized above suggests that the resolution of *wh*-dependencies is a process that involves anticipating potential gap positions, which may turn out to be filled with lexical material, necessitating a continued search for gaps, with the dependency ultimately completed once the actual gap is encountered. Studies investigating subject filled-gap effects emerge revealed mixed evidence, leaving open the question of whether native English speakers anticipate gap positions pre-verbally, using syntactic knowledge. Evidence also suggested that the gap search may be grammatically guided, such that potential gap positions are only predicted in positions allowed by the grammar (i.e., outside islands). In the current study, the processing of *wh*-dependencies is tracked throughout the sentence using EEG, examining both gap prediction and completion of the dependency at the actual gap site, processes which we argue are indexed by distinct brain responses. We expect that the process of predicting gaps in positions which are filled will be associated with N400, which we expect to occur only in grammatically licensed positions if the gap search is grammatically guided, while completion of the dependency at the actual gap site will be associated with P600. In the following section, ERP studies which have examined the processing of *wh*-dependencies are briefly reviewed.

Neurolinguistic approaches to the processing of wh-dependencies

In EEG experiments, participants' brain activity is recorded by electrodes at the scalp. Averaging multiple EEG segments time-locked to the presentation of particular events (e.g., linguistic stimuli) results in event-related potential (ERP) components, brain potentials that vary in terms of their timing, scalp distribution, and polarity (positive or negative voltage shift) (e.g., Kaan, 2007; Kutas, Van Petten, & Kluender, 2006). ERP components have been shown to index different neurocognitive processes, and crucially, distinct types of language processing. EEG data can shed light on the qualitative mechanisms involved in processing linguistic stimuli, which is one important advantage over behavioral measures such as reading times. Self-paced reading, for example, can provide insights into the time-course of sentence processing, but the crucial comparison of reading times can only reveal that a difference in processing has occurred, whereas ERP components may reveal what *type* of processing has occurred. In addition, EEG data is recorded with millisecond-level precision, making this an extremely precise measurement of the *timing* of processing. In summary, EEG provides a multi-dimensional, high-temporal resolution measurement of the dynamics of language processing. In what follows, two key EEG components which have been found to be influenced by gap prediction and dependency resolution, the N400 and the P600 respectively, are reviewed.

The N400 is a negative-going waveform that peaks at about 300-500ms post-stimulus onset. Early ERP studies categorized the N400 as a component sensitive to semantic anomalies (e.g., larger N400s were found at a sentence-final anomalous word, as in, *He spread the warm bread with <u>socks</u>* as compared to a congruent word such as *<u>butter</u>*; Kutas & Hillyard, 1980). However, recent studies have investigated the extent to which the N400 reflects, at least in part, lexical prediction (e.g., Federmeier, 2007; Lau et al., 2008, 2013; Michel, 2014; Van Berkum et al., 2005). This line of research has shown that the amplitude of the N400 is modulated by lexical pre-activation, with more expected words showing reduced N400 amplitude compared to unexpected continuations.

In one such study, Federmeier and Kutas (1999) presented native English speakers with brief story narratives that led to the expectation of a particular noun. For example, the narrative in (11) below contains a semantically constraining context (e.g., *They wanted to make the hotel look more like a tropical resort*), supporting an expectation for a particular noun (*palms*). Responses to the expected noun were compared to unexpected but within the same semantic category (*pines*) and a noun that was unexpected and from outside of the expected category (*tulips*).

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

The expected noun *palms* yielded the smallest N400. Crucially, significant differences between the two unexpected nouns emerged, with the within-category noun (e.g., *pines*) eliciting reduced N400s in comparison to the out-of-category noun (e.g., *tulips*). Federmeier and Kutas (1999) argue that these results support an account in which the pre-activation of the expected lexical item *palms* causes nouns with similar semantic features (e.g., *pines*) to be activated to some extent (see also Federmeier, McLennan, De Ochoa, & Kutas, 2002). Under an integration account, both unexpected nouns are proposed to be equally difficult to integrate into the constraining context, because neither is a canonical choice for conveying a tropical atmosphere. These findings suggest that at least in sentences that are highly-semantically constraining, readers anticipate upcoming lexical material during processing.

As noted previously, the literature investigating the role of prediction in the N400 has primarily focused on lexical-semantic prediction. However, a recent study by Michel (2014) suggests that the N400 may also reflect prediction of gap sites during *wh*-dependency resolution, based on evidence that encountering an actual gap in an unexpected position in the sentence elicits a larger N400 than encountering an actual gap in an expected position. Specifically, Michel compared brain responses at post-verbal gap sites located either within an island, rendering the sentence ungrammatical, or in a licit position. For example, (12a) below contains an island violation because a gap site is located inside a *whether*-island. Brain responses at the gap site following the verb *befriended* (i.e., at *openly*) were compared to the same region in a sentence without an island (12b):

(12) a. ***Who** had the sailor inquired [whether the captain befriended __ openly before the final mutiny hearing]?

b. **Who** had the sailor assumed that the captain befriended __ openly before the final mutiny hearing?

At the gap site, an N400 effect emerged, with island condition (12a) yielding a greater N400 than non-island condition (12b). This effect was attributed to prediction, in that when the parser is in the process of resolving a *wh*-dependency, encountering an island boundary (i.e., at *whether*) forces the parser to revise its prediction that a gap is forthcoming. Thus, when evidence of a gap was encountered within an island, an increased N400 was yielded because this gap site was not expected. Participants in the study were also assessed on a measure of working memory, and a relationship between working memory and the N400 effect emerged, such that individuals with higher reading span scores were found to yield greater negativities within the island conditions. Because high-span readers are argued to be better able to revise their gap prediction in order to avoid positing gaps in illicit positions, high-span readers yielded an increased N400 when evidence for a gap was encountered in the island. From these findings, Michel proposes the Gap Predictability Account of island sensitivity, arguing that the N400 reflects predictability of upcoming gap sites, with those in islands being less predictable.

Another ERP component associated with the processing of filler-gap dependencies is the P600, a late positive-going waveform. This component is typically peaks at around 500-900 ms post-stimulus onset, variation in P600 onset latency has been observed, with some studies reporting an earlier onset beginning at around 300 ms (e.g., Friederici et al., 2001; Gouvea et al., 2010; Kaan et al., 2000; Phillips et al., 2005). Early studies elicited the P600 for syntactic anomalies such as grammatical violations and garden path sentences (e.g., Friederici et al., 1996; Frisch, Schlesewsky, Saddy, & Ackerman, 2002; Osterhout & Holcomb, 1992; for a recent review see Molinaro et al., 2011). However, studies have also shown that P600 can be elicited for well-formed sentences, and has been suggested to index syntactic integration during the processing of long-distance dependencies (e.g., Felser et al., 2003; Gouvea et al., 2010; Kaan et al., 2000; Phillips et al., 2005). For example, Kaan et al. (2000) examined brain responses at a gap-licensing verb in grammatical sentences that contained *wh*-dependencies. In (13a), the verb *imitate* makes available a direct object gap position allowing completion of the dependency, as compared to the verb position in (13b), a sentence without *wh*-extraction.

(13) a. Emily wondered *who* the performers in the concert **imitate** for the audience's amusement.

b. Emily wondered *whether* the performers in the concert **imitate** a pop star for the audience's amusement.

A P600 emerged between 500-900 ms at the verb in (13a) as compared to (13b), which Kaan et al. suggest reflects syntactic integration processes resulting from linking *who* with the verb. Kaan and colleagues further speculate that the amplitude of the P600 can be used as an index of syntactic integration difficulty, with more 'difficult' integration processes linked to larger amplitude P600.

To directly test Kaan et al.'s (2000) claim regarding the P600 amplitude and syntactic integration complexity, Phillips et al. (2005) manipulated the length of *wh*-dependency, including short and long-distance conditions, as shown in (14).

(14) a. The detective hoped that the lieutenant knew *which accomplice* the shrewd witness would **recognize** in the lineup.

b. The lieutenant knew *which accomplice* the detective hoped that the shrewd witness would **recognize** in the lineup.

At the verb (*recognize*), a P600 emerged for both the short- and long-distance conditions, providing further evidence that P600s can be elicited during the processing of well-formed dependencies. Importantly, there were no significant differences in the amplitude of the P600s for the two conditions, suggesting that P600 amplitude may not serve as an index of syntactic integration difficulty, as proposed by Kaan et al. (2000). Note that while there was not a difference in P600 amplitude between short- and long-distance conditions, there were significant differences in the latency of the two components, with the short-distance condition P600 emerging earlier, in the 300-500 ms interval, as compared to the 500-700 ms interval for the long-distance condition. Thus, the length of the *wh*-dependency does affect one aspect of the P600, its latency. Phillips et al. (2005) suggest that the amplitude instead reflects "the syntactic and semantic operations involved in confirming the compatibility of the filler and the verb for thematic role assignment, and compositionally interpreting the verb and its arguments" (p. 425).

In a separate line of research, ERP studies which have set out to directly examine filled-gap effects have shown mixed results (Hestvik, Maxfield, Schwartz, & Shafer, 2007; Hestvik, Bradley, & Bradley, 2012; Jessen, Festman, Boxell, & Felser, 2017; Schremm, 2012, 2013). In two studies, Hestvik and colleagues (2007, 2012) examined responses at an object filled gap in sentences which contained a fronted noun phrase (15a) as compared to sentences without extraction (15b).

(15) a. *The zebra that the hippo kissed **the camel** on the nose ran far away.
b. The zebra said that the hippo kissed **the camel** on the nose and then ran far away.

An early anterior negativity emerged at the object filled gap in (15a) as compared to (15b), which Hestvik et al. (2007) interpreted to be an ELAN (early left anterior negativity) effect yielded by a phrase structure violation. At the same object filled-gap region, Hestvik et al. (2012) reported an anterior negativity followed by a late posterior positivity (P600), suggested to be related to syntactic integration difficulty. In a modified replication of Hestvik and colleagues' studies, Schremm (2012, 2013) observed P600s at the filled gap. However, a limitation of both Hestvik et al. (2007, 2012) and Schremm (2012, 2013) is that the stimuli included sentences which were ultimately ungrammatical. This makes it difficult to make inferences based on the ERP components elicited at these regions, since, unlike grammatical sentences with filled gaps where an actual gap site is ultimately encountered later in the sentence, encountering a filled gap in these sentences serves as an indicator of whether the sentence is well-formed overall. This raises questions regarding to what extent the processing of filler-gap dependencies in Hestvik et al. (2007, 2012) and Schremm (2012, 2013) is representative of how these dependencies are processed during natural comprehension. To address this concern, the current study utilizes grammatical, well-formed sentences in the filled-gap paradigm[3].

Current Study

The current study tracks the processing of *wh*-dependencies across the sentence, examining brain responses at (i) the subject position, (ii) the object position, both inside and

---

[3] Some ERP studies have examined other aspects of *wh*-dependency processing, including initiating the search for a gap at the filler and holding the filler in memory (e.g., Felser et al., 2003; Fiebach, Schlesewsky, & Friederici, 2002; Garnsey et al., 1989; King & Kutas, 1995; Kluender & Kutas 1993a/b). These studies reported anterior negativities (often described as LAN) emerging at a variety of positions within the sentence, including at the onset of the dependency (i.e., the filler) and at the gap site. The negativity has also been shown to be sustained, persisting from the onset of the dependency until resolution at the verb site, and thus these studies have suggested that the anterior negativity reflects storage of the filler in working memory during processing of the rest of the sentence (cf., Phillips et al., 2005).

outside islands, and (iii) the actual gap site. By using ERPs to examine both filled-gap effects at the subject and object positions, and dependency resolution at the actual gap site, we examine whether the processes of gap prediction and dependency resolution are indexed by distinct ERP components, N400 for gap predictability (following Michel, 2014) and P600 for syntactic integration (e.g., Gouvea et al., 2010; Kaan et al., 2000; Phillips et al., 2005). A second goal of this dissertation is to examine to what extent native speakers vary in their processing of *wh*-dependencies, specifically in gap prediction and dependency resolution, and whether this variation can be accounted for by individual differences in cognitive abilities.

To address the first question regarding whether gap predictability is indexed by N400 and whether the process is sensitive to grammatical constraints, we examined filled-gap effects at a pre-verbal subject position and a post-verbal object position. If readers predict potential gap sites at each grammatically licensed position within the sentence, then following Michel's (2014) Gap Predictability Account, an N400 should emerge at both the filled subject and filled object positions. Recall that in Michel's study, an N400 emerged at a region in which a gap site was *not* predicted because it was located within an island; in the present study, if the parser predicts a gap but then finds it is filled with lexical material (i.e., a filled gap position) an N400 should also emerge. We additionally manipulated whether the potential gap is grammatically licensed by including a condition in which the potential gap site is located inside an island. Following previous reading time studies which have observed that readers do not attempt to posit a gap inside an island (e.g., Johnson et al., 2016; Stowe, 1986), no N400 effect is expected in this condition, which would provide evidence that the parser only engages in prediction when licensed by the grammar.

The current study also examines the source of island sensitivity, investigating whether it is the case that native speakers avoid positing gaps in islands due to grammatical constraints or a processing burden. To address this question, we follow Sprouse et al. (2012) and include cognitive measures targeting working memory and attentional control to investigate whether there is a relationship between processing capabilities and gap-positing within islands. Under a grammatical account of islands, no relationship between cognitive abilities and ERP effects within the island should emerge. As argued by Sprouse et al. (2012), under the processing account, individuals with greater cognitive resources should show an increased ability to posit a gap within the island, thus yielding increasingly negative (N400-like) responses inside the island (e.g., Kluender & Kutas, 1993b; Kluender, 2004; Hofmeister & Sag, 2010). This study also includes a measure of offline sensitivity to islands, and for the first time examines the relationship between an individual's offline and online processing of sentences containing islands.

A relationship between the cognitive measures and the size of the N400 effect in filled-gap positions outside of islands (i.e., subject position, licit object position) does not tease apart the two competing accounts. However, given that several studies have reported a relationship between cognitive abilities and prediction in grammatical contexts, suggesting that those with greater cognitive resources may be better able to engage in predictive processing (e.g., Hutchison, 2007; Johnson, 2015; Johnson et al., 2016; Nicenboim et al., 2015), we will also examine whether individuals with greater processing resources show larger filled-gap effects in licit contexts.

Finally, this study also tests whether dependency resolution is indexed by P600 by examining processing at the actual gap site. Following previous research, it is expected that dependency resolution at the actual gap will be indexed by a P600 (e.g., Felser et al., 2003; Gouvea et al., 2009; Kaan et al., 2000; Phillips et al., 2005).

In summary, the current study will track the processing of *wh*-dependencies across the sentence at three positions, including an examination of dependency formation inside and outside of islands, using fully grammatical sentences for the first time in the literature to our knowledge. In the following sections, the experimental design is described in detail.

Methods

Participants

The participants were 44 native English speakers (12 males, mean age 20.25, range 18-30) recruited from the University of Kansas. All were right-handed and had normal or corrected-to-normal vision. All participants provided informed consent to participate, and completed a background questionnaire and handedness inventory (Oldfield, 1971) prior to the experiment. Participants received $10 per hour for participating in one testing session which lasted approximately three hours. The three cognitive tasks were administered first, with the order of the tasks counterbalanced across participants. Following this, the participants took the EEG experiment, which was followed by the offline acceptability judgment task. All tasks were administered using Paradigm presentation software (Tagliaferri, 2005).

EEG Experiment

Stimuli. 160 sets of sentences were constructed. Each set consisted of four target sentence types, crossing the factors Extraction (no extraction, *wh*-extraction) and Island (non-island, island). The target sentences were divided into four Latin-Square lists (n=40 targets per condition in each list), such that every participant read a sentence from every set, but not more than one sentence from a given set. (16) shows the target sentences.

(16) NO EXTRACTION, NON-ISLAND
a. Jamie wondered if **the editor** interviewed **Dave Campbell with** the reporter from the department.

WH-EXTRACTION, NON-ISLAND
b. Jamie wondered who **the editor** interviewed **Dave Campbell with __** from the department.

NO EXTRACTION, ISLAND
c. Jamie wondered if **the editor** [that interviewed **Dave Campbell**] **kissed** the reporter after the meeting.

WH-EXTRACTION, ISLAND
d. Jamie wondered who **the editor** [that interviewed **Dave Campbell**] **kissed __** after the meeting.

All target sentences began with a first name (e.g., *Jamie*) in subject position followed by one of four verbs that take a sentential complement (*wondered*, *questioned*, *revealed*, *asked*). In sentences that did not involve extraction, this verb was followed by the complementizer *if*; in sentences containing *wh*-extraction, the first verb was followed by the *wh*-item *who*. There is a determiner phrase (e.g., *the editor*) in the embedded subject position across conditions. In non-island sentences, a main verb (e.g., *interviewed*) immediately followed the determiner phrase; in island sentences, the main verb was preceded by *that*, beginning a relative clause island. There was a total of 40 main verbs; all 40 verbs were transitive, to ensure that there was a potential object gap position. In addition, the main verbs were also required to be able to take a prepositional phrase with an animate object (e.g., with the reporter) so that they could provide an actual gap site (the prepositional object position) in the non-island condition. The 40 verbs were repeated four times across the 160 sets of sentences.

Each list also included 80 filler sentences that were matched in length to the target sentences: 20 sentences included *wh*-extraction from a subject position (17) and 20 sentences had *wh*-extraction from a direct object position (18), the two positions which contain filled gaps in the target sentences. The remaining 40 fillers were declarative sentences (19).

(17) SUBJECT EXTRACTION
The musician guessed who __ recorded Victoria Johnson at the rock concert.

(18) OBJECT EXTRACTION
The mailman asked who Shane Russell persuaded __ at the business meeting.

(19) DECLARATIVE
The biologist was teaching while the students took notes in the lab.

Participants were instructed to read all sentences carefully and were informed that they would be prompted to answer comprehension questions following some of the sentences. Yes-no comprehension questions followed one third (n=80) of the stimuli to encourage participants to attempt to derive meaning from the sentences as they would during natural language comprehension. The questions did not target the resolution of the *wh*-dependency (e.g., *Was it the editor who interviewed Dave Campbell?*).

Each experimental session began with six practice trials. During each trial, a sentence was presented using rapid serial visual presentation (RSVP), with 450 ms for each word and a 300 ms pause between words. The inter-trial interval was 500-1000 ms, pseudorandomly varied at 50 ms increments to ensure that the ERP does not become time-locked with the stimulus presentation rate. During the main experiment, participants were given a short break after every 20 trials, for a total of 11 breaks. The total length of the EEG experiment was approximately 60 minutes.

Offline Acceptability Judgment Task

The acceptability judgment task used stimuli from Aldosari (2015), who built directly on Sprouse et al. (2012). Aldosari modified the stimuli in several ways in order to address criticisms raised by Hofmeister et al. (2012a,b). First, a context declarative sentence was included prior to every trial in order to provide a natural context for the *wh*-questions. In addition, the target stimuli were revised from a bare *wh*-word to include a lexical *wh*-filler (Hofmeister & Sag, 2010; see also Goodall, 2015). Four island types were tested: *whether* islands, complex noun phrase islands, subject islands, and adjunct islands. The stimuli crossed the factors Island (island, non-island) and *Wh*-Dependency Length (matrix, embedded) in a 2×2 design, as shown in (20), an example set containing an adjunct island.

(20) NON-ISLAND / MATRIX
The helpful worker thinks that the boss left her keys in the car.
a. Which worker __ thinks that the boss left her keys in the car?

NON-ISLAND / EMBEDDED
The worker thinks that the boss left her office keys in the car.
b. Which keys does the worker think that the boss left __ in the car?

ISLAND / MATRIX
The helpful worker worries if the boss leaves her keys in the car.
c. Which worker __ worries [if the boss leaves her keys in the car]?

ISLAND / EMBEDDED
The worker worries if the boss leaves her office keys in the car.
d. *Which keys does the worker worry [if the boss leaves __ in the car]?

In total, 16 sets of sentences were constructed for each of the four conditions; the sets were distributed across four lists in a Latin-Square design. Each participant read 64 target sentences. Participants were first presented with the context sentence, and on the subsequent screen, the target sentence was presented. Below the target sentence, a seven-point rating scale was shown, ranging from ‘totally unnatural’ to ‘perfectly natural.’ Participants were instructed to select a rating based on the naturalness of the target sentence.

The scores entered into the individual difference statistical models were ‘DD scores,’ which were calculated following Sprouse et al. (2012) and Aldosari (2015). DD (differences-in-differences) scores measure sensitivity to island effects. First, each participant’s acceptability judgment ratings were converted into z-scores to account for individual variability in usage of the seven-point rating scale. The first difference score (D1) measures the effect of an island in a long dependency (ratings to 20b subtracted from ratings to 20d, above) and the second difference score (D2) measures the effect of an island structure in a short dependency (ratings to 20a subtracted from ratings to 20c, above). Finally, the D2 score is subtracted from the D1 score, yielding an overall DD score which measures sensitivity to the effect of an island structure in a long-distance dependency as compared to an island structure within a short dependency. Higher DD scores indicate a stronger sensitivity to island effects, with lower scores indicating weaker sensitivity to island effects (i.e., greater acceptance of ungrammatical island sentences).

Cognitive Measures

Working Memory. We utilize two complex span working memory tasks that contain a memory component and a processing component, targeting working memory rather than short-term memory (see Hofmeister et al., 2012b). A non-verbal measure, the Counting Span task (Case, Kurland, & Goldberg, 1982), is included as well as a measure of verbal working memory, the Reading Span task (Daneman & Carpenter, 1980). The reading span task has been used in previous studies to investigate the relationship between working memory and *wh*-dependencies (e.g., Hofmeister, Casasanto, & Sag, 2014; Johnson et al., 2016; Nakano et al., 2002).

In the Counting Span task, following the protocol of Conway et al. (2005), participants are presented with visual arrays with target objects (dark blue circles) and distractor objects (light green circles); participants are asked to count the target objects out loud and remember the total. After between 2 to 6 arrays, participants are prompted to input the total number from each array in the previous set. In the Reading Span task, following Conway et al. (2005), participants read sentences aloud, provide a semantic judgment regarding the sentence, and then say a letter out loud that is presented on the screen that they are expected to remember. After a set of between 2 to 5 sentences, participants are required to input the letters they were presented with in the last set of sentences. Both tasks are scored from 1-100 based on the percentage of letters or numbers recalled, including partial scoring.

Attentional Control. Following Johnson (2015), we measure attentional control via a number Stroop task (Bush et al., 2006). Participants are asked to count the number of words presented on the screen, and to press a button indicating the total number, which ranges from 1 to 4. In congruent trials, participants are presented a screen with 2 to 4 monosyllabic, common animal words (e.g., *cat cat*). Participants are instructed to enter the number of words they saw on the screen (e.g., 2) by pressing the corresponding number on a button box. In the incongruent trials, participants are presented with 2 to 4 monosyllabic number words, and crucially the number words do not match the quantity of words on the screen (e.g., *four four four*). Participants are asked to enter the appropriate quantity of words on the screen, inhibiting the semantic meaning of the words (e.g., enter 3). The speed and accuracy of the participants' responses are measured for each trial.

Data Analysis

EEG Analysis. The electroencephalogram was continuously recorded using an elastic electrode cap (Electro-Cap International, Inc.) containing 32 Ag/AgCl scalp electrodes arranged in a modified 10-20 layout (midline: FPZ, FZ, FCZ, CZ, CPZ, PZ, OZ; lateral: FP1/2, F3/4, F7/8, FT7/8, FC3/4, T3/4, C3/4, TP7/8, CP3/4, T5/6, P3/4, O1/2; AFZ was used as the ground). Three bipolar montage electrode pairs were placed on the outer canthi and above and below each eye to monitor blinks and horizontal eye movements, and an electrode was placed on each mastoid. Impedances for all scalp electrodes were kept below 5 kΩ. Data were sampled at a rate of 1 kHz and referenced online to the left mastoid. Recordings were filtered with a bandpass of 0.1-200 Hz, and amplified with a Neuroscan Synamps 2 amplifier (Compumedics Neuroscan, Inc.).

EEG was re-referenced offline to the average of both mastoids. Trials containing blinks, horizontal eye movements or muscle artifacts were manually rejected, resulting in the exclusion of 16% of target trials. The data were then epoched using a 300 ms prestimulus to 1200 ms post-stimulus interval, time-locked to the presentation of the critical word, baseline corrected using a -300 to 0 ms pre-stimulus interval, filtered with a 30 Hz low-pass filter, and averaged. Electrodes which did not successfully record electricity at the scalp were replaced with interpolated values from other electrodes in their vicinity prior to averaging (Lopez-Calderon & Luck, 2014).

Electrodes were coded for hemisphere (left, midline, right) and anteriority (frontal, central, posterior). The coding scheme created the following groups: Left Anterior (FP1, F7, F3, FT7, and FC3), Right Anterior (FP2, F8, F4, FT8, and FC4), Left Posterior (TP7, CP3, T5, P3, and O1), Right Posterior (TP6, CP4, T6, P4, and O2), Midline Anterior (FPZ, FZ, and FCZ), and Midline Posterior (CPZ, PZ, and OZ). The EEG data was analyzed using linear mixed-effects models with the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) in the R programming environment, with *p*-values calculated using the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2016). For each critical position in the sentence (i.e., subject position, object position, gap site), ERPs were time-locked to the onset of the first word of the phrase. For the subject and object filled gap regions which consist of noun phrases with two words (e.g., *the editor* and *Dave Campbell*), encountering the first word of the phrase provides evidence that the potential gap site is already filled with a noun phrase. Two time-windows were selected for analysis, which together capture processing up to 900 ms after the onset of the target word. The time windows correspond to the two ERP components of interest based on previous research, for the N400, 300-500 ms and for the P600, 500-900 ms.[4]

The dependent variable in each model was the mean amplitude at a given electrode, for a given condition. Model fitting began by including the following fixed factors and all possible interactions: Extraction (no Extraction, *wh*-extraction), Island (non-island, island), Hemisphere (left, midline, right), and Anteriority (anterior, central, posterior). The model included Subjects as random intercepts. The model was then progressively backwards-fit, such that interactions and fixed effects which did not explain a significant portion of the variance were removed. The baseline condition was the no extraction, island condition in midline and central region. The comparisons are all in relation to this condition. Because the variables Hemisphere and Anteriority have three levels each, analyses for these factors tested differences between the baseline level and the two other levels (e.g., for Hemisphere, midline vs. left and midline vs. right). For significant interactions reported below, we also report which level interacts with the

---

[4] Based on previous studies examining *wh*-dependency resolution, it is also possible that the onset of the P600 will emerge as early as the 300-500 ms time window (e.g., Gouvea et al., 2010; Kaan et al., 2000; Phillips et al., 2005).

baseline (e.g., ‘anterior level’ indicates a significant difference between electrodes in the central and anterior regions). For each analysis, the best model was identified using the R package *LMERConvenienceFunctions* (Tremblay & Ransijn, 2015), which utilizes a series of iterative log-likelihood ratio tests to arrive at the simplest model that best fit the data. The final models that best fit the data for each time-window are described in the sections below for each time-window. Only significant main effects and interactions involving the critical factors Extraction and Island are discussed below. We consider $p < 0.05$ to be significant.

Individual Difference Analyses. In the second step of the analysis, scores from the individual difference measures were included in separate models. Each model was conducted following the guidelines outlined above using *LMERConvenienceFunctions* (Tremblay & Ransijn, 2015). Descriptive statistics for results from the three cognitive measures (Count Span, Reading Span, Number Stroop) as well as DD scores from the acceptability judgment task, are shown below in Table 1. Note that on the Number Stroop task two variables were measured, interference effects for reaction times and accuracy.

Table 1: Descriptive statistics for individual difference measures.

	Mean	**Standard Deviation**	**Minimum**	**Maximum**
Count Span (%)	62.85	11.43	36.00	84.44
Reading Span (%)	65.79	14.39	26.25	95
Stroop Reaction Time Interference Effect (congruent – incongruent) (ms)	-73.69	57.17	-217.57	24.37
Stroop Accuracy Interference Effect (congruent – incongruent) (%)	3.41	5.17	-3.75	21.25
Acceptability Judgment Ratings/DD score (z-score)	1.65	0.46	0.60	2.48

Scores on the two working memory tasks were significantly correlated ($r = .40$, $p < .01$), and thus a composite variable called Working Memory was utilized in the individual difference analyses. To create the composite variable, scores from the Count Span and Reading Span were transformed into z-scores and added together.

For the two variables measured for the Number Stroop task, the subtraction of Congruent minus Incongruent conditions to calculate the interference effect yields opposite patterns for the two variables. For the Stroop Reaction Time Interference Effect, higher values indicate better performance on the task (i.e., higher attentional control). For the Stroop Accuracy Interference Effect, higher values indicate that a participant was ‘more strooped’ (i.e., lower attentional control). These two variables were marginally correlated ($r = -.28$, $p = .065$), with the negative correlation value indicating that better performance on the reaction time measure was related to better performance on the accuracy measure. Thus, a composite variable called Attentional Control was created by taking the z-score transformed scores from the two Stroop measures and adding them together. Prior to creating the composite variable, Accuracy Interference scores were multiplied by -1; thus, on the composite Attentional Control variable, higher values reflect better performance on task (i.e., higher attentional control).

Finally, we acknowledge that individual difference analyses in the current study are preliminary given the sample size (n=44), which is relatively smaller than the key studies examining individual differences discussed in the literature review (e.g., Nicenboim et al., 2015, n=71; Johnson, 2015, n=100; Johnson et al., 2016; n=54).

Behavioral Results

Descriptive statistics for accuracy on the comprehension questions in the main EEG task are shown below in Table 2. Across the four target conditions, mean accuracy was 69.0% (SD: 17.4), reflecting the fact that the target sentences were long and complex. A repeated-measures

ANOVA with conditions Island (island vs. non-island) and Extraction (*wh*-extraction vs. no extraction) revealed no significant effects of Extraction (*F*(1, 43) = 2.679, *p* = .11), Island (*F*(1, 43) = 2.320, *p* = .13), or an interaction between Extraction and Island (*F*(1, 43) = .303, *p* = .59) indicating that participants were equally accurate responding to comprehension questions across the target conditions.

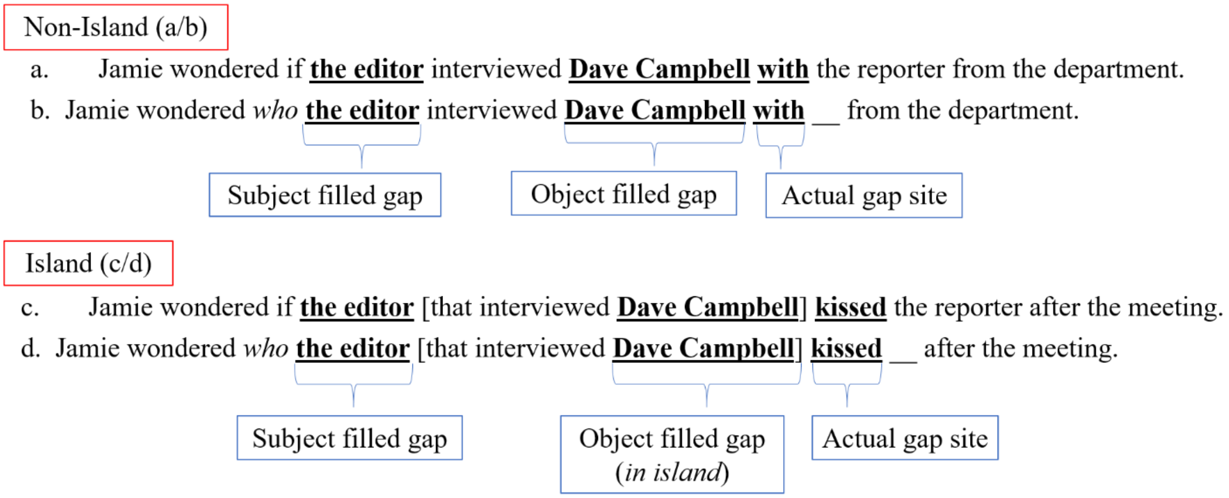
Table 2: Descriptive statistics for accuracy on comprehension questions across conditions.

Condition	**Mean**	**Standard Deviation**	**Minimum**	**Maximum**
Non-Island, No Extraction	70.00	18.17	30	100
Non-Island, *Wh*-Extraction	71.82	15.89	30	100
Island, No Extraction	64.77	19.82	0	100
Island, *Wh*-Extraction	69.32	15.16	40	100
Overall (including fillers)	76.72	9.20	53.57	95.71

EEG Results

Before reporting results from the EEG analyses, Figure 1 is presented below. The target stimuli are repeated from (16) above, and the three regions of analysis are highlighted: subject position, object position (in and outside of the relative clause island), and actual gap site. The models described below follow this order, beginning with the model for the subject filled gap position.

Figure 1: Target stimuli repeated with three regions of analysis highlighted.



Non-Island (a/b)

a. Jamie wondered if **the editor** interviewed **Dave Campbell** **with** the reporter from the department.

b. Jamie wondered *who* **the editor** interviewed **Dave Campbell** **with** __ from the department.

Island (c/d)

c. Jamie wondered if **the editor** [that interviewed **Dave Campbell**] **kissed** the reporter after the meeting.

d. Jamie wondered *who* **the editor** [that interviewed **Dave Campbell**] **kissed** __ after the meeting.

Subject Position

Processing at the subject filled gap position can reveal whether native English speakers engage in pre-verbal gap prediction, predicted to be indexed via an N400 response (e.g., Michel, 2014). Subject filled-gap effects would provide evidence that native English speakers anticipate gap positions using syntactic knowledge, a finding that is mixed in the literature (e.g., Aldwayan et al., 2010; Canales, 2012; Stowe, 1986; Johnson, 2015; Johnson et al., 2016; Lee, 2004). Although analyses for the 500-900 ms time window do not shed light on whether participants predict a gap, given that the predicted N400 is expected to emerge in the first 300-500 ms time window, examining processing in the later time window provides a comprehensive picture of processing the subject filled gap. Models for both time windows are reported below.

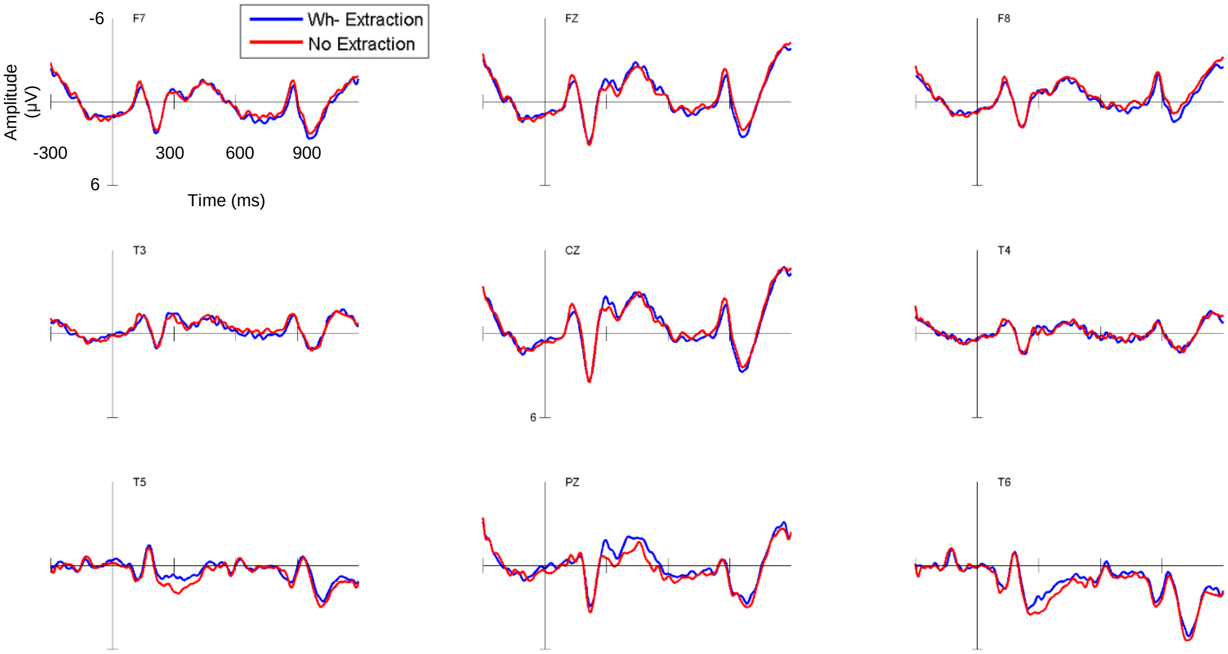
In both the island and non-island conditions, the target *wh*-extraction sentence was lexically identical from the beginning of the sentence through the embedded subject position (e.g., *Jamie wondered who the editor*). Thus, statistical analyses for this region collapsed across the factor Island.

300-500 ms time-window analysis. The results for the overall, best-fitting model are shown below in Table 3. A significant main effect of Extraction emerged, indicating that at the subject position (e.g., *the editor*), the amplitude for the *wh*-extraction condition was more negative than for the baseline no extraction condition. The interaction of Extraction × Anteriority (posterior level) was also significant, which showed that negativity in the posterior region was larger than in the central region. In sum, an N400 emerged in the 300-500 ms time window, which was most prominent at the posterior region; this is shown in Figure 2 below.

Table 3: Linear mixed-effects overall model for subject position from 300-500 ms, with no individual difference measures included.

