L1 AND L2 PROCESSING OF FILLER-GAP DEPENDENCIES IN MANDARIN CHINESE SPEAKERS: EVIDENCE FROM SELF-PACED READING TIMES AND EVENT-RELATED POTENTIALS

SHUZHEN WANG

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the
Doctor of Philosophy in Linguistics

Department of Linguistics
Faculty of Arts
University of Ottawa

Abstract

Language processing is a complex task requiring our brain to rapidly integrate various linguistic information from strings of unfolding words and quickly build comprehension. While previous studies have shown that native speakers (L1) employ both incremental and predictive processing in real time (e.g., Clifton & Frazier, 1989; Altmann & Kamide, 1999), debates persist regarding how second language speakers (L2) process sentences in real time and whether their processing mechanisms align with or differ from those of native speakers. In the field of L2 processing, some scholars believe that L1 and L2 processing share qualitatively similar mechanisms, and differences in L2 processing profiles arise from individual factors or language features (e.g., Dallas et al., 2023; Sabourin et al., 2014; Sabourin & Stowe, 2008; Berghoff, 2022; Grüter & Rohde, 2021). Others argue for qualitative differences between L1 and L2 processing (e.g., Clahsen & Felser, 2006; Pakulak & Neville, 2011), and these differences are especially apparent when the sentence structure includes filler-gap dependencies, a non-local syntactic structure requiring parsers to establish a connection between a displaced filler and its corresponding syntactic gap over distance (e.g., Marinis et al., 2005; Dong et al., 2023). The primary goal of this dissertation thus is to examine how L1 and L2 speakers process filler-gap dependencies in real time and whether they show qualitatively similar or different processing profiles and underlying neural mechanisms in response to both syntactic and semantic cues.

Three experiments were conducted in the dissertation. Experiment 1 used a self-paced reading paradigm to investigate how L1 English speakers and L2 speakers with L1 Mandarin Chinese process filler-gap dependencies with varying dependency lengths (long vs. short) and filler plausibility (plausible vs. implausible). The results showed that both L1 and L2 groups, regardless of their L2 proficiency and years of immersion, exhibited increased reading times at the verb, the position that indicates a potential syntactic gap, yet their sensitivity to filler-verb mismatches was observed only in the second post-verbal position. These findings indicated active and sequential processing of syntactic and semantic information. Then, using the same paradigm and within-subject factor manipulations, Experiment 2 investigated whether native Mandarin Chinese speakers process Chinese topic structures similarly to how English speakers process filler-gap dependencies. The results revealed that native Mandarin Chinese speakers, like English speakers, actively searched for gaps, evidenced by elevated reading times at the first verb they encountered and sensitivity to semantic mismatches shortly thereafter. Previous research has shown that active gap-filling strategy can be interpreted as a response to the characteristics of *wh*-dependencies. (Wagers & Phillips, 2009). Thus, the findings lend support to the existence of movement-derived filler-gap dependencies in Mandarin Chinese, despite its *wh*-in-situ nature. Lastly, Experiment 3 used event-related potentials (ERPs) to examine the brain responses of L1 and L2 speakers with the manipulation of filler plausibility and gap licensor types, which were either a transitive verb or a preposition. The results showed no overall group difference. However, while L2 speakers with high proficiency showed native-like but attenuated and delayed predictive processing, L2 speakers with low proficiency showed no evidence of prediction and experienced syntactic reanalysis.

Overall, this dissertation revealed qualitatively similar processing mechanisms between L1 and L2 speakers, with L2 differences being explained by individual factors. Moreover, the findings lent support to a movement-derived account of Chinese topic structures, despite Chinese being a *wh*-in-situ language.

Acknowledgements

I have conceived this acknowledgment many times during my PhD journey, often finding my eyes welling up with tears, sometimes by the adversities I faced and other times by the overwhelming support and kindness I was fortunate to receive. Now, I have reached this moment, and I can finally express my long-held gratitude to my loved ones.

My deepest gratitude goes to my supervisor, Prof. Sabourin, who guided me through the world of linguistics and helped me discover what I want to dedicate my life to. She pointed me in the right direction every time I felt lost. She is my life mentor and someone I can always count on. Words fall far short of expressing my gratitude. Having Prof. Sabourin as my supervisor is truly the greatest blessing of my PhD journey. I am also grateful to my committee members—Prof. Zamuner, Prof. Francis, Prof. Grant, and Prof. Valenzuela—for their expertise, patience, and support. Their thoughtful feedback helped shape this work into something I am proud of.

I want to give a big thank you to Dr. Gosselin and Dr. Manning for generously sharing their experience in ERP with me. Thank you to Yubin Xing for helping me proofread my stimuli and answering all my questions about statistics. Thank you to Melanie Knezevic and Setareh Dorood for being wonderful partners in running EEG experiments. To all my other lab members, thank you for being part of this journey. You have truly made these years special.

I am also thankful to my loving family. To my parents, Qinghong Ji and Lifa Wang, thank you for supporting my dream, even though it meant watching your only daughter study so far from home for all these years. To my mother-in-law, Meiling Kang, and father-in-law, Guangqing Niu, thank you for respecting my choices and being proud of me. A special thank you to my husband, Jijun Niu. We got married during the pandemic, at a difficult point in my academic journey. He came into my life and changed everything for the better. His decision to move to Canada with me and his unwavering support through every challenge have consistently given me strength. My husband is my most steadfast supporter, and this dissertation could not have been completed without him. I hope to spend the rest of my life returning the love he has given me.

Finally, I want to thank myself for persevering through this odyssey. I have walked a long way and endured many challenges on the path of completing this doctoral dissertation. This dissertation may not be flawless, but I am grateful for it and for this journey that has shaped who I am today. My greatest hope is that someday I can use what I have learned to help others. If I can make this happen, that would make every step worthwhile.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AN	Anterior negativity
ANOVA	Analysis of Variance
AoA	Age of Acquisition
CL	Central-left
CM	Central-midline
CP	Complementizer Phrase
CR	Central-right
DLT	Dependency Locality Theory
DPM	Declarative/Procedural Model
EEG	Electroencephalography
ELAN	Early left anterior negativity
ERP	Event-related potential
FL	Frontal-left
FM	Frontal-midline
FR	Frontal-right
IELTS	International English Language Testing System
L1	First language
L2	Second language
LAN	Left anterior negativity
LN	Late negativity
M	Mean
PL	Posterior-left
PM	Posterior-midline
PR	Posterior-right
PREP	Preposition
RAGE	Reduced Ability to Generate Expectations
ROI	Region of interest
RT	Reading time
SD	Standard deviation
SE	Standard error
SSH	Shallow Structure Hypothesis
TOEFL	Test of English as a Foreign Language
VT	Transitive verb
YoI	Years of Immersion

Chapter 1

Introduction

1.1 Background and Motivation

Human online language processing features the real-time integration of linguistic inputs from strings of words rapidly and effortlessly. One of the central goals of L1 processing is to understand how different aspects of linguistic inputs (e.g., lexical-semantics, syntax, pragmatics) are computed to achieve successful comprehension in real time. This question is equally essential to research of L2 processing. Understanding how a non-native language is processed in real time provides insights into the mechanisms by which our brain manages a late-acquired and less-used language. Additionally, it helps explain why, despite years of formal instruction, L2 learners still exhibit varying levels of L2 performance within and across individuals and often show a less-than-native-like ultimate attainment in L2 (e.g., Bley-Vroman, 1989, 1990). Among various levels of linguistic representation, this dissertation specifically focuses on L2 sentence processing. The goal is to understand how sentence comprehension mechanisms rapidly construct syntactic structures and assign interpretations to incoming inputs in L2 speakers and whether they employ the same parsing mechanisms as native speakers.

Research on L2 sentence processing has expanded rapidly over the past few decades. Much of this work has focused on understanding how second language learners process sentences in real time and whether their parsing mechanisms align with or differ from those of native speakers. However, the research has not yet yielded a conclusive consensus, and there are two competing views in the literature that address the relationship between L1 and L2 parsing mechanisms: 1) L1 and L2 processing depend on shared mechanisms, and native-like processing is attainable (e.g., MacWhinney, 2013; Abutalebi, 2008; Hernandez et al., 2005; Indefrey, 2006); 2) L1 and L2 processing depend on fundamentally distinct mechanisms, and even when native-like performance is achieved, L1 and L2 processing occur in neurocognitively distinct manners (e.g., Bley-Vroman, 1989, 1990; Clahsen & Felser, 2006; Pakulak & Neville, 2011; Dong, 2014).

The first view supports the shared mechanism and thus argues that native-like processing is achievable. This first view posits that native-like processing in L2 is attainable, but both individual factors of the L2 speakers and the nature of the languages must be taken into consideration. Individual factors that improve the likelihood of native-like processing include a high level of proficiency (e.g., Ojima et al., 2005; Dallas et al., 2013), an early age of acquisition (e.g., Sabourin et al., 2014), and extensive L2 exposure (e.g., Dussias & Sagarra, 2007; Pliatsikas & Marinis, 2013; Berghoff, 2022), whereas the increased demand for cognitive resources (e.g., McDonald, 2006; Hopp, 2010), decreased ability to anticipate upcoming information during L2 parsing, known as the Reduced Ability to Generate Expectations Hypothesis (RAGE) (Grüter & Rohde, 2013, 2021) and sensitivity to memory retrieval interference (Cunnings, 2017a, 2017b) seem to make native-like processing harder. On top of the individual factors between speakers, the potential for native-like processing largely depends on the nature of the languages. Sabourin and Stowe (2008) suggest that the similarity of syntactic constructions between L1 and L2 and their rule-governed nature are likely to result in similar L1 and L2 processing. Additionally, Hopp (2017) suggests that lexical coactivation from cognates in experimental items can facilitate the construction of syntactic structures in L2 speakers. A recent review suggests that the tendency of L2 learners to rely less on grammar is more related to the delayed and less robust retrieval of morphosyntactic information in real time, rather than to a fundamental difference between L1 and L2 processing (Hopp, 2022). It is important to note that while the aforementioned accounts suggest that native-like processing is attainable after individual factors are accounted for, we cannot assertively claim a shared or identical processing mechanism unless all observed differences in L1 and L2 processing can be explained.

Given the long-standing question as to why second language acquisition (SLA) tends to be generally less effective or more flawed than the acquisition of the first language, the second view suggests that the parsing mechanisms underlying L2 processing are fundamentally distinct from L1 processing (e.g., Bley-Vroman, 1989; Clahsen & Felser, 2006). In particular, Clahsen and Felser (2006) proposed the Shallow Structure Hypothesis (SSH) arguing that L2 sentence processing can only construct shallow representations that involve lexical-semantic and pragmatic cues but not syntactic representations, whereas L1 speakers are able to generate full-fledged syntactic representations including establishing syntactic dependencies in real time. Supporting evidence comes from findings that L2 learners showed online sensitivity to lexical-semantic cues but lacked disambiguation preferences in the absence of such cues (e.g., Felser et al., 2003; Papadopoulou & Clahsen, 2003) and failed to show filler reactivation before encountering the subcategorizing verb, suggesting a lack of syntactic gap-filling (e.g., Felser & Robert, 2007; Marinis et al., 2005). Under this view, the argument is that individual difference factors are insufficient to account for the discernible differences between L1 and L2 processing and that real-time L2 processing may continue to be non-nativelike even when L2 speakers have ultimate L2 performance (e.g., Pakulak & Neville, 2011). The SSH also draws theoretical support from the Declarative/Procedural model (Ullman, 2001, 2005, 2006). Ullman claims that language processing relies on two distinct memory systems, the declarative and procedural memory systems. In terms of L1 processing, the declarative memory serves the lexicon and is largely accessible through explicit learning, whereas the procedural memory supports rule-based grammar and deals with implicit, unconscious learning. With respect to L2 learning, the tenet of the Declarative/Procedural model is that L2 learners rely on the declarative memory system at least at the initial stage of acquisition while native speakers take advantage of both systems. Moreover, the declarative memory system continues to develop after puberty, whereas the procedural memory system reaches full maturity before that. Restricted by the age limitation, late L2 learners therefore must initially rely solely on the declarative memory system to achieve L2 comprehension. Furthermore, as the capacity of one system grows, the other tends to decline. This model provides a persuasive explanation for why L2 speakers tend to struggle more with syntactic rules than with lexical acquisition. A recent study (Dong et al., 2023) has provided neurophysiological evidence for SSH through an ERP study. Their findings showed that L2 speakers had a reduced sensitivity to structural violations during filler-gap integration, and such non-native pattern remained consistent regardless of L2 proficiency or working memory capacity.

Among proponents of the fundamentally distinct mechanism hypothesis, a key question is whether the L2 processing mechanism can be shaped to become native-like with sufficient training, even if it is initially distinct from the L1 processing mechanism. Recent work advocates for a less dichotomous view of L1 and L2 sentence processing (e.g., Steinhauer et al., 2009, 2014). Instead of treating the two mechanisms as perpetually distinct, there is a call to examine the gradual change of the L2 processing mechanism through training. The idea is that while L1 and L2 processing may start off with fundamentally different mechanisms, L2 processing strategies and neurocognition have the potential to become more native-like overtime as proficiency or exposure increases. The difference between this idea and the initially proposed SSH is that this view admits a dynamic and changing processing mechanism for L2. In other words, although L1 and L2 processing mechanisms are inherently different, it does not exclude the possibility of L2 processing becoming increasingly similar to L1 processing. In fact, Clahsen & Felser (2018) clarified their initially proposed SSH regarding whether L2 processing can become native-like. They contradicted earlier interpretations about the impossibility of L2 learners achieving native-like syntactic parsing capability. Instead, they stated that the type of information prioritized and the processing pathway selected cannot be consciously controlled. Increased exposure to a language may result in more automatic or deeply ingrained grammatical processing patterns. This is in line with the dynamically and gradually changing view of L1-L2 difference.

Regarding the impact of proficiency, Steinhauer et al. (2009) proposed a developmental model suggesting that neuronal activity is distinct in the early stages of L2 acquisition but becomes more native-like with increased proficiency. However, this model is primarily based on results from violation-based paradigms, so it remains uncertain whether these findings are applicable to the processing of grammatical syntactic constructions. Bowden et al. (2013) supported this model through their study that examined the neurocognition of late-learned L2 Spanish in two groups of university foreign-language learners compared to native speakers. One group had one year of college classroom experience (L2 Low), while the other group had over three years of classroom experience and 1-2 semesters of immersion abroad (L2 Advanced). The findings suggest that for semantic violations, L2 speakers at both levels engage similar neurocognitive mechanisms to L1 speakers. However, L2 syntactic processing appears to undergo a developmental shift, with higher proficiency and increased exposure contributing to native-like processing patterns.

In terms of the type of training, Morgan-Short and colleagues (2012) conducted a study where participants learned an artificial language (Brocanto2) through either implicit or explicit instruction methods. The training spanned approximately 10 days over three sessions. Following the training, the participants were tested on gender agreement violations and word order violations using ERP experiments. At high proficiency, both the implicit and explicit groups showed ERP responses similar to those of native speakers. However, at low proficiency, only learners from the implicit group exhibited native-like ERP responses. This indicates that the nature of L2 exposure and training (implicit vs. explicit) can influence L2 processing strategies and neurocognitive development, with implicit training contributing to native-like neurocognitive responses earlier than explicit training (e.g., Morgan-Short et al., 2012). Similarly, a recent study by Dong et al. (2023) acknowledges that L1 and L2 processing of filler-gap sentences are fundamentally distinct, as they observed decreased sensitivity to structural violations during filler-gap integration in L2 speakers, and this non-native-like pattern remained unchanged regardless of proficiency level or working memory capacity. The researchers noted that all participants in their study had received explicit instruction only, with the grammatical rules taught in their native language. They suggested that the lack of implicit instruction involving rich, naturalistic input might have limited learners' ability to develop native-like syntactic rules.

While numerous studies have examined whether and to what extent L2 sentence processing aligns with and differs from L1 processing, the findings remain highly controversial, especially in L2 studies on non-local dependencies, including filler-gap dependencies. Many studies provide evidence that native-like processing of L2 is typically restricted to local dependencies, suggesting that non-local dependencies might differ from local processing (e.g., Clahsen & Felser, 2006; Levy, 2008). Other research indicates that although the processing of *wh*-constructions by L2 learners is not as immediate and efficient as that of early learners and native speakers, potentially due to factors such as late ages of immersion, shorter lengths of residence, and lower proficiency, these adult learners still utilized detailed structural representations early during online reading (Jung, 2017).[1] Given this background, the current study examines L1 and L2 processing of filler-gap dependencies, which provide a prime exemplar for psycholinguistics to investigate the parsing mechanisms of L2 learners for the following reasons. First, such non-local dependencies require parsers to establish a connection between an abstract trace and its corresponding syntactic gap within a complex structural framework. This directly assesses L2 speakers' ability to use structural information during real-time parsing (Dong, 2014). Second, filler-gap dependencies allow the manipulation of the dependency length, the linear length between the filler and the gap, providing a practical method to evaluate Clahsen and Felser's (2006) hypothesis: L2 learners underuse syntactic information during the processing of sentences that contain long-distance filler-gap dependencies. Last, natural language is full of filler-gap dependencies, which appear differently on

[1] In Jung (2017), early learners are defined as individuals whose age of arrival (AoA) in an English-speaking environment was before 12 years old (mean AoA = 7.43), while adult ESL learners had an AoA after 16 years old (mean AoA = 26.29).

the surface in various languages; hence this structure allows investigation into the effects of L1 grammar on L2 processing. Hawkins & Chan (1997) proposed the Failed Feature Hypothesis, which assumes that the initial state of L2 acquisition relies on the L1 grammar and that any target language features not present in the L1 will not be acquired. This hypothesis suggests that L1 interference plays a crucial role in limiting L2 acquisition. However, Dong's (2014) findings challenged this view. In his Experiment 2, Mandarin Chinese speakers[2] of English were successful in processing tense morphology, a feature absent in their native language. This suggests that L1 transfer is not the sole determinant of L2 processing abilities and that learners can overcome L1 interference and acquire an L2 feature that is not instantiated in their native language. The Shallow Structure Hypothesis (SSH) proposes that "L1 transfer [should] influence L2 processing only indirectly, as a consequence of one or more of the knowledge sources that feed the processing system being affected by properties of the L1" (Clahsen & Felser, 2006, p. 118). In a later critical review, the authors claim that they remain neutral regarding the role of L1 influence in L2 parsing. They assert that the SSH, which assumes that L2 learners tend to underuse grammatical information and over-rely on non-grammatical information, is independent of the learners' L1 (Clahsen & Felser, 2018). While the debate on L1 interference often centers on tense morphology, it is crucial to consider this discussion when investigating filler-gap dependencies, a construction requiring *wh*-movement in English but allowing *wh*-in-situ in Mandarin Chinese.

The discussion so far has shown that while numerous studies have investigated how L2 speakers may differ from monolingual natives in sentence processing in real time, the findings remain inconclusive, and two major research gaps can be found. On the one hand, there is limited research that compares the processing of filler-gap dependencies in native English speakers and in Mandarin Chinese speakers who learn English as their L2 (Dallas et al., 2013; Dong et al., 2023;

[2] I use "Mandarin Chinese" throughout to refer to the participant groups in the present dissertation and discuss syntactic structures in this language. When reviewing previous studies, I maintain each author's original terminology.

Covey et al., 2024). This pairing of participant groups is intriguing as it allows investigation into the differences between L1 and L2 processing, and whether these differences stem from L1 interference. Previous research claims that languages like Korean, Mandarin Chinese, and Japanese do not have overt *wh*-movement, and a *wh*-phrase remains in its base-generated position (*wh*-in-situ), thereby not requiring such filler-gap processing procedures (e.g., Jung, 2017; Dallas et al., 2013). However, it is important to clarify that there is still ongoing debate regarding whether Mandarin Chinese involves movement-derived non-local dependencies. Recent studies have increasingly recognized that Mandarin Chinese does support overt movement in topic structures, a structure similar to English filler-gap dependencies (Yang & Tao, 2014; Dong et al., 2021; Huang & Kaiser, 2008). In this dissertation, I designed two self-paced reading experiments to test 1) how native English speakers process English filler-gap sentences, 2) how L1 Mandarin L2 English speakers process English filler-gap sentences, and 3) how native Mandarin Chinese speakers process gap-type topic structures (equivalent to English filler-gap dependencies). These experiments allow me to 1) investigate L1 and L2 processing of filler-gap dependencies, particularly in the context of wh-in-situ languages, 2) to compare native sentence processing in Mandarin and English, and 3) to test potential cross-language transfer by examining how L1 Mandarin-L2 English speakers process their native language, Mandarin, in comparison to their second language, English.

On the other hand, the impact of specific individual factors, particularly proficiency and exposure, on L2 processing remains complex and requires further clarification. As previously discussed, proponents of fundamental L1-L2 processing differences do not necessarily advocate for the existence of two distinct processing systems. There is ongoing debate about whether the L2 processing mechanism is permanently different from L1 or if the L2 processing mechanism can evolve to become more native-like. If L2 processing is indeed dynamic, it is crucial to identify the individual factors that contribute to achieving native-like processing. While many studies highlight the role of proficiency in L2 processing and suggest that L2 processing mechanisms can shift towards L1-like mechanisms with increased proficiency or exposure (e.g., Dallas, 2013; Bowden, 2013), few studies distinguish between proficiency and other factors that may affect L2 processing. Sabourin et al. (2014) investigated the influence of the age of acquisition (AoA) on bilingual lexical processing. Participants included native English speakers and English-French bilinguals of varying proficiency levels. The research used a masked priming paradigm to examine automatic lexical processing at the semantic level. Findings indicated that translation priming effects were significant only in simultaneous and early bilinguals and correlated with AoA, emphasizing the importance of early AoA for L2-to-L1 priming effects, regardless of second language proficiency. Building on this finding, which illuminates the role of AoA on bilingual lexical processing, more research is expected to disentangle different factors that contribute to L2 processing, particularly at the syntax level. In addition to these two major research gaps, it remains unclear whether Mandarin Chinese gap-type topic structures, akin to English filler-gap dependencies, are movement-derived or base-generated, and if they are movement-derived, do these topic structures involve the same parsing mechanisms as used in English filler-gap dependencies.

1.2 Research Questions

To address these research gaps, this dissertation examines how L1 Mandarin Chinese-L2 English speakers process filler-gap dependencies in real time and how their parsing mechanisms may differ from those of native English speakers using self-paced reading and event-related potentials (ERPs). More specifically, the following questions are answered:

1) Is the online processing of filler-gap dependencies in L2 similar to that of L1, with the manipulation of filler plausibility[3], dependency length[4], and gap licensor[5]?
    - How does filler plausibility (plausible vs. implausible) affect the processing of filler-gap dependencies in L1 and L2 English speakers?
    - How does dependency length (long vs. short) affect the processing of filler-gap dependencies in L1 and L2 English speakers?
    - How does the different types of gap licensor (transitive verb vs. preposition) affect the processing of filler-gap dependencies in L1 and L2 English speakers?
2) How do L1 and L2 speakers integrate syntactic and semantic information during filler-gap processing?
    - Do L2 speakers largely rely on lexical-semantic information from subcategorizing verbs to integrate a filler into the gap and underuse syntactic information from prepositions, as claimed by the Shallow Structure Hypothesis?
    - Does the timing and degree of syntactic and semantic information integration differ between L1 and L2 speakers?
3) To what extent can proficiency and immersion effects explain L1 and L2 processing differences?

[3] The filler treated as the direct object of the verb can be either plausible or implausible. For example, in the sentence "The manager knew which *customer* the receptionist called about the problem.," the filler *customer* represents a plausible situation. However, replacing the filler with *spoon* would render the situation implausible. As a result, a processing difficulty is expected at the verb *called* or shortly thereafter.

[4] The dependency length refers to the distance between the displaced filler and the gap. In the short dependency condition, there are only three words separating the filler from the gap, as exemplified by the sentence, "the manager knew which customer the receptionist called ___ about the problem." In contrast, the long dependency condition has a prepositional phrase inserted between the filler and the gap, extending the length by five words, as in, "The manager knew which customer the receptionist with a large cloth bag called ___ about the problem."

[5] The gap licensor refers to an element that licenses the presence of an abstract syntactic gap. Two types of gap licensor are compared in this study: transitive verbs, which provide both syntactic and semantic information, and prepositions, which primarily provide syntactic information. For example, in "Lucy identifies which article the author wrote a column to *describe* ___ in the newspaper," the transitive verb *describe* licenses the gap position, while in "Lucy identifies which article the author wrote a column *about* ___ in the newspaper," the preposition *about* serves as the gap licensor.

Although not falling into the domain of L2 sentence processing, I will also examine the processing of the gap-type Mandarin Chinese topic structure, which is equivalent to English filler-gap dependencies, in native Mandarin Chinese speakers. Dong et al. (2021) investigated the online processing of how Mandarin Chinese speakers fill syntactic gaps in Mandarin Chinese topic structures, and the results demonstrated participants' sensitivity to relative clause islands and complex noun phrase islands. The sensitivity of island constraints is widely considered a signature feature of syntactic movement (Chomsky, 1986; Traxler and Pickering, 1996). In light of these findings, the present study aims to contribute to a more comprehensive real-time parsing profile for the gap-type Mandarin Chinese topic structure, an understudied construction, by examining whether native speakers employ a top-down active strategy when processing this structure. The use of the active strategy by speakers would indicate that they are prepared to handle the unbounded nature of dependencies and the necessity to connect elements across potentially long distances in sentence structures. This would provide evidence for the same movement-derived filler-gap dependency observed in English. Therefore, the following question is also investigated:

4) Does the processing of Mandarin Chinese topic structures involve an active gap-filling strategy?
    - How do dependency length (long vs. short) and filler plausibility (plausible vs. implausible) affect the processing of gap-type Mandarin Chinese topic structures in native speakers?

1.3 Organization of the Dissertation

The organization of the dissertation is structured as follows. The present chapter, Chapter 1, has provided an introduction to major theories and models of L2 sentence processing. In this chapter, I have identified gaps in existing research and formulated specific research questions. Chapter 2 is a literature review focusing on major sentence processing models. It begins with an overview of human sentence processing, with a discussion of incremental processing in L1 and L2, followed by predictive processing in both populations. The chapter then examines various theoretical models of L1 and L2 processing and factors that modulate L2 sentence processing. Chapter 3 focuses specifically on filler-gap processing in L1 and L2 contexts, with particular attention to English and Mandarin Chinese. After introducing the linguistic background of filler-gap dependencies in both languages, this chapter reviews both behavioral and neural approaches to studying these dependencies. Chapter 4 presents two self-paced reading experiments. Experiment 1 examines filler-gap dependency processing in L1 and L2 English speakers, with the manipulations of filler plausibility and dependency length. This experiment also examines how individual differences in L2 proficiency and year of immersion influence their processing patterns. Experiment 2 investigates how L1 Mandarin Chinese speakers process filler-gap dependencies in their native language, with similar manipulations of filler plausibility and dependency length. Chapter 5 presents an ERP study investigating whether L1 and L2 speakers at high and low proficiency levels show similar brain responses during *wh*-dependency processing. This chapter introduces the ERP method and relevant components, followed by methodology, results, and discussion. Finally, Chapter 6 summarizes all experimental results and discusses their implications for L1 and L2 filler-gap processing models, factors influencing L2 processing profiles, and directions for future research.