Number of obs: 2728, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-1.708	0.227	103.8	-7.541	< .001	***
Extraction	-0.338	0.114	2673	-2.975	0.003	**
Hemisphere (left)	1.183	0.156	2673	7.594	< .001	***
Hemisphere (right)	1.537	0.156	2673	9.863	< .001	***
Anteriority (ant)	0.436	0.181	2673	2.411	0.016	*
Anteriority (post)	1.911	0.199	2673	9.600	< .001	***
Extraction × Anteriority (ant)	0.164	0.151	2673	1.085	0.278	
Extraction × Anteriority (post)	-0.436	0.171	2673	-2.554	0.011	*
Hemisphere (left) × Anteriority (ant)	-0.733	0.204	2673	-3.598	< .001	***
Hemisphere (right) × Anteriority (ant)	-1.172	0.204	2673	-5.752	< .001	***
Hemisphere (left) × Anteriority (post)	-0.720	0.226	2673	-3.182	0.0015	**
Hemisphere (right) × Anteriority (post)	-0.499	0.226	2673	-2.203	0.028	*

Figure 2: Subject filled-gap effect at representative electrodes.



500-900 ms time-window analysis. The results from the best-fitting overall model for the second time window are shown below in Table 4. The main effect of Extraction in this model was not significant, and the interaction terms involving Extraction × Anteriority were either not significant (anterior level) or marginal (posterior level). In other words, no significant differences emerged between the amplitude of *wh*-extraction and no extraction conditions at the subject position from 500-900 ms.

Table 4: Linear mixed-effects overall model for subject position from 500-900 ms, with no individual difference measures included.

Number of obs: 2728, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-0.172	0.246	91.2	-0.698	0.487	
Extraction	0.039	0.116	2673	0.335	0.738	
Hemisphere (left)	0.149	0.159	2673	0.942	0.346	
Hemisphere (right)	0.511	0.159	2673	3.221	0.001	**
Anteriority (ant)	0.333	0.184	2673	1.807	0.071	.
Anteriority (post)	0.897	0.203	2673	4.423	< .001	***
Extraction × Anteriority (ant)	0.116	0.154	2673	0.751	0.453	
Extraction × Anteriority (post)	-0.312	0.174	2673	-1.798	0.072	.
Hemisphere (left) × Anteriority (ant)	0.001	0.207	2673	0.003	0.998	

Hemisphere (right) × Anteriority (ant)	-0.514	0.207	2673	-2.479	0.013	*
Hemisphere (left) × Anteriority (post)	-0.285	0.230	2673	-1.236	0.216	
Hemisphere (right) × Anteriority (post)	-0.137	0.230	2673	-0.595	0.552	

To summarize, participants yielded N400 at the subject region (e.g., *the editor*), suggesting that the parser predicted a gap at the pre-verbal subject position. No significant effects emerged in the later time window, in line with our predictions. In the next analysis, processing at a second filled gap site, the object position, is examined.

Object Position

At the post-verbal object position, we crucially manipulated whether the potential object gap was grammatically licensed, as in the non-island condition, or whether it was ungrammatical and located inside of a relative clause island. If the parser only engages in gap prediction when licensed by the grammar, an N400 is expected to emerge at the filled object position in the licit non-island condition only; no N400 effect is expected inside of the island. These predictions concern the N400 time window, 300-500 ms, although we also report findings for the later time window to provide a comprehensive picture of processing a filled object position both inside and outside of an island.

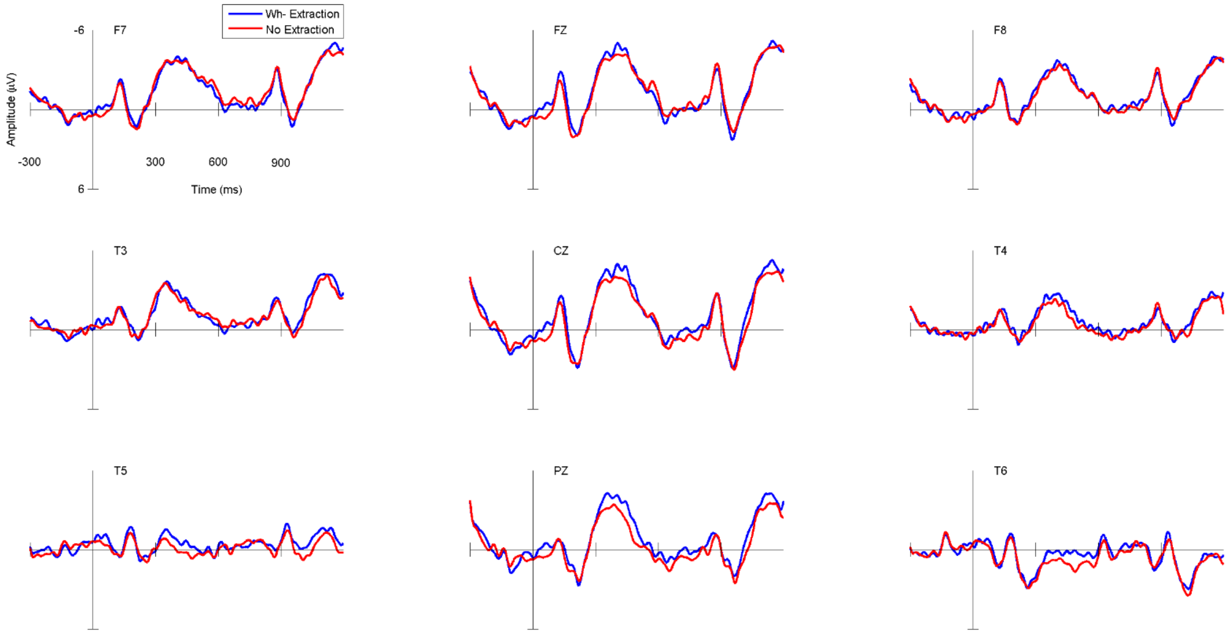
300-500 ms time-window analysis. The best-fitting overall model for the object position for the 300-500 ms (N400) time window was very large; the full *lmer* summary is provided in Appendix A. The critical interaction between Extraction × Island was significant (*t* (5395) = -3.273, *p* < .01). To interpret this interaction, follow-up analyses were carried out by running separate models for the two conditions of the factor Island (non-island, island). As in fitting the initial model, follow-up models were first fit with every possible fixed effect and interaction and then progressively backwards-fit to arrive at the simplest model that best fit the data.

Results for the best-fitting overall non-island model are shown in Table 5 below. The main effect of Extraction was significant, and indicated that the amplitude for the target *wh*-extraction condition was more negative as compared to the baseline no extraction condition. In other words, a broadly distributed N400 emerged at the object filled-gap position (e.g., *Dave Campbell*), shown in Figure 3.

Table 5: Linear mixed-effects overall model for object position, Non-Island condition, from 300-500 ms, with no individual difference measures included.

Number of obs: 2728, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-3.434	0.266	111.2	-12.922	< .001	***
Extraction	-0.472	0.083	2675	-5.675	< .001	***
Anteriority (ant)	-0.103	0.212	2675	-0.485	0.628	
Anteriority (post)	2.233	0.232	2675	9.636	< .001	***
Hemisphere (left)	1.109	0.201	2675	5.527	< .001	***
Hemisphere (right)	1.868	0.201	2675	9.309	< .001	***
Anteriority (ant) × Hemisphere (left)	-0.696	0.263	2675	-2.651	0.008	**
Anteriority (post) × Hemisphere (left)	-0.611	0.292	2675	-2.095	0.036	*
Anteriority (ant) × Hemisphere (right)	-1.128	0.263	2675	-4.297	< .001	***
Anteriority (post) × Hemisphere (right)	-0.779	0.292	2675	-2.670	0.008	**

Figure 3: Object filled-gap effect for Non-Island condition at representative electrodes.

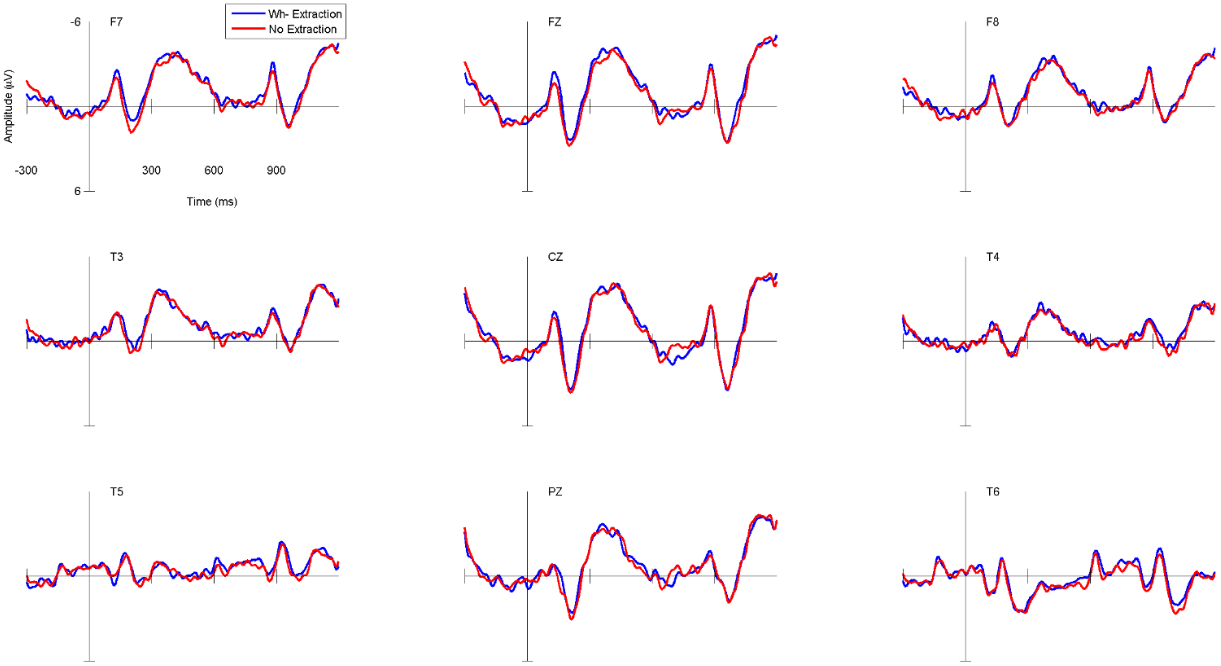


Results for the best-fitting model for the island condition for the 300-500 ms time window are shown in Table 6. During model-fitting, the term Extraction was removed, indicating that this variable did not explain a significant amount of variance in the model. Thus, no significant difference emerged between the *wh*-extraction and no extraction conditions inside the island (i.e., no N400). Figure 4 shows the comparison between the two conditions for the island condition at the object position.

Table 6: Linear mixed-effects overall model for object position, Island condition, from 300-500 ms, with no individual difference measures included.

Number of obs: 2728, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-3.367	0.227	84.2	-14.806	< .001	***
Anteriority (ant)	-0.072	0.162	2676	-0.443	0.658	
Anteriority (post)	1.966	0.178	2676	11.050	< .001	***
Hemisphere (left)	0.896	0.154	2676	5.815	< .001	***
Hemisphere (right)	1.708	0.154	2676	11.085	< .001	***
Anteriority (ant) × Hemisphere (left)	-0.716	0.202	2676	-3.555	< .001	***
Anteriority (post) × Hemisphere (left)	-0.386	0.224	2676	-1.727	0.084	.
Anteriority (ant) × Hemisphere (right)	-1.170	0.202	2676	-5.805	< .001	***
Anteriority (post) × Hemisphere (right)	-0.382	0.224	2676	-1.706	0.088	.

Figure 4: Object position inside island at representative electrodes.



500-900 ms time window analysis. Results from the overall best-fitting model for the second time window are shown in Table 7. The main effect of Extraction was significant, indicating that amplitude for *wh*-extraction condition was more negative as compared to the baseline no extraction condition. Crucially, the interaction of Extraction × Island was removed during model-fitting, indicating that there was no significant difference in the negativity for island and non-island conditions. In sum, a late negativity emerged for both island and non-island conditions at the filled object position.

Table 7: Linear mixed-effects overall model for object position from 500-900 ms, with no individual difference measures included.

Number of obs: 5456, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-0.242	0.175	69	-1.383	0.171	
Extraction	-0.111	0.051	5404	-2.182	0.029	*
Island	0.113	0.090	5404	1.265	0.206	
Hemisphere (left)	-0.178	0.067	5404	-2.645	0.008	**

Hemisphere (right)	0.098	0.067	5404	1.456	0.145	
Anteriority (ant)	-0.380	0.084	5404	-4.510	< .001	***
Anteriority (post)	0.009	0.095	5404	0.094	0.925	
Island × Anteriority (ant)	-0.347	0.119	5404	-2.912	0.004	**
Island × Anteriority (post)	0.483	0.134	5404	3.596	< .001	***

At the filled object position, a complex pattern of results emerged. In the 300-500 ms time window, an N400 emerged for the non-island condition only, suggesting that the parser posited a gap only in the grammatically licensed (non-island) position. No evidence of gap prediction was found for the island condition in the 300-500 ms time window. In the second time window, a late negativity was observed for both non-island and island conditions. Specific predictions for processing in this later time window were not formulated, as it was not directly relevant for our research questions regarding gap prediction and island sensitivity; this late negativity in both conditions was therefore unexpected. In the next set of analyses, processing at the actual gap site is examined.

Actual Gap Site

Processing at the actual gap site shows whether participants resolve the *wh*-dependency and yield P600 at the actual gap site, in line with previous research (e.g., Felser et al., 2003; Gouvea et al., 2009; Kaan et al., 2000; Phillips et al., 2005). The actual gap site was present at different words in the island and non-island conditions (e.g., *with* vs. *kissed*) due to the inclusion of the relative clause in the island condition. To account for this lexical difference, separate models were run for the non-island and island conditions. The first analyses reported below focus on the actual gap site in the non-island condition.

Non-Island: 300-500 ms time-window analysis. Results from the best-fitting model for the non-island condition for the 300-500 ms time window are shown in Table 8. This model revealed a significant main effect of Extraction, indicating that the amplitude for the *wh*-extraction condition was more positive than the no extraction condition. Two significant interactions emerged involving the factor Extraction. Extraction × Hemisphere (left level) indicated that amplitude for electrodes in the left hemisphere were less positive than those in the midline region. The interaction term Extraction × Anteriority (anterior level) indicated that amplitude for electrodes in the anterior region were less positive than those in the central region. In sum, at the actual gap site for the non-island condition there was an early positivity which was attenuated in left anterior sites.

Table 8: Linear mixed-effects overall model for the actual gap site for the Non-Island condition from 300-500 ms, with no individual difference measures included.

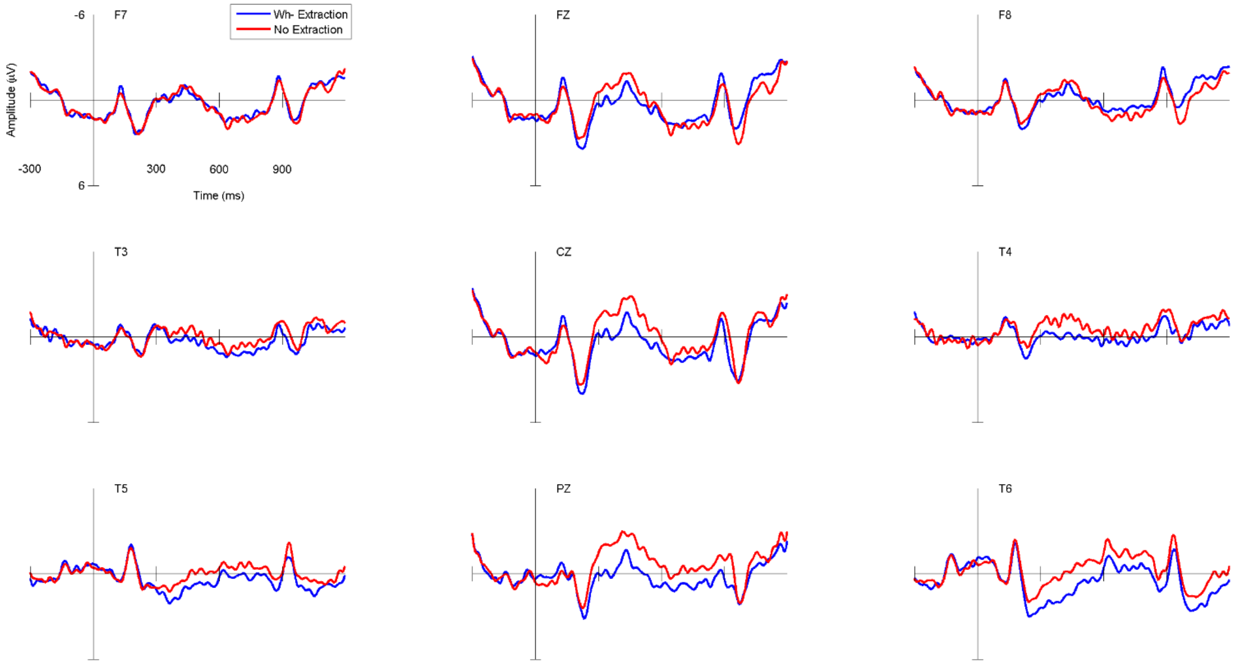
Number of obs: 2728, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-2.126	0.250	110	-8.512	< .001	***
Extraction	1.452	0.173	2671	8.400	< .001	***
Hemisphere (left)	1.357	0.183	2671	7.422	< .001	***
Hemisphere (right)	0.841	0.183	2671	4.601	< .001	***
Anteriority (ant)	0.853	0.186	2671	4.585	< .001	***
Anteriority (post)	0.841	0.205	2671	4.104	< .001	***
Extraction × Hemisphere (left)	-0.453	0.176	2671	-2.570	0.010	*
Extraction × Hemisphere (right)	-0.126	0.176	2671	-0.717	0.474	
Extraction × Anteriority (ant)	-0.601	0.156	2671	-3.860	< .001	***
Extraction × Anteriority (post)	0.202	0.176	2671	1.147	0.252	
Hemisphere (left) × Anteriority (ant)	-0.708	0.210	2671	-3.380	< .001	***
Hemisphere (right) × Anteriority (ant)	-0.758	0.210	2671	-3.616	< .001	***
Hemisphere (left) × Anteriority (post)	-0.509	0.233	2671	-2.186	0.029	*
Hemisphere (right) × Anteriority (post)	-0.115	0.233	2671	-0.496	0.620	

Non-Island: 500-900 ms time-window analysis. The results from the best-fitting model for the 500-900 ms time-window for the Non-Island condition are shown in Table 9 below. The main effect of Extraction was significant, indicating that the amplitude for *wh*-extraction condition was more positive than the no extraction condition. The interaction of Extraction × Anteriority (anterior level) was also significant, and indicated that amplitude for electrodes in the anterior region were less positive than those in the central region. The interaction of Extraction × Anteriority (posterior level) was marginally significant. Thus, in the 500-900 ms time window a positivity emerged that was attenuated at anterior sites. Figure 5 below illustrates this P600 effect for the non-island condition.

Table 9: Linear mixed-effects overall model for the actual gap site for the Non-Island condition from 500-900 ms, with no individual difference measures included.

Number of obs: 2728, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-0.003	0.188	82	-0.015	0.988	
Extraction	0.698	0.117	2677	5.955	< .001	***
Hemisphere (left)	0.018	0.088	2677	0.208	0.835	
Hemisphere (right)	-0.431	0.088	2677	-4.883	< .001	***
Anteriority (ant)	0.858	0.110	2677	7.785	< .001	***
Anteriority (post)	-0.563	0.124	2677	-4.528	< .001	***
Extraction × Anteriority (ant)	-0.893	0.156	2677	-5.732	< .001	***
Extraction × Anteriority (post)	0.318	0.176	2677	1.808	0.071	.

Figure 5: Actual gap site in Non-Island condition at representative electrodes.



Together, both time windows for the non-island condition showed that an extended positivity emerged at the actual gap site which was significant from 300-900 ms. In the earlier time window, the positivity was attenuated in the left hemisphere and the anterior region, and in the later time window the positivity was attenuated in the anterior region. The topography of the positivity is characteristic of the P600 response, which typically has a posterior distribution (e.g., Hagoort & Brown, 2000; Hagoort, Brown, & Osterhout, 1999). The latency is also in line with previous studies reporting P600s with an earlier onset, beginning at 300 ms (e.g., Kaan et al., 2000; Phillips et al., 2005). In summary, the predicted P600 emerged at the actual gap site in the non-island condition; next, results from the island condition are summarized.

Island: 300-500 ms time-window analysis. Results from the best-fitting model for the 300-500ms time window for the island condition at the actual gap site are shown in Table 10 below. No significant effects involving the factor Extraction emerged in this early time window at the actual gap site for sentences containing an island.

Table 10: Linear mixed-effects overall model for the actual gap site for the Island condition from 300-500 ms, with no individual difference measures included.

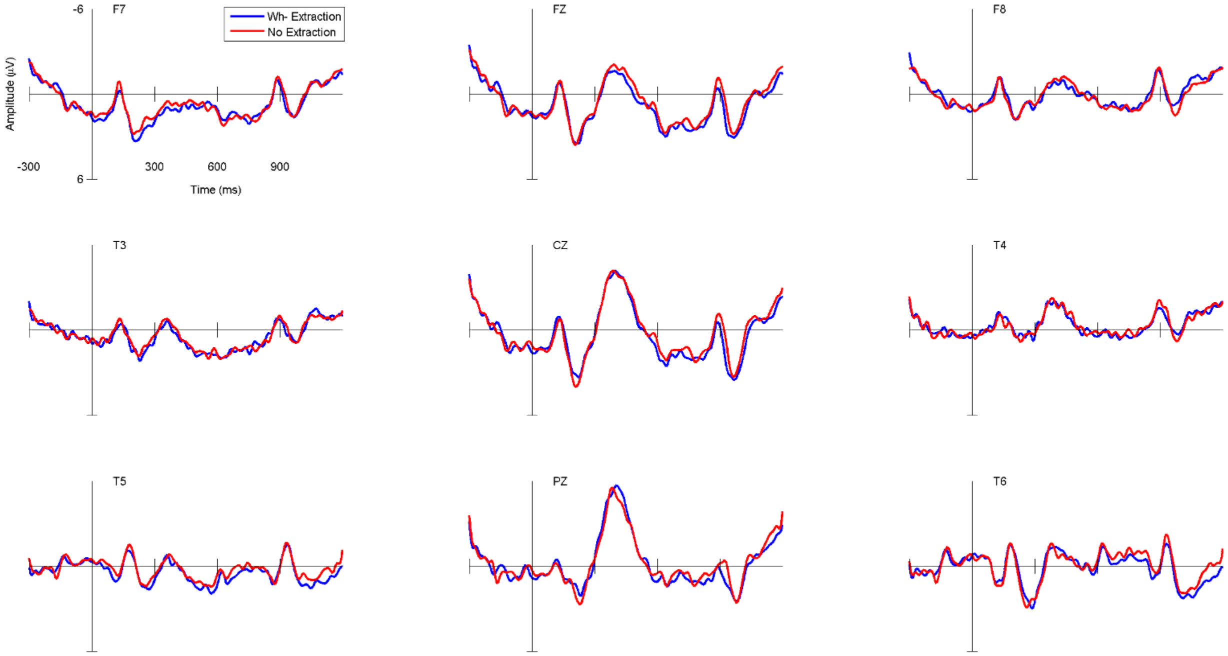
Number of obs: 2728, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-3.194	0.288	96.7	-11.108	< .001	***
Extraction	0.039	0.140	2673	0.278	0.781	
Hemisphere (left)	2.446	0.191	2673	12.798	< .001	***
Hemisphere (right)	1.369	0.191	2673	7.165	< .001	***
Anteriority (ant)	2.110	0.222	2673	9.515	< .001	***
Anteriority (post)	0.826	0.244	2673	3.383	< .001	***
Extraction × Anteriority (ant)	0.276	0.186	2673	1.488	0.137	
Extraction × Anteriority (post)	-0.269	0.209	2673	-1.284	0.199	
Hemisphere (left) × Anteriority (ant)	-1.368	0.250	2673	-5.471	< .001	***
Hemisphere (right) × Anteriority (ant)	-1.429	0.250	2673	-5.717	< .001	***
Hemisphere (left) × Anteriority (post)	-1.299	0.278	2673	-4.678	< .001	***
Hemisphere (right) × Anteriority (post)	-0.373	0.278	2673	-1.343	0.179	

Island: 500-900 ms time-window analysis. Table 11 below shows the results from the best-fitting model for the 500-900 ms time window. The main effect of Extraction was significant, indicating that amplitude for the *wh*-extraction condition was more positive compared to the no extraction condition. Thus, as predicted, a P600 emerged at the actual gap site for the island condition, shown in Figure 6 below.

Table 11: Linear mixed-effects overall model for the actual gap site for the Island condition from 500-900 ms, with no individual difference measures included.

Number of obs: 2728, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	0.838	0.193	68.7	4.340	< .001	***
Extraction	0.191	0.068	2679	2.812	0.005	**
Hemisphere (left)	0.036	0.090	2679	0.403	0.687	
Hemisphere (right)	-0.704	0.090	2679	-7.824	< .001	***
Anteriority (ant)	0.304	0.080	2679	3.827	< .001	***
Anteriority (post)	-0.413	0.090	2679	-4.602	< .001	***

Figure 6: Actual gap site in Island condition at representative electrodes.



Summary

Before reporting results from the individual difference analyses, the key findings of the main ERP analyses are summarized as the following. At the subject position, N400 emerged, suggesting that participants predicted a gap prior to encountering the subcategorizing verb. At the object position, N400 emerged only in the non-island condition; no N400 was yielded in the island condition. This finding provides evidence that participants posited a gap only in the grammatically licensed position, and not inside the island. A late negativity emerged at the filled object position in both the island and non-island conditions, a component which was unexpected. Finally, at the actual gap site, P600 was present in both the non-island and island conditions, indicating that successful dependency resolution took place in both conditions.

Individual Difference Analyses

Individual difference analyses including the cognitive measures and the offline acceptability judgments were conducted for pre-selected time windows based on when significant effects surfaced. For example, at the subject position, the main effect of Extraction was significant only in the 300-500ms time window. Thus, analyses including the individual difference measures are only reported for the 300-500ms time window for the subject position. Note that DD scores from the offline acceptability judgment task were only included in the models for the object position. DD scores, which measured offline island sensitivity, were used to examine processing at the region of the sentence involving an island, the object position. All best-fitting models discussed below are reported in full in Appendix A.

Subject Position: 300-500 ms time-window analysis.

The first analysis examined how individual differences in cognitive abilities impact subject gap prediction, indexed via N400. As described above, several studies have found a relationship between cognitive abilities and prediction, suggesting that those with greater cognitive resources may be better able to engage in predictive processing (e.g., Hutchison, 2007;

Johnson, 2015; Johnson et al., 2016; Nicenboim et al., 2015). We therefore test whether individuals with greater working memory and attentional control resources show larger subject filled-gap effects.

Working Memory. In the overall best-fitting model including working memory scores, the crucial interaction term Extraction × Working Memory was significant (t (2668) = 2.157, $p < .05$), and indicated that higher scores on the working memory composite measure were associated with more positive amplitude for the *wh*-extraction condition, as compared to the no extraction condition. In other words, individuals with higher working memory yielded a reduced N400 at the subject position (i.e., smaller filled gap effect).

Attentional Control. In the best-fitting model including the composite attentional control score, the critical interaction term Extraction × Attentional Control was marginally significant in the model (t (2668) = 1.882, $p = .060$). However, the three-way interaction of Extraction × Anteriority (anterior level) × Attentional Control was significant (t (2668) = -2.289, $p < .05$). This effect indicates that individuals with higher attentional control resources yielded larger negativities in the anterior region.

In summary, the N400 yielded at the subject position was modulated differently by the working memory and attentional control measures. Higher working memory was associated with a smaller subject filled-gap effect: a reduced N400. Better attentional control resources were associated with an increased N400, specifically in the anterior region.

Object Position: 300-500 ms time-window analysis.

Individual difference analyses at the object position address the research question concerning the source of island sensitivity. The main EEG analysis showed that readers did not attempt to posit a gap inside an island, such that N400 emerged only in the licit non-island context and crucially not in the island condition. Next, we examine whether participants avoid positing gaps in islands due to grammatical constraints or a processing burden. The grammatical account of islands predicts no relationship between cognitive abilities and ERP effects within the island (e.g., Sprouse et al., 2012). As argued by Sprouse et al. (2012), the processing account would expect that individuals with greater cognitive resources should show an increased ability to posit a gap within the island, thus yielding increasingly N400-like effects in the island (e.g., Hofmeister & Sag, 2010; Kluender & Kutas, 1993b; Kluender, 2004). A relationship between the cognitive measures and the size of the N400 effect in filled-gap positions outside of islands does not tease apart the two competing accounts. However, as at the subject position, we examine whether individuals with greater processing resources show larger filled-gap effects in the licit context.

Working Memory. In the best-fitting model including the composite working memory score, the critical three-way interaction term Extraction × Island × Working Memory was removed during model-fitting, indicating that it did not explain a significant portion of the variance in the model. The interaction Extraction × Working Memory was also not significant (*t* (5387) = 1.517, *p* = .129). However, the interaction term Island × Working Memory was significant (*t* (5387) = -4.258, *p* < .001), indicating that higher working memory was associated with increasingly negative amplitude for non-island sentences as compared to the baseline island sentences. Without the effect of Extraction involved in the interaction, this indicates that overall, both conditions containing islands (no extraction, *wh*-extraction) were more positive than the two non-island conditions (no extraction, *wh*-extraction). This relationship between the overall amplitude of island vs. non-island conditions at the object position does not address our research questions, and thus is not discussed further.

Attentional Control. In the best-fitting model including attentional control composite scores, the critical interaction Extraction × Island × Attentional Control was significant (t (5382) = -3.004, $p < .01$). To follow-up on this interaction, separate models were run for each level of the factor Island. In the best-fitting non-island model, the critical interaction Extraction × Attentional Control was significant (t (2672) = -3.836, $p < .001$), indicating that with increasing attentional control scores, the amplitude for the *wh*-extraction condition becomes more negative as compared to the baseline no extraction condition. In other words, better attentional control resources are associated with larger N400 (i.e., larger filled-gap effect) in the licit non-island condition. In the best-fitting model for the island condition, the critical interaction Extraction × Attentional Control was marginally significant (t (2666) = 1.787, $p = .085$), indicating that attentional control did not modulate processing inside the island.

Acceptability Judgment Ratings. In the best-fitting model including DD scores, the critical interaction Extraction × Island × DD score was significant (t (5392) = -3.182, $p < .01$). To further investigate this interaction, separate models were run for each level of the factor Island. In the non-island model, main effects and interactions involving the factor DD scores were removed from model-fitting, indicating that DD scores did not explain a significant portion of variance in the non-island overall model. Note that a relationship between DD scores, which measure offline island sensitivity, and processing at the licit object filled gap in the non-island condition is not expected to emerge, in line with the findings for this model. In the best-fitting model for the island condition, the critical interaction of Extraction × DD score was significant (t (2668) = 2.070, $p < .05$). This effect shows that as DD scores increase, amplitude for the *wh*-

extraction condition was less negative as compared to the baseline no extraction condition. In other words, greater island sensitivity in the offline rating task was associated with a less N400-like effect inside the island.

To summarize, in the 300-500 ms time window at the filled object position, working memory did not modulate gap prediction. In contrast, higher attentional control was associated with a larger object filled-gap effect (increased N400) in the licit non-island context. Crucially, neither working memory nor attentional control modulated processing inside of the island, in contrast to Sprouse et al.’s (2012) predictions for the processing account of island sensitivity (e.g., Kluender & Kutas, 1993b; Kluender, 2004; Hofmeister & Sag, 2010). Finally, the analysis including DD scores revealed that individuals with greater offline island sensitivity showed less predictive effects inside the island. That is, the more sensitive an individual was to island violation sentences in the offline acceptability judgment task, the less likely they were to yield a filled-gap effect inside the island.

Object Position: 500-900 ms time-window analysis.

The main EEG analysis revealed a late negative component at the filled object position for both the non-island and island contexts. This late negativity was unexpected; in the following analyses, we explore whether individual differences in cognitive abilities and offline island sensitivity modulate this late negativity, potentially shedding light on the mechanisms underlying this component.

Working Memory. In the best-fitting model including the composite working memory score, the critical three-way interaction term Extraction × Island × Working Memory was removed during model-fitting, indicating that it did not explain a significant portion of the variance in the model. The other crucial interaction terms were also not significant: Extraction × Working Memory (*t* (5388) = 0.478, *p* = .633) and Island × Working Memory (*t* (538) = -0.696, *p* = .487). Thus, working memory did not modulate the late negativity.

Attentional Control. In the best-fitting model including attentional control composite scores, the critical interaction Extraction × Island × Attentional Control was significant (*t* (5398) = -6.394, *p* < .001). To follow-up on this interaction, separate models were run for each level of the factor Island. In the best-fitting non-island model, the critical interaction term Extraction × Attentional Control was significant (*t* (2676) = -7.654, *p* < .001), indicating that with increasing attentional control, the amplitude for the *wh*-extraction condition becomes more negative as compared to the baseline no extraction condition. In other words, higher attentional control was associated with a larger late negativity in the non-island condition. In the best-fitting model for the island condition, all main effects and interaction terms involving the factor Extraction were removed during model fitting. Thus, attentional control did not modulate processing inside the island in the 500-900 ms time window.

Acceptability Judgment Ratings. In the best-fitting model including DD scores, the critical interaction Extraction × Island × DD score was significant (*t* (5398) = -2.677, *p* < .01). To further investigate this interaction, separate models were run for each level of the factor Island. In the non-island model, main effects and interactions involving the factor DD scores were removed from model-fitting; as expected, offline island sensitivity did not modulate processing in the non-island condition. The best-fitting model for the island condition showed a significant interaction of Extraction × DD score (*t* (2678) = 5.837, *p* < .001). This effect shows that as DD scores increase, amplitude for the *wh*-extraction condition was more positive as compared to the no extraction condition. In other words, greater offline island sensitivity was associated with a reduced late negativity in the island condition.

To briefly summarize, the late negativity at the object position was modulated by both attentional control and offline island sensitivity. Interestingly, in the main EEG analysis the late negativity was significant for both the non-island and island condition; this analysis revealed that the late negativities for the two conditions were distinctly modulated by different individual difference measures. Higher attentional control was associated with a larger negativity in the non-island condition. Increased offline sensitivity to islands was associated with a reduced negativity in the island condition. This pattern helps shed light on the interpretation of the late negativity, which we discuss below as being related to the processing of thematic arguments.

Actual Gap Site: 500-900 ms time-window analysis

Processing at the actual gap site showed that participants completed the *wh*-dependency, yielding the expected P600 response in both the non-island and island conditions. To our knowledge, no study to date has investigated how individual differences in cognitive resources modulate the P600 response yielded at the actual gap site of a *wh*-dependency. Following results from the subject and object position, we examine whether greater cognitive abilities are associated with an increased ability to successfully resolve the dependency, yielding larger P600.

Working Memory. As in the main EEG analyses, separate models were run for non-island and island conditions due to lexical differences at the actual gap site. In the best-fitting model for the non-island condition containing the factor Working Memory, the critical interaction Extraction × Working Memory was significant (t (2672) = 2.954, p < .01). This interaction indicates that as working memory scores increase, the P600 becomes larger. Three-way interaction terms involving the factor Anteriority were also significant: Extraction × Anteriority (anterior level) × Working Memory (*t* (2672) = -2.206, *p* < .05) and Extraction × Anteriority (posterior level) × Working Memory (*t* (2672) = 1.874, *p* < .05). This complex pattern reveals that overall, individuals with higher working memory yielded a larger amplitude P600 which was centro-posterior distributed, as the positivity was attenuated at anterior sites and larger at posterior sites.

In the best-fitting model for the island condition at the actual gap site, the critical interaction Extraction × Working Memory was not significant (*t* (2672) = 0.344, *p* = .731). However, the three-way interaction Extraction × Anteriority (posterior level) × Working Memory was significant (*t* (2672) = 1.965, *p* < .05). This interaction indicates that individuals with higher working memory yielded a larger P600 in the posterior region at the actual gap site in sentences containing an island.

Attentional Control. In the overall model for the non-island condition including the factor Attentional Control, a significant interaction between Extraction × Attentional Control emerged (*t* (2674) = 8.973, *p* < .001). This interaction showed that higher attentional control was related to larger P600 amplitude at the actual gap site in the non-island condition.

In the best-fitting island model, all interactions involving the factor Attentional Control were removed during model-fitting. Thus, attentional control did not modulate processing at the actual gap for sentences containing an island.

In summary, both working memory and attentional control modulated P600s at the actual gap site. Higher working memory was associated with larger P600s in both the non-island and island conditions, whereas higher attentional control was associated a larger P600 in the non-island condition only. As expected, greater cognitive abilities were associated with an increased ability to successfully resolve the dependency, yielding a larger P600.

Discussion

The present study tracked the processing of *wh*-dependencies across a grammatical sentence, examining whether distinct ERP components index prediction of a gap and resolution of the dependency, and whether gap prediction is grammatically guided. Our study yielded several key findings, summarized in Table 12 below. Gap prediction at both subject and licit object positions was indexed by N400, and was indeed sensitive to grammatical constraints, as no N400 emerged inside the island. At the actual gap site, P600s emerged, suggested to index successful *wh*-dependency resolution. Finally, both gap prediction and dependency resolution were subject to individual variation in processing related to cognitive abilities and offline island sensitivity. In the following sections, each finding is discussed in turn.

Table 12: Summary of main ERP findings for native English speakers.