Chapter 2

The Nature of Human Sentence Processing

2.1 Introduction

In our daily lives, language processing is a complex task, as it requires real-time integration of various types of linguistic information, such as phonetic, phonemic, lexical, morphological, syntactic, semantic, pragmatic, and discourse representations. All these types of information must be rapidly integrated during language use. This research specifically examines sentence processing, which at least involves analyzing syntactic structures and interpreting semantic information. These two processes are termed *syntactic processes*, which involve generating and selecting syntactic hypotheses that are most likely to match the input, and *interpretive processes*, which involve setting up thematic relations and semantic representations (Omaki, 2010). While typical native speakers perform sentence processing with ease, second language (L2) sentence processing is always less efficient. Then, the question naturally follows: how is this incremental process carried out by both native and second language speakers, and what factors contribute to the differences between L1 and L2 processing?

This chapter provides a comprehensive literature review of theoretical perspectives and some experimental evidence regarding both L1 and L2 processing. It begins by introducing the concept of incremental processing in both L1 and L2, followed by a discussion of predictive processing in the two populations. The chapter then explores various theoretical models of L1 sentence processing, including the modular account as well as the interactive account. When it comes to the models specific to L2 sentence processing, I review some of the predominant models such as the *Shallow Structure Hypothesis (SSH)*, the *Declarative/Procedural Model*, and the *Prediction-by-Production Model*.

A major portion of the chapter is dedicated to examining the ongoing debates in the field of L2 processing research. It presents evidence for both qualitative similarities and differences between L1 and L2 processing. The chapter also discusses the shift in perspective from seeking a binary distinction to viewing L2 processing as a gradual and dynamic process. Following this debate, the chapter focuses on factors that modulate L2 sentence processing. It examines the roles of L2 proficiency, naturalistic exposure and immersion, and L1-L2 similarity in shaping L2 processing patterns. Throughout the chapter, relevant empirical studies are reviewed, including but not restricted to those investigating filler-gap dependencies, which are of particular interest to this dissertation. The chapter concludes with a summary of the main points discussed.

2.2 The Nature of L1 and L2 Sentence Processing

2.2.1 Incremental processing in L1 and L2

One of the major questions in psycholinguistics is how parsers process sentences in real time. Studies on real-time native sentence processing widely accept the idea that sentence processing is incremental; parsers integrate each single word as it unfolds rather than waiting for the entire sentence to be completed (e.g., Frazier, 1987; Altmann & Kamide, 1999; Aoshima, 2003; Aoshima et al., 2004).

Native incremental processing

As parsers incrementally merge each incoming linguistic element into a partial and developing interpretation, they often face difficulties with temporarily ambiguous sentences. This can lead to misanalyses partway through the sentence, requiring reanalysis until successful comprehension is achieved (e.g., Bever, 1970; Clifton et al., 2003; Frazier & Rayner, 1982; Trueswell et al., 1994; Juffs & Harrington, 1996). Reanalysis has been well attested in research on processing garden-path sentences. For example, when processing the sentence in (1), readers are expected to slow down when encountering the phrase *by the principal*.

(1) The teacher observed by the principal received a promotion.

The ambiguity arises at the word *observed,* which can be interpreted either as a past participle introducing a reduced relative clause or as a past tense verb in the main clause. The reading time slowdown observed at *by the principal* suggests that readers initially favor the main clause interpretation and then experience confusion when this phrase rules out the initial interpretation. This phenomenon is commonly referred to as a *garden-path effect* (Bever, 1970).

Non-native incremental processing

For non-native parsers, Juffs and Harrington (1996) provided evidence that L2 learners process temporarily ambiguous garden-path sentences similarly to native speakers, as they also experience misinterpretation and reanalysis. Their study measured word-by-word reading times for L1 Chinese advanced learners and native English speakers using two types of garden-path sentences:

(2) a. Sam warned the student cheated on the exam.

b. Jane knew her mother hated Tom.

(3) a. After Bill drank the water proved to be poisoned.

b. After Sam arrived the guests began to eat and drink.

(Juffs & Harrington, 1996, p. 297)

In (2b), the parser initially analyzed *her mother* as the direct object of *knew*, assigning it a theta role before encountering *hated.* Thus, reanalysis was required to reanalyze *her mother* as the subject of the embedded clause. Note that this reanalysis—reinterpreting *her mother* from the object of *knew* to the subject of *hated*—happens within the same thematic structure initially set up by the verb *knew*. To illustrate, the verb *knew* can take both a direct object (e.g., "Jane knew her mother.") and a complement clause (e.g., "Jane knew that …"). Although *knew* starts to take a complement rather than a direct object after the reanalysis, *her mother* remains within the same theta domain—that is, the set of arguments licensed by the verb *knew*. Because this reanalysis occurs within the same thematic structure, it is considered less effortful. In (2a), however, parsers initially interpret *the student* as the goal of the verb *warned,* assigning it a theta role as the recipient of the warning. Upon encountering *cheated,* they must abandon the first interpretation and take *the student* as the subject of the embedded clause (*the student cheated on the exam*). This reanalysis requires moving the student from one thematic domain—where it is the goal of the warning event—to another, where it is the agent of the cheating event. Because this process involves removing a theta role assigned by one verb and reassigning a new role under a different verb in a separate clause, it is considered a reanalysis across theta domains. As such, this type of garden-path sentence is more processing costly. Moreover, in (3a), the structural reanalysis arises from the need to reassign roles and recognize *the water* as the subject of a new clause after initially interpreting it as the object of *drank.* The verb *drank* can be both transitive and intransitive, and thus allows for multiple possible interpretations. In contrast, (3b) has a clear and simple structure with no need for reanalysis as the parsing follows the expected syntactic pattern without ambiguity. The results showed increased reading times for both native speakers and L2 learners on the second verb in (2a) and (3a) compared to (2b) and (3b). This indicates that L2 learners, like native speakers, utilize verb argument structures for their initial parsing decisions. When subsequent information contradicts the initial interpretation, both groups exhibit longer reading times, reflecting the need for reanalysis. Juffs and Harrington (1996) concluded that the parsing behavior of advanced L2 learners is almost identical to that of native speakers, despite overall slower reading times for the learners.

Reanalysis effects have also been observed in studies investigating speakers' processing of non-canonically ordered, temporarily ambiguous sentences, particularly those containing filler-gap dependencies. A *filler-gap dependency* describes the relationship between a displaced sentence element (the filler) and its original location (the gap), with the verb usually assigning a thematic role to the filler (Clifton & Frazier, 1989). For example, in the sentence "The patient that the doctor treated ___ yesterday," *The patient* is the filler, and the underscore represents the syntactic gap. Jessen and Felser (2019) investigated the processing of filler-gap dependencies in sentences by both native (L1) and L2 speakers, focusing on reanalysis effects. Using event-related potentials (ERPs), a neuroimaging technique that measures brain activity in response to specific events or stimuli, they compared native English speakers with German-speaking L2 learners of English when they read sentences such as "Bill liked the house/woman that Bob built ___ some ornaments for ___ at his workplace" (Jessen & Felser, 2019, p. 288). The study revealed that both groups initially attempted to integrate the filler with the potential lexical licensor *built,* as indicated by a larger N400 effect in the implausible condition (e.g., *woman*) compared to the plausible one (e.g., *house*). However, when later parts of the sentence (e.g., *some ornaments*) necessitated reanalysis, only the L2 speakers showed a P600 effect in response to the plausible-at-verb condition compared to the formerly implausible one. This suggests that L2 speakers engaged in more effortful syntactic reanalysis. In contrast, L1 speakers showed no significant differences between conditions, indicating a more automatic reanalysis process. The authors concluded that L2 speakers experienced greater difficulty recovering from an initial misanalysis than native speakers, consistent with earlier findings indicating that syntactic reanalysis is more costly for L2 speakers (e.g., Dussias & Piñar, 2010; Jessen & Felser, 2019; Roberts & Felser, 2011).

Based on L2 speakers' performance in processing both garden-path and filler-gap sentences, L2 speakers are capable of processing their second language incrementally. The detected differences in L1 and L2 processing may not necessarily indicate an inability for incremental processing. Rather, these differences appear to result from L2 speakers' overall slower reading times and greater challenges they are confronted with during reanalysis. L1 speakers, in contrast, have more automatic and efficient recovery from ambiguity.

2.2.2 Predictive processing in L1 and L2

Multiple studies have argued that during real-time sentence processing, native speakers employ predictive ability, by which they anticipate linguistic elements before they actually appear (Altmann & Kamide, 1999; Altmann & Mirković, 2009; Van Petten & Luka, 2012; Clifton & Frazier, 1989; Levy, 2008; Pickering & Gambi, 2018; Huettig et al., 2022). Predictive processes in language comprehension operate on multiple levels with varying degrees of certainty. These processes range from category predictions that are highly reliable, to probabilistic predictions, and to lexical predictions that show greater variability in their reliability (Phillips & Ehrenhofer, 2015). For example, the determiner "the" reliably signals that a noun will follow soon. Probabilistic predictions occur when certain verbs, like "expect," are often (but not always) followed by an infinitive, as in "expect to do something". Lexical predictions can vary greatly in reliability. For instance, the phrase "the book that the author …" is likely to be continued with the verb "wrote," although this is not guaranteed. Similarly, when hearing "the recipe that the chef...," one might anticipate the verb "created," but other verbs like "modified" or "borrowed" are also plausible.

Native predictive processing

Researchers generally concur that parsers consistently predict upcoming input, yet there is not a consensus on what type of linguistic information precisely constitutes prediction. Altmann and Kamide (1999) demonstrated prediction of semantics using the visual-world paradigm. They investigated how verbs can influence predictive processing during sentence comprehension. Participants viewed a semi-realistic visual scene, such as a boy, a cake, and various distractor objects, while listening to sentences like "the boy will move the cake" or "the boy will eat the cake."

The key finding was that eye movements to the target object (the cake) occurred significantly earlier in the "eat" condition than in the "move" condition. Specifically, saccades to the target were initiated before the spoken word "cake" was heard. They started from the verb "eat", suggesting that the verb "eat" allowed participants to predict the upcoming reference to the cake picture more effectively than the verb "move". This difference in eye movement timing indicated that the verb's semantic constraints enabled participants to anticipate the specific object that would be mentioned next. The study concluded that predictive processing in language comprehension involves using the semantic features associated with the verb to anticipate upcoming linguistic elements.

Using the same paradigm, Kamide et al. (2003) found that both German and English participants used case-marking and verb semantics to predict the most likely subsequent noun phrase before its onset. They thus concluded that syntactic and semantic cues are both integrated quickly to facilitate prediction, though with minor quantitative differences in the timing between the two languages. Ito et al. (2018) found that native speakers also pre-activated the critical noun's phonological form. They had L1 English and L1 Japanese-L2 English participants listening to highly predictable sentences,[6] as in (4), while being presented with a four-object display. The four visual displays comprised a critical object that was highly predictable (e.g., cloud) or an object whose onset was phonologically overlapped with the critical object (e.g., clown), and three unrelated items. As expected, both L1 and L2 participants looked more at the critical object (cloud) than the unrelated items before the word onset, indicating lexical prediction. Interestingly, they found that only L1 speakers also looked at the phonologically-overlapped object more than the unrelated one before the word onset, indicating that native speakers also pre-activated phonological form in advance of encountering the upcoming noun.

[6] The predictability refers to the probability of the target word in the blank. Before the eye-tracking experiment, 20 native English speakers and 12 L1 Japanese-L2 English speakers were asked to complete the experimental sentences with the first word coming to mind. 97.5% L1 speakers and 88.6% L2 speakers filled in the blanks with the intended target words. Although the cloze probability of the L2 group was lower than that of the L1 group, it was still much higher than the 65% reported for L2 participants in Martin et al. (2013).

(4) The tourists expected rain when the sun went behind the _____, but the weather got better later.

(Ito et al., 2018, p.4)

In addition to evidence from the visual-world paradigm, previous literature has employed various other methodologies to investigate predictive processing in native speakers (see Pickering & Gambi, 2018 for a review). For example, Wicha et al. (2004) examined how Spanish speakers use grammatical gender to predict following nouns using event-related potentials (ERPs), which measures brain's electrical activities associated with specific stimuli. They used highly predictive sentences (80 cloze probability) in which the target noun required a specific gender-marked article[7]. When participants read an article that mismatched the predicted noun in gender (e.g., masculine *el* instead of feminine *la* before the predicted feminine noun *maleta* "suitcase"), they showed an increased late positivity (500-700 ms post-stimulus), reflecting processing difficulty associated with gender agreement violations, even before the occurrence of the predicted noun.

Non-native predictive processing

Questions remain regarding to what extent L2 learners engage in predictive processing. In two review articles on L2 predictive processing, both Hopp (2022) and Pickering & Gambi (2018) suggest that L2 speakers do not apply predictive processing as effectively as native speakers do. However, their perspectives on the extent to which L1 and L2 predict upcoming linguistic inputs differ in certain respects. In the following paragraphs, I aim to compare their views on the similarities and differences in L1 and L2 predictive processing.