	Subject filled gap	Object filled gap (non-island)	Object filled gap (island)	Actual gap site (non-island)	Actual gap site (island)
Natives (n=44)	N400	▪ N400 ▪ Late negativity	▪ No N400 ▪ Late negativity	P600	P600

Gap predictability and the N400

N400s emerging in the current study are compatible with a prediction-related account of the N400, as are the N400 effects reported for unexpected actual gap sites in Michel (2014). Previous studies reporting prediction-related modulation of the N400 have examined lexical prediction, or specific predictions for features of particular words (e.g., Federmeier, 2007; Lau et al., 2013; Van Berkum et al., 2005). The current study, together with Michel (2014), suggests that N400 also reflects gap prediction, as indicated by the N400 effects yielded by lexical material appearing where a gap has been predicted (in the current study), as well as encountering an actual gap where it was not predicted (Michel, 2014). The N400 additionally sheds light on the nature of filled-gap effects, since some researchers have hypothesized that these effects reflect prediction, while others have suggested that the reading time slowdowns index a costly structural reanalysis process (Lee, 2004). Given that N400s emerged in filled gap positions, rather than P600s, our findings do not straightforwardly support accounts attributing filled-gap effects solely to structural reanalysis; instead, findings suggest that filled-gap effects may reflect, at least in part, encountering unpredicted lexical material.

Our experimental design allowed us to examine both pre-verbal and post-verbal gap prediction. Filled gaps in both subject and object position (outside of islands) elicited N400s, indicating that native English speakers engaged in prediction of an upcoming gap site before encountering the gap licensing verb (in subject position), and after encountering the verb (in object position). Pre-verbal gap prediction provides strong evidence that gaps can be anticipated using syntactic knowledge, and this study provides converging evidence with Johnson (2015) that native English speakers can generate predictions immediately after encountering a filler, and do not require additional time to be able to do so. As described previously, earlier findings in the literature are variable regarding emergence of subject filled-gap effects in English (e.g., Lee, 2004; Stowe, 1986). It is possible that that the fine-grained temporal sensitivity of EEG allowed us to capture this subtle effect.

Analyses at the group level revealed effects of gap prediction at the subject position, although individual variation in processing was also observed at this region. Specifically, increased working memory was associated with reduced N400 amplitude at the pre-verbal subject position. The influence of working memory on processing is perhaps opposite than what was expected, as other studies that found a relationship between filled-gap effects and working memory reported increased predictive effects, as indexed by longer reading time slowdowns (e.g., Johnson et al., 2016). In the current study, however, individuals with increased working memory capacity showed less of predictive effect (i.e., smaller N400) at the subject position. This suggests that individuals with higher working memory capacity either did not engage in gap prediction, or did so to a lesser degree.

There are several possibilities regarding the nature of the relationship between working memory capacity and gap prediction during *wh*-dependency resolution. One possibility is that working memory is a cognitive resource that is allocated to processes other than gap prediction. Several researchers, for example, have argued that working memory is critical in the storage and maintenance of the *wh*-filler in memory (e.g., Gibson, 1998, 2000; King & Just, 1991; King & Kutas, 1995). This ability is crucial in the processing of long-distance dependencies to be able to ultimately resolve the dependency at the actual gap site. It is possible, then, that at a potential gap site which ultimately turns out to be filled, working memory resources are specifically allocated to storing and maintaining the *wh*-filler in memory while simultaneously processing the encountered noun phrase. Note that this explanation does not account for the fact that working memory did not modulate processing at the object filled gap site. Given that the object position is located further downstream from the *wh*-filler, working memory resources should be particularly important in the storage and maintenance of the *wh*-word in memory upon encountering a filled gap. However, this relationship did not emerge at the object position, although we note that with increasing sample size this interaction may become significant.

A second possibility was suggested by Johnson et al. (2016), one of the only other studies to report a similar relationship between higher working memory and smaller filled-gap effects, albeit for L2 learners. Johnson et al. found that Korean-speaking learners of English with increased working memory yielded smaller object filled-gap effects. To explore this relationship further, learners were split into high and low working memory groups, which revealed that learners with higher working memory showed a reduced filled-gap effect at the spillover region immediately following the filled object position, as compared to the learners in the lower working memory group. The reduced filled gap effect at the spillover region for learners with increased working memory resources suggests that increased working memory resources allow the parser to more quickly recover from the object filled-gap in the preceding region. Our findings for native English speakers are compatible with this explanation, such that increased working memory allows the parser to rapidly recover from the subject filled-gap effect to continue *wh*-dependency formation. A related explanation is that higher working memory individuals may be better able to retain an alternate object gap analysis. Individuals with better working memory resources may be better able to pursue multiple analyses regarding the location of an upcoming gap, leading them to commit less to the subject analysis than individuals with lower working memory resources (e.g., Lee, 2004). In other words, upon encountering the filled subject position, higher working memory individuals would be able to realize that a potential object gap is likely to be forthcoming, allowing them to commit to the subject gap less and continue to search for the actual gap site. Further research directly investigating the contribution of working memory in the processing of subject vs. object filled-gap effects is necessary to tease apart the role of working memory resources in gap prediction during the processing of *wh*-dependencies.

The current study included a measure targeting a second cognitive ability, attentional control. Recall that Johnson (2015) found that increased attentional control was associated with a larger subject filled-gap effect in her self-paced reading study. In the current sample, higher attentional control was associated with a larger negativity in the anterior region at the subject filled gap. In the main EEG analysis (i.e., without individual difference measures included in the model), the subject filled gap N400 was shown to be largest at posterior sites. Thus, the relationship with attentional control in the anterior region reveals that individuals with higher attentional control yielded a more broadly distributed N400. A similar effect emerged at the licit object position: in the non-island condition, greater attentional control was associated with increased N400 amplitude at the filled object position. In line with Johnson (2015), this suggests that attentional control resources play an important role in gap prediction during *wh*-dependency resolution. This finding is also in line with research showing that attentional control modulates prediction in language processing (e.g., Boudewyn et al., 2012; Hutchison, 2007), as well as expectation-driven processing in general (e.g., Bar, 2009; Kane & Engle, 2002). Hutchison (2007), for example, found that individuals with increased attentional control were more sensitive to a context cue, yielding larger prediction effects in a word-pair semantic priming study. In the processing of *wh*-dependencies, attentional control resources may similarly modulate an individual's sensitivity to a cue (i.e., the *wh*-filler) that a potential gap site is forthcoming (e.g., Johnson, 2015). Because the potential gap site was filled with lexical material, it may also be the case that attentional control resources are involved in the recovery from the incorrect prediction at the filled gap. In a recent paper, Hsu and Novick (2016) reported that in a condition in which cognitive (attentional) control was engaged, participants showed an increased ability to recover from an initial misinterpretation when listening to temporarily ambiguous sentences. Attentional control was argued to facilitate the process of adjusting and recovering from syntactic misinterpretations during online processing. In the current study, it may also be the case that increased attentional control allows participants to rapidly recover from the filled-gap effect and continue to search for upcoming potential gap sites in the input.

Island sensitivity

A second line of investigation for this study examined sensitivity to island constraints, specifically investigating whether the parser avoided predicting a gap inside an island, where *wh*-extraction is not licensed by the grammar. In the critical N400 time window, the results indicated that the parser did not attempt to posit a gap within the islands, consistent with the view that the parser is guided by grammatical constraints (e.g., Phillips, 2006, 2013; Sprouse et al., 2012; Wagers & Phillips, 2009). Although previous studies have shown sensitivity at island boundaries such as at a complementizer (e.g., Kluender & Kutas, 1993b; Michel, 2014) and at actual gap positions within islands in ungrammatical sentences (Michel, 2014), this study provides some of the first electrophysiological evidence of island sensitivity at a filled-gap position located within a fully grammatical sentence.

We are also able to address the debate concerning the specific nature of island sensitivity in our analyses including the individual difference measures (e.g., Hofmeister & Sag, 2010; Kluender, 2004; Kluender & Kutas, 1993b; Sprouse et al., 2012). Crucially, neither working memory nor attentional control modulated processing inside the island in the critical N400 time window. As Sprouse et al. (2012) argue, the processing account of islands would expect that increased cognitive abilities would associated with an increasingly N400-like effect, such that higher attentional control or working memory would allow participants to overcome the processing burden to be able to posit a gap inside the island domain. Instead, individuals with increased cognitive abilities did not show evidence of attempting to posit a gap within the island, as might be expected under a processing account of island (e.g., Hofmeister & Sag, 2010). This finding is better accounted for by the grammatical account of island sensitivity, which does not expect such a relationship to emerge.

An additional strength of this study is its ability to compare offline and online island sensitivity within participants. DD scores, an index of offline sensitivity to ungrammatical island violation sentences, were included in the analysis as a predictor of EEG effects inside the island. Recall that higher DD scores are indicative of increased sensitivity to island violations. At the object position, higher DD scores were associated with a less negative ERP response. In other words, individuals who showed greater sensitivity to island violations offline yielded *less* of a prediction-related response in the island, a domain where gap positing is grammatically banned. This provides some of the first evidence linking offline and online island sensitivity in adult native English speakers.

Dependency resolution and the P600

The P600 effect at the actual gap site lends support to accounts which posit that the P600 serves as an index of successful dependency resolution (e.g., Felser et al., 2003; Gouvea et al., 2010; Kaan et al., 2000; Phillips et al., 2005). This is the first study to test for successful dependency resolution in fully grammatical sentences containing islands, and the fact that P600s emerged at the actual gap site indicates that readers eventually resolved the dependency, despite the increased processing burden incurred by processing the island.

We next consider the differences between the P600s elicited at the actual gap site in the two conditions. In the current study, P600s emerged both at the subcategorizing verb in island conditions and at the gap licensing preposition in the non-island conditions. Thus far, P600s related to dependency resolution have only been elicited at a subcategorizing verb (e.g., Kaan et al., 2000; Phillips et al., 2005). While we cannot directly compare the P600 components across conditions due to lexical differences, the positivity elicited by the preposition (non-island condition) had an earlier onset, reaching significance in the 300-500 ms time window, and appeared to be more robust and sustained in comparison to the positivity at the verb (island condition). There are several reasons why these differences may have emerged. As reviewed above, Phillips et al. (2005) found that the onset of the P600 was earlier for a shorter *wh*-dependency as compared to a long-distance dependency. Although the current study did not aim to directly investigate latency differences in the P600 by manipulating *wh*-dependency length, as a consequence of including a relative clause in the island condition, the *wh*-dependency in the island condition is one word longer than in the non-island condition due to the addition of the complementizer *that*. In line with Phillips et al., we do find an earlier onset in the P600 for the non-island condition, which was significant starting from the earlier 300-500ms time window. In contrast, the P600 in the island condition reached significance in the 500-900ms time window.

Another difference between the P600 responses besides timing or onset is amplitude, the size of the effect. Kaan et al. (2000) suggests the P600 amplitude serves as an index of integration cost, with long-distance dependencies yielding increased amplitude as compared to shorter distance dependencies. Phillips et al. (2005) fail to replicate this effect in their experiment, and although our study cannot directly compare ERP amplitude across conditions, our data visually suggest that the opposite pattern is in effect. That is, non-island sentences, which contained a *wh*-dependency that was one word shorter than in the island condition, appeared to yield a more robust ERP effect. Another proposal by Felser et al. (2003) is that the amplitude of the P600 is larger for the depending on the number of integration processes involved. In the non-island condition, there is a larger integration "cost" because the gap site is licensed by a preposition rather than a verb. That is, upon encountering the gap site at the preposition, the parser must semantically integrate the filler with the gap, and additionally, the prepositional phrase must also be linked to its subcategorizing verb to interpret the sentence.

Gouvea and colleagues (2010) put forth a comprehensive proposal regarding both the latency and amplitude of the P600. Their study directly compared P600s yielded for different contexts, including *wh*-dependency resolution, syntactic violations, and garden path sentences. They argue that, "P600 amplitude and duration directly reflect structure-building (and dismantling) operations, whereas the retrieval processes that are needed to initiate structure building are reflected only in the onset latency of the P600" (p. 27). While the present study is unable to directly compare processing at verbal vs. prepositional gap sites, Gouvea et al.'s proposal makes predictions regarding P600 latency and amplitude for dependency resolution in these two conditions. First, in line with Phillips et al. (2005), longer-distance *wh*-dependencies should have later onset P600 based on the fact that longer time is needed to retrieve the more distant *wh*-filler from memory. As noted above, our results support this proposal. Gouvea et al.'s second prediction regarding amplitude of the P600 is similar to Felser et al.'s (2003), in that there are likely more "structure-building operations" involved in dependency resolution taking place at a preposition as compared to a verb. The *wh*-filler can be readily integrated at the verb, the gap site for the island condition, as compared to the non-island condition, where the gap is licensed by a preposition which must be integrated with its licensing verb. In sum, the latency and amplitude differences of the P600 across the non-island and island conditions reveal variation in the P600 response yielded for dependency resolution, and future research which systematically examines this variation may shed light on the cognitive mechanisms underlying this component (e.g., Gouvea et al., 2010).

Our findings additionally address a criticism put forth by Michel (2014) regarding the claim that P600 reflects successful dependency resolution. Michel ultimately argues for a conservative account in which the P600 indexes identification of a gap, rather than completion of the dependency, and points out that most robust dependency resolution P600 effects have been found for studies and conditions in which a discourse linked (d-linked) filler was utilized. For example, Kaan et al. (2000) report a stronger P600 effect for a condition in which the *wh*-item was d-linked (e.g., *which pop-star*) as compared to bare filler (e.g., *who*). Other studies reporting the P600 have also utilized d-linked fillers (e.g., *which accomplice* in Phillips et al., 2005), and Michel suggests that the P600 may be driven by a d-linking manipulation, citing evidence from Hofmeister (2007) who demonstrated that d-linked fillers exhibit processing facilitation compared to bare fillers at the gap site. However, the current study used bare fillers (e.g., *who*) and found robust P600s across non-island and island conditions. This would argue against an account in which the positivity emerges primarily due to the semantically rich nature of the (d-linked) filler.

Additionally, this study is one of the first to examine how cognitive abilities impact an individual's ability to successfully resolve a *wh*-dependency. Both working memory and attentional control were shown to modulate processing at the actual gap site. Across non-island and island sentences, increased working memory was associated with larger P600s at the actual gap site. This relationship is in line with memory-based proposals of *wh*-dependency resolution, and suggests that individuals with increased working memory are better able to successfully resolve the dependency at the actual gap site (e.g., Gibson, 1998, 2000; King & Just, 1991; King & Kutas, 1995). This relationship may emerge due to a stronger ability to maintain the *wh*-filler in memory, an increased ability to reactivate the *wh*-filler at the actual gap site, or some combination of both processes. Further studies are needed to tease apart the specific contribution of working memory in dependency resolution.

Attentional control modulated processing only in the non-island condition, revealing that increasing attentional control was associated with larger P600s. At least for the non-island sentences, attentional control resources are associated with a greater ability to successfully resolve the dependency. There are several possibilities regarding why this relationship did not emerge for the island condition. One possibility relates to the latency and amplitude differences in the P600 amplitude between the two conditions. Although there was a significant positivity emerging in the 500-900ms time window for the both the island and non-island sentences, the non-island P600 emerged earlier and appears to be more robust. Thus, it may be case that less variability is captured in the P600 response in the island condition, making it difficult for significant relationships with the cognitive scores to emerge. Given that working memory scores significantly predicted P600 amplitude in this condition, this possibility remains unlikely. Instead, this difference may stem from the processing of the island itself.

Consider the processing profile at the previous region, the object filled gap site. At this region, gap prediction was in evidence within the non-island condition, yielding an N400 response. By the time the parser reaches the actual gap site in the non-island condition, the parser has already predicted a gap twice, at the embedded subject and object position, neither of which ended up being the actual gap site. Thus, in order to successfully resolve the *wh*-dependency in this condition, the parser must be able to ‘recover’ from failed gap prediction and continue to search throughout the rest of the sentence for the actual gap site. Individuals with increased attentional control likely have a greater ability allocate attention to the continued gap search and subsequent resolution process, yielding a larger P600 at the ultimate gap site. Next, consider the sentences containing an island. At the object filled gap position inside the island, no prediction-related ERP effects were in evidence, suggesting that the parser utilized grammatical knowledge to avoid predicting a gap in this domain. Therefore, upon reaching the actual gap site in the island condition, the parser has only had to recover from one failed prediction at the early subject position. The P600 response at the actual gap site was not modulated by attentional control, which may suggest that increased attentional control is particularly useful in resolving *wh*-dependencies in sentences which involve multiple instances of failed prediction. Note that this possibility does not imply that processing sentences containing islands is ‘easier’ than the non-island comparison sentence. Rather, this difference may stem from the gap prediction process, related to the number of potential gap sites present in a given sentence.

Late negativity at the object position

A negativity emerged from 500-900 ms at the object position, which was significant in both the non-island and island conditions. This late negative component was unexpected, and interestingly, the individual difference analyses revealed that different measures modulated the component in distinct ways. Attentional control modulated the late negativity only in the non-island condition, such that individuals with higher attentional control yielded a larger negativity. Offline island sensitivity, on the other hand, was shown to modulate processing only in the island condition, with increased island sensitivity associated with a smaller negativity. By considering the post-verbal context in which this late negativity emerged, as well as the individual difference findings, we believe that one possble interpretation is that this component indexes processes related to thematic role assignment. Whereas N400s were yielded at both the subject and object positions in the sentence, indicating that the parser predicted a gap at both regions, the late negativity emerged only at the object filled gap position, following the verb. Considering the differences in the type of information available to the parser at the subject vs. object positions may shed light on why the late negativity was only in evidence at the object filled gap. To be able to predict a gap at the subject position, the only available cue that a gap is forthcoming is the presence of the *wh*-filler itself; the subcategorizing verb has not yet been encountered. The object position, where the late negativity emerged, directly follows the subcategorizing verb (e.g., *interviewed*) which licenses thematic arguments. One possibility, then, is that the late negativity is related to thematic role assignment driven by the verb. Several studies have reported an increased negativity in sentences with argument-induced conflicts (e.g., Frenzel, Schlesewsky, & Bornkessel-Schlesewsky, 2011; Frisch & Schlesewsky, 2001, 2005). This line of research has shown an enhanced negativity in sentences with thematic conflicts, such as when two animate NPs which can grammatically serve as the same thematic argument (i.e., two grammatical subjects). Frisch and Schlesewsky (2001), for example, suggest that this negativity is linked to the use of animacy information during the processing of two potential referents. In the current study, the filled object gap position indeed involves thematic role assignment difficulty. As soon as the verb is encountered, the *wh*-filler can be integrated directly into the argument structure and assigned a theta role. Therefore, the noun phrase that is next encountered in the object position (e.g., *Dave Campbell*) cannot be integrated into the argument structure, yielding an increased late negativity. Although increased negativities for thematic argument conflicts have been reported in earlier time windows (e.g., 400-550 ms), unlike in our study (500-900 ms), these studies utilized a different linguistic manipulation, such as contexts with ‘double case’ ungrammatiality (e.g., Frisch & Schlesewsky, 2005).

One study by Hestvik and colleagues (2012) has reported a similar late negativity argued to be related to thematic role assignment difficulty, utilizing a filled-gap paradigm. In an earlier paper, Hestvik et al. (2007) formulated specific predictions for particular ERP responses which should emerge at a filled gap, based on Friederici and colleagues’s model of syntactic parsing (e.g., Friederici, 1995, 2002; Friederici, Hahne, & Mecklinger, 1996; Friederici, Hahne, & von Cramon, 1998). They argue that if the first ERP component to emerge at a filled gap site became significant in the 300-500 ms time window (i.e., either N400 or LAN), the time-course that Friederici’s model associates with argument structure building, this would indicate that ungrammatical filled gaps are treated as an argument structure violation. In their follow-up study to Hestvik et al. (2007), Hestvik et al. (2012) found an anterior[5] negativity at a filled gap site, which they interpret as a left anterior negativity (LAN). As in the current study, however, this negativity became significant in a later time window, 500-800 ms. Hestvik et al. suggest that this negativity is yielded from processing argument structure operations, and the latency of this component is in line with the late negativity we observed at the object filled gap.

If the late negative component yielded at the object filled gap is related to thematic role assignment, it is important to explain why this component was yielded in both the non-island and

---

[5] Hestvik et al.’s (2012) statistical analysis did not involve an omnibus ANOVA where all electrodes and regions were included in the analysis. Instead, electrodes in the left and right anterior inferior region were selected based on visual inspection.

island conditions. In the island condition sentence (e.g., *Jamie wondered who the editor [that interviewed Dave Campbell] kissed after the interview*), the potential subcategorizing verb *interviewed* is crucially located inside the relative clause island. If native English speakers are able to utilize syntactic island constraints in online processing, the parser should not attempt to link the *wh*-filler with this verb and assign it a theta role, because *wh*-extraction from within islands is banned in English. However, the late negativity emerging at the filled object region inside the island indicates that there was a thematic argument conflict between the *wh*-filler *who* and the object NP *Dave Campbell*. This suggests that upon encountering a verb within an island domain, thematic role assignment to the *wh*-filler has taken place, despite the fact that this is grammatically unlicensed in English. Crucially, DD scores measuring offline island sensitivity were shown to modulate the late negativity in the island condition. Individuals with greater offline sensitivity to island violations showed a reduced late negativity. That is, with increasing offline sensitivity to islands, native English speakers show less interference from the *wh*-filler when assigning thematic roles to the verb inside the island (i.e., yielding reduced negativities or no effects). We take this pattern to suggest that individuals with greater island sensitivity offline show an increased sensitivity to island violations online. Importantly, cognitive abilities were not shown to modulate this component in the island condition, such that it was not the case that individuals with greater processing resources showed evidence of being better able to link arguments inside the island (i.e., yielding a larger late negativity), which would be in line with a processing-based account of island sensitivity (e.g., Hofmeister & Sag, 2010). Instead, attentional control was shown to increase the late negativity in the non-island condition only, indicating that greater attentional control resources are associated with a greater ability to link thematic arguments only in grammatically licensed positions.

Conclusions

This chapter presented results from a study tracking the processing of *wh*-dependencies across a well-formed sentence, examining gap prediction as well as dependency resolution. We used a filled-gap paradigm to examine pre- and post-verbal filled-gap effects, both inside and outside of islands. Results revealed subject and object filled-gap effects indexed by an N400. In a condition containing a relative clause island, no object filled-gap effect (N400) emerged, suggesting that the parser avoided positing gaps within islands. At the actual gap site, P600s were yielded, indicating that the parser successfully completed the *wh*-dependency. These results support an account of the N400 as an index of prediction, and an account which argues that the P600 indexes not only ungrammaticality or syntactic reanalysis, but also successful syntactic processing.

Chapter 3: Processing of *wh*-dependencies by L2 learners

Introduction

The second part of this dissertation examines how Chinese-speaking learners of English process *wh*-dependencies, addressing two central debates in the L2 psycholinguistics literature regarding (i) whether learners can utilize grammatical knowledge in the course of processing a sentence, and (ii) whether predictive processing is possible for L2 learners. A large body of L2 sentence processing research has attempted to directly test predictions of the prominent L2 theory Shallow Structure Hypothesis (Clahsen & Felser, 2006). This hypothesis makes direct predictions for non-native sentence processing, and specifically makes claims about the type of information available to a learner in real time. The processing of *wh*-dependencies provides a strong test case for this proposal, as it is possible to examine whether learners, like native speakers, are sensitive to grammatical information during *wh*-dependency formation, or whether learners instead employ a “shallow” processing mechanism, relying instead on semantic/pragmatic information.

There is also much evidence to suggest that native speakers predict during language processing, including in this dissertation in Chapter 2. During the processing of *wh*-dependencies in particular, it has been argued that native speakers actively predict upcoming syntactic structure in order to try to resolve the dependency (e.g., Clifton & Frazier, 1989; Frazier, 1987; Frazier & Clifton, 1989; Frazier & Flores D’Arcais, 1989). However, a debate continues in SLA regarding whether adult learners similarly predict during comprehension generally, as well as during *wh*-dependency formation. One SLA theory at the forefront of this debate proposes that L2ers have a reduced ability to generate expectations as compared to natives, and thus predicts that L2 learners are unlikely to show effects of prediction during processing (e.g., Grüter et al., 2012, 2017). In contrast, several researchers have proposed that predictive processing is possible, but that the ability to predict may be modulated by a range of factors, including proficiency and cognitive abilities (Hopp, 2013; Kaan, 2014). The current study examines L2 prediction in the processing of *wh*-dependencies in two contexts, shedding new light on this debate by examining whether the type of information available in the sentence influences an L2ers' ability to predict. We also examine how individual differences in proficiency impact L2 prediction.

To address these two SLA debates, this study builds on a body of psycholinguistics literature investigating the processing of *wh*-dependencies in natives and learners using a filled-gap paradigm. Using EEG enables us to examine whether natives and learners use qualitatively similar processing mechanisms and whether processing unfolds on a similar time-course, allowing for a more precise comparison between the two populations and shedding new light on the possibilities and limitations of adult L2 acquisition. Additionally, this study stands to be one of the first L2 ERP studies to not rely on a violation paradigm, allowing us to examine sentence processing under more natural circumstances, as opposed to examining simply whether learners can detect grammatical anomalies.

The structure of Chapter 3 is as follows. First, studies examining island sensitivity in L2 learners are reviewed, with a focus on research testing predictions of the Shallow Structure Hypothesis (Clahsen & Felser, 2006). The L2 prediction debate is examined next, and studies investigating prediction in the processing of *wh*-dependencies are briefly summarized. The current study is then introduced, and specific predictions for each critical region are outlined; this study builds directly on the study of native speakers reported Chapter 2. Following the predictions, the L2 results are summarized, and the chapter concludes with a discussion that addresses the two SLA debates and compares native and non-native findings.

Literature Review

Island sensitivity in L2 learners

For the past 30 years, many studies in the field of second language acquisition investigated whether adult L2 learners can acquire morphosyntactic features that are not present in their native language, addressing a debate between ‘representational’ and ‘full access’ theories. Representational accounts posit a syntactic deficit in late L2 learners (e.g., Hawkins, 2009; Hawkins & Chan, 1997; Hawkins & Casillas, 2008; Tsimpli & Dimitrakopoulou, 2007). Proponents of representational accounts argue that there is a critical period for the acquisition of abstract features, such that after the close of the critical period, features which are not present in the inventory of the learner’s L1 cannot be acquired to native-like levels. In contrast, a second group of theories known as a full access accounts argue that features absent from the L1 grammar can ultimately be acquired by adult learners to native-like levels (e.g., Lardiere, 2009; Prévost & White, 2000; Schwartz & Sprouse, 1994, 1996).

The present study examines acquisition of island constraints, investigating whether late L2 learners whose L1 grammar does not select for syntactic island constraints (i.e., the [+*wh*] feature) have access to such features. Many studies directly investigating this question have examined sensitivity to island constraints using offline acceptability judgments, testing adult L2 learners whose native language does not instantiate overt *wh*-movement. Findings in this literature are somewhat mixed, with some researchers arguing that L2 acquisition is constrained by a critical period, based on evidence that late learners perform differently than native speakers in rejecting sentences containing island violations (e.g., Johnson & Newport, 1991; Hawkins & Chan, 1997; Schachter, 1990), and others showing that late learners are indeed sensitive to island violations (e.g., Aldosari, 2013, 2015; Li, 1998; Martohardjono, 1993). In one recent study that addressed this debate, Aldosari (2015) examined whether native speakers of Najdi Arabic, a *wh*-in-situ language, show sensitivity to island constraints on *wh*-movement. Aldosari built directly on Sprouse et al. (2012)[6], testing a large sample of native English speakers and Arabic learners of English in an offline acceptability task targeting four island types (*whether* islands, complex noun phrase islands, subject islands, and adjunct islands). The experiment implemented a 2×2 design crossing the factors Island and Dependency Length. In (21) below, (repeated from (20) above), an example set containing an adjunct island is shown.

(21) NON-ISLAND / MATRIX
The helpful worker thinks that the boss left her keys in the car.
a. Which worker __ thinks that the boss left her keys in the car?

NON-ISLAND / EMBEDDED
The worker thinks that the boss left her office keys in the car.
b. Which keys does the worker think that the boss left __ in the car?

ISLAND / MATRIX
The helpful worker worries if the boss leaves her keys in the car.
c. Which worker __ worries [if the boss leaves her keys in the car]?

ISLAND / EMBEDDED
The worker worries if the boss leaves her office keys in the car.
d. *Which keys does the worker worry [if the boss leaves __ in the car]?

Aldosari replicated Sprouse et al.’s findings, showing that native English speakers rated the ungrammatical island violation sentence (21d) the lowest of the four conditions. This was also

---

[6] Aldosari’s (2015) modifications of Sprouse et al. (2012) are discussed in Chapter 2, Methods, “Offline Acceptability Judgment Task.”

true of the Arabic learners, who showed similarly low acceptability judgments as the native speakers. This finding was taken to suggest that the Arabic learners had successfully acquired syntactic island constraints in their L2.

Like Sprouse et al. (2012), Aldosari (2015) also investigated the nature of islands by examining the relationship between island sensitivity and processing resources. As outlined in Chapter 2, probing the role of individual differences in processing capabilities can help tease apart predictions of the grammatical account and processing account of islands. Under a resource-limitation account of island effects, low acceptability judgment violation sentences are a consequence of a processing burden posed by the island structure in the embedded clause. As Sprouse et al. outline, such an account would likely predict that individuals with increased processing resources should be able to overcome the processing burden and posit a gap inside an island, yielding higher acceptability ratings in this condition. Aldosari tested this prediction by including a measure of working memory capacity (operation span task) in his study. Crucially, no relationship emerged between working memory and acceptability ratings for either the native speakers or the learners, suggesting that for both groups, island sensitivity is grammatically-guided rather than related to limited processing resources.

A related body of literature in SLA regards whether L2 learners show sensitivity to island constraints during online processing, which is the focus of the current study. The Shallow Structure Hypothesis makes direct claims regarding the role of syntactic constraints in L2 processing (Clahsen & Felser, 2006). This hypothesis is in line with representational accounts in arguing that late learners are unable to fully acquire the grammar of their L2, and makes the additional claim that L2 sentence processing is qualitatively different than native processing. In contrast to representational accounts, however, the Shallow Structure Hypothesis predicts that all adult L2 learners regardless of their L1 process sentences non-natively. Specifically, this hypothesis proposes that L2 learners are unable to utilize abstract syntactic information during processing and instead must rely on non-syntactic information, such as semantic or pragmatic information, to resolve syntactic dependencies (e.g., Felser & Roberts, 2007; Marinis et al., 2005). This makes straightforward predictions regarding how L2 learners should process *wh*-dependencies: if learners do not have access to abstract syntactic information during processing, then they will not be able to posit syntactic gaps. In contrast, if non-native speakers must rely on semantic/pragmatic information during processing, that the L2 processing of *wh*-dependencies should be verb-driven, similar to the mechanism proposed in the Direct Association Hypothesis (see Ch. 2; Pickering & Barry, 1991); the Shallow Structure Hypothesis specifically proposes that L2 learners do not posit gaps during processing but instead attempt to link thematic arguments, integrating the filler with a subcategorizing verb.

Clahsen and Felser (2006) present findings from several studies to support their proposal, but most relevant to the current study is Marinis et al. (2005), who built upon findings from a self-paced reading study by Gibson and Warren (2004) to examine whether L2 English learners utilize *intermediate gaps* during the processing of filler-gap dependencies. It has been proposed that in sentences with long-distance cyclic movement, intermediate landing sites at clause boundaries provide an intermediate gap site where the filler can be reactivated (Chomsky, 1973). For example, (22a) below involves *wh*-extraction from a complement clause, and it has been proposed that there is an intermediate gap at the clause boundary (e.g., after *inspired*). In (22b), however, no intermediate gap is present because the *wh*-extraction occurs across a complex NP.

(22) a. The actress *who* the journalist suggested __ that the talented writer had **inspired** __ will go on stage tonight.

b. The actress *who* the journalist's suggestion about the talented writer had **inspired** **__** will go on stage tonight.

Gibson and Warren (2004) provided evidence that native English speakers indeed posit gaps at intermediate landing sites, showing faster reading times at the actual gap site (e.g., *inspired*) for the intermediate gap condition (22a) as compared when no intermediate gap was present (22b). This facilitation was taken to indicate that positing an intermediate gap (i.e., reactivating the filler) makes it easier to ultimately integrate the filler at the actual gap site.

Marinis et al. (2005) tested native English speakers and L2 learners from a variety of L1 backgrounds (Greek, German, Chinese, and Japanese). While the native English speakers replicated Gibson and Warren's (2004) findings, learners did not show evidence of positing an intermediate gap, showing no reading time difference at the actual gap site across conditions (22a-b). This was true across L1 groups, with learners whose L1 does not instantiate *wh*-movement (Chinese, Japanese) showing insensitivity to the intermediate gap structure similarly to learners whose L1 instantiates *wh*-movement (German, Greek). The lack of an intermediate gap effect is taken to suggest that late L2 learners build 'shallow' structures during processing (i.e., not positing syntactic gaps) in comparison to native speakers. This finding also led Clahsen and Felser (2006) to argue against a strong role for the L1 in L2 processing, predicting that regardless of the native language, all L2 learners are similarly unable to utilize abstract syntactic information in the course of processing (see also Williams, Möbius. & Kim, 2001).

The Shallow Structure Hypothesis also makes direct claims regarding whether learners show sensitivity to islands. Because this theory proposes that learners cannot use grammatical information in the course of processing *wh*-dependencies, learners should not show sensitivity to syntactic islands online. Psycholinguistic studies which have examined L2 learners' sensitivity to islands have reported mixed findings, with some studies observing that L2 learners are sensitive to island constraints online (e.g., Aldwayan et al., 2010; Cunnings, Batterham, Felser, & Clahsen, 2010; Johnson et al., 2016; Omaki & Schulz, 2011), and others which have questioned whether natives and non-natives process islands in a similar way (Boxell & Felser, 2016; Felser et al., 2012; Kim et al., 2015).

One study that reported findings in support of the Shallow Structure Hypothesis was Felser et al. (2012), who proposed that the crucial difference between native and non-native sensitivity to island constraints lies in the time-course of processing, with L2ers showing delays as compared to natives (see also Boxell & Felser, 2016). Felser et al. tested German-speaking learners of English in two eye-tracking experiments, one using a plausibility mismatch design (Exp. 1) and the other, a filled-gap paradigm (Exp. 2). In both experiments, Felser and colleagues examined whether L2 learners were sensitive to island constraints by examining whether natives and learners posited a gap within an island. Results for both experiments showed qualitative similarities in native and L2 processing, with both groups showing sensitivity to relative clause islands during online processing. However, important timing differences emerged between natives and L2 learners across experiments. In Experiment 1, which utilized a plausibility mismatch design (see (23) below), the German learners showed the predicted interaction between plausibility and presence of island in their first-pass reading times at the critical verb region (e.g., *read*). These results provide evidence that learners posited a gap in the non-island condition (23a), and correctly did not attempt to resolve the *wh*-dependency within the relative clause island in (23b). The native English speakers also showed evidence of island sensitivity, although the predicted interaction between plausibility and presence of an island emerged only at the spillover region (e.g., *extensively*) in natives’ regression path and rereading times. Thus, the learners showed evidence of island sensitivity earlier during processing than the native speakers, with the L2 effect emerging at the critical verb as compared to the spillover region, and captured by measures of earlier processing (i.e., first-pass reading times vs. rereading times). Felser et al. (2012) take these results to suggest that learners immediately utilize plausibility information during online processing of filler-gap dependencies, whereas native speakers do so at later stages of processing.

(23) a. Everyone liked the magazine/shampoo that the hairdresser **read** extensively and with such enormous enthusiasm about __ before going to the salon.

b. Everyone liked the magazine/shampoo that the hairdresser [who **read** extensively and with such enormous enthusiasm] bought __ before going to the salon.

In the filled-gap experiment, the predicted interaction (in this design, between *wh*-extraction × presence of island) emerged for the L2 group in rereading times in the spillover region following the target filled-gap region. This interaction demonstrated that learners attempted to resolve the *wh*-dependency in the non-island condition, and did not do so within the island. This effect also surfaced for the native speaker group, although crucially it emerged in first-pass reading times at the critical region. Felser et al. (2012) propose that this timing difference between natives and learners suggests that L2ers were delayed in utilizing syntactic structure information (i.e., an empty argument position) because effects emerged in the spillover region as well as in a relatively late eye-movement measure. Felser et al. (2012) argue that these findings support the Shallow Structure Hypothesis (Clahsen & Felser, 2012), pointing out that while learners were able to use semantic information to constrain early processing in the plausibility mismatch experiment (Experiment 1), they were delayed in doing so in the filled gap design (Experiment 2). This is broadly compatible with the Shallow Structure Hypothesis, which predicts that learners over-rely on semantic/pragmatic information during processing and fail to utilize abstract grammatical information on par with natives.

Omaki and Schulz (2011) also examined online sensitivity to island constraints in L2 learners, although they ultimately argued against claims of the Shallow Structure Hypothesis. Omaki and Schulz built on Traxler and Pickering's (1996) influential study (reported in Chapter 2) which used a plausibility mismatch manipulation. Spanish-speaking learners of English read sentences such as (24) below. In the 2×2 design, the filler was manipulated to be either a plausible argument of the first verb (e.g., *book*) or an implausible argument (e.g., *city*). The second factor manipulated was the presence of an island structure, which in this experiment was a relative clause. In (24c-d), the critical verb is located inside the relative clause, allowing Omaki and Schulz to examine whether learners attempt to posit a gap inside an island domain.

(24) NON-ISLAND / IMPLAUSIBLE
a. *The city* that the author **wrote** regularly about __ was named for an explorer.

NON-ISLAND / PLAUSIBLE
b. *The book* that the author **wrote** regularly about __ was named for an explorer.

ISLAND / IMPLAUSIBLE
c. *The city* that the author [who **wrote** regularly] saw __ was named for an explorer.

ISLAND / PLAUSIBLE
d. *The book* that the author [who **wrote** regularly] saw __ was named for an explorer.

In the self-paced reading experiment, a plausibility mismatch effect emerged for both native English speakers and Spanish L2 learners of English. At the spillover region (e.g., *regularly*) immediately following the critical verb, there were reading time slowdowns for the implausible non-island condition (24a) as compared to in the plausible non-island condition (24b). This effect demonstrates that natives and learners attempted to posit a gap at the verb *wrote* in the grammatically licit conditions. In the island conditions, on the other hand, no plausibility mismatch was in evidence for either the native English speakers or the Spanish learners: there were no reading time differences at the critical verb *wrote* or at the spillover region *regularly* in the two island conditions (24c-d). This demonstrates that natives and learners did not attempt to posit a gap inside the island, which suggests that learners, like native speakers, are able to utilize abstract syntactic information (i.e., island constraints) when processing filler-gap dependencies. Omaki and Schulz (2011) therefore argue against claims of the Shallow Structure Hypothesis that learners cannot rely on structural information during processing. This study also provides evidence that learners are not delayed their use of syntactic knowledge online, counter to Felser et al.'s (2012) findings.

Kim et al. (2015) raise the possibility that sensitivity to island constraints online may be dependent on the learners' native language (L1). Their study tested Spanish and Korean-speaking learners of English who were matched on their L2 English proficiency. Importantly, while Spanish is a language which instantiates overt *wh*-movement, Korean does not exhibit overt *wh*-movement and therefore does not instantiate island constraints in the same way English does. Results from an offline grammaticality judgment task showed that both L2 groups displayed sensitivity to island violations, rating sentences in the ungrammatical island condition significantly lower than those in a grammatical condition. To examine online island sensitivity, Kim et al. also included a stop-making-sense task, adapted from materials used by Traxler and Pickering (1996), containing a plausibility mismatch manipulation. In this task, participants read sentences one region at a time and pressed the spacebar to reveal the subsequent region. Participants were instructed to press a stop key at the point in the sentence where they felt it no longer made sense. This task provides two dependent measures, implausibility detection rates and reading times, that can reveal whether readers show a plausibility mismatch effect and attempt to posit a gap both in non-island, grammatically licit positions, as well as within an island, an unlicensed position.

Results for the implausibility detection rates (i.e., whether a participant pressed the 'stop-making-sentence' button) showed that at the critical region, all groups showed the crucial Plausibility × Island manipulation. Follow-up tests revealed that participants were more likely to press the stop key in the implausible as compared to the plausible sentence, in the non-island condition. In the condition containing a relative clause island, there was no significant difference in stop rates between the implausible and plausible sentences. Thus, Spanish and Korean L2 learners, like native English speakers, were sensitive to the plausibility manipulation, attempting to posit a gap crucially only in grammatically licensed positions and not inside an island. However, reading-time results showed a slightly different pattern. While the Spanish learners patterned with the native English group, showing no plausibility mismatch effect (i.e., reading time slowdown) within the island domain, the Korean learners, whose L1 does not instantiate overt *wh*-movement, did not show the expected Plausibility × Island interaction. In contrast, a main effect of Plausibility indicated that the Korean learners attempted to posit a gap both in the predicted non-island condition, and inside the island domain as well.

Kim et al. (2015) acknowledge that their findings are more in line with Felser et al.'s (2012) proposal about the time-course of L2 sentence processing. Note that Korean speakers demonstrated island sensitivity in the offline grammaticality judgment task as well as in the stop-make-sense judgments. In reading times, which provide a better measure of the initial stages of processing, Korean learners did not appear to be able to utilize structural information about the relative clause to constrain processing, similar to the delay Felser et al. reported in their filled-gap experiment. Unlike Felser et al., who tested German learners of English, Kim et al. suggest that *wh*-dependency resolution may be modulated by the L1, in that only learners whose L1 exhibits overt *wh*-movement may show sensitivity to islands online. Note that this proposal differs from the Shallow Structure Hypothesis, which expects a lack of sensitivity to islands for all non-native speakers regardless of L1 (Clahsen & Felser, 2006). Thus, the role of the L1 in the processing of *wh*-dependencies remains an important open question.

One recent study is uniquely poised to address several of the proposals raised here, including the Shallow Structure Hypothesis (Clahsen & Felser, 2006), claims about L2ers' delay of island sensitivity (e.g., Boxell & Felser, 2016; Felser et al., 2012), and proposals regarding the role of the L1 (e.g., Kim et al., 2015). In a self-paced reading study, Johnson, Fiorentino, and Gabriele (2016) tested advanced Korean learners of English using a filled-gap design. Johnson et al. examined the processing of *wh*-dependencies in two types of sentences, those that did not contain islands, as in (25) below, and sentences which contained a relative clause island shown in brackets in (26).

(25) a. The instructor wondered *who* **Chris** will film **Tom** with **__** at the reception.

b. The instructor wondered if **Chris** will film **Tom** with Susan at the reception.

(26) a. My father asked *who* **the actress** [that married **Tyler** last summer] kissed **__** during the rehearsal.

b. My father asked if **the actress** [that married **Tyler** last summer] kissed the director during the rehearsal.

To examine island sensitivity, the filled object position was examined. In the non-island condition (25), the filled object position (*Tom*) is a potential gap site, whereas in the island condition (26), the object position is located inside of the relative clause island. Importantly, the

Shallow Structure Hypothesis predicts that because of learners' reliance on semantic/pragmatic information, learners may appear to resolve *wh*-dependencies in a post-verbal object position by associating the *wh*-element with a subcategorizing verb in the assignment of thematic roles. Thus, post-verbal object filled-gap effects for L2 learners are possible according to a shallow processing account, and may appear in both grammatically licit and illicit positions.

Results for both the native English speakers and Korean learners revealed a significant object filled-gap effect for the Non-Island sentences (25), with reading time slowdowns at the critical filled object region (e.g., *Tom*). No filled-gap effect emerged for either group within the relative clause island in (26), indicating that learners, like natives, did not posit a gap within the relative clause island. In line with Omaki and Schulz (2011), Johnson et al.'s findings are inconsistent with claims of the Shallow Structure Hypothesis regarding island sensitivity. Johnson et al. also did not find that learners were delayed in utilizing syntactic knowledge compared to native speakers, with effects emerging in the same time window for both groups. It is also important to note that, in testing L1 Korean L2 English learners, Johnson and colleagues are able to test Kim et al.'s (2015) proposal about the role of the L1. Although in their study, Kim et al. found that Korean learners showed evidence of gap positing inside an island, the Korean learners in Johnson et al.'s study did not yield reading time slowdowns in the island. This difference between L2 performance may be explain by several factors, including the different tasks used in the two studies (stop-making-sense task, self-paced reading) and/or differences in L2 proficiency of the Korean learners, which was measured with different tests (Kim et al., 2015: cloze test; Johnson et al., 2016, Michigan Listening Comprehension Test).

A second region in the sentence that Johnson et al. (2016) examined was the embedded subject position. As discussed in Chapter 2, subject filled-gap effects provide evidence that gap positions are anticipated using syntactic knowledge, although studies examining whether native English speakers yield subject filled-gap effects have thus far provided mixed evidence (e.g., Aldwayan et al., 2010; Canales, 2012; Johnson, 2015; Lee, 2004; Stowe, 1986). Johnson et al. included a filled subject position in both the non-island (*Chris*) and island (*the actress*) conditions, which served as a licit potential gap site and allowed the researchers to probe pre-verbal gap prediction. The Shallow Structure Hypothesis predicts that because L2 processing is verb-driven, learners should not show filled-gap effects at the subject position because at that point in the sentence, the verb has not yet been encountered. Thus, subject filled-gap effects provide insight into whether L2ers are able to generate structural predictions about gap sites prior to encountering the verb. Results for this analysis showed that only the native speakers showed a subject filled-gap effect, which was only significant in the island condition. Given that few studies have reported consistent subject filled-gap effects in native English speakers, Johnson et al. (2016) note that the asymmetrical emergence of the subject filled-gap effect for the natives is in line with these mixed findings.

Examining whether L2 learners show subject filled-gap effects can also be informative with regards to a second debate focusing on the role of prediction in L2 processing. This literature is briefly reviewed in the following section.

Predictive Processing in L2 Learners

Although a body of psycho- and neurolinguistic research has suggested that native speakers engage in predictive processing in a variety of contexts, the role of prediction in L2 processing is still under debate. Grüter and colleagues (Grüter et al., 2012, 2017) have proposed that adult L2 learners have a *Reduced Ability to Generate Expectations* (RAGE Hypothesis) compared to native speakers, such that learners are unlikely to show effects of prediction during language processing in general. Indeed, in the L2 psycho- and neurolinguistics literature several studies have failed to show predictive effects for L2 learners in contexts in which native speakers predict (e.g., Lew-Williams & Fernald, 2010; Grüter et al., 2012, 2017; Martin et al., 2013; Mitsugi & MacWhinney, 2016). However, several studies have provided evidence of native-like predictive processing in L2 learners (e.g., Foucart, Martin, Moreno, & Costa, 2014; Hopp, 2013; Johnson, 2015; Leal, Slabakova, & Farmer, 2017). Some researchers have therefore suggested that the ability to predict in the L2 may be dependent on a range of factors, including proficiency and individual differences in cognitive abilities (Hopp, 2013; Kaan, 2014). That is, it may be the case that only learners with advanced proficiency and sufficient cognitive resources will show evidence of prediction.

Although L2 prediction has been investigated in a variety of domains, including examining whether learners can predict using grammatical gender (e.g., Grüter et al., 2012; Hopp, 2013), case marking (e.g., Mitsugi & MacWhinney, 2018), and discourse information (e.g., Grüter et al., 2017), relatively few studies have focused on syntactic prediction in particular. The filled-gap design used in Johnson et al. (2016) and the present study is particularly useful for investigating not only *whether* learners predict during *wh*-dependency formation, but is also providing insight into the *type* of information that a learner can use to predict. Subject filled-gap effects, for example, provide insight into whether L2ers are able to generate structural predictions about gap sites prior to encountering the verb. That is, this position can reveal whether learners are able to utilize syntactic information (i.e., the presence of a *wh*-element) to generate a prediction that a gap is forthcoming. Thus, subject filled-gap effects provide a clear method to address the debate as to whether predictive processing in L2 learners is possible. There have been relatively few investigations of this phenomenon in the L2 literature, however. As reviewed previously, there is mixed evidence in the native processing literature with respect to whether subject filled-gaps emerge (Johnson, 2015; Lee, 2004; Stowe, 1986). To our knowledge, there is only one L2 study which has directly examined this issue, Johnson (2015).

Johnson's (2015) self-paced reading experiment was described in detail in Chapter 2, as it provides an important contribution to the native processing literature as one of the first large scale (n=110) studies taking a systematic approach to examining subject filled-gap effects. A primary research question for the study was to examine whether Korean-speaking learners of English (n=100) would show evidence of predicting gaps in the subject position. At the filled subject position *Diana* in (27) below, the native English speakers yielded a reading time slowdown in the *wh*- extraction sentence (27a) as compared to the baseline no extraction sentence (27b). This slowdown was also in evidence for the Korean learners of English, suggesting that both groups predicted a gap immediately after encountering the *wh*-filler.

(27) a. The principal questioned *who* **Diana** will put the girl near **__** for the exam.
b. The principal questioned if **Diana** will put the girl near for the exam.

These findings provide some of the strongest evidence that L2 learners engage in gap prediction pre-verbally. Importantly, subject filled-gap effects reveal that learners can utilize syntactic information to engage in predictive processing, contrary to proposals which suggest that L2 predictive processing is limited (e.g., Grüter et al., 2017) or that L2 sentence processing is 'shallow' or verb-driven (e.g., Clahsen & Felser, 2006). The current study builds on Johnson

(2015) to test for subject filled-gap effects using the EEG. In the following section, ERP studies examining L2 processing of filler-gap dependencies are reviewed.

Neurolinguistic approaches examining L2 processing of filler-gap dependencies

Although there is an active research agenda in L2 psycholinguistics examining the processing of *wh*-dependencies, there are relatively few studies which have utilized electrophysiological approaches. Two recent EEG studies have examined L2 processing of filler-gap dependencies using a plausibility mismatch paradigm. For example, Dallas, DeDe, and Nicol (2013) built on work by Garnsey et al. (1989), who measured ERP responses to sentences which either contained a filler that was a plausible object of a verb or an implausible one. Garnsey et al. reported an N400 effect at the critical verb when the filler was an implausible direct object, indicating that the native English speakers had difficulty semantically integrating the filler with the verb. Dallas et al. investigated whether Chinese learners of English would yield also yield an N400 effect in their study, in which participants read sentences like (28) below.

(28) a. The umpire asked which *player* the coach **threatened __** before the game.
b. The umpire asked which *football* the coach **threatened __** before the game.

Dallas et al. (2013) reported a significant N400 at the verb (e.g., *threatened*) for the native English speakers, replicating Garnsey et al. (1989), but no significant effects emerged for the Chinese learners. This was true within the N400 time-window examined (300-450 ms), as well as in later time windows (450-700 ms, 700-900 ms). Thus, the L2 learners did not show evidence of attempting to integrate the filler with the verb. However, in a follow-up analysis, L2 participants' English proficiency scores (collapsed across three measures) were shown to be significantly correlated with the magnitude of the ERP effect in the 300-450 ms time window. Individuals with higher English proficiency showed increasingly negative ERP effects, or larger N400 effects. Dallas and colleagues argue that this proficiency effect suggests that native-like processing of filler-gap dependencies is possible with increasing L2 proficiency.

Building on Dallas et al. (2013), Jessen and Felser (2018) utilized a similar plausibility manipulation to examine whether native speakers of English and German-speaking learners of English would yield an N400 effect when attempting to integrate an implausible filler at the verb (e.g., *built*), as in (29b) below. One major difference in Jessen and Felser's stimuli is that the sentences were ultimately plausible in both conditions. For example, in (29) below, the actual gap site after the preposition (*for*) renders the verb phrase plausible for both fillers.

(29) a. Bill liked *the house* that Bob **built** some ornaments for __ at his workplace.
b. Bill liked *the women* that Bob **built** some ornaments for __ at his workplace.

Both natives and learners yielded an N400 at the target verb region, although this effect was left-lateralized and longer lasting in the native speakers (300-650 ms), compared to the L2 learners, whose effect emerged later (400-550ms) and peaked in right frontal electrodes. Despite the differences in scalp distribution between the groups, Jessen and Felser (2018) interpret both negativities as an N400. Thus, in contrast to Dallas et al. (2013), who did not find a group-level N400 effect for their Chinese L2 learners, Jessen and Felser report a plausibility mismatch N400 effect for German L2 learners, suggesting that both natives and learners evaluate the plausibility of the filler upon encountering a potential subcategorizing verb. One possibility regarding why an N400 effect only emerged for highly proficient learners in Dallas et al.'s study is the role of the L1 – whereas Chinese does not instantiate overt *wh*-movement, German is closely related to

English and does exhibit *wh*-movement. Jessen and Felser dismiss this possibility, citing results from Williams et al. (2001), who found plausibility mismatch effects for Chinese, Korean, and German L2 learners of English. Instead, the authors attribute the L1 difference to the high proficiency of their German learners.

In addition to the plausibility mismatch paradigm, several recent studies have also implemented a filled-gap design to examine L2 processing of *wh*-dependencies. In the native processing literature reviewed in Chapter 2, work by Hestvik and colleagues was mentioned as one of only a few studies investigating filled gap effects using ERPs (Hestvik et al., 2007, 2012). Two subsequent L2 studies have since attempted to replicate Hestvik et al. utilizing the filled-gap paradigm. For example, in a master's thesis, Schremm (2012) tested 14 Swedish learners of English in an auditory EEG experiment where participants heard sentences such as (30) below, designed based on Hestvik et al.'s (2007) stimuli. The filled gap condition (30a) crucially has an extra argument in the direct object position of the main verb; the sentence in this condition is ultimately ungrammatical. The grammatical object sentence (30b) is the control condition which also contains a filled object position, but does not involve a filler-gap dependency and is fully grammatical.

(30) a. FILLED GAP
*The receptionist that the painter scared **the reporter** by accident answered the phone.

b. GRAMMATICAL OBJECT
The receptionist said that the painter scared **the reporter** by accident and then answered the phone.

At the critical object region (e.g., *the reporter*), a significant positivity emerged between 850-1000 ms for the filled gap condition (30a) in comparison to the grammatical object condition

(30b). Following Hestvik et al. (2007), Schremm (2012) interprets this late positivity as a P600 response associated with syntactic reanalysis and repair after learners attempt to integrate the filler noun phrase at the filled object position. A related possibility is that the P600 emerging at the filled gap reflects a phrase structure violation yielded by an attempt at positing a gap in a position that is already filled with a noun phrase, resulting in two immediately adjacent NPs. It is important to further consider that in Schremm and Hestvik et al.'s stimuli, sentences in the filled gap condition are ultimately ungrammatical and are not 'rescued' by an actual gap site later in the sentence (as they are in the current study). Thus, throughout the experiment participants could become aware that encountering a filled object gap always renders the sentence ungrammatical, and if this is the case, learners may no longer continue to try to search for potential gaps throughout the sentence or integrate the *wh*-filler at potential gap sites.

In a follow-up study, Schremm (2013) conducted the same experiment with native English speakers, who also yielded a P600 response at the object filled gap site. The onset of the positivity for natives emerged earlier, beginning at 650 ms as compared to 850 ms for the Swedish learners. Based on this finding, Schremm suggests that syntactic processing may be delayed in learners as compared to natives, which is compatible with predictions of Felser et al. (2012). Schremm notes, however, that native speakers and L2 learners appeared to use qualitatively similar parsing mechanisms in the processing of filler-gap dependencies, which is incompatible with the Shallow Structure Hypothesis (Clahsen & Felser, 2006).

A dissertation by Dong (2014) also built on Hestvik et al. (2007), testing Chinese speaking learners of English in an auditory EEG experiment in which participants heard sentences like (31) below. Unlike in Schremm (2012, 2013), the baseline sentence in (31b) also contains filler-gap dependency, such that both the target filled gap sentence in (31a) and the comparison condition (31b) involve a gap search. In the target ungrammatical sentence (31a), the verb *kissed* is a potential subcategorizer for the animate filler (e.g., *the zebra*), making the critical object region a potential gap site. A filled-gap effect is expected for the target condition (31a) only, and not for the inanimate version (31b), in which the filler-gap dependency involves adjunct extraction.

(31) a. UNGRAMMATICAL (FILLED GAP)
*The zebra that the hippo kissed **the camel** on the nose ran far away.

b. ADJUNCT
The weekend that the hippo kissed **the camel** on the nose **__**, it was humid.

At the filled gap region, a central negativity emerged for the Chinese learners of English in the filled gap condition as compared to the baseline adjunct sentence, lasting from 250-1000 ms and peaking in the 500-600 ms time window. A late, posterior positivity also emerged in the 900-1000 ms time window. Dong (2014) argues that the negative ERP effect is best characterized as a N400 effect, whereas the positive effect emerges too late to be categorized as a P600. Note, however, that the P600 emerging in Schremm (2012) was similarly late, becoming significant in the 850-1000ms time window. Dong proposes then that the presence of an N400 and the absence of a P600 suggest that the Chinese L2 learners processed the filler-gap dependency using semantic rather than syntactic information, and ultimately argues in support of the Shallow Structure Hypothesis. Further support for this claim comes from a separate proficiency analysis, which revealed that increasing L2 proficiency was not associated with increasingly P600-like effects; instead, higher proficiency was related to a larger N400 effect. Dong compares this L2 pattern of results to Hestvik et al.'s (2007) findings for native English speakers (Dong did not test native speakers), noting that if natives and L2ers employed qualitatively similar processing mechanisms, higher proficiency should be linked with increasingly native-like ERP effects (P600).

One final L2 ERP study to employ a filled-gap paradigm was conducted by Jessen et al. (2018). In their experiment, the filled gap region was realized as a resumptive indirect object phrase, as in (32a) below. As in Schremm (2012) and Dong (2014), the filled gap (here, the resumptive phrase) rendered the sentence ungrammatical. The baseline comparison sentence in (32b) did not contain a filler-gap dependency but rather included an adjunct clause (e.g., after *while*).

(32) a. *Sarah tickled *the monkey* for which Peter arranged some classes **for it** after the vacation.

b. Sarah tickled the monkey while Peter arranged some classes **for it** after the vacation.

Native English speakers and upper-intermediate German-speaking L2 learners of English heard sentences like (32). At the critical filled-gap region, both groups yielded a late positivity emerging at 600-800ms. Thus, both groups showed a P600 response, and this response emerged on the same time-course for both natives and learners. Jessen et al. (2018) argue that there are multiple possible interpretations of the P600 at the filled gap. One possibility is that participants interpret the indirect object prepositional phrase (e.g., *for it*) as resumptive, with the filler being the antecedent. Under this interpretation, the P600 would be related to a successful syntactic integration process, as the sentence would be interpretable. On the other hand, in line with studies examining processing at a direct object filled gap, the P600 emerging in Jessen et al. may reflect a phrase structure violation resulting from the parser dealing with two noun phrases. The interpretation of the P600 therefore remains an important open question.

The current study builds on L2 ERP studies investigating the processing of filler-gap dependencies using a filled-gap manipulation (e.g., Dong, 2014; Jessen et al., 2018; Schremm, 2012). We additionally investigate L2 island sensitivity using ERPs, examining whether learners utilize grammatical knowledge to constrain online processing. Before turning to the present study, it is worth mentioning that to our knowledge, no ERP study has directly examined the processing of islands in L2 learners. Island sensitivity has been a focus in the psycholinguistics literature, as it provides a strong test case of proposals like the Shallow Structure Hypothesis. Examining the online processing of islands using ERPs is valuable as it is possible that offline acceptability judgments do not capture variability that may emerge during the course of processing the island itself. ERPs allow for a precise measurement of the time-course of processing, given that it is not related to a behavioral response on each word (as in self-paced reading) and it unfolds in real-time. ERPs also allow us to draw inferences about the qualitative processing mechanisms employed during processing based on the timing or distribution of the waveform. Additionally, the majority of L2 ERP studies have relied on a ‘violation paradigm’ where ERP responses are measured to syntactic or semantic anomalies. Roberts and colleagues (2018) point out that this has important consequences for interpreting native and L2 processing, given that natural language comprehension rarely involves encountering such anomalies. The current study is one of the first to use fully grammatical sentences throughout the experiment, allowing us to examine language processing under more natural circumstances.

Current Study

The current study tracks the dynamics of language processing *wh*-dependencies across a grammatical sentence, taking a systematic approach to examine brain responses at three critical regions which allow us to examine gap prediction both inside and outside of island and *wh*-dependency resolution. Building on the native speaker study, we use distinct ERP components, the N400 and P600, to examine these processes. A primary goal of the L2 investigation is to address two critical debates regarding the role of grammatical knowledge and prediction in non-native sentence processing.

Before turning to our specific research questions, we return to the issue of the role of the L1 in the processing of *wh*-dependencies. The current study tests Mandarin Chinese-speaking learners of English. Mandarin Chinese is a language which does not instantiate overt *wh*-movement (Sohn, 1980; Sohn, 1999), thus allowing us to address proposals regarding the role of the transfer in processing of *wh*-dependencies and island sensitivity. For example, the Shallow Structure Hypothesis proposes that L1 background does not influence the state of L2 sentence processing, such that all learners regardless of language background are similarly unable to utilize abstract syntax during processing (e.g., Clahsen & Felser, 2006; Marinis et al., 2005). In contrast, Kim et al. (2015) reported L1 differences in their study, with Korean learners of English showing evidence of attempting to posit a gap within an island. Kim et al. suggest that only learners whose L1 instantiates overt *wh*-movement (e.g., Spanish learners) may show sensitivity to island constraints during the initial stages of processing, with learners whose L1 does not exhibit overt *wh*-movement showing delays in using syntactic information online. However, it is important to point out that certain phenomenon in *wh*-in-situ languages, such as topicalization, may arguably involve movement (e.g., Lin, 2006; Qu, 1994). Thus, it is difficult to completely rule out potential transfer of grammatical constraints from the L1 if L2 learners are successful. Nevertheless, given the differences between L1 groups that emerged in Kim et al. (2015), Mandarin Chinese-speaking learners present an interesting test case for whether native-like processing is possible.

The study has three main research questions:

RQ1: Are learners sensitive to island constraints in the processing of *wh*-dependencies?

RQ2: Is predictive processing possible for learners?

RQ3: Do learners successfully resolve *wh*-dependencies in a native-like way?

To address the first question regarding island sensitivity, we examine processing inside of a relative clause island, investigating whether learners, like native speakers, avoid positing gaps in islands. In Chapter 2, native English speakers were shown to predict a gap in only grammatically licit positions, yielding an N400. In contrast, in sentences containing islands, no gap was predicted within a relative clause island, and an N400 did not emerge. For L2 learners, the Shallow Structure Hypothesis predicts that learners are not able to use grammatical constraints on the same time-course as native speakers and thus, learners will not show sensitivity to island constraints, yielding an N400 in both grammatically licit and unlicensed/illicit domains. Kim et al. (2015) make a similar prediction because the L2 learners in the present study are native speakers of a language which does not instantiate overt *wh*-movement. In line with Felser et al. (2012), it is also possible that L2ers will demonstrate sensitivity, but will yield a delayed N400 in the grammatically licit conditions. In contrast, several researchers have argued that native-like processing is indeed possible for very advanced L2 learners even in cases where the L1 and L2 differ with respect to whether or not there is overt *wh*-movement, as has been shown in a range of self-paced reading studies examining *wh*-dependencies and island sensitivity (e.g., Aldwayan et al., 2010; Johnson et al., 2016; Omaki & Schulz, 2011). These accounts predict a native-like pattern of ERP results for advanced learners with success being potentially modulated by L2 proficiency.

To address the second research question regarding prediction, we test for subject filled-gap effects which can provide insight into whether L2ers are able to generate structural predictions about gap sites prior to encountering the verb. In line with our native data, if an N400 effect emerges at the pre-verbal filled subject position it would provide evidence that learners can make structural predictions in the processing of *wh*-dependencies as opposed to simply relying on a thematic relationship between the *wh*-item and the verb, as predicted by the Shallow Structure Hypothesis (Clahsen and Felser, 2006). Accounts which propose that predictive processing is limited in L2 learners would predict that an N400 would not emerge at the subject filled-gap position (Grüter et al., 2012, 2017; Martin et al., 2013). In contrast, if predictive processing in both natives and L2 learners is modulated by proficiency, as proposed by Hopp (2013) and Kaan (2014), a relationship between proficiency and N400 responses to the subject filled-gap subject position may emerge. These analyses will allow us to examine whether variability in predictive processing is modulated by L2 proficiency or whether predictive processing is simply beyond the capacities of L2 learners.

Finally, the third research question about the resolution of the *wh*-dependency is tested by examining processing at the actual gap site. We expect a P600 to emerge at the actual gap site if learners successfully resolve the *wh*-dependency, as was shown for natives in our study and elsewhere (e.g., Kaan et al., 2000; Phillips et al., 2005). This study is one of the first to examine successful *wh*-dependency resolution in L2 learners, and testing for completion of the dependency in sentences containing a relative clause island may be particularly informative as to how and whether learners successfully process a complex island structure.

Methods

Participants

The participants were 23 native speakers of Mandarin Chinese (7 males, mean age 24.7, range 18-36) recruited from the University of Kansas. All participants were right-handed, had normal or corrected-to-normal vision, provided informed consent to participate. All participants considered themselves native speakers of Mandarin, although many participants also spoke another dialect of Chinese, including Cantonese (n=2), Sichuan (n=2), Henan, Ningbo, Wu, Fuyang, Taiwanese, Shanghaiese, Shandong, and Hebei.[7] Table 13 shows descriptive statistics for four main variables collected on a language background questionnaire. The first column shows the age at which the participants began English classes in China, considered to be the age of first exposure. The related variable, age of arrival, differed greatly, as all learners did not arrive to the U.S. until after puberty. The following column shows the number of years spent in the U.S., which was relatively low. Finally, the last variable shows participants' self-rating of their overall English proficiency; the mean rating of 3.9 corresponds to a rating of "somewhat good" on the questionnaire, with 5 corresponding to "very good" and 1 corresponding to "somewhat poor." The learners were also asked to estimate how often they used English in their

---

[7] With the current sample, it is difficult to directly investigate the influence of participants' knowledge of other Chinese dialects on their processing of *wh*-dependencies. With increasing sample size, we will address this question by including language factors into the statistical models, and/or grouping learners based on knowledge of a particular dialect. It will also be important to investigate whether it is the case that every Chinese dialect spoken by participants does not exhibit overt *wh*-movement, as in Mandarin; currently, published research on the syntax of each dialect reported above is limited, and we were unable to confirm the status of *wh*-movement in each language.

daily life. The majority of participants indicated that they used English daily (n=17) or a few times a week (n=5); one participant answered that they used English at least once a week.

Table 13: Descriptive statistics for L2 learners' background information.

	Age of first English exposure	Age of arrival to U.S.	Years in U.S.	Overall proficiency self-rating (0-5)
Mean	9.74	22.26	2.43	3.91
SD	2.96	4.36	2.27	0.51
Min	4	15	0	3
Max	15	36	6	5

All experimental methods and procedures were identical to those described in Chapter 2[8]. L2 learners took two independent measures of English proficiency after all experimental tasks were complete, described below in more detail.

Proficiency measures. The first measure is the Lexical Test for Advanced Learners of English (LexTALE; Lemhöfer & Broersma, 2012). The LexTALE uses vocabulary knowledge to assess language proficiency. The LexTALE stimuli include words ranging from high to low frequency, as well as orthographically well-formed and pronounceable nonwords. In the task, participants are instructed to identify whether the visually presented word is a real word in English or not. The final score takes into account responses to both words and nonwords, with correct and incorrect identifications. This measure is available online and takes approximately five minutes to complete. The second proficiency measure is a listening comprehension test that targets English grammar. The Examination for the Certificate of Proficiency in English from the

---

[8] Note that while the L2 learners also completed the cognitive tasks described in Chapter 2, the current study only reports results for the main EEG experiment and an exploratory analysis including scores from the offline acceptability judgment task due to the small sample size (n=23).

University of Michigan (2003) is widely used in L2 studies and takes approximately 5-10 minutes to complete. Results from the two English proficiency measures (LexTALE and Michigan Listening Comprehension Test) were collapsed into one composite variable given their conceptual relatedness. The average score on the composite proficiency measure was 81.23 out of a possible 100 points (*SD* = 5.38, min = 69.79, max = 90.90).

Data Analysis

EEG Analysis. The experimental design, data collection processes, and statistical approach for the L2 learner study was the same as in the native study reported in Chapter 2. For each region in the sentence, two overall models are reported which correspond to the two time windows for analysis, 300-500 ms and 500-900 ms. The following fixed factors and all possible interactions were included in the initial overall model: Extraction (no extraction, *wh*-extraction), Island (non-island, island), Hemisphere (left, midline, right), Anteriority (anterior, central, posterior), and Proficiency composite score. The models included Subjects as random intercepts. The model was progressively backwards-fit, and the final best-fitting model that fit the data are reported for below for each time window.

***Individual Difference Analysis*.** A second step of the analysis for the L2 learners included DD scores from the acceptability judgment task into the models for the object position. With this analysis, we directly examine the relationship between online and offline sensitivity to islands.

Behavioral Results

Behavioral results for the comprehension questions in the main EEG task reveal whether participants were successful in interpreting the complex sentences for meaning. Note that the comprehension questions did not target resolution of the *wh*-dependency, and were included to keep participants on task and ensure that accurate comprehension for meaning. Table 14 below shows accuracy on the comprehension questions for the target conditions and overall. A repeated-measures ANOVA with conditions Island (island vs. non-island) and Extraction (*wh*-extraction vs. no extraction) revealed a significant main effect of Island ($F(1, 22) = 5.823$, $p < .05$) which indicated that participants were more accurate responding to comprehension questions in the non-island conditions as compared to the island conditions. The main effect of Extraction was not significant ($F(1, 22) = 0.018$, $p = .894$), nor was the interaction between Extraction and Island ($F(1, 22) = .282$, $p = .601$). Together, this suggests that learners had difficulty with comprehension questions in sentences that contained a relative clause island, although crucially there was no difference in their accuracy in the *wh*-extraction vs. no extraction sentences in the island condition.

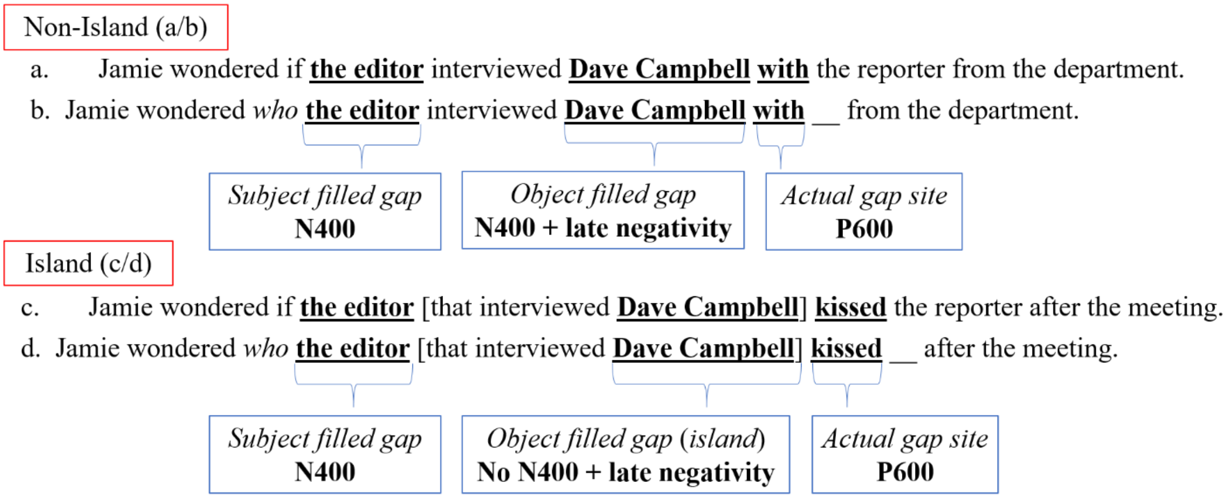
Table 14: Descriptive statistics for accuracy on comprehension questions across conditions.

Condition	**Mean**	**Standard Deviation**	**Minimum**	**Maximum**
Non-Island, No Extraction	66.96	15.79	40	100
Non-Island, *Wh*-Extraction	65.65	21.71	30	100
Island, No Extraction	73.48	17.99	30	100
Island, *Wh*-Extraction	75.65	20.85	30	100
Overall (including fillers)	76.11	9.98	60	92.14

EEG Results

Prior to reporting results from the EEG analysis, Figure 7 below repeats the target stimuli for reference, and includes a summary of the native findings in bold. We begin first with the subject filled gap region.

Figure 7: Target stimuli repeated with native English speaker findings summarized.



Non-Island (a/b)

a. Jamie wondered if **the editor** interviewed **Dave Campbell** **with** the reporter from the department.

b. Jamie wondered *who* **the editor** interviewed **Dave Campbell** **with** __ from the department.

Island (c/d)

c. Jamie wondered if **the editor** [that interviewed **Dave Campbell**] **kissed** the reporter after the meeting.

d. Jamie wondered *who* **the editor** [that interviewed **Dave Campbell**] **kissed** __ after the meeting.

Subject Position

If learners predict a gap in the subject position, then encountering a filled subject position (e.g., *the editor*) is expected to yield N400, as was found for the native English speakers. This would provide evidence that L2 learners anticipate gaps using syntactic knowledge, a finding which is mixed in the L2 psycholinguistic literature (e.g., Aldwayan et al., 2010; Canales, 2012; Johnson, 2015; Johnson et al., 2016). Subject filled-gap effects provide a strong test case of proposals which argue that learners are unable to utilize syntactic information in the course of processing, such as the Shallow Structure Hypothesis (Clahsen & Felser, 2006). We are also able to test predictions of the *RAGE Hypothesis*, which expects that learners are unlikely to show predictive effects on par with native speakers (Grüter et al., 2012, 2017). Both Shallow Structure and *RAGE* expect that learners will not show a subject filled-gap effect, yielding no N400 at this position.

300-500 ms time-window analysis. Due to the lexical overlap between the island and non-island conditions from the beginning of the sentence through the embedded subject position, analyses collapsed across the factor Island. The results for the best-fitting overall model are shown below in Table 15. The main effect of Extraction was significant, indicating that the amplitude at the subject position (e.g., *the editor*) was more positive the *wh*-extraction condition as compared to the no extraction condition. The interaction of Extraction × Hemisphere (left hemisphere level) was also significant, indicating the amplitude for electrodes in the left hemisphere were less positive than those at the midline. These effects reveal that in the 300-500 ms time window at the subject position (e.g., *the editor*), a positivity emerged for the *wh*-extraction condition, which was attenuated in the left hemisphere.

Table 15: Linear mixed-effects overall model for subject position from 300-500 ms.