[7] In Spanish, the definite article *el* is used for masculine nouns while *la* is used for feminine nouns.

Both the above-mentioned reviews acknowledge that predictive processing in L2 learners is less robust compared to L1 speakers. Additional empirical evidence comes from a recent study by Covey et al. (2024) who investigated the processing of filler-gap dependencies in English by native speakers and Mandarin Chinese learners of English by examining their brain responses at filled gap sites within grammatically licit and illicit structures during sentence processing.[8] Native speakers showed an N400 effect only at licit gap sites, indicating predictive processing, whereas learners exhibited a P600 effect in the same condition, indicating syntactic integration difficulty. The study concluded that advanced Mandarin Chinese-speaking learners of English show sensitivity to grammatical constraints both online and offline, indicating that second language processing is guided by grammatical constraints. However, the P600 observed suggests that predictive processing in resolving *wh*-dependencies may be limited, as learners showed syntactic integration difficulty rather than gap prediction violation. Other studies have reported P600s at a filled-gap position for L2 learners, as well as native speakers (e.g. Dong, 2014; Jessen et al., 2017; Hestvik et al., 2012). In general, P600s at filled gaps have been argued to reflect syntactic integration difficulty. Covey and colleagues (2024) attributed the lack of gap expectation partially to the difference between participants' first and second languages. Specifically, while English instantiates overt *wh*-movement, Mandarin Chinese is a *wh*-in-situ language and does not necessarily embody overt *wh*-movement. The impact of L1-L2 difference on predictive processing was also shown in van Bergen and Flecken (2017)'s study, who investigated how differences in placement verbs between languages affect predictive processing. Specifically, in German and Dutch, different verbs are used depending on the position of the object being placed; there is one verb for putting something down in a standing position (put. STAND) and a different verb for putting something down in a lying position (put.LIE). In contrast, English and French leave the

[8] The grammatically illicit structure refers to a syntactic island, from which *wh*-extraction is prohibited under the framework of generative grammar.

position of the object unspecified; a single verb “put” is used regardless of the object’s position. Their study involved advanced L2 learners of Dutch with German, English, and French as their L1. The researchers found that L2 learners of Dutch, whose L1 (German) shares this similar placement verb distinction with Dutch, exhibited predictive eye movements similar to native Dutch speakers. However, French and English L2 learners of Dutch, whose L1s do not encode the same placement verb distinction, did not show such predictive effects despite their proficiency in Dutch. The findings from van Bergen and Flecken (2017) provided evidence that L1-L2 similarity modulates how L2 learners predict upcoming information in their second language.

Both Hopp (2022) and Pickering and Gambi (2018) also suggest that predictive processing is effortless for L2 users and that non-native speakers' predictions may be similar to native speakers' predictions at the semantic level (e.g., Hopp, 2015; Trenkic et al., 2014). However, non-native speakers may be less likely to predict other types of cues than native speakers. Evidence has been shown from L2 speakers' limited/reduced prediction in processing at the filled gap positions (Covey et al., 2024), at the article "a/an" (Martin et al., 2013), using grammatical gender (Foucart et al., 2014, 2016; Lew-Williams & Fernald, 2010; Dussias et al., 2013; Hopp, 2013), using case-marking (Mitsugi & MacWhinney, 2016; Hopp, 2015) and using classifiers (Grüter et al., 2020). For instance, Martin et al. (2013) investigated predictive processing in English monolinguals and Spanish-English bilinguals using ERPs. Participants read English sentences where the critical noun began with either a vowel or consonant, and the preceding article ("a" or "an") was manipulated to either match or mismatch the expected phonological form of the noun. For example, in "the day was breezy so the boy went outside to fly a kite/an airplane," the article *an* signals that a noun beginning with a vowel is likely to follow, thereby ruling out *kite* as the upcoming word. The results showed that English monolinguals exhibited an N400 response when the article did not match the expected noun, indicating prediction violation at the phonological level. However, bilinguals did not show this effect, suggesting that they are less able to make prediction using the phonological agreement markers like articles compared to monolinguals. Moreover, L2 speakers' predictive processing using morphological case marking has also been found to be limited. Mitsugi and MacWhinney (2016) explored how native Japanese speakers and L2 learners use case markers for predictive processing. While native speakers effectively utilized case markers to anticipate theme arguments and showed anticipatory looks before the theme appeared, L2 learners only showed increased looks at the theme after the onset of the theme noun. This indicates that non-native speakers' predictions at the morphological level are less robust compared to native speakers. Similarly, Grüter et al. (2020) examined whether L2 listeners prioritize semantic over grammatical cues of classifiers to predict upcoming words. Using a visual world paradigm, they tested listeners' fixation patterns in response to classifiers that preceded nouns in Mandarin Chinese. Some pictures matched the classifier grammatically but not semantically (e.g., *dog* with the classifier *tiáo* "long, thin"), while others matched semantically but were not grammatically licensed. Native speakers mostly fixated on nouns that were grammatically appropriate for the classifier, indicating greater reliance on grammatical form class during predictive processing. In contrast, L2 learners were more affected by the semantic information of the classifiers, often fixating on items that fit the meaning of the classifier even when they were not grammatically compatible. This suggests that while L2 speakers may effectively use semantic information for creating predictions, they are less likely to utilize grammatical form-class membership as efficiently as native speakers.

I have so far discussed both the incremental and predictive processing during real-time language comprehension. While early studies show language comprehension is incremental, with interpretations being built progressively as each word is encountered (e.g., Frazier, 1987; Altmann & Kamide, 1999), more recent accounts tend to demonstrate the parsers' ability to anticipate upcoming linguistic information before it is actually encountered (e.g., Levy, 2008; Huettig et al., 2022; Pickering & Gambi, 2018). A particular concern lies in the interpretation of research findings related to predictive or incremental processing. Predictive processing occurs before the input is received, while incremental processing happens as the input is being received. In addition, predictive processing is primarily top-down, using prior knowledge to generate expectations. Incremental processing is generally bottom-up, immediately integrating the upcoming information into the existing construction, which may sometimes require reanalysis. Crucially, these processes are not mutually exclusive; they often work together for efficient language comprehension. For example, Traxler and Pickering (1996, pp. 458–459) conducted a self-paced reading task where participants read sentences containing a filler-gap dependency (e.g., "That's the pistol/garage with which the heartless killer shot the hapless man yesterday afternoon."). When participants encountered *garage* followed by *shot,* they were predicted to experience processing difficulty as the two concepts do not have a logical instrument-action relationship. The researchers found that native English speakers experienced processing difficulty, reflected by a reading time slowdown at verbs *shot* when the preceding filler was implausible (e.g., *garage*) compared to when the filler was plausible (e.g., *pistol*). The researchers concluded that the increased reading time at *shot* was due to the prediction of an empty category, the trace, at this position. Since *shot* and *garage* are not semantically compatible, participants experienced a prediction violation when they encountered the verb. However, Pickering and Gambi (2018) pointed out that interpreting this effect as evidence for predictive processing is not definitive and depends on the theoretical framework adopted. According to theories that posit a trace, readers might predict an empty category before encountering the verb. In this case, the reading time slowdown at the verb *shot* suggests that readers were predicting an empty category, which would support predictive processing. In contrast, theories that does not assume empty categories would suggest that the filler (e.g., *with which*) is directly associated with the verb (e.g., *shot*) without the need for prediction. Under this view, the observed effect would reflect incremental processing, where comprehension builds up progressively as the sentence unfolds. Therefore, whether such findings are interpreted as evidence for predictive processing depends largely on the linguistic assumptions about syntax made by the researchers.

In summary, research has indicated that while L2 learners can utilize predictive processing, they do so less efficiently than native speakers, particularly with non-semantic cues. Native speakers effortlessly integrate both syntactic and semantic information to anticipate upcoming linguistic elements, whereas L2 learners rely more heavily on semantic cues, compensating for their difficulties with syntactic prediction, such as a forthcoming syntactic gap. In the following section, I will explore relevant theoretical models that explain L1 and L2 sentence processing, providing a theoretical framework for understanding some observed disparities.

2.2.3 Models of L1 sentence processing

Now the question turns to how parsers deal with various sources of linguistic information incrementally and predictively. Controversial questions thus arise: Is all information processed simultaneously, or is some information prioritized while other types are delayed? What factors drive individual differences in processing efficiency? A range of models have been proposed to explain how this process may occur. In the field of sentence processing, there are two different types of accounts regarding the how various kinds of linguistic information come into play. *Modular accounts* postulate that different modules are responsible for processing different linguistic cues, and they typically suggest that syntax guides the initial stages of sentence processing. In contrast, *interactive accounts* suggest that all linguistic representations are activated simultaneously.

Modular accounts posit that the human mind consists of specialized modules that operate on specific types of information, such as syntax, semantics and pragmatics, independently. These accounts suggest that language processing is not a holistic process but instead having a series of discrete stages operated by different modules (e.g., Fodor, 1983). Within the modular framework, syntactic information is integrated and analyzed before other types of linguistic information in sentence processing. This syntax-driven approach is a key feature of the garden-path model, which argues that syntactic cues are superior to other linguistic cues such as semantics and contextual cues in analyzing temporarily ambiguous sentences (e.g., Frazier, 1987). According to this model, sentence parsing involves two stages: an initial phase where the parser relies solely on syntactic information to construct the simplest possible structure using principles including Minimal Attachment, Late Closure, and Active Gap-Filling, and a subsequent phase where non-syntactic information may prompt reanalysis if the initial parse proves incorrect (see van Gompel, 2013 for a review).

Minimal Attachment stipulates that parsers prefer to build the most straightforward structure with the fewest number of tree structure nodes. For example, in the sentence "The horse raced past the barn fell," the parser initially interprets *raced* as a main verb rather than a past participle due to its simpler structure (Bever, 1970, p. 316). Reading difficulties arises when parsers encounter *fell*, and they have to proceed with reanalysis. In cases where there exist two analyses for an ambiguous structure, and both analyses have the same number of nodes, the principle of Late Closure enters the picture. Late Closure predicts that an ambiguous phrase is preferred to be attached to the most recent phrase. For instance, in the sentence "While Suzy was eating the pizza was delivered," when the parser encounters *the pizza*, it initially attaches it to *was eating*, following the Late Closure principle. Thus, *the pizza* is first interpreted as the object of *was eating*, and only upon reading *was delivered* does the parser realize the need of reanalyzing the sentence to correct the initial misinterpretation. Of particular relevance to this dissertation is the third principle within the modular account framework: Active Gap-Filling. This principle posits that during the processing of unbounded dependencies, such as relative clauses and *wh*-dependencies, parsers tend to fill gaps as early as possible based on the syntactic cues of unfolding sentence (Frazier & Clifton, 1989). For example, in the sentence "Who did the housekeeper from Germany urge the guests to consider ___?", parsers would prematurely fill a gap at *urge* based on the initial structural cues, resulting in processing difficulty when they encounter *the guests,* where the gap is expected to be (van Gompel & Pickering, 2012, p. 291).

Contrary to the modular accounts which endorse the existence of domain-specific modules that operate independently and sequentially, the interactive accounts suggest that multiple sources of information are activated immediately and processed simultaneously during sentence processing. These accounts assume that multiple sources of information, such as word meaning, syntactic structure cues, and contextual information, are employed in parallel for sentence comprehension and ambiguity resolution (MacDonald, 1999; Trueswell et al., 1994). For example, Trueswell et al. (1994) conducted an eye-tracking study to investigate how thematic role information affects the resolution of syntactic ambiguities. The study revealed that readers immediately utilize semantic information to guide their parsing decisions, often resolving ambiguities based on the plausibility of thematic roles. This research provides robust evidence for the interactive accounts, highlighting the simultaneous and parallel processing of syntactic and semantic information during sentence processing. However, Clifton et al. (2003) used the materials by Trueswell et al. (1994) and found that even with meaning-based cues (animate vs. inanimate), readers still struggled with ambiguous sentences. They thus suggested that while meaning-based thematic role information can influence understanding, it does not entirely prevent structural parsing difficulties. This challenges the idea that semantic information guides initial sentence understanding and suggests a more nuanced interaction between syntax and semantics than what proposed by interactive accounts.

It is important to note that interactive accounts form a framework under which there are several specific models that share the same core assumption: multiple sources of linguistic information operate simultaneously during sentence processing. For example, the *constraint-based model* assumes that during real-time sentence comprehension, multiple sources of information are activated in parallel as constraints for resolving ambiguities (e.g., MacDonald et al., 1994; McRae & Matsuki, 2013). The model assumes minimal delay in the activation of any particular type of information and thus predicts that even highly sensitive methods like eye-tracking would fail to detect delays in integrating non-structural information. Another notable model within the interactive account framework is proposed by Townsend and Bever (2001). They suggest that an initial "quick-and-dirty parse" is elicited based on lexical information, phrase segmentation, and possible argument relations between the heads of major phrases, and this initial, shallow representation is then refined into a complete structure that checks and revises the initial hypothesis if necessary (Townsend & Bever, 2001, p. 163). That is, non-structural information plays a crucial role in forming the initial parse, and it can be understood before a complete syntactic representation is fully computed. The quick-and-dirty model aligns with the Good-Enough Representations (GER) theory, which suggests that during language comprehension, our brains do not always create a detailed and precise syntactic structure. Instead, we tend to form a quick and superficial interpretation of sentences that is sufficient for the current context (Ferreira et al., 2002; Ferreira & Patson, 2007). Introduced by Ferreira and colleagues (2002), the concept of GER is that the cognitive system prioritizes efficiency and resource conservation, thus resulting in representations that are "good enough" for immediate needs. Ferreira and Patson (2007) further developed this concept and suggested that such representation is sufficient for effective communication without excessive burden on cognitive resources. All these specific interactive models reject the serial modular processing and endorse the rapid and parallel activation of multiple linguistic cues during real-time sentence processing. Interestingly, both the quick-and-dirty model and GER appear to support the idea that lexical-semantic cues may be weighted differently from syntactic cues. Such cue-weighting differences are central to the Shallow Structure Hypothesis (Clahsen & Felser, 2006), which will be discussed in detail in the following section.