Number of obs: 1426, groups: Subject, 23						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-1.702	0.308	45.3	-5.526	< .001	***
Extraction	0.658	0.164	1392	4.008	< .001	***
Hemisphere (left)	1.103	0.215	1392	5.136	< .001	***
Hemisphere (right)	1.145	0.215	1392	5.333	< .001	***
Anteriority (ant)	-0.253	0.198	1392	-1.276	0.202	
Anteriority (post)	1.788	0.217	1392	8.228	< .001	***
Extraction × Hemisphere (left)	-0.451	0.207	1392	-2.180	0.029	*
Extraction × Hemisphere (right)	0.000	0.207	1392	0.000	0.9999	
Hemisphere (left) × Anteriority (ant)	-0.373	0.246	1392	-1.514	0.130	
Hemisphere (right) × Anteriority (ant)	-0.694	0.246	1392	-2.819	0.005	**
Hemisphere (left) × Anteriority (post)	-0.509	0.273	1392	-1.861	0.063	.
Hemisphere (right) × Anteriority (post)	-0.326	0.273	1392	-1.192	0.233	

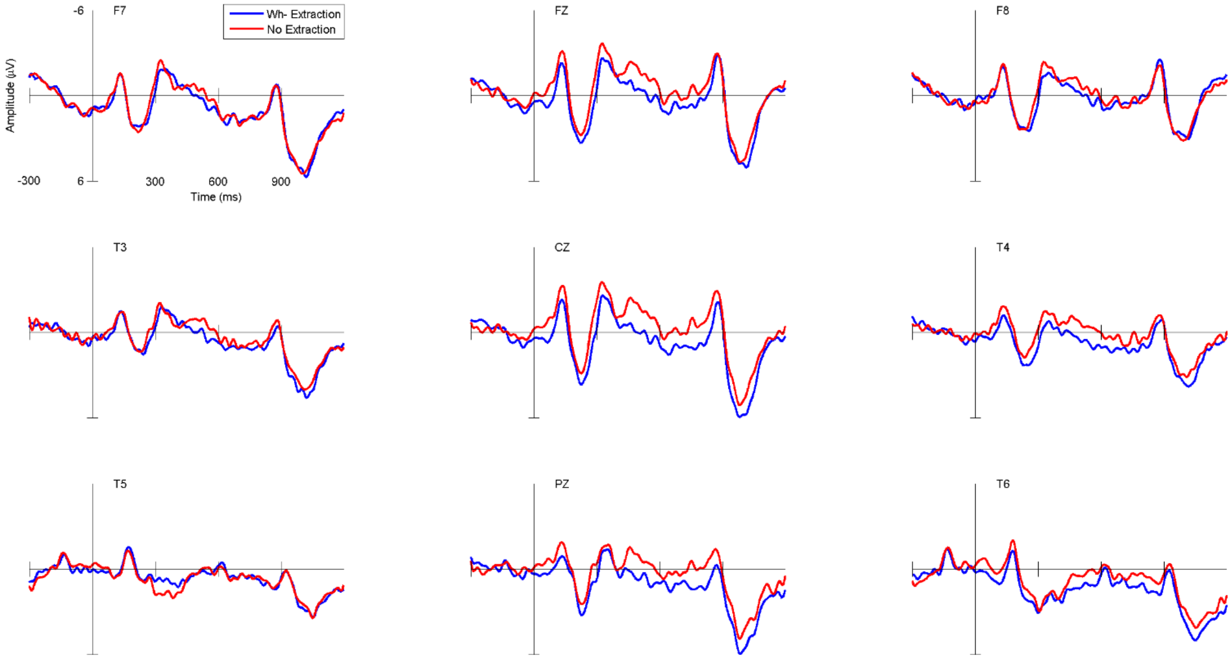
500-900 ms time-window analysis. The results from the overall model for the 500-900 ms time window are shown below in Table 16. The main effect of Extraction in this model was significant, indicating that the amplitude at the subject position (e.g., *the editor*) was more positive for the *wh*-extraction condition.

Table 16: Linear mixed-effects overall model for subject position from 500-900 ms.

Number of obs: 1426, groups: Subject, 23						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-0.322	0.312	43.6	-1.030	0.309	
Extraction	0.622	0.086	1394	7.268	< .001	***
Hemisphere (left)	0.551	0.206	1394	2.671	0.008	**
Hemisphere (right)	0.536	0.206	1394	2.601	0.009	**
Anteriority (ant)	-0.016	0.217	1394	-0.073	0.942	
Anteriority (post)	0.487	0.238	1394	2.046	0.041	*
Hemisphere (left) × Anteriority (ant)	0.059	0.270	1394	0.218	0.827	
Hemisphere (right) × Anteriority (ant)	-0.381	0.270	1394	-1.413	0.158	
Hemisphere (left) × Anteriority (post)	-0.505	0.300	1394	-1.684	0.092	.
Hemisphere (right) × Anteriority (post)	-0.316	0.300	1394	-1.056	0.291	

In summary, learners yielded an extended positivity, P600, at subject filled gap. In the earlier 300-500 ms, the P600 was attenuated in the left hemisphere, and in the later time window it was broadly distributed. Figure 8 below shows the subject filled-gap P600 effect. Note that the factor Proficiency did not become significant in either of the models, and did not significantly interact with the critical factor Extraction, indicating that proficiency scores did not modulate the P600 emerging at the subject position.

Figure 8: Subject filled-gap effect at representative electrodes.



Next, we turn to the analysis examining the filled object position, testing for effects of gap prediction in the licit, non-island condition and island sensitivity in the relative clause island condition.

Object Position

At the object position, we examine whether learners show sensitivity to island constraints in the online processing of *wh*-dependencies. The native English speakers yielded N400 in the non-island condition only, suggesting that the parser only engages in gap prediction when licensed by the grammar. Crucially, no N400 effect is expected inside of the island if learners are able to utilize grammatical knowledge in the course of processing *wh*-dependencies, in contrast to predictions of the Shallow Structure Hypothesis (Clahsen & Felser, 2006).

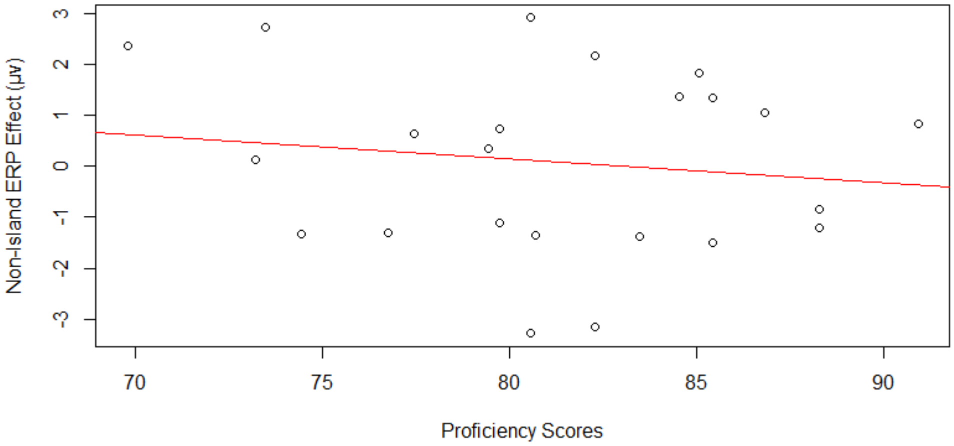
300-500 ms time-window analysis. The best-fitting overall model for the object position for the 300-500 ms time window was very large, and is reported in full in the Appendix B. Although the critical interaction between Extraction × Island was not significant (*t* (2803) = 0.665, *p* = .506), a higher-order interaction involving both Extraction and Island was significant: Extraction × Island × Anteriority (anterior level) × Proficiency (*t* (2803) = -2.291, *p* < .05). To investigate this interaction further, follow-up models for the island and non-island conditions were conducted separately.

Results for the best-fitting model for the non-island condition are shown in Table 17. The main effect of Extraction was not significant, indicating that there was no significant difference in amplitude between the *wh*-extraction condition as compared to the baseline no extraction condition. However, a significant interaction between Extraction × Proficiency emerged, indicating that with increasing proficiency, amplitude for the *wh*-extraction condition became more negative. In other words, learners with higher proficiency yielded an emerging negativity (N400) in the 300-500 ms time window at the filled object position (e.g., *Dave Campbell*). Figure 7 demonstrates this effect, plotting proficiency scores on the x-axis and the N400 effect size (*wh*-extraction minus no extraction) on the y-axis.

Table 17: Linear mixed-effects model for object position, Non-Island condition, from 300-500 ms.

Number of obs: 1426, groups: Subject, 23						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-1.891	0.377	25.8	-5.020	< .001	***
Extraction	0.088	0.091	1397	0.972	0.33097	
Anteriority (ant)	-0.809	0.106	1397	-7.605	< .001	***
Anteriority (post)	1.617	0.120	1397	13.479	< .001	***
Hemisphere (left)	0.124	0.120	1397	1.032	0.302	
Hemisphere (right)	0.898	0.120	1397	7.469	< .001	***
Proficiency	-0.054	0.069	21.7	-0.781	0.443	
Extraction × Proficiency	-0.046	0.017	1397	-2.684	0.007	**

Figure 9: Relationship between proficiency and Non-Island ERP effect size at the object filled gap from 300-500 ms.



Results for the best-fitting model for the island condition are shown in Table 18 below. The main effect of Extraction was removed during model-fitting, indicating it did not explain a significant portion of variance in the model. The interaction term Extraction × Proficiency was also removed during model-fitting, meaning that proficiency did not modulate processing at the filled object position inside the island in the 300-500 ms time window. Thus, no significant effects involving the critical factor Extraction emerged at the object position inside the island.

Table 18: Linear mixed-effects model for object position, Island condition, from 300-500 ms.

Number of obs: 1426, groups: Subject, 23						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-1.783	0.366	25.2	-4.871	< .001	***
Anteriority (ant)	-0.657	0.105	1397	-6.24	< .001	***
Anteriority (post)	1.620	0.119	1397	13.636	< .001	***
Hemisphere (left)	-0.147	0.119	1397	-1.238	0.216	
Hemisphere (right)	0.835	0.119	1397	7.008	< .001	***
Proficiency	0.005	0.068	22.5	0.079	0.938	
Anteriority (ant) × *Proficiency*	-0.078	0.020	1397	-3.89	< .001	***
Anteriority (post) × *Proficiency*	0.004	0.023	1397	0.175	0.861	

500-900 ms time-window analysis. Results from the overall best-fitting model for the 500-900 ms time-window are shown in Table 19. The crucial interaction of Extraction × Island was significant, and thus follow-up models investigating the factor Island were conducted separately for the island and non-island conditions.

Table 19: Linear mixed-effects overall model for object position from 500-900 ms.

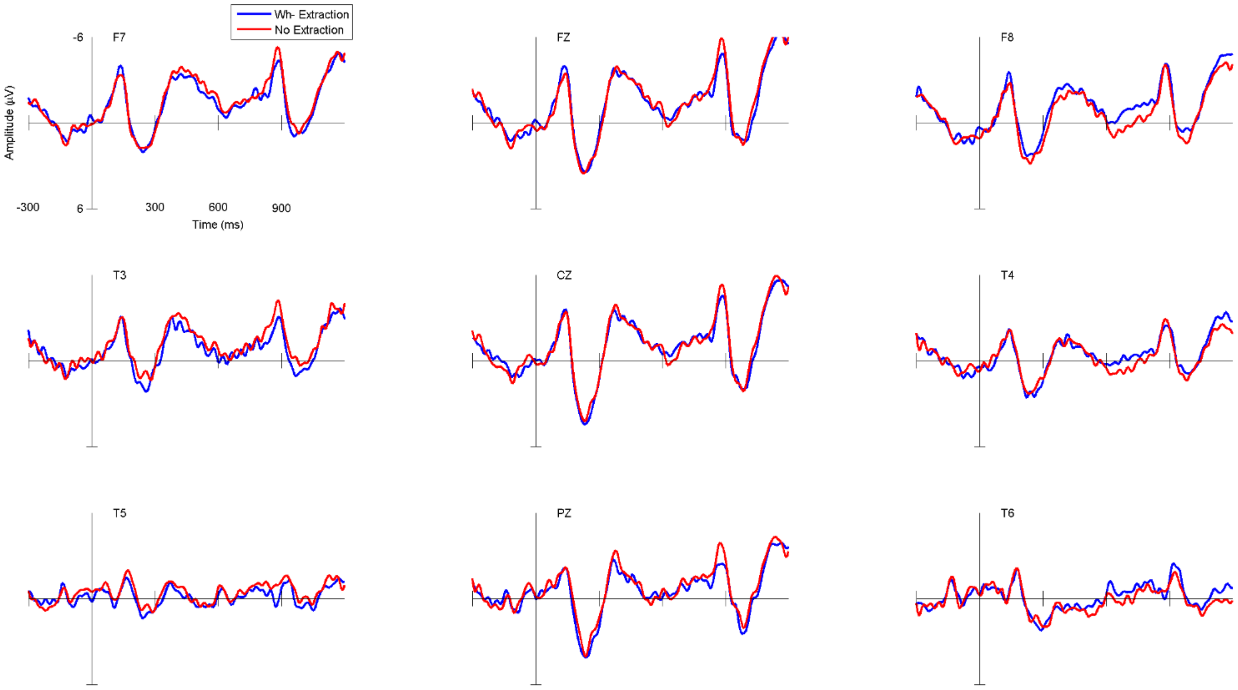
Number of obs: 2852, groups: Subject, 23						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-1.211	0.444	29.5	-2.727	0.011	*
Extraction	-0.603	0.178	2813	-3.379	< .001	***
Island	-0.131	0.108	2813	-1.205	0.228	
Hemisphere (left)	-0.004	0.211	2813	-0.017	0.986	
Hemisphere (right)	0.882	0.211	2813	4.187	< .001	***
Anteriority (ant)	-0.255	0.195	2813	-1.311	0.190	
Anteriority (post)	0.940	0.213	2813	4.410	< .001	***
Proficiency	-0.029	0.079	22.2	-0.373	0.713	
Extraction × Island	0.428	0.153	2813	2.794	0.005	**
Extraction × Hemisphere (left)	0.330	0.203	2813	1.626	0.104	
Extraction × Hemisphere (right)	-0.080	0.203	2813	-0.397	0.692	
Hemisphere (left) × Anteriority (ant)	-0.281	0.242	2813	-1.164	0.245	
Hemisphere (right) × Anteriority (ant)	-0.169	0.242	2813	-0.700	0.484	
Hemisphere (left) × Anteriority (post)	-0.210	0.268	2813	-0.783	0.434	
Hemisphere (right) × Anteriority (post)	-0.721	0.268	2813	-2.687	0.007	**
Extraction × Proficiency	-0.051	0.015	2813	-3.521	< .001	***
Anteriority (ant) × Proficiency	-0.066	0.017	2813	-3.855	< .001	***
Anteriority (post) × Proficiency	0.012	0.019	2813	0.646	0.518	

Results for the best-fitting overall model for the non-island condition are shown in Table 20 below. The main effect of Extraction was not significant, indicating there was not a significant difference in amplitude the *wh*-extraction condition as compared to sentences with no extraction. However, the interaction between Extraction × Proficiency was significant: with increasing proficiency, the amplitude for the *wh*-extraction condition became increasingly negative. In other words, learners with higher proficiency yielded an emerging negativity in the 500-900 ms time window at the filled object position (e.g., *Dave Campbell*). Figure 10 shows the comparison between *wh*-extraction and no extraction conditions at the object position, for the non-island condition.

Table 20: Linear mixed-effects overall model for object position, Non-Island condition, from 500-900 ms.

Number of obs: 1426, groups: Subject, 23						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-1.222	0.471	27.1	-2.592	0.015	*
Extraction	0.017	0.215	1393	0.080	0.936	
Anteriority (ant)	-0.573	0.120	1393	-4.780	< .001	***
Anteriority (post)	0.635	0.135	1393	4.698	< .001	***
Hemisphere (left)	-0.047	0.192	1393	-0.245	0.807	
Hemisphere (right)	0.743	0.192	1393	3.879	< .001	***
Proficiency	-0.042	0.086	22.8	-0.487	0.631	
Extraction × Hemisphere (left)	0.175	0.271	1393	0.648	0.517	
Extraction × Hemisphere (right)	-0.423	0.271	1393	-1.561	0.119	
Extraction × Proficiency	-0.067	0.019	1393	-3.457	< .001	***
Anteriority (ant) × Proficiency	-0.048	0.023	1393	-2.130	0.033	*
Anteriority (post) × Proficiency	0.011	0.026	1393	0.447	0.655	

Figure 10: Object filled-gap effect for Non-Island condition at representative electrodes.

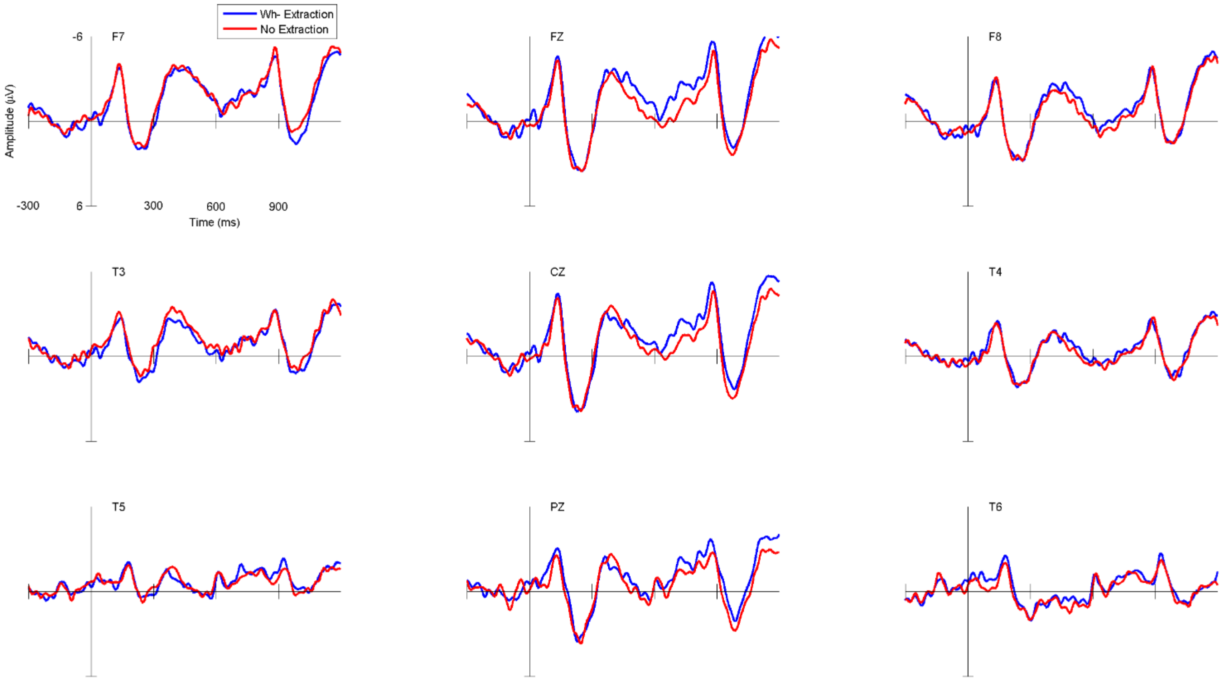


Results for the best-fitting overall model for the island condition are reported in Table 21. The main effect of Extraction was significant, indicating that the amplitude at the object position (e.g., *Dave Campbell*) was more negative for sentences containing *wh*-extraction than for sentences with no extraction. The interaction between Extraction × Proficiency was also significant and indicated that with increasing proficiency, the amplitude for Wh-Extraction sentences became increasingly negative. In other words, learners with higher proficiency yielded a larger negativity in the 500-900 ms time window at the filled object position (e.g., *Dave Campbell*). Figure 11 shows the comparison between *wh*-extraction and no extraction conditions for the island condition at the filled object site.

Table 21: Linear mixed-effects overall model for object position, Island condition, from 500-900 ms.

Number of obs: 1426, groups: Subject, 23						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-1.183	0.451	26.9	-2.623	0.014	*
Extraction	-0.371	0.182	1391	-2.041	0.041	*
Anteriority (ant)	-0.134	0.171	1391	-0.784	0.433	
Anteriority (post)	0.522	0.193	1391	2.707	0.007	**
Hemisphere (left)	-0.088	0.137	1391	-0.642	0.521	
Hemisphere (right)	0.621	0.137	1391	4.543	< .001	***
Proficiency	0.003	0.084	24.4	0.040	0.969	
Extraction × Anteriority (ant)	-0.328	0.242	1391	-1.359	0.174	
Extraction × Anteriority (post)	0.007	0.272	1391	0.024	0.981	
Extraction × Proficiency	-0.076	0.035	1391	-2.193	0.028	*
Anteriority (ant) × Proficiency	-0.132	0.032	1391	-4.063	< .001	***
Anteriority (post) × Proficiency	0.015	0.037	1391	0.401	0.689	
Extraction × Anteriority (ant) × Proficiency	0.098	0.046	1391	2.132	0.033	*
Extraction × Anteriority (post) × Proficiency	-0.003	0.052	1391	-0.049	0.961	

Figure 11: Object position inside island at representative electrodes.



To summarize results from the object position, in the 300-500 ms time window a native-like pattern of responses emerged. In the N400 time window, higher proficiency was associated with an increasingly negative, or N400-like, ERP effect. Crucially, this was only the case in the non-island condition, where gap prediction is grammatically licensed; proficiency did not modulate processing inside the island. In the later time window, higher proficiency was also associated with an increasingly negative ERP effect, which was the case for both the non-island and island contexts. Thus, as was found for the native speakers, learners with higher L2 proficiency yielded robust negativities in both conditions in the 500-900 ms time window. The final analysis examines processing at the actual gap site, the position where the *wh*-dependency is completed.

Actual Gap Site

Processing at the actual gap site sheds light on whether learners, like natives, successfully complete the *wh*-dependency in both non-island and island conditions. We expect a P600 to emerge at the actual gap site if learners successfully resolve the *wh*-dependency, as shown for natives in our study and elsewhere (e.g., Kaan et al., 2000; Phillips et al., 2005).

Non-Island: 300-500 ms time-window analysis. Given the lexical differences at the actual gap site (e.g., *with* vs. *kissed*), separate models were run for the island and non-island conditions; we begin with analyses examining the actual gap site in the Non-Island condition (e.g., *with*). Results from the overall model for the 300-500 ms time window that best fit the data are shown in Table 22. The main effect of Extraction was significant, indicating that the amplitude at the gap site for *wh*-extraction sentences was more positive than for no extraction sentences. The interaction of Extraction × Anteriority (anterior level) was also significant, and indicated that amplitude for electrodes in the anterior region was significantly more positive than those at the midline. In other words, a positivity emerged at the gap site for the non-island condition (e.g., at *with*) which was most robust at anterior sites.

Table 22: Linear mixed-effects overall model for the actual gap site for the Non-Island condition from 300-500 ms.

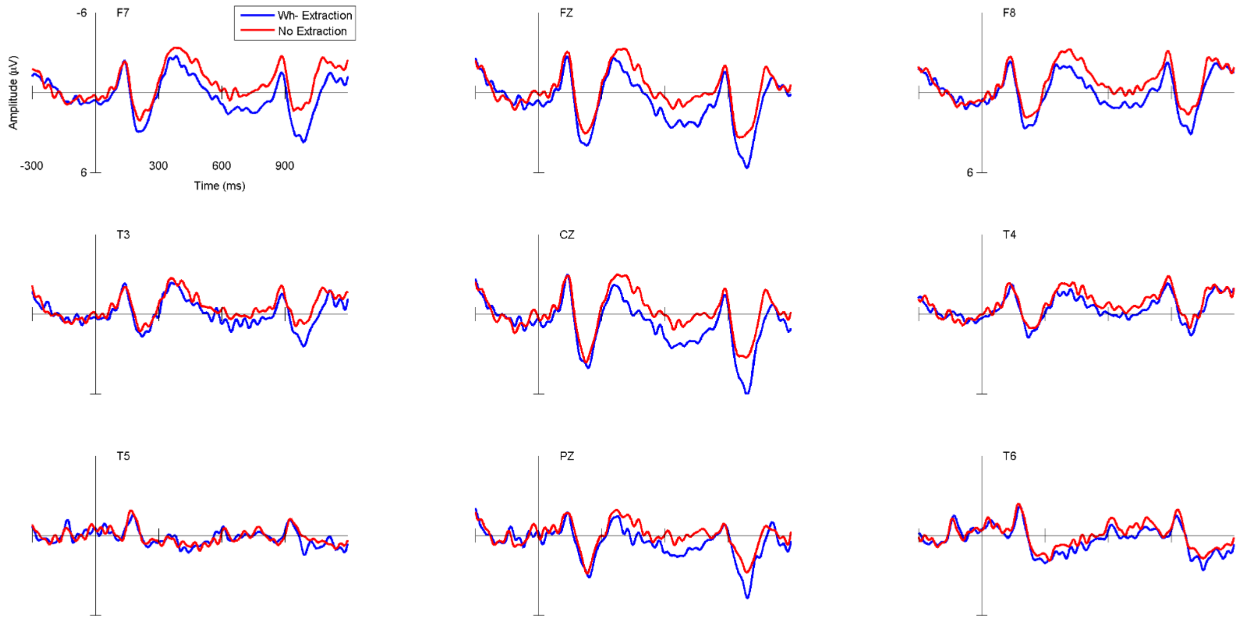
Number of obs: 1426, groups: Subject, 23						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-1.501	0.325	26	-4.614	< .001	***
Extraction	0.748	0.161	1393	4.655	< .001	***
Anteriority (ant)	-0.957	0.151	1393	-6.337	< .001	***
Anteriority (post)	1.973	0.170	1393	11.579	< .001	***
Proficiency	0.027	0.062	26	0.434	0.668	
Extraction × Anteriority (ant)	0.452	0.214	1393	2.116	0.035	*
Extraction × Anteriority (post)	-0.356	0.241	1393	-1.476	0.140	
Extraction × Proficiency	0.015	0.031	1393	0.483	0.629	
Anteriority (ant) × Proficiency	0.054	0.029	1393	1.897	0.058	.
Anteriority (post) × Proficiency	-0.050	0.032	1393	-1.555	0.120	
Extraction × Anteriority (ant) × Proficiency	-0.075	0.041	1393	-1.848	0.065	.
Extraction × Anteriority (post) × Proficiency	0.037	0.046	1393	0.817	0.414	

Non-Island: 500-900 ms time-window analysis. Results from the overall model for the 500-900 ms time window that best fit the data are shown in Table 23. The main effect of Extraction was significant, revealing a positivity for the *wh*-extraction sentences at the actual gap site as compared to the no extraction sentences. The interaction of Extraction × Anteriority (anterior level) was also significant, and as in the previous time window, revealed that this positivity was larger in anterior sites. Additionally, a significant interaction between Extraction × Anteriority (anterior level) × Proficiency emerged. This three-way interaction indicates that as learners' proficiency increases, the positivity at the anterior sites is attenuated as compared to the positivity at the baseline central region. In other words, learners with higher proficiency yield a smaller positivity at anterior sites, instead yielding a centro-posterior distributed positivity. In sum, at the group level learners yielded a P600 at the actual gap site for the 500-900 ms time window; for learners with increased proficiency, this effect was smaller at anterior sites. Figure 12 demonstrates this effect.

Table 23: Linear mixed-effects overall model for the actual gap site for the Non-Island condition from 500-900 ms.

Number of obs: 1426, groups: Subject, 23						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	0.022	0.316	31.5	0.069	0.946	
Extraction	0.798	0.160	1391	4.972	< .001	***
Hemisphere (left)	-0.247	0.121	1391	-2.048	0.041	*
Hemisphere (right)	-0.607	0.121	1391	-5.023	< .001	***
Anteriority (ant)	-0.206	0.151	1391	-1.365	0.172	
Anteriority (post)	0.272	0.170	1391	1.596	0.111	
Proficiency	0.020	0.058	26.9	0.349	0.730	
Extraction × Anteriority (ant)	0.641	0.213	1391	3.002	0.003	**
Extraction × Anteriority (post)	-0.354	0.241	1391	-1.469	0.142	
Extraction × Proficiency	0.012	0.030	1391	0.399	0.690	
Anteriority (ant) × Proficiency	0.047	0.029	1391	1.637	0.102	
Anteriority (post) × Proficiency	-0.017	0.032	1391	-0.54	0.590	
Extraction × Anteriority (ant) × Proficiency	-0.119	0.041	1391	-2.941	0.003	**
Extraction × Anteriority (post) × Proficiency	0.086	0.046	1391	1.874	0.061	.

Figure 12: Actual gap site in non-island condition at representative electrodes.



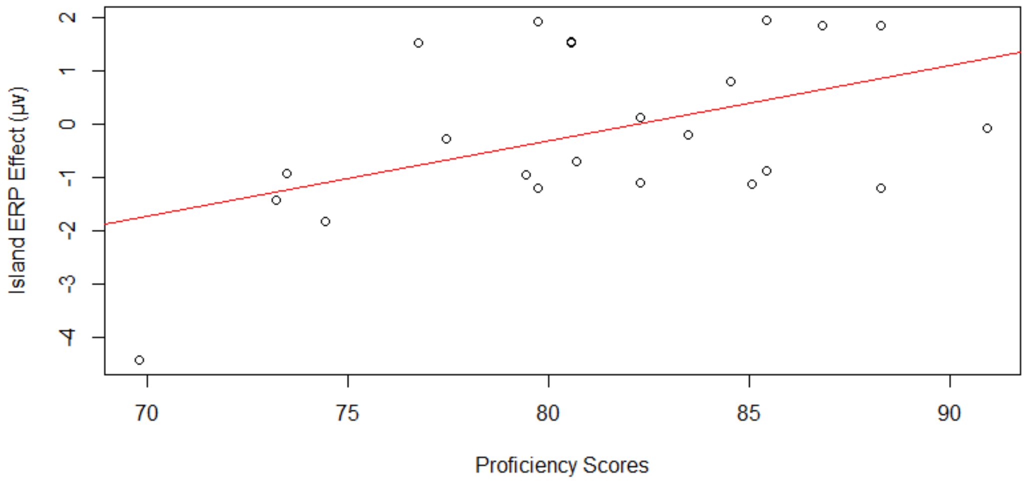
Island: 300-500 ms time-window analysis. Next, we examine the actual gap site in the Island condition (e.g., *kissed*). Results from the overall model for the 300-500 ms time window that best fit the data are shown in Table 24. In this model, the main effect of Extraction was not significant. However, the interaction of Extraction × Proficiency was significant, and indicated that as L2 proficiency increases, the amplitude for *wh*-extraction condition becomes more positive, in comparison with the baseline no extraction condition. In other words, learners with higher proficiency yield a larger positivity at the gap site in the island condition in the 300-500 ms time window, as shown in Figure 13.

Table 24: Linear mixed-effects overall model for the actual gap site for the Island condition from 300-500 ms.

Number of obs: 1426, groups: Subject, 23					
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**
(Intercept)	-0.440	0.350	41.6	-1.258	0.216
Extraction	-0.237	0.183	1389	-1.300	0.194
Hemisphere (left)	0.092	0.239	1389	0.387	0.699
Hemisphere (right)	0.374	0.239	1389	1.566	0.117
Anteriority (ant)	0.349	0.220	1389	1.583	0.114

Anteriority (post)	0.506	0.241	1389	2.096	0.036	*
Proficiency	-0.055	0.058	23.9	-0.949	0.352	
Extraction × Hemisphere (left)	0.379	0.230	1389	1.652	0.099	.
Extraction × Hemisphere (right)	-0.152	0.230	1389	-0.664	0.507	
Hemisphere (left) × Anteriority (ant)	0.109	0.274	1389	0.400	0.689	
Hemisphere (right) × Anteriority (ant)	-0.374	0.274	1389	-1.366	0.172	
Hemisphere (left) × Anteriority (post)	-0.298	0.304	1389	-0.982	0.326	
Hemisphere (right) × Anteriority (post)	0.361	0.304	1389	1.189	0.235	
Extraction × Proficiency	0.142	0.016	1389	8.587	< .001	***
Anteriority (ant) × Proficiency	-0.069	0.019	1389	-3.576	< .001	***
Anteriority (post) × Proficiency	0.002	0.022	1389	0.075	0.940	

Figure 13: Relationship between proficiency and ERP effect size at actual gap site for Island condition, 300-500 ms.



Island: 500-900 ms time-window analysis. Results from the overall model for the 500-900 ms time window that best fit the data are shown in Table 25. The factor Extraction did not reach significance in this model. However, as in the previous time window, the interaction between Extraction and Proficiency was significant, indicating that learners with higher proficiency yielded an emerging P600 at the actual gap site for the Island condition. Figure 14 below shows the group-level ERP effects at the actual gap site in the Island condition. Figure 15 shows the relationship between proficiency and the P600 effect size, with higher proficiency associated with larger positivities.

Table 25: Linear mixed-effects overall model for the actual gap site for the Island condition from 500-900 ms.

Number of obs: 1426, groups: Subject, 23						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	0.995	0.258	39	3.866	< .001	***
Extraction	0.243	0.175	1393	1.387	0.166	
Hemisphere (left)	-0.174	0.156	1393	-1.116	0.265	
Hemisphere (right)	-0.434	0.156	1393	-2.777	0.006	**
Anteriority (ant)	0.922	0.098	1393	9.442	< .001	***
Anteriority (post)	-0.802	0.110	1393	-7.278	< .001	***
Proficiency	-0.012	0.044	25.9	-0.274	0.786	
Extraction × *Hemisphere (left)*	0.120	0.221	1393	0.545	0.586	
Extraction × *Hemisphere (right)*	-0.412	0.221	1393	-1.864	0.063	.
Extraction × *Proficiency*	0.072	0.016	1393	4.545	< .001	***
Anteriority (ant) × *Proficiency*	-0.061	0.019	1393	-3.299	< .001	***
Anteriority (post) × *Proficiency*	0.008	0.021	1393	0.367	0.714	

Figure 14: Actual gap site in island condition at representative electrodes.

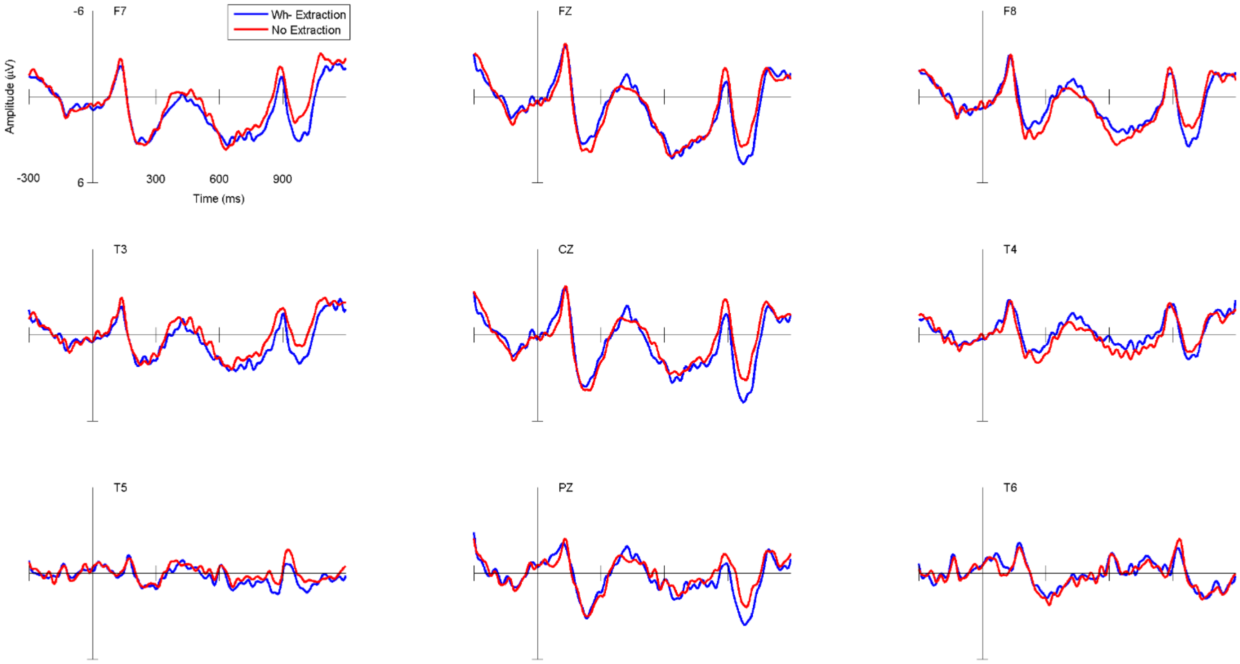
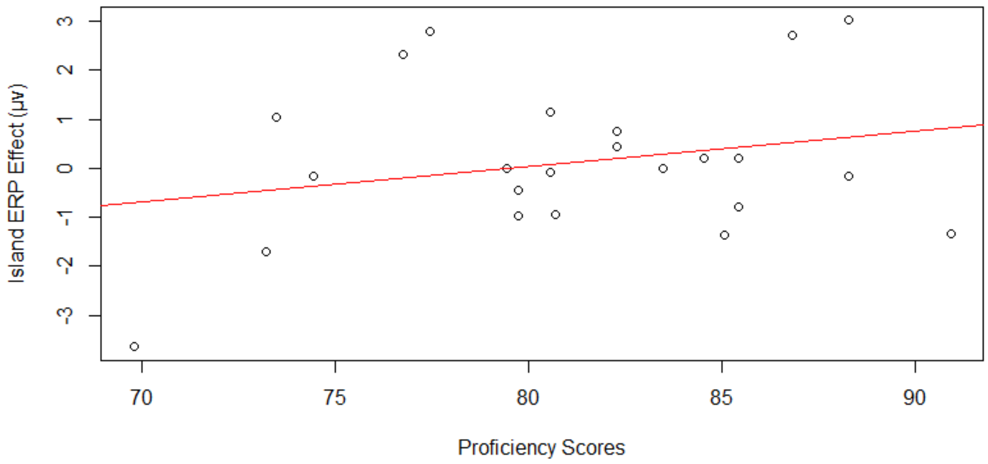


Figure 15: Relationship between proficiency and ERP effect size at actual gap site for Island condition, 500-900 ms.



Summary of EEG results

At the subject filled gap site, a P600 emerged which was not modulated by proficiency. At the object filled gap site, higher proficiency was related to an increasingly negative, or N400-like, ERP effect. In the later 500-900 ms time window, a negativity emerged which was significant only inside the island. In this time window, higher L2 proficiency was related to increasingly more negative effects, and this was true both in the island and non-island sentences. Finally, at the actual gap site, in the non-island sentences, an extended P600 emerged. Increasing proficiency made this positivity larger in posterior sites and smaller in the anterior region, making the effect most prominent at centro-posterior sites. At the actual gap site in sentences containing an island, the overall analysis revealed no significant positivity in the late time window. However, in both the early and late time windows, higher proficiency was related to increasingly positive ERP effects. In the following section, we report findings from an exploratory analysis examining the role of offline island sensitivity (DD scores) at the object position.

Individual difference analysis examining offline island sensitivity

Object Position: 500-900 ms time-window analysis.

The main EEG analysis revealed a late negative component at the filled object position, which was significant only in the island condition. However, increasing proficiency was associated with a larger negativity in both the island and non-island conditions. By including DD scores into the model for the 500-900 ms time window, we explore whether individual differences in offline island sensitivity modulate this late negativity, as was observed in the results for native English speakers.

In the best-fitting model including DD scores, the critical interaction Extraction × Island × DD score was not significant (t (2787) = 0.538, $p = .591$). However, a four-way interaction between Extraction × Island × Proficiency × DD score was significant (t (2787) = -3.010, $p < .01$). To further investigate this interaction, separate models were run for each level of the factor Island. In the non-island model, the crucial interaction Extraction × DD score was significant (t (1373) = 2.477, $p < .05$). This interaction indicates that increased offline sensitivity was associated with a reduced late negativity in the non-island condition. In the best-fitting model for the island condition, the critical interaction of Extraction × DD score was significant (t (1378) = 1.997, $p < .05$). As in the non-island model, this effect shows that as DD scores increase, amplitude for the *wh*-extraction condition was less negative as compared to the no extraction condition. In other words, greater offline island sensitivity was associated with a reduced late negativity in both the island and non-island conditions.

Discussion

The present study investigated whether L2 learners process *wh*-dependencies in a qualitatively similar way to native speakers, examining whether learners predict gaps and successfully resolve the dependency, as well as whether dependency formation is grammatically guided. In examining brain responses at three critical regions across the span of the sentence, we addressed three main research questions. Processing at the subject position tested whether L2 learners engaged in gap prediction pre-verbally, similar to native speakers, who showed an N400 at the subject filled-gap region. This investigation addressed a theoretical debate regarding whether L2 learners are able to engage in prediction during online processing. The second research question concerned whether learners were sensitive to syntactic island constraints during the processing of *wh*-dependencies, which we examined at the object filled gap position. In examining island sensitivity, we address proposals of the Shallow Structure Hypothesis, which argues that learners are unable to utilize abstract grammatical information, i.e., island constraints, in the course of processing. Finally, at the actual gap site, we investigate whether learners, like native speakers, successfully resolve the *wh*-dependency in both non-island and island conditions.

In what follows, we discuss each sentence position and make direct comparisons between the native and non-native findings, addressing our research questions in turn. To summarize main findings across groups, Table 26 shows native speakers’ and learners’ overall results.

Table 26: Comparison of native and L2 key findings.

	Subject filled gap	Object filled gap (non-island)	Object filled gap (island)	Actual gap site (non-island)	Actual gap site (island)
Natives (n=44)	N400	▪ N400 ▪ Late negativity	▪ No N400 ▪ Late negativity	P600	P600
Learners (n=23)	P600	▪ Emerging N400 for higher proficiency L2ers ▪ Emerging late negativity for higher proficiency L2ers	▪ No N400 ▪ Late negativity, (more robust for higher proficiency L2ers)	P600	Emerging P600 for higher proficiency L2ers

Subject Position

Subject filled-gap effects provide insight into whether the parser makes predictions about gap sites using structural information, prior to encountering the verb. Native speakers in the current study yielded an N400 at this position, suggesting that natives predict gaps using syntactic information. Examining L2 processing at this position addresses two SLA debates. The Shallow Structure Hypothesis (Clahsen & Felser, 2006) proposes that learners cannot utilize abstract syntactic information in the course of processing, and instead must rely on semantic/pragmatic information. Shallow Structure therefore predicts that L2 learners would be unable to make structural predictions (i.e., posit a gap) prior to encountering subcategorizing information at the verb. The Reduced Ability to Generate Expectations (RAGE) Hypothesis (Grüter et al., 2017) proposes that predictive processing is limited in L2 learners, and thus both theories expect that prediction-related effects (i.e., N400) would not emerge at the subject filled-gap prediction. Consistent with both hypotheses, there was no evidence to suggest that the learners engaged in syntactic prediction at the subject position. The Chinese learners of English showed a qualitatively different response than native-like N400 at the subject position, yielding a P600 instead.

In the native chapter, the discussion of the P600 focused on studies reporting P600 for *wh*-dependency resolution, which was argued to index successful syntactic integration of the *wh*-filler at the actual gap site (e.g., Kaan et al., 2000; Phillips et al., 2005). However, a large body of ERP research has shown that the P600 is also yielded in contexts in which syntactic integration is unsuccessful, such as phrase structure violations (e.g., Friederici et al., 1996; Frisch et al., 2002; Osterhout & Holcomb, 1992; see Molinaro et al., 2011 for a review). The P600 emerging at the filled subject position may suggest that the learners processed the subject filled gap as a syntactic anomaly. First, consider the role of the L1 in processing the subject filled gap. As discussed previously, Mandarin Chinese does not instantiate overt *wh*-movement, and thus *wh*-elements remain in-situ (e.g., Sohn, 1980; Sohn, 1999). For example, if Mandarin-speaking natives were to encounter a *wh*-word in Mandarin such as *shéi* (who) in the subject position of a given sentence, it would uniformly be the case that *shéi* is the subject of the clause, as *wh*-elements in Mandarin are generated in-situ. If Mandarin learners utilize an L1-like analysis in processing *wh*-dependencies in the L2, then the Mandarin-speaking learners of English may initially assume that upon encountering *who* in the embedded subject position, that *who* will be interpreted as the subject of the embedded clause. If the learners were to maintain this analysis, then they would be surprised to encounter the immediately following noun phrase at the subject filled-gap position (e.g., who *the editor*), and may interpret the sequence of noun phrases as a syntactic anomaly. In this case, the parser would have to process two subject noun phrases back-to-back, yielding a phrase structure violation under this particular analysis; as is discussed above, phrase structure violations have been shown to yield a P600 response in both native speakers (e.g., Friederici et al., 1996) and L2 learners (e.g., Hahne, 2001).

Two L2 ERP studies have reported a similar late positivity emerging at a filled gap site for L2 learners, argued to be related to syntactic integration difficulty (Dong, 2014; Schremm, 2012). One difference between these studies and the current study is that our study uses fully grammatical sentences across the experiment, such that every sentence with a *wh*-dependency contains an actual gap site which renders the sentence grammatical. In Dong (2014) and Schremm (2012), it was the case that the filled gap position rendered the sentence ungrammatical, as there was not an actual gap located further downstream in the sentence. Because all sentences in the present study were fully grammatical, learners were not required to integrate and interpret the two subject NPs at the subject position, given that this position was ultimately always followed by an actual gap site. Further investigation is needed to directly test whether and how processing a filled potential gap site is affected by the overall grammaticality and congruity of the sentence.

Another consideration is that the P600 in the current study is not modulated by proficiency, such that all participants yielded a robust positivity at this region. Researchers like Kaan (2014) propose that predictive processing may be possible for learners with higher proficiency, although this is not what we observed at this pre-verbal position. If predictive processing were modulated by proficiency, increasing proficiency should have been related to increasingly native-like effect, N400. With increasing sample size, we will continue to investigate whether proficiency modulates processing at the subject position. In the current sample, however, we observe that L1 Chinese L2 English learners did not utilize syntactic information to generate a prediction for an upcoming empty argument position, and instead show processing difficulty at the immediately following subject NP.

At the subject position, the Shallow Structure Hypothesis (Clahsen & Felser, 2006) would expect that learners cannot utilize syntactic information during processing and thus would not be able to posit a gap pre-verbally. The lack of a predictive N400 effect at the subject position shows that learners were unable to predict a gap using grammatical information, in line with this proposal. However, it is important to note that while the learners did not yield the prediction-related N400 at the filled subject position, the P600 response that was elicited here is not strictly compatible with predictions of the Shallow Structure Hypothesis. As discussed, the P600 is a syntactic component, elicited at this position likely due to a syntactic anomaly: the subject P600 demonstrates that the learners are in the process of structure building, and encounter syntactic difficulty when encountering two adjacent potential subject NPs. If learners were constructing a shallow structure during processing, relying on semantic and pragmatic information, it is likely that processing at this position would pose a *semantic* integration problem as compared to syntactic difficulty, resulting from processing two arguments that play the same thematic role. Thus, the P600 elicited at the filled subject position is not compatible with predictions of the Shallow Structure Hypothesis. Additionally, describing the learners' P600 as being related to the L1 Mandarin as a *wh*-in-situ language is not predicted by the Shallow Structure Hypothesis, although this is an interpretation that needs to be tested by including an additional L1 group. This hypothesis instead argues that all adult learners regardless of L1 should process *wh*-dependencies similarly, relying on thematic argument matching. In sum, while it is the case that learners do not show evidence of gap prediction in the pre-verbal subject position, the P600 emerging at this region is inconsistent with predictions of the Shallow Structure Hypothesis (Clahsen & Felser, 2006).

Object Position

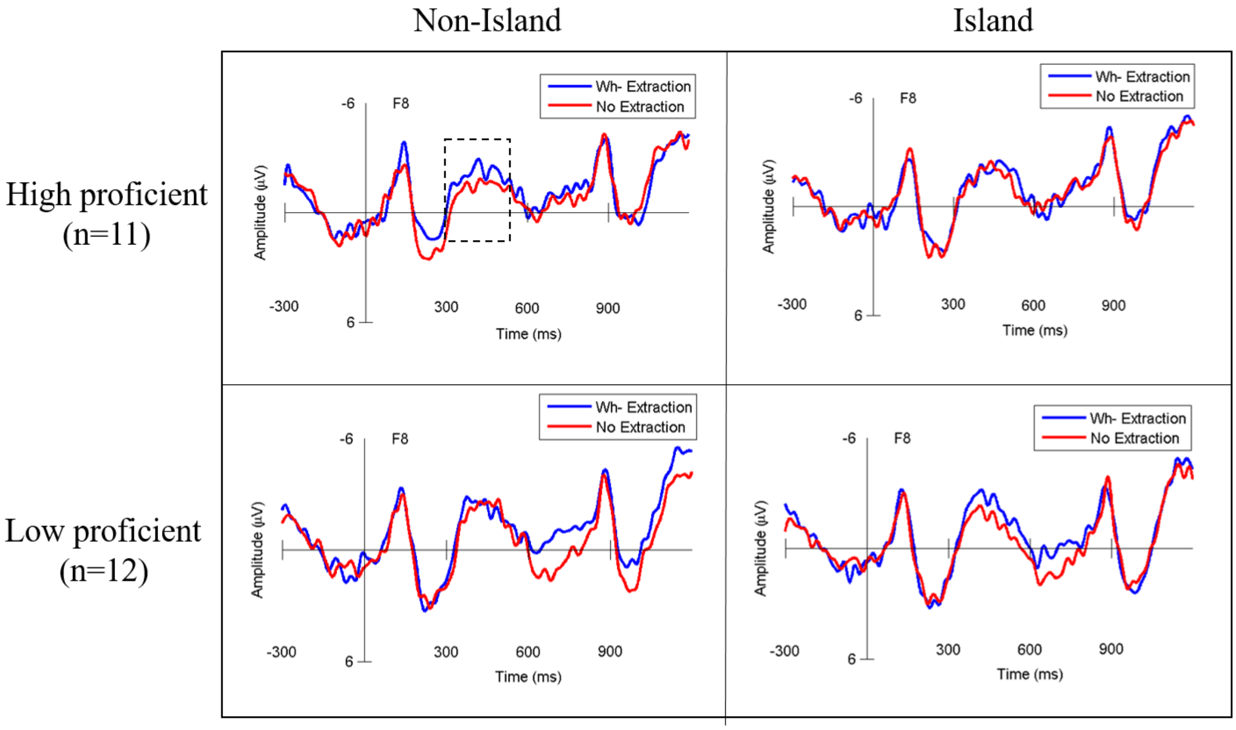
At the object filled-gap position, native speakers demonstrated that they were sensitive to syntactic island constraints, yielding an N400 in the licit, non-island context only. That is, gap prediction was in evidence only in the grammatically licensed position, and there was no evidence that a gap was posited inside the island. Examining L2 processing at this region can shed light on several important issues. First, object filled-gap effects can reveal whether L2 learners predict gaps after encountering the subcategorizing verb. Note however, that the object position does not provide unambiguous evidence of syntactic prediction as in the subject position, given that it is possible for filled-gap effects emerge at the object site due to a direct association between the *wh*-filler and the subcategorizing verb (e.g., Pickering & Barry, 1991). In fact, this is what the Shallow Structure Hypothesis predicts, such that learners may show evidence of gap prediction post-verbally due to simple thematic argument matching rather than utilization of syntactic information (Clahsen & Felser, 2006). Accounts such as Grüter et al.'s (2017) RAGE Hypothesis do not make such fine-grained predictions regarding the contexts in which learners can predict, instead proposing that learners are generally less able to engage in predictive processing compared to native speakers. In contrast, other researchers have suggested that the ability to predict in the L2 may depend on a range of factors, including proficiency level (e.g., Hopp, 2013; Kaan, 2014).

Processing at the object filled-gap site also provides an excellent test of island sensitivity. We directly tested whether learners, like native speakers, avoid positing gaps inside an island, a grammatically unlicensed domain which does not allow *wh*-extraction. If learners were to demonstrate sensitivity to syntactic island constraints, this would provide evidence that L2ers can utilize grammatical knowledge in the course of sentence processing. The Shallow Structure Hypothesis predicts that learners are not able to use grammatical constraints on the same time-course as native speakers and thus, learners should not show sensitivity to island constraints, yielding gap prediction effects N400 in both the non-island and island contexts (Clahsen & Felser, 2006). Kim et al. (2015) make a similar prediction because the L2 learners in the present study are native speakers of Mandarin Chinese, a language which does not instantiate overt *wh*-movement. In contrast, several researchers have argued that native-like processing is indeed possible for very advanced L2 learners even in cases where the L1 and L2 differ with respect to whether or not there is overt *wh*-movement, as has been shown in a range of self-paced reading studies examining *wh*-dependencies and island sensitivity (e.g., Aldwayan et al., 2010; Johnson et al., 2016; Omaki & Schulz, 2011). These accounts predict a native-like pattern of ERP results for advanced learners with success being potentially modulated by L2 proficiency.

The overall picture of L2 processing at the object filled gap position was strikingly similar to the native-like pattern. In the N400 time window, highly proficient learners showed an emerging N400 in the non-island condition, consistent with behavioral studies which report native-like object filled-gap effects for learners (e.g., Aldwayan et al., 2010; Johnson et al., 2016) and in line with proposals which suggest that predictive processing for L2 learners may depend on proficiency (Hopp, 2013; Kaan, 2014). No N400 was observed in the island condition, and importantly proficiency did not modulate processing in the island context for the N400 time window. Thus, there is no evidence to suggest that learners posited a gap inside the illicit island domain. Together, findings suggest that at least highly proficient L2 learners are able to predict gaps post-verbally, as well as utilize grammatical constraints in the course of processing to correctly avoid positing a gap inside the island, similar to native speakers. The relationship with proficiency is not compatible with predictions of the Shallow Structure Hypothesis, which expects that all adult learners regardless of proficiency level are similarly unable to utilize abstract syntactic information during online processing (Clahsen & Felser, 2006). This finding is also in contrast to Kim et al.'s (2015) proposal regarding the role of the L1, which would predict that only learners whose L1 instantiates overt *wh*-movement would show sensitivity to islands online. The findings for the Mandarin Chinese learners of English, whose L1 does not instantiate overt *wh*-movement, suggest that learners of *wh*-in-situ languages can show online sensitivity to islands, in line with recent findings for Korean and Arabic learners (e.g., Aldwayan et al., 2010; Johnson et al., 2016).

In order to better understand the role of proficiency in processing the object filled gap position, Figure 16 below visually demonstrates ERP effects at a representative electrode for each condition using a median split in proficiency scores. The lower proficiency group scored below the overall mean of 80.8 (n=12, *M* = 77.14, *SD* = 3.63) and the higher proficiency group scored above the overall mean (n=11, *M* = 85.7, *SD* = 2.68). Note that all statistical analyses reported above were not conducted via a median split, and proficiency scores were treated as a continuous variable in each model. It is also worth mentioning that the 'low' proficiency group still scored very high on the proficiency composite score, likely putting them into a high-intermediate or a low-advanced English proficiency range.

Figure 16: Median split by proficiency comparing ERP effects at the object filled-gap site at representative electrode F8.



In the non-island condition, an N400 is visually apparent in the 300-500 ms time window for the high proficiency group only. In the island condition, no N400 is in evidence for the highly proficient group. This pattern is qualitatively similar to the native responses emerging at the object position, and suggest that with increasing proficiency, L2 learners engage in gap prediction that is guided by grammatical knowledge. In contrast to predictions of the Shallow Structure Hypothesis (Clahsen & Felser, 2006), high proficiency learners demonstrate an ability to utilize syntactic information, such as the presence of a relative clause island, to constrain online processing.

A second ERP component emerged in the later 500-900 ms time window at the object position. At the group level, the late negativity was significant only in the island condition, with no significant effects emerging in this time window for the non-island condition. However, higher proficiency associated with larger negativities in *both* the island and non-island contexts. This pattern for highly proficiency learners was in line with what was observed for the native speakers, who showed a late negativity at the object filled gap which was significant in both island and non-island conditions. In Chapter 2, we proposed that this component captures the processing of thematic role assignment, yielded by thematic argument conflict of the *wh*-filler and filled object NP (e.g., Frenzel et al., 2011; Frisch & Schlesewsky, 2001, 2005). Importantly, for the native speakers the late negative component was reduced for individuals with higher DD scores. Thus, with increasing offline sensitivity to island violations, native English speakers showed evidence of not attempting to assign thematic roles inside the island (i.e., yielding reduced negativities or no effects), a domain from which *wh*-extraction is banned and argument linking should not take place. We conducted an exploratory analysis to test whether this relationship was also in evidence for the L2 learners. In the late negativity time window, 500-900 ms, higher DD scores were associated with a reduced negativity. For the learners, though, this relationship was significant in the island context, similar to the native English speaker results, and also in the non-island context. It is unclear why increased offline sensitivity to island violations would modulate processing in the non-island context; we plan to continue to investigate this relationship with increasing sample size, examining whether, like native speakers, DD scores modualte processing only in the island condition.

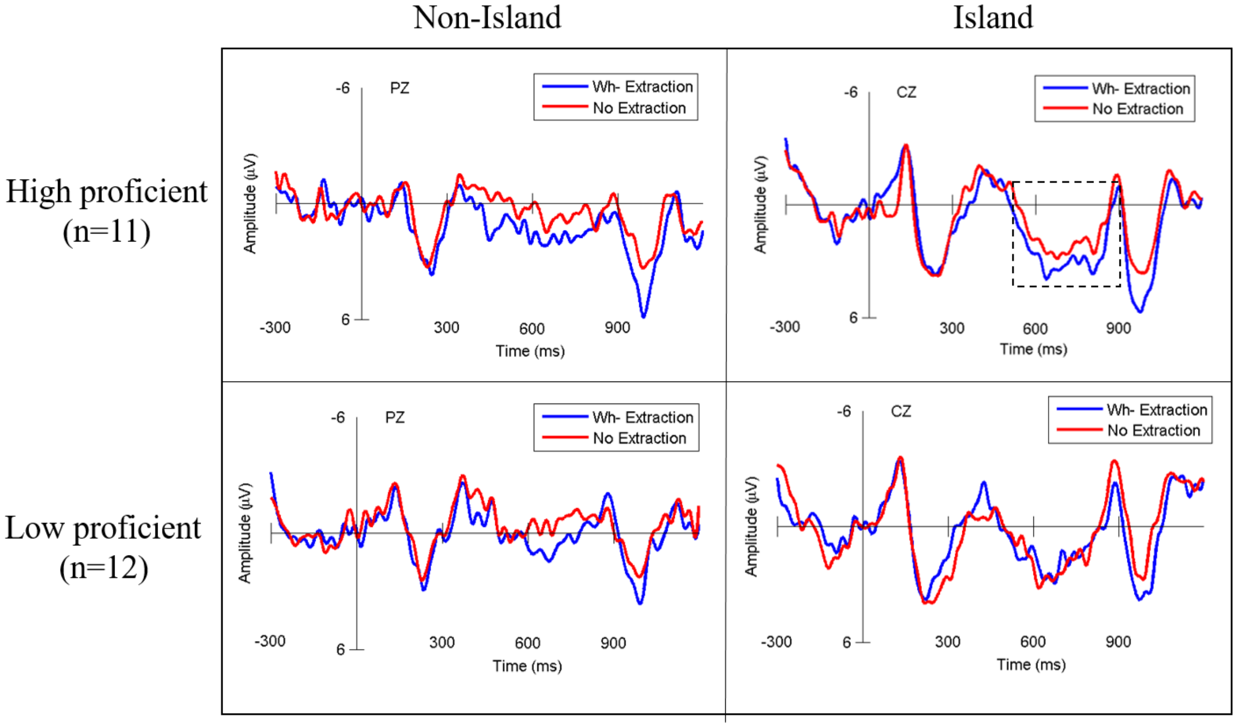
Overall, what is highlighted by the L2 results is that qualitatively native-like processing is in evidence for the highest proficiency learners. In the earlier time window, learners with higher L2 proficiency yielded N400-like effects in the non-island condition, indicating that these learners predicted a gap in the post-verbal licit object position. Crucially, higher L2 proficiency was not associated with N400 effects inside the island domain. This provides evidence that that higher proficiency is not simply associated with an increased ability to link thematic arguments, but rather, with increased sensitivity to syntactic island constraints during online processing. In the later time window, higher proficiency was associated with a native-like pattern as well, with late negativities emerging across conditions. In sum, findings at the object position are inconsistent with predictions of the Shallow Structure Hypothesis (Clahsen & Felser, 2006), and instead suggest that with increasing proficiency, L2 learners are sensitive to island constraints during online processing.

Actual Gap Site

At the actual gap site, P600s were predicted to emerge in both non-island and island conditions, indexing successful *wh*-dependency resolution, as was found for the native speakers in the current study and in previous research (e.g., Gouvea et al., 2010; Kaan et al., 2000; Phillips et al., 2005). In the overall analysis for L2 learners, a significant P600 response emerged in the non-island context, whereas the positivity in island sentences was not statistically significant. At the group level, results indicated that learners were able to successfully resolve the *wh*-dependency in sentences that did not involve islands, but struggled to do so in sentences containing relative clause islands. Learners may have had difficulty processing the complex relative clause island and were thus unable to complete the long-distance dependency at the actual gap site. This is supported by learners' behavioral responses, which showed that learners were less accurate at answering comprehension questions in conditions containing a relative clause island (M = 66.3%) as compared to the non-island sentences (M = 74.6%). Crucially, in both conditions proficiency modulated the ERP effect, with higher proficiency associated with larger positivities. In other words, individuals with higher L2 proficiency showed a greater amplitude P600 in the non-island condition, and evidence of an emerging P600 in the island condition.

To illustrate the role of proficiency at the actual gap site, Figure 17 below visually demonstrates ERP effects at a representative electrode for each condition using a median split in proficiency scores. This figure shows that the positivity emerging for the non-island condition is robust across proficiency level, with a positivity visually apparent for both the lower and higher proficiency groups. In the island condition, on the other hand, the positivity is visually larger in the high proficiency group, revealing the important role of proficiency at the actual gap site in this condition.

Figure 17: Median split by proficiency comparing ERP effects at the actual gap site at representative electrode Pz for Non-Island condition and Cz for Island condition.



The relationship between proficiency and ERP effect size within the island condition shows that some learners were able to successfully resolve the dependency in the island condition in a native-like way, yielding a P600. This is particularly important, showing that L2 learners with higher proficiency were able to overcome the processing burden posed by the relative clause island while continuing to search for upcoming gap sites to complete the unresolved dependency. Although processing at this region does not provide a direct test case of the Shallow Structure Hypothesis, it does suggest that L2 dependency resolution is native-like. To our knowledge, this study presents some of the first ERP evidence of successful *wh*-dependency resolution in L2 learners.

The P600 has been shown to be elicited in a variety of contexts, as both an index of ungrammaticality and successful syntactic processing (for a review see Gouvea, 2015). It is important to carefully consider the interpretation of the P600 in the current study, given that this component emerged at the actual gap site as well as at the subject filled-gap position. In our study, the positivity at the subject position became significant in the first time window at midline and right hemisphere electrodes, and in the second time window was robust across all scalp electrodes. The positivity at the actual gap site in the non-island condition was similarly sustained, and was significant across all regions in both time windows. One distinguishing factor between the two P600s is the role of proficiency, which was not significant at the subject position but crucially increased the size of the positivity at the gap site (in both non-island and island conditions). Proficiency is expected to modulate L2 processing such that learners with higher L2 proficiency are predicted to show more native-like effects. Indeed, this was observed at the object filled gap position, where learners with higher proficiency showed a greater ability to predict gaps, yielding an emerging N400 in the non-island condition. At the actual gap site, higher proficiency learners also yielded larger P600s, the component found for native speakers, indicating that they were increasingly able to successfully resolve the *wh*-dependency. However, this same relationship did not emerge at the subject position, where a group-level P600 was robust. Thus, higher proficiency learners showed evidence of being better able to develop the P600 in only one context, which may suggest that different types of syntactic processing are engaged at the filled subject region and actual gap site.

We suggested that the learners' P600 at the subject position was related syntactic integration difficulty at the subject filled gap, where two potential subject NPs were processed. The P600 at the actual gap site, on the other hand, is argued to index successful *wh*-dependency resolution, as shown for native speakers in the present study and elsewhere (e.g., Felser et al., 2003; Gouvea et al., 2010; Kaan et al., 2000; Phillips et al., 2005). We propose that the P600s emerging for the learners at both the subject position and actual gap site are in line with proposals which characterize the P600 as a component that broadly reflects syntactic processing. For example, Phillips and colleagues (2005) write, “The P600 reflects structure-building processes in congruous sentences, and structure re-building and rechecking processes in garden path and ungrammatical sentences” (p. 425). Although all sentences in the current study are congruous and grammatical, the learners process the filled subject position as involving a temporary syntactic anomaly, utilizing ‘structure rechecking’ computations of the P600. By the end of the sentence at the actual gap site, learners are able to successful resolve the *wh*-dependency, utilizing ‘structure building’ operations of the P600. Because L2 proficiency modulated only the P600 emerging at the actual gap site, and not the P600 at the subject position, this may suggest that we are indeed capturing different computations of the P600 response, related to syntactic integration as a whole (e.g., Phillips et al., 2005). Although we leave this hypothesis as an interesting open question, we note that the present study is one of the first to capture P600 responses related to both structure-rechecking and successful structure-building within the same grammatical sentence, demonstrating that, at least for high proficiency L2 learners, integration difficulty at the beginning of the sentence can still result in successful syntactic integration at the end of the dependency.

Conclusions

This chapter presented results from an L2 ERP study investigating the processing of *wh*-dependencies by Mandarin Chinese-speaking learners of English. The present study tracked the dynamics of dependency resolution across a grammatical sentence to investigate whether learners, like native speakers, engage in gap prediction during *wh*-dependency formation and successfully resolved the long-distance dependencies. We also examined whether learners are sensitive to island constraints in real-time, testing predictions of the prominent Shallow Structure Hypothesis. Overall, the results present a complex picture of what is and what is not possible in L2 processing by adult learners. Learners did not appear to be able to utilize abstract, grammatical information to generate a structural gap prediction at the subject position, in line with accounts which argue that prediction is limited for L2 learners. However, at the direct object position, highly proficient learners showed evidence of predicting a gap, yielding an increasingly native-like N400 response. Highly proficient learners also showed island-sensitive processing, with no predictive effects surfacing inside the island. At the actual gap site, highly proficient learners yielded larger effects of dependency resolution, in line with native speakers.

Together, these results reveal that proficiency plays an important role in the processing of *wh*-dependencies, modulating a learner's ability to use grammatical information and their ability to successfully integrate the *wh*-filler at the actual gap site. Findings suggest that with increasing proficiency, L2 processing of *wh*-dependencies becomes increasingly native-like, although subject filled-gap effects present processing difficulty for all learners that is possibly related to the L1. With increasing sample size, we will continue to investigate L2 processing of *wh*-dependencies, as well as specifically examining whether Mandarin Chinese learners ever predict a gap at the subject position utilizing abstract, syntactic information. In the final chapter, future directions for the L2 and native speaker experiments are discussed, as well as limitations of the current study.

Chapter 4: Overall Discussion

This dissertation tracked the dynamics of *wh*-dependency resolution across a grammatical sentence, using distinct ERP components to examine both prediction and integration by native English speakers and Mandarin Chinese-speaking learners of English. It also investigated whether native and non-native processing was grammatically guided, and examined the extent to which the use of grammatical knowledge during online processing and the ability to engage in predictive processing is modulated by proficiency in L2 learners, and performance on a range of cognitive measures in native speakers. This study is one of the first ERP studies to examine individual differences in the processing of *wh*-dependencies using fully grammatical sentences, and findings suggest an important role for both attentional control and working memory in *wh*-dependency formation for native speakers. A primary research question of this dissertation was whether L2 learners processed *wh*-dependencies in native-like way, yielding qualitatively similar ERP responses. Table 27, repeated from Chapter 3, provides a comparison of the main findings for the two participant groups. In this discussion we review results at each region of the sentence, discussing the broader implications of the findings and as well as identifying the major contributions of this study to both the native and L2 processing literatures.

Table 27: Comparison of native and L2 key findings.

	Subject filled gap	Object filled gap (non-island)	Object filled gap (island)	Actual gap site (non-island)	Actual gap site (island)
Natives (n=44)	N400	▪ N400 ▪ Late negativity	▪ No N400 ▪ Late negativity	P600	P600
Native individual difference results	▪ Higher WM: smaller N400 ▪ Higher AC: larger anterior negativity	▪ Higher AC: larger N400 + larger late negativity	▪ Higher DD: less negative ERP response in both time windows	▪ Higher WM: larger P600 Higher AC: larger P600	Higher WM: larger P600
Learners (n=23)	P600	▪ Emerging N400 for higher proficiency L2ers	▪ No N400 ▪ Late negativity (more robust for	P600	Emerging P600 for higher

		▪ Emerging late negativity for higher proficiency L2ers	higher proficiency L2ers)		proficiency L2ers

Note: WM = working memory, AC = attentional control, DD = differences-in-differences score, a measure of offline island sensitivity; in all individual difference measures, higher scores indicate better task performance.

Prediction in language processing

A broad goal of this dissertation was to examine the role of prediction in language processing by both native speakers and L2 learners. This study adds to both the native and L2 psycholinguistics literatures, examining not only whether native English speakers and Mandarin-speaking learners of English engage in prediction during *wh*-dependency formation, but further investigating the contexts in which natives and learners form predictions. To address this research question, we examined the processing of filled-gap effects in a pre-verbal subject position, as well as a post-verbal object position. To be able to predict a gap at the subject position, the parser must utilize grammatical information to generate a syntactic prediction for an upcoming gap site, because the subcategorizing verb has not yet been encountered. Evidence for subject filled-gap effects for native English speakers is mixed in the psycholinguistic literature (e.g., Aldwayan et al., 2010; Canales 2012; Johnson, 2015; Johnson et al., 2016; Lee, 2004). Our results provide some of the first EEG evidence that native English speakers predict gaps pre-verbally, yielding N400 at the filled subject position. This finding is in contrast to the Direct Association Hypothesis, which argues that filled-gap effects are a consequence of a thematic argument relationship in which the *wh*-filler is directly associated as an argument of the licensing verb (e.g., Pickering & Barry, 1991), and does not expect subject filled-gap effects to emerge. Instead, our results are more in line with predictions of gap-based accounts (e.g., Frazier & Clifton, 1989; Stowe, 1986), and provide strong evidence that native English speakers anticipate syntactic gaps. More broadly, our results are in line with recent research which suggests that native speakers actively form predictions during language processing (e.g., Altmann & Kamide, 1999; Altmann & Mirkovic, 2009; Jaeger & Snider, 2013; Kamide et al., 2003; Levy, 2008; MacDonald, 2013; Pickering & Garrod, 2011, 2013; Van Berkum et al., 2005; Van Petten & Luka, 2012).

To our knowledge, only one self-paced reading study has reported robust subject filled-gap effects for L2 learners (Johnson, 2015). Johnson showed that both native English speakers and Korean learners of English yielded reading time slowdowns at a subject filled gap, which emerged on the same time-course for both groups. While self-paced reading can provide insights into the time-course of sentence processing, the comparison of reading times can only reveal that a difference in processing has occurred. In the current study, results at subject position indeed showed a significant difference between the *wh*-extraction condition and the baseline no extraction condition in the 300-500 ms time window for both native speakers and L2 learners. This result highlights an important strength of utilizing EEG to track the dynamics of processing: ERP components provide a multi-dimensional measurement that can shed light on the *type* of processing that has occurred. We observed that the two groups yielded different ERP components at the subject filled gap, with natives showing N400 and learners showing P600, suggesting that the two populations processed the subject filled gap in a qualitatively different way. The nature of the L2 effect (P600) indicates that learners failed to predict a gap at the subject position.

An important follow-up question is why learners were not able to generate a syntactic prediction in the subject position. One possibility is that our sample size of n=23 learners, which is reduced compared to L2 studies reporting subject filled-gap effects for learners (e.g., Aldwayan et al., 2010; Canales, 2012; Johnson, 2015), was too small to allow variability to emerge for the L2 learner group. That is, it is possible that with increasing sample size, proficiency may modulate ERPs in a similar way as it has been shown to do at the object position and actual gap site, with higher proficiency individuals showing an increasingly native-like N400 effect. In support of this possibility is that the only study to show robust native-like subject filled-gap effects for L2 learners was Johnson (2015), who tested 100 Korean speaking learners of English; other studies which reported inconsistent findings across experiments tested about half the number of participants (Aldwayan et al., 2010, n=40; Canales, 2012, n=64). However, given that proficiency was shown to significantly modulate processing at the two later sentence regions, it is unlikely that the qualitative differences emerging at the subject region can be explained due to sample size, although this is something we will explore as we continue to add Mandarin-speaking learners to the study.

A second possibility regarding why learners did not show effects of syntactic gap prediction at the subject position relates to the type of information available at the subject position in the sentence. At the subject position, the only cue in the input that a gap is forthcoming is that the parser has just encountered a *wh*-filler. Therefore, the parser must be able to use syntactic information to be able to predict a subject gap. At the object position, however, the parser has just encountered a subcategorizing verb (e.g., *Jamie wondered who the editor interviewed* ***Dave Campbell****…*). Thus, at the post-verbal object position, both syntactic and thematic information is available to the parser to be able to predict that a gap is forthcoming. While native English speakers in the present study were able to utilize abstract syntactic information to generate a prediction for an upcoming gap, the Chinese learners were unable to do so.

The second context where predictive processing was examined was the licit filled object position. At this position, native English speakers yielded N400, showing that object gap prediction occurred. In the L2 analysis, a group-level N400 was not observed; however, an interaction with the factor Proficiency revealed that individuals with higher L2 proficiency showed an emerging N400 at the filled object position. In the post-verbal context, higher proficiency learners showed an emerging prediction effect. This relationship is in line with predictions of researchers such as Kaan (2014) and Hopp (2013), who suggest that the ability to predict in the L2 may be modulated by individual differences in proficiency and cognitive abilities.

Overall, results from the subject and object positions reveal a complex picture of predictive processing for L2 learners, allowing us to address an active L2 debate regarding whether learners can engage in prediction online. Several studies have reported that predictive processing is limited for L2 learners, in line with Grüter et al.'s (2017) Reduced Ability to Generate Expectations (RAGE) Hypothesis (e.g., Grüter et al., 2012; Lew-Williams & Fernald, 2010; Martin et al., 2013; Mitsugi & MacWhinney, 2018). In contrast, other studies find native-like effects of prediction for L2 learners, (e.g., Foucart et al., 2014; Hopp, 2013; Johnson, 2016; Leal et al., 2017). This debate has proceeded across several domains, with researchers arguing L2 prediction is either limited or on par with natives based on findings from syntactic, lexical-semantic, and discourse contexts. Examining findings from various linguistic manipulations and L2 learner groups, it is difficult to come to a comprehensive understanding of L2 predictive processing. In line with researchers like Hopp (2013), we propose that learners' ability to engage in prediction may be dependent on the kind of information a learner is able to use to engage in prediction. In our results, Mandarin-speaking learners were able to engage in prediction in a post-verbal position using thematic/semantic information, whereas at the subject position, L2ers were unable to generate a syntactic expectation for a subject gap. A nuanced investigation of the role of prediction in L2 processing which formulates clear predictions for the contexts in which predictive processing is expected to be limited vs. native-like is an important avenue for future research.

Island constraints in the processing of *wh*-dependencies

This dissertation also investigated whether native English speakers and Mandarin Chinese learners of English were sensitive to island constraints during online processing. A line of psycholinguistic studies in the native literature has shown that filled-gap effects are not yielded inside island domains (e.g., Canales, 2012; Johnson et al., 2016; Stowe, 1986; see also Phillips, 2006; Traxler & Pickering, 1996). The grammatical account of island sensitivity argues that the search for potential gap sites is grammatically constrained (e.g., Phillips, 2006, 2013; Sprouse et al., 2012; Wagers & Phillips, 2009). On the other hand, the processing account has argued that the parser avoids positing gaps inside islands due to increased processing difficulty of the island structure itself (Hofmeister & Sag, 2010; Hofmeister, Casasanto, & Sag, 2012a/b, 2013; Kluender, 2004; Kluender & Kutas, 1993b). In order to tease apart the grammatical and processing view, both of which expect no filled-gap effects to emerge in islands, we examined the role of individual differences in processing capabilities. Sprouse et al. (2012) argue that processing accounts predict that individuals with higher processing abilities (e.g., greater working memory) should be more likely to posit gaps within islands as they have greater resources available to attempt to resolve the *wh*-dependency inside complex island structures. Grammatical accounts, on the other hand, do not predict such a relationship between individual differences in processing resources and island sensitivity, given that all native speakers of a language should be similarly sensitive to a grammatical constraint regardless of their processing abilities.