2.2.4 Models of L2 sentence processing

Native speakers process language incrementally and predictively, making rapid use of both bottom-up and top-down information during real-time sentence processing (e.g., Altmann & Kamide, 1999; Frazier & Rayner, 1982). A question naturally arises in the field of second language processing: is this also the case for non-native speakers? As discussed in the introduction of the dissertation, there is an ongoing debate about whether L2 sentence processing fundamentally differs from that of native speakers. Building on this debate, Hopp (2022) indicates that the field is moving towards a more nuanced understanding of L2 sentence processing. Instead of viewing L2 processing as an imperfect form of L1 processing, current research is shifting beyond taking L1 as the standard to elucidating the unique features of L2 processing. Therefore, this section will explore key theoretical models that seek to explicate the characteristics of real-time L2 processing.

Shallow Structure Hypothesis

One of the most influential theories explaining the disparities between L1 and L2 processing is the Shallow Structure Hypothesis (SSH) (Clahsen & Felser, 2006; Clahsen & Felser, 2018). The SSH states that while native speakers construct comprehensive hierarchical syntactic representations using both top-down and bottom-up strategies, L2 learners tend to rely on shallower processing as they "compute predict-argument structure representations of the input that capture thematic roles and other aspects of lexical-semantic structure, but which lack hierarchical detail and more abstract elements of syntactic structure" (Clahsen & Felser, 2006, p.32). Marinis et al. (2005) provided evidence for SSH by conducting a self-paced reading task to examine the online processing of long-distance *wh*-dependencies. The study involved one L1 English group and four groups of L2 English learners with diverse L1 backgrounds (Mandarin Chinese, Japanese, German, and Greek). The experimental sentences were designed as follows:

(5) a. The manager who$_i$ the consultant claimed _t_i_ [+ intermediate site] that the new proposal had pleased _t_i_ [base position] will hire five workers tomorrow. (+extraction, + intermediate gap)

b. The manager who$_i$ the consultant's claim _t_i_ [- intermediate site] about the new proposal had pleased _t_i_ [base position] will hire five workers tomorrow. (+extraction, -intermediate gap)[9]

(Marinis et al., 2005, p. 58)

Marinis et al. (2005) found that both L1 and L2 groups showed elevated reading times at the canonical position for both (5a) and (5b), compared to sentences that do not involve movement. However, while native speakers were sensitive to the intermediate site, as shown in (5a), reflected by elevated reading times, L2 learners did not show sensitivity to the intermediate syntactic gap. The study suggested that while native speakers were sensitive to syntactic gaps, L2 learners often linked the fronted elements directly to their corresponding verbs, or lexical subcategorizers. These findings lent support to the idea that L2 learners, as suggested by the SSH, may underutilize syntactic information.

Further evidence comes from Felser and Roberts (2007), who used a cross-modal picture priming task to investigate whether advanced L1 Greek-L2 English speakers could demonstrate a priming effect at syntactically defined gap positions, similar to native English speakers as shown in Roberts et al. (2007). In the task, participants listened to sentences with filler-gap dependencies and responded to visual probes (e.g., deciding if the probe was alive) presented at specific positions, such as a gap position (where reactivation of the filler is expected) and a control position (where no reactivation should occur). An example stimulus is shown in (6). The priming effect expected by the researchers refers to faster responses to visual prompts semantically linked to a previously

9 According to the copy theory of movement (Chomsky, 1995), when a word or phrase is moved, a copy of that element is left at the base position. This empty category, or *copy*, is unpronounced but identical to the moved element. Government and Binding Theory treats this empty category as a *trace* (Chomsky, 1981). Given that prior psycholinguistic studies frequently associate the gap position with the trace (e.g., Nakano et al., 2002; Omaki & Schulz, 2011; Dong et al., 2023), the present dissertation uses the label t in example sentences to mark the position of the empty category.

mentioned filler (e.g., *the squirrel*). This faster response indicates that the filler is being reactivated at a specific point in the sentence. The study yielded intriguing results. Native English speakers, particularly those with high working memory capacity, demonstrated a priming effect only at the syntactically relevant gap site. In contrast, both high and low working memory capacity L2 learners showed a priming effect not only at the gap position but also at the control position. Based on these findings, Felser and Roberts (2007) concluded that L2 learners process sentences differently from native speakers. Instead of using syntactic structure to guide filler reactivation, L2 learners seem to maintain the filler active throughout the entire sentence rather than specifically reactivating it at the gap site. Their conclusion added weight to the Shallow Structure Hypothesis. However, Phillips and Ehrenhofer (2015) raised a common concern with studies supporting or opposing SSH: the evidence typically centers on whether L2 learners are sensitive to structural cues that native speakers rely on during parsing, but it fails to address whether L2 learners actually tend to prioritize semantic information over structural cues. This gap in the research is addressed in Experiment 1 of Chapter 4.

(6) Fred chased the squirrel to which the nice monkey explained the game's _[control position]_ difficult rules _[gap position]_ in class last Wednesday.

(Felser & Roberts, 2007, p. 20)

The conclusions by Marinis et al. (2005) and Felser and Robert (2007) have been questioned by later studies (e.g., Dekydtspotter et al., 2006; Rodriguez, 2008; Pliatsikas & Marinis, 2013). Upon closer examination of the reading time data from Marinis et al.'s study, Dekydtspotter et al. (2006) found that participants from the German and Japanese L2 groups showed spillover effects in the regions immediately following the intermediate site where native speakers would typically show the presence of an intermediate trace. Specifically, these L2 groups had longer reading times in the spillover region for sentences containing an intermediate trace compared to those without it. This delayed response suggested that L2 learners were indeed processing intermediate traces, albeit with a lag compared to native speakers. This finding undermined Marinis et al.'s conclusion and cast doubt on the Shallow Structure Hypothesis, indicating that L2 learners can process abstract syntactic information, though with delayed processing compared to native speakers. The difference in L1 and L2 parsing thus appears to be quantitative rather than qualitative. Notably, German exhibits overt *wh*-movement while Japanese does not, providing indirect evidence that such syntactic processing can occur regardless of whether the parser's L1 instantiates this structure. Further, using the same experimental sentences as in Marinis et al. (2005), Pliatsikas and Marinis (2013) conducted a self-pace reading study to investigate whether naturalistic exposure to L2 affects the processing of syntactic structures, specifically the intermediate gap in *wh*-movement, in a manner similar to native speakers. The study compared two groups of advanced Greek learners of English—one with significant naturalistic exposure and the other with primarily classroom exposure—alongside a control group of native English speakers. The results showed that the naturalistic exposure group processed intermediate gaps similarly to native speakers, showing facilitation at the final gap when an intermediate gap was present. In contrast, the classroom exposure group did not exhibit evidence of processing intermediate gaps, which suggested that they relied on lexical rather than syntactic information. This study suggested that naturalistic exposure can mitigate the limitations observed by Marinis et al. (2005), enabling L2 learners to achieve more native-like syntactic processing. These findings challenge the SSH, which posits that L2 learners, even with high proficiency, underuse abstract syntactic structures and rely primarily on lexical-semantic information when processing their L2. Their findings are consistent with growing arguments that the difference between L1 and L2 sentence processing is gradual and dynamic instead of qualitative and unchangeable.

Since the initial proposal of the SSH, numerous studies have shown that L2 speakers can utilize syntactic information from input during processing, albeit with a consistent delay (e.g., Omaki & Schulz, 2011; Pliatsikas & Marinis, 2013; Fernandez et al., 2018; Boxell & Felser, 2017). In light of these findings, Clahsen and Felser (2018) revisited and clarified some misunderstandings of the original SSH. Notably, Clahsen and Felser (2018) restated the SSH, arguing that the SSH has often been misinterpreted as a deficit hypothesis focusing on L2 learners' fundamental incapability of using grammatical information effectively. This interpretation is not accurate. Instead, the SSH suggests that L2 speakers might rely more on semantic, pragmatic, or surface-level information compared to L1 speakers, not because they lack syntactic knowledge, but because they have different preferences for utilizing linguistic information. Furthermore, the SSH does not claim that L2 learners can never achieve native-like syntactic parsing. Rather, it suggests that achieving native-like parsing depends on acquiring a native-like grammar, and that with increased exposure, grammatical processing routines can become more automatic. The updated SSH also outlines some approaches to further develop the SSH. Two of these approaches are particularly relevant to the present dissertation. On the one hand, they introduced the idea of "variability in grammatical constraints", suggesting that grammatical knowledge in L2 may be less robust than in L1, and that such L1-L2 differences might reflect gradient rather than categorical distinctions (Clahsen & Felser, 2018, pp. 698–699). This idea of gradient differences contrasts with the traditional binary view, which assumes that grammatical knowledge is either known or not known by a speaker. On the other hand, the SSH has been extended to address not just the types of information used during sentence processing, but also the timing at which grammatical and non-grammatical cues are utilized by L1 and L2 speakers. The updated SSH posits that L2 speakers might exhibit delays, rather than an absence, in processing grammatical information compared to L1 speakers, which can lead to differences in real-time language comprehension.

A recent study by Dong et al. (2023) provides evidence for the updated SSH. They examined how Chinese learners of English process filler-gap dependencies in real time using ERPs, with proficiency and working memory as between-participant variables. The study employed four conditions: Filled-Gap, Adjunct, Object, and Trace, as shown in (7).[10] In both the Filled-Gap and Adjunct conditions, a relativized object filler prompts parsers to search for a possible gap. The main difference is that in the Adjunct condition, the dependency continues past the post-verbal noun phrase without interruption, while in the Filled-Gap condition, the dependency is disrupted by an unexpected noun phrase after the verb, at a position where a gap is expected, resulting in a filled-gap effect. The Object and Trace conditions serve as controls, ensuring that only half of the object relative clauses contain a filled gap. The findings revealed that while L2 learners' behavioral responses were similar to native speakers, their brain responses differed, showing a prefrontal-central positivity rather than the P600 found in native speakers. This suggests that L2 learners showed reduced sensitivity to syntactic violation during gap-filling, which is consistent with the hypothesis proposed in the updated SSH that L2 learners may be able to apply structural information but with less sensitivity than L1 speakers.

(7) a. Filled-Gap condition:

*The zebra that the hippo kissed the camel on the nose ran far away.

b Adjunct condition:

The weekend that the hippo kissed the camel on the nose, it was humid.

c. Object condition:

The zebra said that the hippo kissed the camel on the nose and then ran far away.

d. Trace condition:

The zebra that the hippo kissed on the nose ran far away.

(Dong et al., 2023, p. 9)

[10] The stimuli were originally developed by Hestvik et al. (2012) and were used by Dong et al. (2023) in their ERP study.

The SSH has sparked abundant debate within the field of bilingual processing, leading to a substantial body of supporting and opposing evidence. In Section 3.4, I will thoroughly discuss additional empirical studies on filler-gap processing that both endorse and challenge the updated SSH.

Declarative/Procedural Model and Developmental Model

The updated SSH echoes the Declarative/Procedural Model (DPM) proposed by Ullman (2001) in that both theories attribute late L2 learners' shallow processing to their reliance on lexical-semantic information and their limited capacity for full syntactic parsing. The DPM posits that language relies on two distinct memory systems: the *declarative memory* system, which is responsible for lexical information, and the *procedural memory* system, which handles grammatical rules. Crucially, research has shown that procedural memory reaches full functionality before puberty (Casey et al., 2005; Janacsek et al., 2012). For instance, Janacsek et al. (2012) investigated implicit sequence learning, which is closely tied to procedural memory, across different age groups (from 4 to 85 years old). Their results revealed that children aged 4 to 12 exhibited stronger learning effects, particularly in their sensitivity to raw probabilities in a sequence learning task, compared to older participants. This supports the idea that procedural memory matures relatively early in life, typically before puberty. Regarding L2 sentence processing, earlier learners and late learners differ in how they use the two memories for parsing. Unlike child L2 learners who can effectively use procedural memory, adult learners who begin learning after puberty tend to rely more heavily on declarative memory for sentence processing as their procedural memory becomes less accessible while declarative memory remains more robust. It is important to note that Ullman also mentions that increasing experience with L2 would lead adult learners to lean towards procedural memory, and this shift would result in high proficiency

(Ullman, 2011). In fact, Clahsen and Felser (2018) initially interpreted their findings of L1/L2 difference in morphological processing as supporting evidence of DPM, but they later developed a more refined theory that accounts for more nuanced patterns in L2 processing, such as L2 speakers' native-like processing of lexical constraints but reduced sensitivity to morphological constraints.

The DPM is also partially in line with the Critical Period Hypothesis, which suggests that there is a limited window during which language acquisition occurs most naturally and efficiently (Lenneberg, 1967; Penfields & Roberts, 1959). After this period, reduced brain plasticity makes achieving native-like processing in a second language, especially in syntax, more challenging. However, unlike the Critical Period Hypothesis that suggests a gradual decline in language learning ability after a biologically determined period, the DPM acknowledges a possible path to native-like proficiency through increased practice. Similarly, Steinhauer et al. (2009, 2014) proposed a hierarchy of developmental stages and summarized a series of phases that learners go through as they become more proficient in their L2. This model argues against rigid constraints on L2 learners' abilities to process abstract syntactic information. Instead, it suggests that learners progress through discrete developmental stages and that, with increasing proficiency, L2 learners can develop processing mechanisms that are qualitatively similar to those of native speakers (Steinhauer et al., 2014). It is important to note that Steinhauer et al.'s model does not directly address the Critical Period Hypothesis. Rather, it implies that L2 learners can achieve native-like syntactic processing abilities through these developmental stages, regardless of the age at which they begin learning the language. This model emphasizes the role of proficiency over age-related factors, suggesting that even adult learners can achieve advanced levels of syntactic processing. Steinhauer and colleagues' developmental model aligns with DPM in that both challenge the biologically determined critical period beyond which native-like attainment is unlikely, and both acknowledge that L2 processing can be modulated. The difference is that DPM suggests increased practice leads to native-like use of procedural memory, reflected by high proficiency, whereas Steinhauer's developmental model emphasizes that higher proficiency drives L2 speakers to move forward to a higher developmental stage. However, this raises a potential problem of circular reasoning: if higher proficiency is both the cause and consequence of native-like processing, it becomes difficult to determine the direction of causality. Further research is needed to understand the relationship between proficiency and native-like processing, specifically whether proficiency causes native-like processing or native-like processing is simply manifested as high proficiency. Another difference is that DPM acknowledges the importance of the age of exposure while Steinhauer did not directly address age-related constraints.

Recent empirical evidence comes from Wolpert et al.'s (2024) study, in which they examined the real-time processing of adjective placement in English and Mandarin using event-related potentials (ERPs). Experiment 1 was conducted in English with native English speakers who had high proficiency in Mandarin. They reported a mean age of onset for Mandarin learning at 16 years. This experiment was designed to test if speakers' L2 would influence the processing of their L1. Experiment 2 investigated L1 and L2 processing of Mandarin adjective placement with two participant groups: L1 Mandarin speakers and the same group of native English speakers from Experiment 1. The experimental sentences were designed to be either grammatically correct (adjective before noun) or incorrect (noun before adjective), as shown in (8). Both languages use prenominal adjectives, or adjective-noun word order. A key difference is that Mandarin typically has a particle DE being inserted between the adjective and the noun, as in *haochi DE tiandian* ("tasty DE dessert"). The absence of DE is permitted with mono-syllabic adjectives, and thus the absence of DE would lead to the adjective-noun pair functioning as a single word, as in *mei (DE-absent) nü* ("beautiful woman/beauty"). In (8b), the particle DE is absent between the adjective and the noun; thus, *xin shu* may function as a single word rather than an adjective-noun phrase. Therefore, despite surface similarities, subtle differences exist between the languages, and the researchers aimed to understand if or how native and second language speakers figure out the subtle underlying differences even when the surface structures look alike, and whether knowledge of one language helps or interferes with processing the other.

(8) a. English stimuli

Grammatical: Tom puts the white vase on the table.

Ungrammatical: Tom puts the vase white on the table.

b. Mandarin stimuli

Grammatical:

fuqin	*zuijin*	*xiangkan*	*xin*	*shu*	*le*
Father	recently	wants to read	new	books	Perf[11]

Ungrammatical:

fuqin	*zuijin*	*xiangkan*	*shu*	*xin*	*le*
Father	recently	wants to read	books	new	Pref

(Wolpert et al., 2024, p. 4)

Both native English and Mandarin speakers displayed a biphasic N400-P600 response to adjective placement violations. However, the L2 Mandarin learners exhibited an N400 effect for ungrammatical adjective-noun pairs in the Mandarin stimuli but did not show a subsequent P600 effect for the 500-1000ms time window. This absence of the P600 in L2 learners suggests that L2 Mandarin learners may have not yet fully acquired the procedural memory required for reanalysis or repair of sentence structure. The authors suggest that this reflects an incomplete grammaticalization of Mandarin in these L2 learners, where they rely more on lexical processing (indexed by the N400) rather than on syntactic processing (indexed by the P600) that is more common in native speakers. This finding aligns with the idea that higher proficiency in a second

[11] Perf is an aspect marker to indicate the completion, equivalent to the prefect tense in English.

language is necessary to trigger native-like syntactic processing mechanisms, such as those reflected by the P600 component. The findings are thus consistent with the declarative/procedural model proposed by Ullman (2001) and the developmental stages model outlined by Steinhauer et al. (2009, 2014).

Prediction-by-Production Model

Recently, Pickering and Gambi (2018) proposed a model of prediction based on two key mechanisms: *prediction-by-production* and *prediction-by-association*. In their model, the prediction-by-production mechanism involves comprehenders covertly imitating the speaker's utterance as they hear it, using their own language production system to generate predictions about what comes next. This process is optional and depends on the comprehenders' available cognitive resources and time. In contrast, prediction-by-association is a “non-optional” mechanism demanding relatively less resources (Pickering & Gambi, 2018, p. 1015). It functions through the automatic spreading of activation between related concepts stored in the comprehenders' memory. For example, upon hearing the word "doctor," related words like "nurse" or "hospital" might be automatically activated. This mechanism rapidly activates a wide network of related items, but this activation decays quickly and may therefore play only a limited role in predicting subsequent words.

The model also accounts for differences between L1 and L2 speakers in prediction. While L1 speakers can use prediction-by-production automatically, L2 speakers may rely more on prediction-by-association and less on the more demanding prediction-by-production, leading to slower or less accurate comprehension compared to their L1 counterparts. A recent study by Corps et al. (2023) used the visual-world eye-tracking paradigm to explore whether L2 prediction involves two distinct stages, as suggested by the prediction-by-production model. Thirty-two highly proficient L2 English speakers with various native language backgrounds including Chinese, Italian, Spanish, German, Dutch, Japanese, Russian, Greek and Malay participated in the study. The experimental materials consisted of experimental sentences paired with visual displays of four colored objects. These sentences were divided into gendered sentences, containing predictive verbs associated with stereotypically masculine or feminine objects, and gender-neutral filler sentences. For example, in one of the gendered sentences, a male speaker might say, "I would like to wear the nice…" followed by visual options such as a tie (masculine target), a dress (feminine target), a drill (masculine distractor), and a hairdryer (feminine distractor). Participants' eye movements were tracked as they listened to the sentences and viewed the images. The study aimed to determine whether participants made predictions based on associations (e.g., anticipating a tie after hearing "wear") or based on consistency (e.g., predicting a tie when a male speaker was talking). Both groups showed significantly more fixations to the semantically associated target compared to distractors, with L2 speakers showing this effect around 527ms after verb onset and L1 speakers around 519ms after verb onset. This suggested that both groups quickly engaged in associative predictions. However, for consistent predictions that required the consideration of the speaker's gender, L2 speakers were 251ms slower than L1 speakers in fixating the target (L1: 641ms; L2: 957ms). The researchers attributed the slower consistent prediction in L2 speakers to the increased cognitive demands of L2 comprehension, which left them limited cognitive resources for making predictions that integrate speaker characteristics.

It is important to note that evidence of prediction-by-production in the processing of filler-gap dependencies is relatively rare, as most studies in the field do not involve a potential interaction. According to the prediction-by-production model, the experimental design requires an intermediate site between the filler and the actual gap to suggest predictive processing. For example, Traxler and Pickering (1996, pp. 458–459) addressed the difficulty readers experience at the verb *shot* in the sentence "That's the garage with which the heartless killer shot the hapless man yesterday afternoon," as readers anticipate the syntactic structure before it fully unfolds. This aligns with the prediction-by-production model. However, if the first potential gap of the filler *with which* is the verb *shot*, without any intermediate site, then the observed difficulty at *shot* does not necessarily indicate predictive processing but may instead denote incremental processing.

2.3 Issues in L2 sentence processing

As briefly discussed in Chapter 1, a central question in L2 processing research is whether the observed differences between L1 and L2 sentence processing are primarily quantitative or qualitative in nature. Quantitative differences would reflect the same underlying mechanisms operating with reduced efficiency in L2 speakers, while qualitative differences would suggest that L2 learners employ fundamentally different cognitive strategies. This question has driven extensive debate and research, leading to various theories that attempt to explain the similarities and differences between L1 and L2 sentence processing. In this section, I will review the current debates and empirical evidence concerning whether L2 sentence processing is qualitatively similar to or different from L1 processing. It is important to note that the syntactic structures examined in this section vary; empirical studies related specifically to the processing of filler-gap dependencies will be discussed in Chapter 3. In the final subsection, I will consider various factors that are argued to modulate L2 processing, such as L2 proficiency, extent of naturalistic exposure and immersion, as well as the structural similarity between L1 and L2.

2.3.1 Debate: Qualitatively similar but quantitatively different?

The distinction between quantitative and qualitative differences in L1 and L2 sentence processing is not always clear-cut. The distinction to me seems to be that quantitative differences in L1 and L2 processing refer to variations in the degree or extent of certain behaviors, such as timing, strength, or the likelihood of making predictions. In contrast, qualitative differences refer to differences in the types of cues used (e.g., semantic or syntactic) or the nature of processing mechanisms employed by L1 and L2 speakers. To illustrate, if L1 and L2 speakers exhibit the same ERP component in response to a specific input, but with subtle variations in timing or strength, this would suggest a quantitative difference. Conversely, if a fundamentally different ERP component is observed, even if the two groups achieve similar accuracy on the surface, this would suggest a qualitative difference in the underlying processing mechanisms. Researchers have commented on the initially proposed SSH suggesting that the shallow representations or grammatical rules might not only be a characteristic of L2 processing but can, under certain circumstances, also influence native speakers' processing, though this effect is more pronounced in L2 speakers (Sabourin, 2006; Clashen & Felser, 2018) Therefore, although L2 processing appears to be slower or weaker, this can be considered quantitative differences rather than qualitative ones (Sabourin, 2006; Schlenter, 2023).

Numerous studies have demonstrated that L2 learners exhibit qualitative similarities to native speakers in syntactic processing. One such study by Williams et al. (2001) employed a stop-making-sense judgment task during self-paced reading to assess how advanced English learners from diverse L1 backgrounds process sentences with plausible and implausible verb-object relationships. Participants were presented with sentences where the plausibility of the filler, *which girl* and *which river,* was manipulated in relation to the main verb *push*, as shown in (9). In this stop-making-sense task, participants read through a sentence word-by-word and pressed a “reject” button when the sentence no longer made sense. The reaction times at the rejection point were recorded to indicate where semantic implausibility was first identified. The study found that L2 learners, regardless of their native language background, processed these structures similarly to native speakers by identifying the earliest possible gap site after *push*. However, the L2 learners were slower in reanalyzing the syntactic structure, particularly in implausible conditions (e.g., *push the river*). This indicated qualitatively similar mechanisms but less efficient processing.

(9) a. Which girl did the man push __ the bike into __ late last night? (plausible at V)

b. Which river did the man push __ the bike into __ late last night? (implausible at V).

(Williams et al., 2001, p. 517)

The importance of cognitive load and task complexity in L2 processing is also highlighted by the Continuity Hypothesis, as advocated by Hopp (2010) and McDonald (2006). According to this hypothesis, L2 learners consistently employ the same universal parsing strategies as native speakers but may struggle due to resource limitation. This limitation becomes more pronounced as sentence complexity increases, such as in long-distance non-local dependencies, leading to greater processing differences between L1 and L2 learners. Kaan (2014) suggests that the Continuity Hypothesis is compatible with a predictive approach, as discussed in section 2.2.2. Due to the "resource deficits" (Kaan, 2014, p. 271), L2 learners are less effective at predicting upcoming information as task complexity increases, resulting in performance differences between L1 and L2 processing.

The *Reduced Ability to Generate Expectations* (RAGE) hypothesis proposed by Grüter and Rohde (2013, 2021) aligns with the idea of resource deficits leading to reduced predictive ability. According to this hypothesis, while L2 learners can integrate semantic and morphosyntactic information during online processing, their ability to use this information to predict upcoming sentence structure is weakened compared to native speakers. This weakened ability makes L2 learners more susceptible to processing difficulties in complex syntactic environments, such as those found in island structures. Evidence for the RAGE hypothesis can be seen in Canales (2012), who examined whether Spanish learners of English can assess abstract syntactic representations during the online processing of *wh*-dependencies, as shown in (10). If L2 sentence processing is syntax-driven, they should be aware of the island constraint, a structure that hinders *wh*-extraction, just as native speakers do. In this case, they should not attempt to associate *who* with a gap inside the relative clause island. Therefore, there should be no difference in reading times at *Jacob* between sentences like (10a) and (10b).

(10) a. My teacher wondered **if** the principal that suspended Jacob last spring disappointed the parents with the news.

b. My teacher wondered **who** the principal that suspended Jacob last spring disappointed with the news.