Results indicated that native English speakers indeed avoided positing a gap inside an island, yielding N400 only in the licit non-island condition and no filled-gap N400 effect in the island condition. Crucially, neither working memory nor attentional control were found to modulate processing inside the island in the critical N400 time window. Thus, individuals with increased cognitive abilities did not show evidence of attempting to posit a gap within the island, as might be expected under a processing account of island (e.g., Hofmeister & Sag, 2010). This finding is better accounted for by the grammatical account of island sensitivity, which does not expect such a relationship to emerge (e.g., Sprouse et al., 2012). In addition, we examined the relationship between an individual's offline sensitivity to island violations and their online island sensitivity, as measured by ERP inside the island. This analysis revealed that individuals who showed greater sensitivity to island violations offline yielded *less* of a prediction-related response in the island. That is, individuals with increased offline island sensitivity were less likely to show a filled-gap N400 effect inside the island, providing evidence linking offline and online island sensitivity in adult native English speakers.

The examination of island sensitivity in L2 learners addressed a theoretical debate regarding whether L2 learners can utilize grammatical knowledge in the course of sentence processing. Much work in the behavioral literature has investigated this research question, testing predictions of proposals which argue for qualitative differences between native and non-native speakers' grammatical knowledge (e.g., Hawkins, 2009; Hawkins & Chan, 1997; Hawkins & Casillas, 2008; Tsimpli & Dimitrakopoulou, 2007) and online processing (e.g., Clahsen & Felser, 2006; Felser & Roberts, 2007; Marinis et al., 2005). This dissertation aimed to test claims of the prominent Shallow Structure Hypothesis, which argues that that L2 sentence processing is qualitatively different than native processing because learners are unable to utilize abstract syntactic information during processing (Clahsen & Felser, 2006). Because this theory proposes that learners cannot use grammatical information in the course of processing *wh*-dependencies, learners are not expected to show sensitivity to syntactic islands online. In the literature, results for L2 island sensitivity are mixed (e.g., Aldwayan et al., 2010; Boxell & Felser, 2016; Cunnings et al., 2010; Felser et al., 2012; Johnson et al., 2016; Kim et al., 2015; Omaki & Schulz, 2011).

Like native speakers, learners with increased proficiency showed increasingly N400-like effects at the object position in the licit, non-island condition. The relationship with proficiency was island sensitive, such that it was not the case that there were also increasingly N400-like effects emerging inside the island. In other words, both learners and native speakers showed effects of gap prediction at the post-verbal object site in the grammatically licensed position. This pattern of results suggests that highly proficient learners are able to engage in gap prediction in a post-verbal position, and are able to utilize grammatical knowledge about island constraints in the course of processing a *wh*-dependency. Our findings for the highly proficient L2 learners are in line with studies who show island sensitivity in L2 online processing (e.g., Aldwayan et al., 2010; Johnson et al., 2016; Omaki & Schulz, 2011), and do not provide support for the Shallow Structure Hypothesis, which proposes that all adult learners are unable to utilize abstract syntactic knowledge during processing.

***Wh*-dependency resolution**

In line with previous studies examining processing at the gap site in sentences with filler-gap dependency, we observed P600 at the actual gap site in both the non-island and island conditions for native speakers (e.g., Felser et al., 2003; Gouvea et al., 2010; Kaan et al., 2000; Phillips et al., 2005). The non-island condition in our stimuli involved two potential gap sites in the sentence which were ultimately filled with lexical material. Thus, by the time the parser reaches the actual gap site in the non-island condition, gap prediction has already taken place twice, at the embedded subject and object position, neither of which were the actual gap site. It is particularly impressive that native English speakers successfully resolve the *wh*-dependency in this context, where the parser had recovered from incorrect gap prediction at two previous regions. This is also the first study to test for successful dependency resolution in fully grammatical sentences containing islands, and the fact that P600s emerged at the actual gap site indicates that readers eventually resolved the dependency, despite the increased processing burden incurred by processing the island.

At the actual gap site, highly proficient learners showed native-like effects of dependency resolution as well, showing emerging P600s in both conditions. The present study is one of the first to directly examine successful *wh*-dependency resolution in L2 learners, providing evidence that dependency resolution proceeds in a native-like manner as proficiency increases. At the group level, learners only showed P600 at the actual gap site in the non-island condition, However, proficiency was shown to modulate processing at the gap site in both condition, such that highly proficient learners yielded larger P600s in the non-island condition and showed an emerging P600 in the island condition. This suggests that Mandarin-speaking learners with higher proficiency are better able to resolve the *wh*-dependency in both conditions. It further reveals that proficiency is particularly important in successfully resolving the dependency in the island condition, given that this sentence involves processing a complex relative clause island.

Future Research

The next stage of this study will focus on increasing the sample size of the L2 learner group. In doing so, we will be able to include scores from the cognitive measures in individual difference analyses, examining whether and how individual differences in cognitive abilities impact *wh*-dependency formation for L2 learners. This research will address proposals by Hopp (2013) and Kaan (2014), who suggest that the ability to predict in the L2 may be affected by individual differences in cognitive abilities. These analyses will also allow us to examine whether variability in processing is similarly modulated by cognitive abilities in both native speakers and learners. A future direction for this research includes testing a second L2 participant group whose L1 exhibits overt wh-movement, such as Spanish learners of English. Testing a second group of learners will reveal how and whether transfer from the L1 impacts wh-dependency resolution, address proposals which debate about the role of transfer in island sensitivity (e.g., Clahsen & Felser, 2006; Marinis et al., 2005; Kim et al., 2015). In addition, given the difference in native (N400) and non-native (P600) responses at the subject position, testing a second L2 group will allow us to investigate whether the ability to engage in pre-verbal syntactic prediction is restricted to learners whose L1 instantiates overt wh-movement.

An important next step for this research is to examine processing at other regions within the sentence. For example, we previously mentioned the importance of examining processing at the subcategorizing verb. Examining processing at the verb can further shed light on the late negativity that emerged at the object position for both the natives and learners, which we suggested was related to thematic argument conflict related to the verb. At the verb itself, it is possible to capture thematic role assignment processes: in the non-island condition, the verb can straightforwardly be associated with the *wh*-filler, assigning it a theta role. In contrast attempting to assign a thematic role to the *wh*-filler should not occur in the island, in which the verb is located inside the relative clause and is thus not a potential licensor for the filler. Another region that we plan to examine is the onset of the *wh*-dependency, at the *wh*-filler itself. Previous studies have reported sustained anterior negativities (typically described as LAN) emerging at the filler, persisting from the onset of the dependency until resolution at the verb site (e.g., Felser et al., 2003; Fiebach et al., 2002; Garnsey et al., 1989; King & Kutas, 1995; Kluender & Kutas 1993a/b). These studies have suggested that the anterior negativity reflects storage of the filler in working memory during processing of the rest of the sentence, a proposal that we will be able to directly test in an individual difference analysis including the working memory composite score.

Exploring the nature of islands themselves is an interesting avenue of research. The present study examines relative clause islands, which are considered to be “strong” islands in that extraction out of these positions is categorically prohibited. In contrast, extraction out of “weak” islands (e.g. *wh*-islands), is sometimes allowed. Comparing the processing of strong versus weak islands by native speakers and L2 learners may shed light on whether the parser categorically avoids positing gaps in islands during the processing of *wh*-dependencies or whether, in certain island contexts, *wh*-dependencies are in fact established (e.g., Belikova & White, 2009; Martohardjono, 1991).

Conclusions

The current study is unique in its systematic approach to examining syntactic prediction and *wh*-dependency formation, investigating the qualitative similarities and differences between native and L2 sentence processing. Overall, the results present a complex picture of *wh*-dependency resolution in native speakers and L2 learners. By tracking the dynamics of dependency resolution across the sentence, the study investigated whether native speakers and L2 learners of English engage in prediction during the processing of *wh*-dependencies, and furthermore, examined the type of information that natives and learners can use to predict. Results showed that the contexts in which learners engage in predictive processing differed from native speakers, such that learners were unable to predict gaps using abstract syntactic information at a pre-verbal subject position. Gap prediction for native speakers was grammatically guided, an emerging finding that was also observed for L2 learners with higher proficiency. The present study is one of the first studies to show that individual differences in cognitive abilities and offline island sensitivity modulate gap prediction and dependency resolution in grammatical sentences for native English speakers, a line of research that will be continued in the L2 learner population.

References

Aldosari, S. (2013). *The acquisition of Wh-movement in English by Najdi Arabic speakers.* Unpublished manuscript, Department of Linguistics, University of Kansas, Lawrence, Kansas.

Aldosari, S. (2015). *The role of individual differences in the acceptability of island violations in native and non-native speakers*. Doctoral dissertation. University of Kansas.

Aldwayan, S., Fiorentino, R., & Gabriele, A. (2010). Evidence of syntactic constraints in the processing of wh-movement: A study of Najdi Arabic learners of English. In B. Van Patten and J. Jegerski (Ed.), *Research in Second Language Processing and Parsing*, 65-86. Amsterdam: John Benjamins Publishing Company.

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

Altmann, G., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive science*, *33*(4), 583-609.

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*(3), 191-238.

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

Bar, M. (2009). The proactive brain: memory for predictions. *Philosophical Transactions of the Royal Society of Biological Sciences* 364: 1235-1243.

Bartek, B., Lewis, R. L., Vasishth, S., & Smith, M. R. (2011). In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1178.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. doi: arXiv:1406.5823.

Belikova, A., & White, L. (2009). Evidence for the fundamental difference hypothesis or not? Island constraints revisited. *Studies in Second Language Acquisition*, *31*(2), 199-223.

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*(6), 774-806.

Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2012). Cognitive control influences the use of meaning relations during spoken sentence comprehension. *Neuropsychologia*, *50*(11), 2659-2668.

Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2013). Effects of working memory span on processing of lexical associations and congruence in spoken discourse. *Frontiers in psychology*, *4*, 60.

Bourdages, J. S. (1992). Parsing complex NPs in French. In *Island constraints* (pp. 61-87). Springer, Dordrecht.

Boxell, O. & Felser, C. (2016). Sensitivity to parasitic gaps inside subject islands in native and non-native sentence processing. *Bilingualism: Language and Cognition* 20(3): 494-511.

Bush, G., Whalen, P. J., Shin, L. M., & Rauch, S. L. (2006). The counting Stroop: a cognitive interference task. *Nature protocols*, *1*(1), 230.

Canales, A. (2012). *Online processing of wh-dependencies in English by native speakers of Spanish.* Doctoral dissertation. University of Kansas.

Case, R., Kurland, M. D., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology* 33: 386-404.

Chomsky, N. (1973). Conditions on transformations. *A festschrift for Morris Halle* (pp. 232-286). New York: Holt, Rinehart, & Winston.

Chomsky, N. (1981). Principles and parameters in syntactic theory. In N. Hornstien & D. Lightfoot (Eds.), *Explanation in Linguistics* (pp. 32-75). Longman: Group Limited.

Chomsky, N. (1986): *Barriers*. Cambridge, MA: MIT Press.

Clahsen, H. & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics* 27(1): 3-42.

Clifton, C., Jr., & De Vincenzi, M. (1990). Comprehending sentences with empty elements. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension Processes in Reading*, 265-283. Hillsdale, NJ: Erlbaum.

Clifton, C. Jr. & Frazier, L. (1989). Comprehending sentences with long- distance dependencies. In G. Carlson and M. Tanenhaus (Ed.), *Linguistic Structure in Language Processing,* 94-128. Dordrecht: Kluwer.

Conway, A., Kane, M., Bunting, M., Hambrick, D., Wilhelm, O., & Engle, R. (2005). Working memory span tasks: a methodological review and user's guide. *Psychonomic Bulletin and Review* 12: 769-786.

Crain, S., & Fodor, J. D. (1985). How can grammars help parsers? Natural language parsing: psycholinguistic, computational, and theoretical perspectives, ed. by David Dowty, Lauri Kartunnen, and Arnold M. Zwicky, 94-128.

Cunnings, I., Batterham, C., Felser, C., & Clahsen, H. (2010). Constraints on L2 learners' processing of wh-dependencies. *Research in Second Language Processing and Parsing*, 87-112.

Dallas, A., DeDe, G., & Nicol, J. (2013). An event-related potential (ERP) investigation of filler-gap processing in native and second language speakers. *Language Learning* 63(4): 766-799.

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

de Vincenzi, M. (1991). *Syntactic Parsing Strategies in Italian*. Dordrecht: Kluwer.

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

Dong, Z. R. (2014). *Processing of tense morphology and filler-gap dependencies by Chinese second language speakers of English.* Doctoral dissertation, University of Delaware.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, *24*(6), 409-436.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19-23.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*(4), 469-495.

Federmeier, K. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology, 44*(4), 491-505.

Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*(2), 133-146.

Felser, C. & Roberts, L. (2007). Processing wh- dependencies in a second language: A cross-modal priming study. *Second Language Research* 23(1): 9-36.

Felser, C., Clahsen, H., & Münte, T. F. (2003). Storage and integration in the processing of filler-gap dependencies: An ERP study of topicalization and *wh*-movement in German. *Brain and Language* 87: 345-354.

Felser, C., Cunnings, I., Batterham, C., & Clahsen, H. (2012). The timing of island effects in nonnative sentence processing. *Studies in Second Language Acquisition* 34: 67-98.

Fiebach, C. J., Schlesewsky, M, & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language* 47: 250-272.

Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1461.

Frazier, L. (1987). Processing syntactic structures: Evidence from Dutch. *Natural Language and Linguistic Theory, 5*, 519-559.

Frazier, L. & Clifton, C. Jr. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes* 4(2): 93-126.

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

Frenzel, S., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2011). Conflicts in language processing: A new perspective on the N400–P600 distinction. *Neuropsychologia*, *49*(3), 574-579.

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

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78–84.

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(5), 1219–1248.

Friederici, A. D., Hahne, A., & von Cramon, D. Y. (1998). First-pass versus second-pass parsing processes in a Wernicke's and a Broca's aphasic: Electrophysiological evidence for a double dissociation. *Brain and Language*, *62*, 311–341.

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

Frisch, S., Schlesewsky, M., Saddy, D., & Alpermann, A. (2002). The P600 as an indicator of syntactic ambiguity. *Cognition*, *85*(3), B83-B92.

Frisch, S., & Schlesewsky, M. (2001). The N400 reflects problems of thematic hierarchizing. *NeuroReport*, *12*(15), 3391-3394.

Frisch, S., & Schlesewsky, M. (2005). The resolution of case conflicts from a neurophysiological perspective. *Cognitive Brain Research*, *25*(2), 484-498.

Garnsey, S. M., Tanenhaus, M. K., & Chapman, R. M. (1989). Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research* 18(1): 51-60.

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

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. *Image, Language, Brain*, 95-126.

Gibson, E., Hickok, G., & Schütze, C. T. (1994). Processing empty categories: A parallel approach. *Journal of Psycholinguistic Research* 23: 381-405.

Gibson, E., & Warren, T. (2004). Reading-Time Evidence for Intermediate Linguistic Structure in Long-Distance Dependencies. *Syntax*, *7*(1), 55-78.

Goodall, G. (2015). The D-linking effect on extraction from islands and non-islands. *Frontiers in Psychology*, doi: 10.3389/fpsyg.2014.01493

Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes* 25(2): 149-188.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive Science*, *29*(2), 261-290.

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research* 28(2): 191-215.

Grüter, T., Rohde, H., & Schafer, A.J. (2017). Coreference and discourse coherence in L2: The roles of grammatical aspect and referential form. *Linguistic Approaches to Bilingualism* 7(2): 199-229.

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia*, *38*, 1531-1549.

Hagoort, P., Brown, C. M., & Groosthusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, *8*(4), 439-483.

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research*, *30*(3), 251-266.

Hawkins, R. (2009) Statistical learning and innate knowledge in the development of second language proficiency: Evidence from the acquisition of gender concord. In Benati, A. G. (Ed.). *Issues in Second Language Proficiency*. Bloomsbury Publishing.

Hawkins, R., & Casillas, G. (2008). Explaining frequency of verb morphology in early L2 speech. *Lingua, 118*(4), 595-612.

Hawkins, R., & Chan, C. Y. (1997). The partial availability of Universal Grammar in second language acquisition: The 'failed functional features hypothesis.' *Second Language Research, 13*(3), 187–226.

Hestvik, A., Maxfield, N., Schwartz, R., & Shafer, V. (2007). Brain responses to filled gaps. *Brain and Language* 100: 301-316.

Hestvik, A., Bradley, E., & Bradley, C. (2012). Working memory effects of gap-predictions in normal adults: An event-related potentials study. *Journal of Psycholinguistic Research* 41: 425-438.

Hofmeister, P., Casasanto, L. S., & Sag, I. A. (2012a). How do individual cognitive differences relate to acceptability judgments? A reply to Sprouse, Wagers, and Phillips. *Language* 88(2): 390-40.

Hofmeister, P., Casasanto, L. S., & Sag, I. A. (2012b). Misapplying working memory tests: A reduction ad absurdum. *Language* 88(2): 408-409.

Hofmeister, P., Casasanto, L. S., & Sag, I. A. (2013). Islands in the grammar? Standards of evidence. In J. Sprouse and N. Hornstein (Ed.), *Experimental Syntax and Island Effects,* 42-63. Cambridge: Cambridge University Press.

Hofmeister, P., Casasanto, L. S., & Sag, I. A. (2014). Processing effects in linguistic judgment data:(super-) additivity and reading span scores. *Language and Cognition*, *6*(1), 111-145.

Hofmeister, P. & Sag, I. A. (2010). Cognitive constraints and island effects. *Languag*e 86(2): 366-415.

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research* 29: 33-56.

Hsu, N. S., & Novick, J. M. (2016). Dynamic engagement of cognitive control modulated recovery from misinterpretation during real-time language processing. *Psychological Science*, *27*(4), 572-582.

Hutchison, K. (2007). Attentional control and the relatedness proportion effect in semantic priming. *Journal of Experimental Psychology* 33(4): 645-662.

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, *127*(1), 57-83.

Jessen, A. & Felser, C. (2018). Reanalysing object gaps during non-native sentence processing: Evidence from ERPs. *Second Language Research*. doi: https://doi.org/10.1177%2F0267658317753030

Jessen, A., Festman, J., Boxell, O., & Felser, C. (2017). Native and non-native speakers' brain responses to filled indirect object gaps. *Journal of Psycholinguistic Research*, *46*(5), 1319-1338.

Johnson, A. (2015). *Individual differences in predictive processing: Evidence from subject filled-gap effects in native and non-native speakers of English*. Doctoral dissertation. University of Kansas.

Johnson, A., Fiorentino, R., & Gabriele, A. (2016). Syntactic constraints and individual differences in native and non-native processing of wh-movement. *Frontiers in Psychology* 7:549. doi: 10.3389/fpsyg.2016.00549.

Johnson, J. S., & Newport, E. L. (1991). Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language. *Cognition*, *39*(3), 215-258.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review*, *99*(1), 122.

Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996). Brain activation modulated by sentence comprehension. *Science*, *274*(5284), 114-116.

Kaan, E. (2007). Event-related potentials and language processing: A brief overview. *Language and Linguistics Compass*, *1*(6), 571-591.

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism* 4: 257-282.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes* 15: 159-201.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and language*, *49*(1), 133-156.

Kane, M. & Engle, R. (2002). The role of prefrontal cortex in working memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin and Review* 9(4): 637-671.

Kim, E., Baek, S., & Tremblay, A. (2015). The role of island constraints in second language sentence processing. *Language Acquisition* 4: 384-416.

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

King, J. W. & Kutas, M. (1995). Who did what and when? Using word-and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience* 7(3), 376-395.

Kluender, R. (2004). Are subject islands subject to a processing account? In A. Rodríguez, V. Chand, A. Kelleher, and B. Scheiser (Ed.), *Proceedings of West Coast Conference on Formal Linguistics* 23: 475-499. Somerville, MA: Cascadilla Press.

Kluender, R. & Kutas, M. (1993a). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience* 5: 196-214.

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

Kutas, M., DeLong, K. A., & Smith, N.J. (2011). A look at what lies ahead: Prediction and predictability in language processing. In: M. Bar (Ed.), *Predictions in the Brain: Using Our Past to Generate a Future* (pp. 190-202). Oxford University Press.

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

Kutas, M. & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307(5947): 161-163.

Kutas, M., Van Petten, C. K., & Kluender, R. (2006). Psycholinguistics electrified II (1994–2005). In *Handbook of Psycholinguistics (Second Edition)* (pp. 659-724).

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package ‘lmerTest’. R package version, 2-0.

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience, 25*(3), 484-502.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: [De]constructing the N400. *National Review of Neuroscience, 9*, 920–933.

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research, 25*(2), 173-227.

Leal, T., Slabakova, R., & Farmer, T. A. (2017). The fine-tuning of linguistic expectations over the course of L2 learning. *Studies in Second Language Acquisition*, *39*(3), 493-525.

Lee, M. W. (2004). Another look at the role of empty categories in sentence processing (and grammar). *Journal of Psycholinguistic Research* 33, 51-73.

Lemhöfer, K. & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods* 44(2): 325-343.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126-1177.

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, *63*(4), 447-464.

Li, X (1998). Adult L2 accessibility to UG: An issue revised. In: S. Flynn, G. Martohardjono & W. A. O'Neil (Eds.), *The generative study of second language acquisition* (pp. 89-110). Mahwah: Erlbaum.

Lin, T.-H. (2006). Syntactic structures of complex sentences in Mandarin Chinese. *Nanzan Linguistics* 3: 63-97.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*, 226. DOI: 10.3389/fpsyg.2013.00226

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition* 27: 53-78.

Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, *189*(4198), 226-228.

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

Martin, C. D., Thierry, G., Kuipers, J.-R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language* 69: 574-588.

Martohardjono, G. (1993). *Wh-movement in the acquisition of a second language: a crosslinguistic study of three languages with and without overt movement.* Doctoral dissertation, Cornell University.

Michel, D. (2014). *Individual cognitive measures and working memory accounts of syntactic island phenomena*. Doctoral dissertation. University of California, San Diego.

Mitsugi, S., & MacWhinney, B. (2016). The use of case marking for predictive processing in second language Japanese. *Bilingualism: Language and Cognition*, *19*(1), 19-35.

Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex, 47*(8), 908-930.

Nakano, Y., Felser, C., & Clahsen, H. (2002). Antecedent priming at trace positions in Japanese long-distance scrambling. *Journal of Psycholinguistic Research* 31(5): 531-571.

Nicenboim, B., Vasishth, S., Gattei, C., Sigman, M., & Kliegl, R. (2015). Working memory differences in long-distance dependency resolution. *Frontiers in Psychology* 6:312. doi: 10.3389/fpsyg.2015.00312.

Nicol, J. L. (1993). Reconsidering reactivation. In *Cognitive models of speech processing: The second Sperlonga meeting* (pp. 321-350). Erlbaum: Hillsdale, NJ.

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, *18*(1), 5-19.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* 9(1), 97-113.

Omaki, A. & Schulz, B. (2011). Filler-gap dependencies and island constraints in second language sentence processing. *Studies in Second Language Acquisition* 33: 563-588.

Omaki, A., Lau, E., Davidson White, I., Dakan, M., Apple, A., & Phillips, C. (2015). Hyper-active gap filling. *Frontiers in Psychology, 6*, 384. doi:10.3389/fpsyg.2015.00384

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*(6), 785-806.

Phillips, C. (2006). The real-time status of island phenomena. *Language* 82: 795-823.

Phillips, C. (2013). On the nature of island constraints I: Language processing and reductionist accounts. In J. Sprouse and N. Hornstein (Ed.), *Experimental Syntax and Island Effects*, 64-111. Cambridge: Cambridge University Press.

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research* 22(3): 407-428.

Pickering, M. (1993). Direct association and sentence processing: A reply to Gorrell and to Gibson and Hickok. *Language and Cognitive Processes*, *8*(2), 163-196.

Pickering, M. J. (1994). Processing local and unbounded dependencies: A unified account. *Journal of Psycholinguistic Research*, *23*(4), 323-352.

Pickering, M. J., & Barry, G. D. (1991). Sentence processing without empty categories. *Language and Cognitive Processes, 6*(3), 229–259.

Pickering, M. J., & Garrod, S. (2011). The use of prediction to drive alignment in dialogue. *Grounding sociality. Neurons, Mind and Culture,* 175-191.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329-347.

Pickering, M. J., & Traxler, M. J. (2001). Strategies for processing unbounded dependencies: Lexical information and verb–argument assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(6), 1401-1410.

Pickering, M. J., & Traxler, M. J. (2003). Evidence against the use of subcategorisation frequency in the processing of unbounded dependencies. *Language and Cognitive Processes*, *18*(4), 469-503.

Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second language research*, *16*(2), 103-133.

Qu, Y. (1994). *Object noun phrase dislocation in Mandarin Chinese*. Doctoral dissertation. University of British Columbia.

Roberts, L., Gonzalez Alonso, J., Pliatsikas, C., & Rothman, J. (2018). Evidence from neurolinguistic methodologies: Can it actually inform linguistic/language acquisition theories and translate to evidence-based applications? *Second Language Research, 34*(1), 125-143.

Ross, J. (1967). *Constraints on variables in syntax*. Doctoral dissertation. MIT.

Schachter, J. (1990). On the issue of completeness in second language acquisition. *Second Language Research*, *6*(2), 93-124.

Schremm, A. (2012). *Processing filler-gap dependencies in an L2: An event-related potential study*. Unpublished master's thesis. Lund University.

Schremm, A. (2013). *Event-related brain potentials and the processing of filled gaps in English relative clauses.* Unpublished master's thesis. Lund University.

Schwartz, B. D., & Sprouse, R. (1994). Word order and nominative case in nonnative language acquisition: a longitudinal study of (L1 Turkish) German interlanguage. *Language Acquisition Studies in Generative Grammar*, *31*(4), 71-89.

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, *12*(1), 40-72.

Sohn, H. M. (1980). Theme prominence in Korean. *Korean Linguistics* 2: 2-19.

Sohn, H. M. (1999). *The Korean Language*. Cambridge: Cambridge University Press.

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

Steedman, M. J. (1989). Gapping as constituent coordination. *Linguistics and Philosophy*, *13*(2), 207-263.

Stowe, L. (1986). Parsing wh- constructions: Evidence for online gap location. *Language and Cognitive Processes* 1: 227-245.

Stowe, L. A., Tanenhaus, M. K., & Carlson, G. N. (1991). Filling gaps on-line: Use of lexical and semantic information in sentence processing. *Language and Speech*, *34*(4), 319-340.

Tagliaferri, B. (2005). Paradigm. Perception Research Systems, Inc. www.perceptionresearchsystems.com.

Tanenhaus, M. K., Boland, J., Garnsey, S. M., & Carlson, G. N. (1989). Lexical structure in parsing long-distance dependencies. *Journal of Psycholinguistic Research*, *18*(1), 37-50.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632-1634.

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

Traxler, M. J., Bybee, M. D., & Pickering, M. J. (1997). Influence of connectives on language comprehension: eye tracking evidence for incremental interpretation. *The Quarterly Journal of Experimental Psychology Section A*, *50*(3), 481-497.

Tremblay, A. & Ransijn, J. (2015). LMERConvenienceFunctions: A suite of functions to back-fit fixed effects and forward-fit random effects, as well as other miscellaneous functions. R package version 2.5.

Tsimpli, I. M., & Dimitrakopoulou, M. (2007). The interpretability hypothesis: Evidence from wh-interrogatives in second language acquisition. *Second Language Research*, *23*(2), 215-242.

Tyler, L. K., & Marslen-Wilson, W. D. (1977). The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(6), 683-692.

Van Berkum, J., Brown, C., Zwisterlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 443-467.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176-190

Vasishth, S., & Drenhaus, H. (2011). Locality in german. *Dialogue & Discourse*, *2*(1), 59-82.

Wagers, M. & Phillips, C. (2009). Multiple dependencies and the role of the grammar in real-time comprehension. *Journal of Linguistics* 45(2): 395-433.

Williams, J. N. (2006). Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition*, *9*(1), 71-88.

Williams, J. N., Möbius, P., & Kim, C. (2001). Native and non-native processing of English wh-questions: Parsing strategies and plausibility constraints. *Applied Psycholinguistics*, *22*(4), 509-540.

Appendix A: Statistical Tables, Native Speaker study

Table 28: Linear mixed-effects best-fitting overall model for object position from 300-500 ms, with no individual difference measures included (native speakers)

Number of obs: 5456, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-3.435	0.224	103	-15.348	< .001	***
Extraction	-0.014	0.136	5395	-0.106	0.915	
Island	0.161	0.136	5395	1.186	0.236	
Hemisphere (left)	1.002	0.132	5395	7.614	< .001	***
Hemisphere (right)	1.788	0.132	5395	13.580	< .001	***
Anteriority (ant)	0.019	0.178	5395	0.108	0.914	
Anteriority (post)	2.101	0.197	5395	10.678	< .001	***
Extraction × Island	-0.629	0.192	5395	-3.273	0.001	**
Extraction × Anteriority (ant)	-0.200	0.181	5395	-1.107	0.268	
Extraction × Anteriority (post)	-0.028	0.204	5395	-0.139	0.890	
Island × Anteriority (ant)	-0.308	0.181	5395	-1.703	0.089	.
Island × Anteriority (post)	-0.004	0.204	5395	-0.019	0.985	
Hemisphere (left) × Anteriority (ant)	-0.706	0.172	5395	-4.101	< .001	***
Hemisphere (right) × Anteriority (ant)	-1.149	0.172	5395	-6.671	< .001	***
Hemisphere (left) × Anteriority (post)	-0.499	0.191	5395	-2.607	0.009	**
Hemisphere (right) × Anteriority (post)	-0.580	0.191	5395	-3.033	0.002	**
Extraction × Island × Anteriority (ant)	0.591	0.256	5395	2.309	0.021	*
Extraction × Island × Anteriority (post)	0.058	0.288	5395	0.200	0.841	

Table 29: Linear mixed-effects best-fitting overall model for subject position from 300-500 ms, with Working Memory included in model (native speakers)

Number of obs: 2728, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-1.708	0.224	101	-7.612	< .001	***
Extraction	-0.338	0.113	2668	-3.008	0.003	**
Hemisphere (left)	1.183	0.154	2668	7.677	< .001	***
Hemisphere (right)	1.536	0.154	2668	9.972	< .001	***
Anteriority (ant)	0.436	0.179	2668	2.437	0.015	*
Anteriority (post)	1.911	0.197	2668	9.705	< .001	***
Working Memory	-0.224	0.116	56.9	-1.932	0.058	·
Extraction × Anteriority (ant)	0.164	0.150	2668	1.097	0.273	
Extraction × Anteriority (post)	-0.436	0.169	2668	-2.582	0.010	**
Hemisphere (left) × Anteriority (ant)	-0.733	0.202	2668	-3.637	< .001	***
Hemisphere (right) × Anteriority (ant)	-1.172	0.202	2668	-5.815	< .001	***
Hemisphere (left) × Anteriority (post)	-0.720	0.224	2668	-3.217	0.001	**
Hemisphere (right) × Anteriority (post)	-0.499	0.224	2668	-2.227	0.026	*
Extraction × Working Memory	0.145	0.067	2668	2.157	0.031	*
Anteriority (ant) × Working Memory	-0.177	0.063	2668	-2.799	0.005	*
Anteriority (post) × Working Memory	0.224	0.071	2668	3.150	0.002	**
Extraction × Anteriority (ant) × Working Memory	0.205	0.089	2668	2.292	0.022	*
Extraction × Anteriority (post) × Working Memory	-0.130	0.101	2668	-1.291	0.197	

Table 30: Linear mixed-effects best-fitting overall model for subject position from 300-500 ms, with Attentional Control included in model (native speakers)

Number of obs: 2728, groups: Subject, 44						
	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	-1.708	0.228	99.2	-7.494	< .001	***
Extraction	-0.338	0.113	2668	-2.985	0.003	**
Hemisphere (left)	1.183	0.155	2668	7.621	< .001	***
Hemisphere (right)	1.536	0.155	2668	9.898	< .001	***
Anteriority (ant)	0.436	0.180	2668	2.419	0.016	*
Anteriority (post)	1.911	0.198	2668	9.634	< .001	***
Attentional Control	-0.127	0.124	56.5	-1.030	0.307	
Extraction × Anteriority (ant)	0.164	0.151	2668	1.089	0.276	
Extraction × Anteriority (post)	-0.436	0.170	2668	-2.563	0.010	*
Hemisphere (left) × Anteriority (ant)	-0.733	0.203	2668	-3.611	< .001	***
Hemisphere (right) × Anteriority (ant)	-1.172	0.203	2668	-5.772	< .001	***
Hemisphere (left) × Anteriority (post)	-0.720	0.226	2668	-3.193	0.001	**
Hemisphere (right) × Anteriority (post)	-0.499	0.226	2668	-2.210	0.027	*
Extraction × Attentional Control	0.133	0.071	2668	1.882	0.060	
Anteriority (ant) × Attentional Control	0.225	0.067	2668	3.374	< .001	.
Anteriority (post) × Attentional Control	0.019	0.075	2668	0.252	0.801	
Extraction × Anteriority (ant) × Attentional Control	-0.216	0.094	2668	-2.289	0.022	.
Extraction × Anteriority (post) × Attentional Control	0.143	0.106	2668	1.349	0.177	

Table 31: Linear mixed-effects best-fitting overall model for object position from 300-500 ms, with Working Memory included in model (native speakers)

Number of obs: 5456, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-3.435	0.225	98	-15.280	< .001	***
Extraction	-0.014	0.135	5387	-0.107	0.915	
Island	0.161	0.135	5387	1.192	0.233	
Hemisphere (left)	1.002	0.131	5387	7.653	< .001	***
Hemisphere (right)	1.788	0.131	5387	13.649	< .001	***
Anteriority (ant)	0.019	0.177	5387	0.109	0.914	
Anteriority (post)	2.101	0.196	5387	10.732	< .001	***
Working Memory	-0.080	0.120	63	-0.668	0.507	
Extraction × Island	-0.629	0.191	5387	-3.289	0.001	**
Extraction × Anteriority (ant)	-0.200	0.180	5387	-1.112	0.266	
Extraction × Anteriority (post)	-0.028	0.203	5387	-0.139	0.889	
Island × Anteriority (ant)	-0.308	0.180	5387	-1.711	0.087	.
Island × Anteriority (post)	-0.004	0.203	5387	-0.019	0.985	
Hemisphere (left) × Anteriority (ant)	-0.706	0.171	5387	-4.121	< .001	***
Hemisphere (right) × Anteriority (ant)	-1.149	0.171	5387	-6.705	< .001	***
Hemisphere (left) × Anteriority (post)	-0.499	0.190	5387	-2.620	0.009	**
Hemisphere (right) × Anteriority (post)	-0.580	0.190	5387	-3.049	0.002	**
Extraction × Working Memory	0.087	0.057	5387	1.517	0.129	
Island × Working Memory	-0.138	0.032	5387	-4.258	< .001	***
Hemisphere (left) × Working Memory	-0.079	0.043	5387	-1.830	0.067	.
Hemisphere (right) × Working Memory	0.095	0.043	5387	2.200	0.028	*
Anteriority (ant) × Working Memory	0.103	0.054	5387	1.921	0.055	*
Anteriority (post) × Working Memory	0.142	0.061	5387	2.342	0.019	*
Extraction × Island × Anteriority (ant)	0.591	0.255	5387	2.321	0.020	*

Extraction × Island × Anteriority (post)	0.058	0.287	5387	0.201	0.841	
Extraction × Anteriority (ant) × Working Memory	-0.149	0.076	5387	-1.960	0.050	*
Extraction × Anteriority (post) × Working Memory	0.033	0.086	5387	0.382	0.702	

Table 32: Linear mixed-effects best-fitting overall model for object position from 300-500 ms, with Attentional Control included in model (native speakers)

Number of obs: 5456, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-3.435	0.225	97	-15.292	< .001	***
Extraction	-0.014	0.134	5382	-0.107	0.914	
Island	0.161	0.134	5382	1.200	0.230	
Hemisphere (left)	1.002	0.130	5382	7.703	< .001	***
Hemisphere (right)	1.788	0.130	5382	13.739	< .001	***
Anteriority (ant)	0.019	0.176	5382	0.109	0.913	
Anteriority (post)	2.101	0.195	5382	10.802	< .001	***
Attentional Control	-0.013	0.131	74	-0.102	0.919	
Extraction × Island	-0.629	0.190	5382	-3.311	< .001	***
Extraction × Anteriority (ant)	-0.200	0.179	5382	-1.120	0.263	
Extraction × Anteriority (post)	-0.028	0.202	5382	-0.140	0.888	
Island × Anteriority (ant)	-0.308	0.179	5382	-1.723	0.085	.
Island × Anteriority (post)	-0.004	0.202	5382	-0.019	0.985	
Hemisphere (left) × Anteriority (ant)	-0.706	0.170	5382	-4.149	< .001	***
Hemisphere (right) × Anteriority (ant)	-1.149	0.170	5382	-6.749	< .001	***
Hemisphere (left) × Anteriority (post)	-0.499	0.189	5382	-2.638	0.008	**
Hemisphere (right) × Anteriority (post)	-0.580	0.189	5382	-3.069	0.002	**
Extraction × Attentional Control	0.124	0.084	5382	1.480	0.139	
Island × Attentional Control	-0.010	0.084	5382	-0.115	0.908	
Hemisphere (left) × Attentional Control	-0.084	0.045	5382	-1.883	0.060	.
Hemisphere (right) × Attentional Control	0.033	0.045	5382	0.730	0.465	
Anteriority (ant) × Attentional Control	-0.057	0.079	5382	-0.726	0.468	
Anteriority (post) × Attentional Control	0.320	0.089	5382	3.591	< .001	***
Extraction × Island × Anteriority (ant)	0.591	0.253	5382	2.336	0.020	*
Extraction × Island × Anteriority (post)	0.058	0.285	5382	0.202	0.840	
Extraction × Island × Attentional Control	-0.357	0.119	5382	-3.004	0.003	**
Extraction × Anteriority (ant) × Attentional Control	0.102	0.112	5382	0.916	0.360	
Extraction × Anteriority (post) × Attentional Control	-0.143	0.126	5382	-1.138	0.255	
Island × Anteriority (ant) × Attentional Control	0.011	0.112	5382	0.095	0.924	
Island × Anteriority (post) × Attentional Control	-0.049	0.126	5382	-0.392	0.695	
Extraction × Island × Anteriority (ant) × Attentional Control	-0.113	0.158	5382	-0.715	0.474	
Extraction × Island × Anteriority (post) × Attentional Control	0.296	0.178	5382	1.662	0.097	.