(Canales, 2012, p. 53)

The findings showed no significant differences in reading times at the critical region *Jacob* for either native speakers or learners, suggesting that learners, like native speakers, are able to access detailed syntactic representations while processing theses sentences in real time. However, an interesting finding was observed at *suspended* only in the L2 group. The learners exhibited significantly longer reading times at *suspended* for the illicit *wh*-extraction condition, as in (10b), compared to the same region in the non-extraction condition, as in (10a). The author explained that learners may have expected a verb to immediately follow the subject *the principal*, as in a sentence like "My teacher wondered who the principal fired last spring" (Canales, 2012, p. 73) They were surprised when they encountered a relative clause introduced by that, and this unexpected input likely caused the longer reading times observed at the region following that in the illicit wh-extraction condition. The author noted that L1 speakers may also experience this kind of surprise but are better at recovering from such surprises. Thus, Canales (2021) concluded that L2 parsing may not be qualitatively different from L1 parsing, as both are guided by syntax. The different processing patterns observed at the verb *suspended* between L1 and L2 groups appear to reflect L2 learners' reduced ability to recover from unexpected input. According to the RAGE hypothesis, there is no need to stipulate a qualitative difference; instead, it is L2 speakers' reduced predictive ability that results in the surprise at the unexpected input. However, a methodological concern here is that no comprehension questions were reported in Canales' study (2012). Without measuring L2 speakers' comprehension, we cannot determine whether the comparable reading times between the two groups reflect similar processing mechanisms or simply result from L2 learners not fully engaging in comprehension of these complex structures.

Further evidence for qualitative similarity between L1 and L2 sentence processing comes from challenging existing hypotheses that suggest fundamental differences, for example, the SSH. Rodríguez (2008) recruited two groups of non-native speakers of English: a group of native Spanish speakers and a group of native Chinese speakers, along with a native English speaker group as the baseline. The participants were asked to perform a self-paced reading comprehension task in English where the materials included intermediate gaps and garden path sentences, as shown in (11). The study aimed to assess the SSH proposed by Clahsen and Felser (2006). An analysis of the residual reaction times (RTs) for these sentences across different groups revealed that the patterns of peaks and variations are comparable between learners and native speakers. All participants exhibited increased RTs at the complementizer *that* or the preposition *about.* Additionally, all groups showed increased RTs at the gap site *will not*, for all sentence types. Rodríguez then concluded that both native and non-native speakers exhibit similar parsing profiles, suggesting that the differences in processing may not be as fundamental as the SSH proposes.

(11) a. Intermediate gap:

The man who | the lawyer determined | **that** | the illegal contract | **had confused** ___ | will not | go to prison.

b. Garden-path:

The man who | the lawyer understood | **that** | the illegal contract | **had confused** ___ | will not | go to prison.

c. Nominalization:

The man who | the lawyer's determination | **about** | the illegal contract | **had confused** ___ | will not | go to prison.

(Rodríguez, 2008, p. 71)

Similarly, a recent study by Song et al. (2022) investigated the processing of Mandarin Chinese base-generated topic sentences in native and second language speakers (e.g*., Shuiguo wo zui ai chi xiangjiao* "As for fruit, I love eating bananas most"). Using an eye-tracking experiment, the researchers examined the extent to which L1-Korean speakers can achieve native-like processing of Mandarin Chinese base-generated topic sentences. Participants included native Mandarin Chinese speakers and L1-Korean learners, who were divided into intermediate and advanced groups based on their proficiency. The experimental design included four distinct conditions, as illustrated in (12). The results showed that both the intermediate and advanced Korean learners showed syntactic reanalysis when they first processed the topic, the first noun, and then encounter the subject *wo ("I")* as in (12a), (12b), and (12c). This suggested their sensitivity to syntactic cues in L2 sentence processing. For the processing of the object noun phrase, however, there was a clear difference between proficiency groups. While both native speakers and advanced learners showed increased reading times at the unacceptable object noun phrase, *shuiguo* ("fruit") in (12b) and *xiangjiao* ("banana") in (12c), compared to the acceptable object noun phrase *xiangjiao* ("banana") in (12a), this similar processing pattern was absent in the intermediate group. The researchers suggested that the intermediate learners' lack of sensitivity to the semantic constraint—the topic and its related NP should form a superordinate-hyponym relation—may be due to their reliance on the verb for prediction and their limited capacity to abandon the initial interpretation. These findings challenged the SSH from two aspects. First, Korean learners of Mandarin Chinese demonstrated the ability to use structural information, contrary to SSH's claim of shallow syntactic processing in L2. Moreover, only advanced learners showed native-like sensitivity to semantic constraints, contrary to SSH's assumption that L2 learners prioritize and rely on semantic information during sentence processing.

(12) a. Base-generated sentence: Acceptable sentences where the topic *fruit* and its related noun phrase *banana* form a superordinate-hyponym relation.

Shuiguo wo zui ai chi xiangjiao, suoyi wo jingchang mai xiangjiao.

fruit I most love eat **banana**, so I often buy banana.

'As for fruits, I love eating bananas most, so I often buy bananas.'

b. Base-generated sentence: Unacceptable sentences where the topic banana and its related NP fruit form a hyponym-superordinate relation.

Xiangjiao wo zui ai chi shuiguo, suoyi wo jingchang mai shuiguo.

banana I most love eat **fruit**, so I often buy fruit.

'As for bananas, I love eating fruits most, so I often buy fruits.'

c. Base-generated sentence: Unacceptable sentences where the topic apple and its related NP banana form a sisterhood relation.

Pingguo wo zui ai chi xiangjiao, suoyi wo jingchang mai xiangjiao.

apple I most love eat **banana**, so I often buy banana.

'As for apples, I love eating bananas most, so I often buy bananas.'

d. Non base-generated sentence: Acceptable control sentences beginning with an adverb before.

Yiqian wo zui ai chi xiangjiao, suoyi wo jingchang mai xiangjiao.

before I most like eat **banana**, so I often buy banana.

'I used to love eating bananas most, so I often buy bananas.'

(Song et al., 2022, p. 7)

2.3.2 Debate: Qualitatively different?

As discussed in the previous section, the subtle differences observed in L1 and L2 processing may be interpreted as quantitative rather than qualitative, as they can be attributed to factors such as delayed sensitivity to linguistic cues in L2, reduced predictive ability in L2, and the increased cognitive load required by L2. Although Clahsen and Felser (2018) noted in their discussion of the Shallow Structure Hypothesis that an absolute dichotomy was not what they intended, they did not explicitly support the notion of qualitative similarity. Instead, they highlighted a gradual L1/L2 difference. This suggests that the differences between L1 and L2 processing are not stark or absolute but rather exist on a continuum. However, these gradual differences are nuanced and not as clear-cut as a black-and-white distinction. Clahsen and Felser (2018) suggest that the efficient use of non-grammatical information can enable learners to achieve native-like attainment. Thus, to argue for qualitative differences, sensitive experimental techniques would need to be employed. For example, if electroencephalogram (EEG) data shows that specific linguistic input triggers completely different components in native and non-native speakers, we could argue for qualitative differences in L1 and L2 processing. In the following section, I will review studies that explore the qualitative distinctions between L1 and L2 sentence processing.

One piece of evidence for qualitative differences between L1 and L2 processing comes from empirical studies that show the absence of predictive use of case marking in non-native speakers, regardless of proficiency. For example, both Hopp (2015) and Mitsugi and MacWhinney (2016) investigated how L2 learners use morphosyntactic information to predict upcoming arguments in comparison to native speakers. Hopp (2015) examined how L1 and L2 speakers use morphosyntactic cues, specifically case marking, to predict syntactic structures in German. The experimental sentences were in both subject-initial and object-initial orders. For instance, in *Den Wolf tötet sogleich der Jäger…* ("The-ACC wolf kills soon the-NOM hunter…"), the case markers on *Den Wolf* ("accusative wolf") and *der Jäger* ("nominative hunter") signal their roles as the object and subject, respectively (Hopp, 2015, p. 287). L1 speakers of German were able to use these morphosyntactic cues to anticipate that *der Jäger* ("nominative hunter") would be the agent performing the action. However, the study found that L2 speakers of German, whose native language was English, did not use the morphosyntactic cues to anticipate the subsequent agent or patient regardless of their proficiency in German. Instead, they relied more on the semantic information provided by the verb *tötet* ("killed") and the linear order of the words. As such, L2 speakers showed more looks towards *der* ("nominative") in the example, suggesting that they encountered difficulty interpreting *der Jäger* ("nominative hunter") as the subject of the sentence and the doer of the verb. This lack of using case marking for prediction was not modulated even by high proficiency. The absence of using morphosyntactic cues and the reliance on lexical and semantic cues illustrate how L1 and L2 speakers differ in predictive processing, at least in morphosyntactic prediction.

Similarly, Mitsugi and MacWhinney's (2016) eye-tracking study examined if native Japanese speakers and intermediate-level English learners of Japanese could use case markers to predict what comes next in a sentence in a verb-final structure. The experiment compared participants' looks with three types of sentence structures, as shown in (13).

(13) a. Canonical Order: A sentence with a typical S-IO-DO order

School-LOC serious student-NOM strict teacher-DAT quietly test-ACC handed over

"At school, the serious student quietly handed over the test to the strict teacher."

b. Scrambled Order: A sentence with a reversed IO-S-DO order

School-LOC strict teacher-DAT serious student-NOM quietly test-ACC handed over

c. Accusative Condition: A sentence with a monotransitive construction (S-DO) that rules out the presence of an additional argument

School-LOC serious student-NOM strict teacher-ACC quietly teased

"At school, the serious student quietly teased the strict teacher."[12]

(Witsugi & MacWhinney, 2016, p. 23)

[12] The abbreviations used here are: S (subject), IO (indirect object), DO (direct object), NOM (nominative case), DAT (dative case), ACC (accusative case), and LOC (locative case).

The researchers aimed to investigate whether L1 and L2 Japanese speakers would show evidence of prediction at the adverb *quietly* before they encounter the theme *test-ACC.* To achieve this, the researchers analyzed the proportion of participants' eye fixations to the theme compared to other items and measured from which region would participants show significantly more looks towards the theme *test-ACC.* The results showed that in the canonical and scrambled conditions, L1 speakers revealed increasing looks to the theme *test-ACC* starting from the onset of the adverb *quietly,* whereas L2 speakers only began to fixate on the theme *test-ACC* after its occurrence. They thus suggested that L2 speakers did not use morphological case markers for prediction. Schlenter (2023), in a recent review on prediction in bilingual sentence processing, suggests that this consistent qualitative difference in the predictive use of case marking may be due to unfamiliarity with case marking from their L1. However, if the similarity between L1 and L2 processing relies on the existence of similar structures in the participants' L1, it indirectly supports the argument for a qualitative difference.

In section 2.2.4, I reviewed two ERP studies that provided evidence for qualitative differences between L1 and L2 processing. To briefly recap, Wolpert et al. (2024) examined adjective placement in English and Mandarin and found that native speakers of both languages showed a biphasic N400-P600 response to ungrammatical adjective-noun order, while L1 English-L2 Mandarin bilinguals showed only an N400. The absence of a P600 was attributed to structure-specific, rather than general, proficiency. However, this explanation seems problematic because the experimental sentences used in the study avoided the distinctions in adjective placement between English and Mandarin. Specifically, the experimental sentences neither included an intervening degree adverb *(e.g., Zhang really smart)* nor the intervening particle *DE* (e.g., *red DE socks*). Therefore, due to the similarity of the adjective-noun order in both English and Mandarin, the bilinguals would be expected to exhibit at least a delayed or slight P600. The complete absence of the P600 was inconsistent with the ERP profile of native speakers and thus provided evidence for the qualitative difference between L1 and L2 processing of adjective placement.

Using the same technique, Dong et al. (2023) presented a more straightforward conclusion regarding the qualitative difference between L1 and L2 filler-gap processing (see 2.2.4). The study found that while L2 learners' behavioral responses were similar to those of native speakers, their brain responses were qualitatively different. Specifically, native speakers exhibited a canonical central-parietal P600, associated with syntactic violations and reanalysis, whereas L2 learners showed a prefrontal-central positivity. This non-native-like ERP pattern was not modulated by their proficiency or working memory capacity and thus provided evidence for the qualitative differences between L1 and L2 filler-gap processing.

2.3.3 Factors modulating sentence processing in L2

Previous research has indicated that factors affecting native sentence processing, such as working memory (e.g., Jung, 2017; Cunning, 2017), executive functioning (e.g., Novick et al., 2014; Brothers et al., 2021), and child-adult age differences (Trueswell et al., 1999), may also play a role in L2 sentence processing. Additionally, there are several factors being found particularly relevant to L2 sentence processing. These include the age of acquisition (AoA) (e.g., Weber-Fox & Neville, 1996, Hernandéz & Li, 2007; Sabourin et al., 2014), L2 proficiency (e.g., Ojima et al., 2005; Rossi et al., 2006; Morgan-Short et al., 2007; Pakulak & Neville, 2010; Dallas et al., 2013; Brothers et al., 2021), age of immersion (e.g., Weber-Fox & Neville, 1999; Tanner et al., 2014), years of exposure (Pliatsikas & Marinis, 2013; Berghoff, 2022), types of exposure (implicit vs. explicit) (e.g., Morgan-Short et al., 2012), and L1-L2 similarity (e.g., Sabourin & Stowe, 2008; Dowens et al., 2010; Dowens et al., 2011; van Bergen & Flecken, 2017; Andersson et al., 2019). Scholars arguing against qualitative differences between L1 and L2 processing suggest that native-like processing can be achieved if relevant individual differences are adequately addressed. In contrast, proponents of qualitative differences argue that even when individual differences are well addressed, processing differences remain detectable. In this section, I aim to review how differences in L2 proficiency, years of immersion, and L1-L2 similarity influence patterns of L2 processing.