Table 33: Linear mixed-effects best-fitting model, Non-Island condition, for object position from 300-500 ms, with Attentional Control included in model (native speakers)

Number of obs: 2728, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-3.434	0.266	106	-12.921	< .001	***
Extraction	-0.472	0.082	2672	-5.732	< .001	***
Anteriority (ant)	-0.103	0.209	2672	-0.490	0.624	
Anteriority (post)	2.233	0.229	2672	9.732	< .001	***
Hemisphere (left)	1.109	0.199	2672	5.582	< .001	***
Hemisphere (right)	1.868	0.199	2672	9.402	< .001	***
Attentional Control	-0.061	0.139	52.6	-0.439	0.663	

Anteriority (ant) × Hemisphere (left)	-0.696	0.260	2672	-2.678	0.007	**
Anteriority (post) × Hemisphere (left)	-0.611	0.289	2672	-2.116	0.034	*
Anteriority (ant) × Hemisphere (right)	-1.128	0.260	2672	-4.340	< .001	***
Anteriority (post) × Hemisphere (right)	-0.779	0.289	2672	-2.697	0.007	**
Extraction × Attentional Control	-0.198	0.052	2672	-3.836	< .001	***
Anteriority (ant) × Attentional Control	-0.051	0.060	2672	-0.850	0.395	
Anteriority (post) × Attentional Control	0.348	0.068	2672	5.119	< .001	***

Table 34: Linear mixed-effects best-fitting model, Island condition, for object position from 300-500 ms, with Attentional Control included in model (native speakers)

Number of obs: 2728, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-3.360	0.234	90.6	-14.342	< .001	***
Extraction	-0.014	0.111	2666	-0.130	0.897	
Anteriority (ant)	0.028	0.177	2666	0.160	0.873	
Anteriority (post)	1.980	0.195	2666	10.162	< .001	***
Hemisphere (left)	0.896	0.152	2666	5.876	< .001	***
Hemisphere (right)	1.708	0.152	2666	11.203	< .001	***
Attentional Control	-0.007	0.134	64.3	-0.049	0.961	
Extraction × Anteriority (ant)	-0.200	0.148	2666	-1.352	0.177	
Extraction × Anteriority (post)	-0.028	0.167	2666	-0.169	0.866	
Anteriority (ant) × Hemisphere (left)	-0.716	0.199	2666	-3.593	< .001	***
Anteriority (post) × Hemisphere (left)	-0.387	0.222	2666	-1.745	0.081	.
Anteriority (ant) × Hemisphere (right)	-1.170	0.199	2666	-5.866	< .001	***
Anteriority (post) × Hemisphere (right)	-0.382	0.222	2666	-1.724	0.085	.
Extraction × Attentional Control	0.124	0.070	2666	1.787	0.074	.
Anteriority (ant) × Attentional Control	-0.058	0.065	2666	-0.881	0.379	
Anteriority (post) × Attentional Control	0.320	0.074	2666	4.329	< .001	***
Hemisphere (left) × Attentional Control	-0.115	0.052	2666	-2.201	0.028	*
Hemisphere (right) × Attentional Control	0.047	0.052	2666	0.897	0.370	
Extraction × Anteriority (ant) × Attentional Control	0.102	0.093	2666	1.106	0.269	
Extraction × Anteriority (post) × Attentional Control	-0.143	0.104	2666	-1.374	0.170	

Table 35: Linear mixed-effects best-fitting overall model for object position from 300-500 ms, with DD scores included in model (native speakers)

Number of obs: 5456, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-3.547	0.713	51	-4.978	< .001	***
Extraction	-0.755	0.311	5392	-2.430	0.015	
Island	-0.431	0.311	5392	-1.387	0.166	
Hemisphere (left)	1.002	0.132	5392	7.620	< .001	***
Hemisphere (right)	1.788	0.132	5392	13.590	< .001	***
Anteriority (ant)	0.019	0.177	5392	0.108	0.914	
Anteriority (post)	2.101	0.197	5392	10.685	< .001	***
DD scores	0.068	0.410	48	0.166	0.868	
Extraction × Island	0.627	0.439	5392	1.428	0.153	**
Extraction × Anteriority (ant)	-0.200	0.181	5392	-1.108	0.268	
Extraction × Anteriority (post)	-0.028	0.204	5392	-0.139	0.890	
Island × Anteriority (ant)	-0.308	0.181	5392	-1.704	0.088	.
Island × Anteriority (post)	-0.004	0.204	5392	-0.019	0.985	
Hemisphere (left) × Anteriority (ant)	-0.706	0.172	5392	-4.103	< .001	***
Hemisphere (right) × Anteriority (ant)	-1.149	0.172	5392	-6.676	< .001	***

Hemisphere (left) × Anteriority (post)	-0.499	0.191	5392	-2.609	0.009	**
Hemisphere (right) × Anteriority (post)	-0.580	0.191	5392	-3.035	0.002	**
Extraction × DD scores	0.448	0.169	5392	2.651	0.008	**
Island × DD scores	0.358	0.169	5392	2.120	0.034	*
Extraction × Island × Anteriority (ant)	0.591	0.256	5392	2.311	0.021	*
Extraction × Island × Anteriority (post)	0.058	0.288	5392	0.200	0.841	
Extraction × Island × DD scores	-0.761	0.239	5392	-3.182	0.001	**

Table 36: Linear mixed-effects best-fitting model, Non-Island condition, for object position from 300-500 ms, with DD scores included in model (native speakers)

Number of obs: 2728, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-3.434	0.266	111.2	-12.92	< .001	***
Extraction	-0.472	0.083	2675	-5.675	< .001	***
Anteriority (ant)	-0.103	0.212	2675	-0.485	0.628	
Anteriority (post)	2.233	0.232	2675	9.636	< .001	***
Hemisphere (left)	1.109	0.201	2675	5.527	< .001	***
Hemisphere (right)	1.868	0.201	2675	9.309	< .001	***
Anteriority (ant) × Hemisphere (left)	-0.696	0.263	2675	-2.651	0.008	**
Anteriority (post) × Hemisphere (left)	-0.611	0.292	2675	-2.095	0.036	*
Anteriority (ant) × Hemisphere (right)	-1.128	0.263	2675	-4.297	< .001	***
Anteriority (post) × Hemisphere (right)	-0.779	0.292	2675	-2.67	0.008	**

Table 37: Linear mixed-effects best-fitting model, Island condition, for object position from 300-500 ms, with DD scores included in model (native speakers)

Number of obs: 2728, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-3.406	0.783	57	-4.349	< .001	***
Extraction	-0.854	0.421	2668	-2.030	0.042	
Anteriority (ant)	-0.046	0.421	2668	-0.110	0.913	
Anteriority (post)	1.841	0.473	2668	3.893	< .001	***
Hemisphere (left)	0.896	0.154	2668	5.834	< .001	***
Hemisphere (right)	1.708	0.154	2668	11.123	< .001	***
DD scores	0.028	0.452	54.7	0.061	0.952	
Extraction × Anteriority (ant)	-0.588	0.560	2668	-1.050	0.294	
Extraction × Anteriority (post)	0.988	0.631	2668	1.565	0.118	
Anteriority (ant) × Hemisphere (left)	-0.716	0.201	2668	-3.567	< .001	***
Anteriority (post) × Hemisphere (left)	-0.386	0.223	2668	-1.732	0.083	.
Anteriority (ant) × Hemisphere (right)	-1.170	0.201	2668	-5.824	< .001	***
Anteriority (post) × Hemisphere (right)	-0.382	0.223	2668	-1.712	0.087	.
Extraction × DD scores	0.508	0.246	2668	2.070	0.039	*
Anteriority (ant) × DD scores	0.045	0.231	2668	0.195	0.845	
Anteriority (post) × DD scores	0.084	0.261	2668	0.323	0.747	
Extraction × Anteriority (ant) × DD scores	0.235	0.327	2668	0.718	0.473	
Extraction × Anteriority (post) × DD scores	-0.615	0.368	2668	-1.670	0.095	.

Table 38: Linear mixed-effects best-fitting overall model for object position from 500-900 ms, with Working Memory included in model (native speakers)

Number of obs: 5456, groups: Subject, 44					
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**
(Intercept)	-0.181	0.182	87	-0.994	0.323

Extraction	-0.151	0.089	5388	-1.689	0.091	.
Island	0.031	0.132	5388	0.238	0.812	
Hemisphere (left)	-0.218	0.095	5388	-2.293	0.022	*
Hemisphere (right)	0.036	0.095	5388	0.373	0.709	
Anteriority (ant)	-0.464	0.103	5388	-4.503	< .001	***
Anteriority (post)	0.062	0.116	5388	0.536	0.592	
Working Memory	0.263	0.109	87	2.421	0.018	*
Island × Hemisphere (left)	0.079	0.135	5388	0.591	0.555	
Island × Hemisphere (right)	0.125	0.135	5388	0.931	0.352	
Extraction × Anteriority (ant)	0.164	0.119	5388	1.380	0.168	
Extraction × Anteriority (post)	-0.112	0.134	5388	-0.833	0.405	
Island × Anteriority (ant)	-0.344	0.119	5388	-2.891	0.004	**
Island × Anteriority (post)	0.489	0.134	5388	3.639	< .001	***
Extraction × Working Memory	0.026	0.053	5388	0.478	0.633	
Island × Working Memory	-0.055	0.079	5388	-0.696	0.487	
Hemisphere (left) × Working Memory	-0.006	0.057	5388	-0.099	0.921	
Hemisphere (right) × Working Memory	-0.118	0.057	5388	-2.071	0.038	*
Anteriority (ant) × Working Memory	0.064	0.061	5388	1.044	0.296	
Anteriority (post) × Working Memory	-0.182	0.069	5388	-2.619	0.009	**
Island × Hemisphere (left) × Working Memory	-0.073	0.080	5388	-0.907	0.365	
Island × Hemisphere (right) × Working Memory	0.111	0.080	5388	1.379	0.168	
Extraction × Anteriority (ant) × Working Memory	-0.159	0.071	5388	-2.236	0.025	*
Extraction × Anteriority (post) × Working Memory	0.072	0.080	5388	0.901	0.368	
Island × Anteriority (ant) × Working Memory	-0.095	0.071	5388	-1.344	0.179	
Island × Anteriority (post) × Working Memory	0.116	0.080	5388	1.446	0.148	

Table 39: Linear mixed-effects best-fitting overall model for object position from 500-900 ms, with Attentional Control included in model (native speakers)

Number of obs: 5456, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-0.231	0.178	69	-1.300	0.198	
Extraction	-0.133	0.072	5398	-1.853	0.064	.
Island	0.092	0.102	5398	0.896	0.370	
Hemisphere (left)	-0.178	0.067	5398	-2.662	0.008	**
Hemisphere (right)	0.098	0.067	5398	1.465	0.143	
Anteriority (ant)	-0.380	0.084	5398	-4.537	< .001	***
Anteriority (post)	0.009	0.095	5398	0.095	0.925	
Attentional Control	-0.069	0.106	57	-0.653	0.516	
Extraction × Island	0.043	0.101	5398	0.425	0.671	
Island × Anteriority (ant)	-0.347	0.119	5398	-2.930	0.003	**
Island × Anteriority (post)	0.483	0.134	5398	3.618	< .001	***
Extraction × Attentional Control	0.070	0.045	5398	1.573	0.116	
Island × Attentional Control	0.121	0.045	5398	2.707	0.007	**
Hemisphere (left) × Attentional Control	0.006	0.042	5398	0.154	0.878	
Hemisphere (right) × Attentional Control	0.095	0.042	5398	2.268	0.023	*
Extraction × Island × Attentional Control	-0.402	0.063	5398	-6.364	< .001	***

Table 40: Linear mixed-effects best-fitting model, Non-Island condition, for object position from 500-900 ms, with Attentional Control included in model (native speakers)

Number of obs: 2728, groups: Subject, 44					
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**
(Intercept)	-0.106	0.213	71.7	-0.496	0.622

Extraction	-0.239	0.122	2676	-1.951	0.051	.
Anteriority (ant)	-0.905	0.115	2676	-7.874	< .001	***
Anteriority (post)	0.498	0.130	2676	3.841	< .001	***
Hemisphere (left)	-0.139	0.092	2676	-1.508	0.132	
Hemisphere (right)	0.161	0.092	2676	1.749	0.080	.
Attentional Control	0.091	0.119	45	0.768	0.446	
Extraction × Anteriority (ant)	0.359	0.163	2676	2.209	0.027	*
Extraction × Anteriority (post)	-0.007	0.183	2676	-0.036	0.971	
Extraction × Attentional Control	-0.332	0.043	2676	-7.654	< .001	***

Table 41: Linear mixed-effects best-fitting model, Island condition, for object position from 500-900 ms, with Attentional Control included in model (native speakers)

Number of obs: 2728, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-0.256	0.187	64.2	-1.372	0.175	
Anteriority (ant)	-0.382	0.080	2678	-4.784	< .001	***
Anteriority (post)	0.006	0.090	2678	0.071	0.943	
Hemisphere (left)	-0.218	0.090	2678	-2.417	0.016	*
Hemisphere (right)	0.036	0.090	2678	0.394	0.694	
Attentional Control	-0.053	0.112	54.7	-0.472	0.639	
Hemisphere (left) × Attentional Control	0.014	0.056	2678	0.245	0.807	
Hemisphere (right) × Attentional Control	0.136	0.056	2678	2.419	0.016	*

Table 42: Linear mixed-effects best-fitting overall model for object position from 500-900 ms, with DD scores included in model (native speakers)

Number of obs: 5456, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	1.069	0.628	55	1.702	0.094	.
Extraction	-1.563	0.269	5398	-5.805	< .001	***
Island	-0.276	0.279	5398	-0.988	0.323	
Hemisphere (left)	-0.178	0.067	5398	-2.654	0.008	**
Hemisphere (right)	0.098	0.067	5398	1.460	0.144	
Anteriority (ant)	-0.600	0.231	5398	-2.601	0.009	**
Anteriority (post)	-0.606	0.260	5398	-2.327	0.020	*
DD scores	-0.787	0.365	54	-2.158	0.035	*
Extraction × Island	1.026	0.381	5398	2.693	0.007	**
Island × Anteriority (ant)	-0.347	0.119	5398	-2.922	0.003	**
Island × Anteriority (post)	0.483	0.134	5398	3.608	< .001	***
Extraction × DD scores	0.866	0.157	5398	5.512	< .001	***
Island × DD scores	0.223	0.157	5398	1.417	0.157	
Anteriority (ant) × DD scores	0.133	0.130	5398	1.023	0.306	
Anteriority (post) × DD scores	0.372	0.147	5398	2.536	0.011	*
Extraction × Island × DD scores	-0.595	0.222	5398	-2.677	0.007	**

Table 43: Linear mixed-effects best-fitting model, Non-Island condition, for object position from 500-900 ms, with DD scores included in model (native speakers)

Number of obs: 2728, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-0.106	0.213	74.6	-0.497	0.621	
Extraction	-0.239	0.124	2677	-1.930	0.054	.
Anteriority (ant)	-0.905	0.116	2677	-7.791	< .001	***

Anteriority (post)	0.498	0.131	2677	3.801	< .001	***
Hemisphere (left)	-0.139	0.093	2677	-1.492	0.136	
Hemisphere (right)	0.161	0.093	2677	1.730	0.084	.
Extraction × Anteriority (ant)	0.359	0.164	2677	2.186	0.029	*
Extraction × Anteriority (post)	-0.007	0.185	2677	-0.036	0.972	

Table 44: Linear mixed-effects best-fitting model, Island condition, for object position from 500-900 ms, with DD scores included in model (native speakers)

Number of obs: 2728, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	0.859	0.646	47	1.329	0.190	
Extraction	-1.563	0.254	2678	-6.147	< .001	***
Anteriority (ant)	-0.382	0.079	2678	-4.81	< .001	***
Anteriority (post)	0.006	0.090	2678	0.071	0.943	
Hemisphere (left)	-0.218	0.090	2678	-2.43	0.015	*
Hemisphere (right)	0.036	0.090	2678	0.396	0.692	
DD scores	-0.635	0.374	45.5	-1.697	0.097	.
Extraction × DD scores	0.866	0.148	2678	5.837	< .001	***

Table 45: Linear mixed-effects best-fitting model, Non-Island condition, for actual gap site from 500-900 ms, with Working Memory included in model (native speakers)

Number of obs: 2728, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-0.003	0.190	78.6	-0.014	0.988	
Extraction	0.698	0.116	2672	5.995	< .001	***
Hemisphere (left)	0.018	0.088	2672	0.209	0.834	
Hemisphere (right)	-0.431	0.088	2672	-4.916	< .001	***
Anteriority (ant)	0.858	0.110	2672	7.836	< .001	***
Anteriority (post)	-0.563	0.124	2672	-4.558	< .001	***
Working Memory	-0.052	0.107	62.1	-0.488	0.627	
Extraction × Anteriority (ant)	-0.893	0.155	2672	-5.770	< .001	***
Extraction × Anteriority (post)	0.318	0.175	2672	1.820	0.069	.
Extraction × Working Memory	0.205	0.069	2672	2.954	0.003	**
Anteriority (ant) × Working Memory	0.099	0.065	2672	1.514	0.130	
Anteriority (post) × Working Memory	-0.201	0.074	2672	-2.726	0.006	**
Extraction × Anteriority (ant) × Working Memory	-0.204	0.092	2672	-2.206	0.027	*
Extraction × Anteriority (post) × Working Memory	0.195	0.104	2672	1.874	0.061	.

Table 46: Linear mixed-effects best-fitting model, Island condition, for actual gap site from 500-900 ms, with Working Memory included in model (native speakers)

Number of obs: 2728, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	0.823	0.199	76.1	4.142	< .001	***
Extraction	0.221	0.119	2672	1.851	0.064	.
Hemisphere (left)	0.036	0.090	2672	0.404	0.686	
Hemisphere (right)	-0.704	0.090	2672	-7.846	< .001	***
Anteriority (ant)	0.348	0.112	2672	3.100	0.002	**
Anteriority (post)	-0.426	0.127	2672	-3.366	< .001	***
Working Memory	-0.070	0.112	60.8	-0.623	0.536	
Extraction × Anteriority (ant)	-0.087	0.159	2672	-0.547	0.585	
Extraction × Anteriority (post)	0.026	0.179	2672	0.143	0.886	

Extraction × Working Memory	0.024	0.071	2672	0.344	0.731	
Anteriority (ant) × Working Memory	-0.054	0.067	2672	-0.812	0.417	
Anteriority (post) × Working Memory	-0.113	0.075	2672	-1.500	0.134	
Extraction × Anteriority (ant) × Working Memory	-0.145	0.095	2672	-1.527	0.127	
Extraction × Anteriority (post) × Working Memory	0.210	0.107	2672	1.965	0.049	*

Table 47: Linear mixed-effects best-fitting model, Non-Island condition, for actual gap site from 500-900 ms, with Attentional Control included in model (native speakers)

Number of obs: 2728, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-0.003	0.186	79.7	-0.015	0.988	
Extraction	0.698	0.115	2674	6.067	< .001	***
Hemisphere (left)	0.018	0.087	2674	0.212	0.832	
Hemisphere (right)	-0.431	0.087	2674	-4.975	< .001	***
Anteriority (ant)	0.858	0.108	2674	7.931	< .001	***
Anteriority (post)	-0.563	0.122	2674	-4.613	< .001	***
Attentional Control	-0.362	0.105	53.9	-3.446	0.001	**
Extraction × Anteriority (ant)	-0.893	0.153	2674	-5.840	< .001	***
Extraction × Anteriority (post)	0.318	0.173	2674	1.842	0.066	.
Extraction × Attentional Control	0.366	0.041	2674	8.973	< .001	***
Anteriority (ant) × Attentional Control	-0.045	0.048	2674	-0.944	0.345	
Anteriority (post) × Attentional Control	0.197	0.054	2674	3.662	< .001	***

Table 48: Linear mixed-effects best-fitting model, Island condition, for actual gap site from 500-900 ms, with Attentional Control included in model (native speakers)

Number of obs: 2728, groups: Subject, 44						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	0.838	0.193	68.7	4.340	< .001	***
Extraction	0.191	0.068	2679	2.812	0.005	**
Hemisphere (left)	0.036	0.090	2679	0.403	0.687	
Hemisphere (right)	-0.704	0.090	2679	-7.824	< .001	***
Anteriority (ant)	0.304	0.080	2679	3.827	< .001	***
Anteriority (post)	-0.413	0.090	2679	-4.602	< .001	***

Appendix B: Statistical Tables, L2 learner study

Table 49: Linear mixed-effects best-fitting overall model for object position from 300-500 ms, with proficiency included (L2 learners)

Number of obs: 2852, groups: Subject, 23						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-1.900	0.363	29.6	-5.231	0.000	***
Extraction	0.101	0.209	2803	0.480	0.631	
Island	-0.008	0.166	2803	-0.049	0.961	
Hemisphere (left)	-0.140	0.125	2803	-1.121	0.262	
Hemisphere (right)	0.938	0.125	2803	7.508	0.000	***
Anteriority (ant)	-0.446	0.156	2803	-2.857	0.004	**
Anteriority (post)	1.776	0.176	2803	10.082	0.000	***
Proficiency	0.014	0.067	26.1	0.209	0.836	
Extraction × Island	0.156	0.235	2803	0.665	0.506	
Extraction × Hemisphere (left)	0.257	0.177	2803	1.454	0.146	
Extraction × Hemisphere (right)	-0.143	0.177	2803	-0.809	0.418	
Extraction × Anteriority (ant)	-0.417	0.221	2803	-1.887	0.059	.
Extraction × Anteriority (post)	-0.304	0.249	2803	-1.219	0.223	
Island × Anteriority (ant)	-0.151	0.221	2803	-0.682	0.495	
Island × Anteriority (post)	-0.100	0.249	2803	-0.402	0.688	
Extraction × Proficiency	-0.017	0.032	2803	-0.546	0.585	
Island × Proficiency	-0.065	0.032	2803	-2.068	0.039	*
Anteriority (ant) × Proficiency	-0.109	0.030	2803	-3.687	0.000	***
Anteriority (post) × Proficiency	0.006	0.033	2803	0.170	0.865	
Extraction × Island × Anteriority (ant)	-0.012	0.312	2803	-0.039	0.969	
Extraction × Island × Anteriority (post)	0.177	0.352	2803	0.504	0.615	
Extraction × Island × Proficiency	-0.001	0.045	2803	-0.028	0.978	
Extraction × Anteriority (ant) × Proficiency	0.063	0.042	2803	1.505	0.132	
Extraction × Anteriority (post) × Proficiency	-0.003	0.047	2803	-0.073	0.942	
Island × Anteriority (ant) × Proficiency	0.105	0.042	2803	2.510	0.012	*
Island × Anteriority (post) × Proficiency	-0.008	0.047	2803	-0.163	0.870	
Extraction × Island × Anteriority (ant) × Proficiency	-0.136	0.059	2803	-2.291	0.022	*
Extraction × Island × Anteriority (post) × Proficiency	0.014	0.067	2803	0.208	0.836	

Table 50: Linear mixed-effects best-fitting overall model for object position from 300-500 ms, with both Proficiency and DD scores included (L2 learners)

Number of obs: 2852, groups: Subject, 23						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-1.958	0.356	27	-5.504	< .001	***
Extraction	0.101	0.206	2791	0.487	0.627	
Island	-0.008	0.164	2791	-0.050	0.960	
Hemisphere (left)	-0.108	0.124	2791	-0.868	0.386	
Hemisphere (right)	0.928	0.124	2791	7.484	< .001	***
Anteriority (ant)	-0.433	0.155	2791	-2.805	0.005	**
Anteriority (post)	1.821	0.174	2791	10.449	< .001	***
Proficiency	0.036	0.068	25.2	0.526	0.604	

Proficiency	-0.854	0.549	22.2	-1.555	0.134	
Extraction × Island	0.156	0.231	2791	0.675	0.500	
Extraction × Hemisphere (left)	0.257	0.174	2791	1.475	0.140	
Extraction × Hemisphere (right)	-0.143	0.174	2791	-0.821	0.412	
Extraction × Anteriority (ant)	-0.417	0.218	2791	-1.914	0.056	.
Extraction × Anteriority (post)	-0.304	0.246	2791	-1.236	0.216	
Island × Anteriority (ant)	-0.151	0.218	2791	-0.692	0.489	
Island × Anteriority (post)	-0.100	0.245	2791	-0.407	0.684	
Extraction × Proficiency	-0.022	0.031	2791	-0.698	0.485	
Island × Proficiency	-0.058	0.031	2791	-1.872	0.061	.
Hemisphere (left) × Proficiency	0.008	0.017	2791	0.473	0.637	
Hemisphere (right) × Proficiency	0.006	0.017	2791	0.366	0.714	
Anteriority (ant) × Proficiency	-0.102	0.030	2791	-3.464	< .001	***
Anteriority (post) × Proficiency	-0.013	0.033	2791	-0.378	0.706	
Extraction × DD score	0.232	0.107	2791	2.161	0.031	*
Island × DD score	-0.353	0.107	2791	-3.292	0.001	**
Hemisphere (left) × DD score	0.015	0.143	2791	0.103	0.918	
Hemisphere (right) × DD score	0.371	0.143	2791	2.594	0.010	**
Anteriority (ant) × DD score	-0.525	0.126	2791	-4.148	< .001	***
Anteriority (post) × DD score	0.328	0.143	2791	2.297	0.022	*
Proficiency × DD score	0.107	0.101	21.4	1.066	0.298	
Extraction × Island × Anteriority (ant)	-0.012	0.308	2791	-0.039	0.969	
Extraction × Island × Anteriority (post)	0.177	0.347	2791	0.511	0.610	
Extraction × Island × Proficiency	-0.001	0.044	2791	-0.028	0.977	
Extraction × Anteriority (ant) × Proficiency	0.063	0.041	2791	1.527	0.127	
Extraction × Anteriority (post) × Proficiency	-0.003	0.047	2791	-0.074	0.941	
Island × Anteriority (ant) × Proficiency	0.105	0.041	2791	2.547	0.011	*
Island × Anteriority (post) × Proficiency	-0.008	0.047	2791	-0.166	0.868	
Hemisphere (left) × Proficiency × DD score	-0.060	0.027	2791	-2.258	0.024	*
Hemisphere (right) × Proficiency × DD score	0.019	0.027	2791	0.703	0.482	
Anteriority (ant) × Proficiency × DD score	-0.024	0.023	2791	-1.016	0.310	
Anteriority (post) × Proficiency × DD score	-0.082	0.026	2791	-3.098	0.002	**
Extraction × Island × Anteriority (ant) × Proficiency	-0.136	0.059	2791	-2.324	0.020	*
Extraction × Island × Anteriority (post) × Proficiency	0.014	0.066	2791	0.211	0.833	

Table 51: Linear mixed-effects best-fitting overall model for object position from 500-900 ms, with both Proficiency and DD scores included (L2 learners)

Number of obs: 2852, groups: Subject, 23						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-1.411	0.448	27.2	-3.152	0.004	**
Extraction	-0.612	0.176	2787	-3.469	< .001	***
Island	-0.054	0.155	2787	-0.347	0.729	
Hemisphere (left)	0.050	0.208	2787	0.240	0.810	
Hemisphere (right)	0.910	0.208	2787	4.381	< .001	***
Anteriority (ant)	-0.133	0.212	2787	-0.626	0.531	
Anteriority (post)	0.942	0.233	2787	4.041	< .001	***
Proficiency	-0.002	0.083	22.8	-0.023	0.982	
DD score	0.154	0.694	22.8	0.221	0.827	

Extraction × Island	0.503	0.153	2787	3.295	< .001	***
Extraction × Hemisphere (left)	0.330	0.199	2787	1.654	0.098	.
Extraction × Hemisphere (right)	-0.080	0.199	2787	-0.404	0.686	
Island × Anteriority (ant)	-0.223	0.179	2787	-1.248	0.212	
Island × Anteriority (post)	0.088	0.202	2787	0.438	0.661	
Hemisphere (left) × Anteriority (ant)	-0.281	0.237	2787	-1.184	0.236	
Hemisphere (right) × Anteriority (ant)	-0.169	0.237	2787	-0.713	0.476	
Hemisphere (left) × Anteriority (post)	-0.210	0.264	2787	-0.797	0.426	
Hemisphere (right) × Anteriority (post)	-0.721	0.264	2787	-2.734	0.006	**
Extraction × Proficiency	-0.043	0.021	2787	-2.053	0.040	*
Island × Proficiency	-0.016	0.030	2787	-0.543	0.587	
Hemisphere (left) × Proficiency	0.037	0.020	2787	1.881	0.060	.
Hemisphere (right) × Proficiency	0.021	0.020	2787	1.058	0.290	
Anteriority (ant) × Proficiency	-0.068	0.025	2787	-2.773	0.006	**
Anteriority (post) × Proficiency	-0.001	0.028	2787	-0.053	0.958	
Extraction × DD score	0.525	0.175	2787	3.001	0.003	**
Island × DD score	-0.288	0.250	2787	-1.152	0.250	
Hemisphere (left) × DD score	-0.613	0.164	2787	-3.742	< .001	***
Hemisphere (right) × DD score	-0.145	0.164	2787	-0.883	0.377	
Anteriority (ant) × DD score	-0.480	0.205	2787	-2.342	0.019	*
Anteriority (post) × DD score	0.029	0.231	2787	0.126	0.900	
Proficiency × DD score	0.290	0.129	22.8	2.253	0.034	*
Extraction × Island × Proficiency	-0.056	0.030	2787	-1.871	0.061	.
Island × Anteriority (ant) × Proficiency	0.019	0.035	2787	0.542	0.588	
Island × Anteriority (post) × Proficiency	0.006	0.039	2787	0.146	0.884	
Extraction × Island × DD score	0.133	0.247	2787	0.538	0.591	
Island × Anteriority (ant) × DD score	0.039	0.290	2787	0.136	0.892	
Island × Anteriority (post) × DD score	-0.086	0.327	2787	-0.264	0.792	
Extraction × Proficiency × DD score	0.016	0.032	2787	0.486	0.627	
Island × Proficiency × DD score	0.019	0.046	2787	0.410	0.682	
Hemisphere (left) × Proficiency × DD score	-0.099	0.030	2787	-3.250	0.001	**
Hemisphere (right) × Proficiency × DD score	-0.051	0.030	2787	-1.689	0.091	.
Anteriority (ant) × Proficiency × DD score	0.028	0.038	2787	0.731	0.465	
Anteriority (post) × Proficiency × DD score	-0.104	0.043	2787	-2.429	0.015	*
Extraction × Island × Proficiency × DD score	-0.138	0.046	2787	-3.010	0.003	**
Island × Anteriority (ant) × Proficiency × DD score	-0.097	0.054	2787	-1.803	0.072	.
Island × Anteriority (post) × Proficiency × DD score	0.039	0.061	2787	0.644	0.520	

Table 52: Linear mixed-effects best-fitting model, Non-Island condition, for object position from 500-900 ms, with both Proficiency and DD scores included (L2 learners)

Number of obs: 1426, groups: Subject, 23						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-1.469	0.489	25.4	-3.004	0.006	**
Extraction	0.262	0.263	1373	0.995	0.320	
Anteriority (ant)	-0.355	0.169	1373	-2.105	0.035	*
Anteriority (post)	0.710	0.190	1373	3.736	< .001	***
Hemisphere (left)	0.004	0.189	1373	0.022	0.982	
Hemisphere (right)	0.758	0.189	1373	4.004	< .001	***

Proficiency	-0.022	0.093	23.4	-0.240	0.813	
DD score	-0.310	0.777	23.4	-0.399	0.694	
Extraction × Anteriority (ant)	-0.362	0.238	1373	-1.518	0.129	
Extraction × Anteriority (post)	-0.081	0.269	1373	-0.301	0.764	
Extraction × Hemisphere (left)	0.169	0.266	1373	0.634	0.526	
Extraction × Hemisphere (right)	-0.430	0.266	1373	-1.616	0.106	
Extraction × Proficiency	-0.103	0.035	1373	-2.939	0.003	**
Anteriority (ant) × Proficiency	-0.063	0.033	1373	-1.927	0.054	.
Anteriority (post) × Proficiency	0.020	0.037	1373	0.546	0.585	
Hemisphere (left) × Proficiency	0.043	0.026	1373	1.625	0.104	
Hemisphere (right) × Proficiency	0.030	0.026	1373	1.150	0.250	
Extraction × DD score	0.719	0.290	1373	2.477	0.013	*
Anteriority (ant) × DD score	-0.442	0.273	1373	-1.618	0.106	
Anteriority (post) × DD score	0.082	0.308	1373	0.266	0.791	
Hemisphere (left) × DD score	-0.459	0.218	1373	-2.101	0.036	*
Hemisphere (right) × DD score	0.063	0.218	1373	0.288	0.773	
Proficiency × DD score	0.322	0.144	23.4	2.235	0.035	*
Extraction × Anteriority (ant) × Proficiency	0.028	0.046	1373	0.606	0.545	
Extraction × Anteriority (post) × Proficiency	-0.031	0.052	1373	-0.595	0.552	
Extraction × Anteriority (ant) × DD score	0.014	0.386	1373	0.036	0.971	
Extraction × Anteriority (post) × DD score	-0.260	0.435	1373	-0.597	0.551	
Extraction × Proficiency × DD score	-0.181	0.054	1373	-3.372	< .001	***
Anteriority (ant) × Proficiency × DD score	-0.153	0.051	1373	-3.020	0.003	**
Anteriority (post) × Proficiency × DD score	-0.041	0.057	1373	-0.725	0.469	
Hemisphere (left) × Proficiency × DD score	-0.088	0.040	1373	-2.169	0.030	*
Hemisphere (right) × Proficiency × DD score	-0.021	0.040	1373	-0.522	0.601	
Extraction × Anteriority (ant) × Proficiency × DD score	0.169	0.072	1373	2.362	0.018	*
Extraction × Anteriority (post) × Proficiency × DD score	-0.045	0.081	1373	-0.561	0.575	

Table 53: Linear mixed-effects best-fitting model, Island condition, for object position from 500-900 ms, with both Proficiency and DD scores included (L2 learners)

Number of obs: 1426, groups: Subject, 23						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	
(Intercept)	-1.353	0.444	24.4	-3.047	0.005	**
Extraction	-0.371	0.179	1378	-2.074	0.038	*
Anteriority (ant)	-0.149	0.169	1378	-0.878	0.380	
Anteriority (post)	0.578	0.191	1378	3.030	0.002	**
Hemisphere (left)	-0.028	0.136	1378	-0.208	0.835	
Hemisphere (right)	0.665	0.136	1378	4.879	< .001	***
Proficiency	0.026	0.086	24.4	0.301	0.766	
DD score	0.270	0.720	24.4	0.375	0.711	
Extraction × Anteriority (ant)	-0.328	0.238	1378	-1.381	0.168	
Extraction × Anteriority (post)	0.007	0.268	1378	0.025	0.980	
Extraction × Proficiency	-0.087	0.034	1378	-2.529	0.012	*
Anteriority (ant) × Proficiency	-0.122	0.033	1378	-3.730	< .001	***
Anteriority (post) × Proficiency	0.005	0.037	1378	0.131	0.896	
Hemisphere (left) × Proficiency	0.031	0.027	1378	1.184	0.237	
Hemisphere (right) × Proficiency	0.012	0.027	1378	0.434	0.665	

Extraction × DD score	0.582	0.292	1378	1.997	0.046	*
Anteriority (ant) × DD score	-0.235	0.275	1378	-0.856	0.392	
Anteriority (post) × DD score	-0.256	0.310	1378	-0.825	0.410	
Hemisphere (left) × DD score	-0.767	0.221	1378	-3.474	< .001	***
Hemisphere (right) × DD score	-0.352	0.221	1378	-1.595	0.111	
Proficiency × DD score	0.314	0.131	22.5	2.403	0.025	*
Extraction × Anteriority (ant) × Proficiency	0.108	0.046	1378	2.350	0.019	*
Extraction × Anteriority (post) × Proficiency	-0.013	0.052	1378	-0.259	0.796	
Extraction × Anteriority (ant) × DD score	-0.500	0.388	1378	-1.288	0.198	
Extraction × Anteriority (post) × DD score	0.552	0.437	1378	1.262	0.207	
Anteriority (ant) × Proficiency × DD score	0.027	0.036	1378	0.750	0.454	
Anteriority (post) × Proficiency × DD score	-0.105	0.041	1378	-2.572	0.010	*
Hemisphere (left) × Proficiency × DD score	-0.110	0.041	1378	-2.677	0.008	**
Hemisphere (right) × Proficiency × DD score	-0.081	0.041	1378	-1.990	0.047	*