L2 proficiency

Driven by the Critical Period Hypothesis (Lenneberg, 1967; Penfields & Roberts, 1959), early studies primarily focused on how the age of acquisition (AoA) influences brain plasticity after a critical period for language learning and how it impacts the sentence processing mechanism of non-native speakers (e.g., Weber-Fox & Neville, 1996). Since Weber-Fox and Neville's (1996) influential work, a rich body of research has demonstrated that AoA is a reliable predictor for ultimate attainment in L2 (e.g., DeKeyser, 2012; Abrahamsson, 2012) and the reliance on similar neurocognitive mechanisms as native speakers (e.g., Hahne & Friederici, 2001). However, as AoA is typically associated with longer exposure and higher proficiency, the trajectory leading to the ultimate attainment in L2 or similar neurocognitive mechanisms can vary considerably (Steinhauer, 2014). In other words, it remains unclear which factor specifically contributes to native-like processing. To address this issue, subsequent studies have sought to disentangle these factors by focusing on the specific impact of L2 proficiency on syntactic processing in individuals with consistent AoA (e.g., Friederici, Steinhauer & Pfeifer, 2002; Rossi et al., 2006; Ojima et al., 2005; Brothers et al., 2021).

Ojima et al. (2005) investigated how L2 proficiency affects brain responses during language processing. Using ERPs, the researchers compared native English speakers with Japanese learners of English at high and intermediate proficiency levels. Participants read English sentences containing either semantic violations or subject-verb agreement errors while their brain activities were recorded. A semantic pair can be “This house has ten rooms/cities in total” while a syntactic pair can be “Turtles move/moves slowly” (Ojima et al., 2005, p. 1213). The results revealed that all groups showed N400 responses to semantic violations, but L2 learners exhibited delayed peaks compared to native speakers, with high-proficiency learners showing earlier peaks than low-proficiency learners. For syntactic processing, both native English speakers and high-proficiency learners demonstrated a left-lateralized negativity (LAN) in response to agreement errors, while learners with low proficiency did not. These findings suggested that L2 processing can approximate L1 processing as proficiency increases, particularly in syntactic processing. Note that only native speakers showed an additional P600 to subject-verb agreement errors, whereas nonnative speakers did not. The absence of the P600 in nonnative speakers was attributed to the lack of agreement features like person and number in their L1.

While most previous studies treat L2 proficiency as a categorical factor, meaning that L2 participants are divided into different groups depending on their L2 proficiency, several recent studies have taken this variable as a continuous factor in their experimental design (e.g., Fromont et al., 2020; Brothers et al., 2021; Dallas et al., 2013). For example, Dallas et al. (2013) conducted a study where participants were grouped by proficiency. They investigated whether Chinese learners of English would exhibit an N400 effect in respond to implausible filler-verb integrations when processing filler-gap dependencies. The results showed that native English speakers showed a significant N400 effect at the verb, indicating difficulty in integrating word meaning into a sentence. A set of experimental sentences is shown in (14). Native speakers showed an N400 effect at *threatened* in sentence (14b), as *football* is an implausible object for the verb. However, Chinese learners as whole did not exhibit this effect. Then, the researchers plotted the amplitude difference between implausible and plausible conditions at the CZ electrode against the continuous proficiency z-scores and found that higher English proficiency was correlated with a stronger N400 response. Then, they conducted an additional analysis by splitting the L2 participants into higher and lower proficiency groups based on the median score. The group data revealed no significant difference between L1 speakers and L2 speakers with higher proficiency, whereas the low proficiency group showed significantly more negative mean amplitude than the other two groups. Therefore, they concluded that increased proficiency leads to more native-like processing.

(14) 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, p. 776)

A recent eye-tracking study by Brothers et al. (2021) investigated how Mandarin Chinese-English bilingual speakers and English monolinguals process temporary syntactic ambiguities in real time. The study used sentences containing temporary object-subject ambiguities to examine readers' sensitivity to verb subcategorization information. The experimental sentences were shown in (15). The findings revealed that native English speakers showed sensitivity to verb transitivity, with longer reading times and more regressions for intransitive verbs compared to transitive verbs at the ambiguous noun phrase *the retired admiral*. Bilinguals as a group, however, did not show sensitivity to verb transitivity cues during online processing. Importantly, L2 proficiency, driven by L2 experience, appeared to be a key predictor of sensitivity to verb subcategorization. More precisely, the authors suggested that with more experience of the target language, speakers' proficiency would increase. Once the proficiency level was accounted for, no parsing difference was found between native and non-native speakers. These findings aligned with the findings of Dallas et al. (2013) and suggested that increased L2 proficiency contributes to more native-like processing patterns.

(15) a. Transitive: After visiting the retired admiral received a call on his cellphone.

b. Intransitive: After arriving the retired admiral received a call on his cellphone.

c. Control: After dinner the retired admiral received a call on his cellphone.

(Brothers et al., 2021, p. 615)

Fromont et al. (2020) used ERPs to compare the brain responses from late L2 French learners to those of native speakers as they read sentences containing various linguistic errors. The study controlled the effect from AoA. Native speakers exhibited the typical N400-P600 pattern, indicative of processing difficulties caused by ungrammatical structures. In contrast, L2 learners exhibited only an N400 response, indicative of semantic processing difficulties. Importantly, they found that L2 proficiency and exposure, rather than their AoA, were the primary predictors of L2 learners' native-like brain responses.

Based on a large body of ERP studies that consider proficiency as an independent variable, Steinhauer et al. (2009) proposed a proficiency-based hierarchy development model. This model suggests that low-proficiency learners may initially show N400 effects. As learners reach higher levels of proficiency, their brain responses may shift toward a more native-like pattern, exhibiting large P600 or biphasic LAN-P600 responses, which are typically only seen with native speakers. This model was further extended to the Convergence Hypothesis (Steinhauer et al., 2014), which suggests that as individuals become more skilled in their L2, their brain responses change in predictable ways. When late L2 learners reach very high levels of proficiency, their brain activity patterns, measured by ERPs, would become nearly identical to those of native speakers. This suggests that with enough practice and exposure, even those who start learning a language later in life can process L2 in a way similar to L1. Based on this model, Tanner et al. (2014) suggested that in early-stage L2 learners, there is significant variability in brain response types and magnitudes. However, as proficiency increases, this variability tends to decrease, and highly proficient bilinguals start to show more consistent, native-like neural responses.

However, even at an advanced L2 level, some differences in processing persist (e.g., Díaz et al., 2016; Dong et al., 2014, 2022; Wolpert et al., 2024). Two main factors have been proposed to contribute to these persistent L1-L2 processing differences: structure-specific proficiency and L1-L2 similarity. On the one hand, Wolpert et al. (2024) found that while native speakers of English and native speakers of Mandarin both showed a biphasic N400-P600 response to adjective placement violations, L2 Mandarin learners only exhibited an N400 effect without the subsequent P600 effect for ungrammatical structures in Mandarin. This finding supported the idea that structure-specific L2 proficiency, rather than overall L2 proficiency, predicts the native-like ERP patterns. In other words, ERP patterns are not determined by overall L2 proficiency but by the specific proficiency level for the L2 structure under investigation (Steinhauer et al., 2009; Steinhauer, 2014). On the other hand, Díaz et al.(2016) examined how Spanish learners of Basque process grammatical features that are either shared between their L1 and L2 or unique to their L2. They manipulated both the group difference (native vs. early bilinguals with high proficiency vs. late learners with intermediate proficiency) and L1-L2 similarity (shared vs. L2-specific grammatical features). Notably, for the linguistic feature present only in Basque but not in Spanish (ergative case), native-like processing was absent even with early age of acquisition and high proficiency. This result suggested that L1-L2 similarity is an important factor for achieving native-like ERP components. The effect of L1-L2 similarity on L2 processing is discussed in detail in the following section.

L1-L1 similarity

The Competition Model posits that language comprehension is driven by the competition between various linguistic cues, such as word order, case marking, and agreement (Bates & MacWhinney, 1989). These cues compete to influence real-time language processing. Initially, the model was used to explain how monolingual children acquire their first language by learning to identify the most reliable cues from strings of linguistic inputs. In the field of second language acquisition, this model suggests that learners initially transfer the cue weights from their L1 to their L2, which can cause both facilitation and interference effects in L2 processing. MacWhinney (2005) later refined this concept and introduced the Unified Competition Model based on extensive

ERP research on crosslinguistic transfer in L2 grammatical processing. This refined model introduces three types of transfer: positive transfer (when L1 and L2 features are similar), negative transfer (when L1 and L2 features conflict), and no transfer (for L2-specific structures). The Unified Competition Model emphasizes that when grammatical features are instantiated in both languages, learners can effectively apply their L1 knowledge to L2 processing. This model also accounts for the online predictive processing in L2, shedding light on how learners employ reliable cues to predict upcoming information in their L2.

Numerous studies have explored the impact of L1-L2 similarity on L2 syntactic processing across various grammatical features (e.g., tense-marking violation: Tokowicz & MacWhinney, 2005; grammatical gender: Sabourin & Stowe, 2008; Foucart & Frenck-Mestre, 2011; verb agreement: Zawiszewski et al., 2015, Díaz et al., 2016; word order constraint: Andersson et al., 2019; Mickan & Lemhöfer, 2020; adjective placement: Wolpert et al., 2024). A recent study by Mickan and Lemhöfer (2020) investigated how L1-L2 similarity influences the development of L2 syntactic processing. Participants were German-speaking learners of Dutch at three proficiency levels: beginners, intermediate, and advanced. A group of native Dutch speakers served as the control group. The researchers conducted both a grammaticality judgement task and an ERP experiment to examine how German L2 learners process Dutch sentences with syntactic violations in conditions that were either similar or different between Dutch and German. In the conflict condition, sentences contained double infinitive constructions where German and Dutch require opposite verb orders (e.g., "want-buy" in Dutch vs. "buy-want" in German). The researchers wondered if German L2 learners would wrongly accept “buy-want” in Dutch because of L1 interference. In the non-conflict condition, the sentences contained coordinating or subordinating conjunctions that follow the same word order rules in both Dutch and German. For example, subordinating conjunctions like *omdat* in Dutch and *weil* in German (“because”) require the verb-final order, while coordinating conjunctions like *want* in Dutch and *denn* in German (“thus”) require the standard SVO word order. This allowed German speakers to apply their L1 rules to processing Dutch sentences with these conjunctions. The behavioral results revealed that advanced learners showed native-like accuracy in both conflict and non-conflict conditions, while the other two proficiency groups with lower proficiency showed lower accuracy for both conditions. Regarding the ERP results, only advanced and intermediate L2 learners exhibited native-like P600 to word-order violations for the non-conflict structure. In the conflict conditions, advanced and intermediate L2 learners showed delayed, weakened P600 compared to that of native speakers, whereas beginners showed a N400 effect. The findings provide evidence for persistent L1 influence on L2 processing, even for high proficiency learners whose behavioral performance appear to be native-like.

Despite the growing body of research on the factor of L1-L2 similarity, the findings on whether a learner's L1 can be transferred to facilitate L2 processing remain inconclusive, particularly in the study of L2 processing of *wh*-dependencies. Of particular importance to this study, several studies may have based their discussion on the view that Mandarin Chinese lacks overt *wh*-movement, which is flawed and may have weakened their interpretations of Mandarin Chinese learners' performance in L2 English processing (e.g., Juffs, 2005; Covey et al., 2024). Juffs (2005, p. 129), for example, reported that Mandarin Chinese native speakers were sensitive to both subject extractions (e.g., "*Who* does the nurse know ___ saw the patient at the hospital?") and object extractions (e.g., "*Who* does the nurse know the doctor saw ___ in his office?"). Their reading times were found to be similar to native English speakers and Spanish learners of English. Based on these findings, Juffs concluded that there was no L1 transfer effect, as both Spanish learners, whose L1 has *wh*-movement, and Chinese learners, whose L1 was thought not to have overt *wh*-movement, demonstrated similar processing patterns. Similarly, Covey et al. (2024) attributed non-native-like ERP components in Mandarin Chinese learners of English to the presumed absence of overt *wh*-movement in Mandarin Chinese. However, recent research by Dong et al. (2021, see detailed review in Chapter 3) has provided evidence that Mandarin Chinese topic structures are, in fact, movement-derived and allow overt *wh*-movement. Based on this research, we might need to reconsider earlier assumptions about the lack of overt *wh*-movement in Mandarin Chinese.

Naturalist Exposure and Immersion

Kaan (2014) argued that frequency information is a crucial source of predictive processing in native speakers. The frequency information refers to the stored knowledge about how likely certain linguistic elements are to occur in incoming language input. This information is accumulated through years of language exposure. For example, native English speakers know that certain verbs, like “claim,” are more frequently followed by a clause (e.g., “She claimed that...”) than by a noun phrase (e.g., “She claimed the prize”). The subcategorization preferences for certain verbs help native speakers predict what might come next in a sentence. L2 speakers, due to their lack of exposure and immersion in authentic language input of the target language, may differ from native speakers in their frequency information, thus leading to differences in predictive processing. In line with this idea, previous research has suggested that when L2 speakers are consistently exposed to substantial input of the target language, the length of immersion would positively influence L2 speakers’ sensitivity to various grammatical features (e.g., relative clause attachment: Frenck-Mestre, 2002; Dussias, 2003; long-distance dependency: Pliatsikas & Marinis, 2013; Berghoff, 2022; Bowden et al., 2013; gender-agreement: Dowens et al., 2010).

Pliatsikas and Marinis (2013) investigated how naturalistic exposure to L2 affects the processing of long-distance dependencies in L2 English learners using a self-paced reading task. The study included three participant groups: Greek-English speakers with naturalistic exposure to English, Greek-English speakers with only classroom exposure, and native English speakers as the baseline. Participants read sentences with *wh*-dependencies containing intermediate gaps, gaps posited at intermediate clause boundaries in the long-distance dependencies. According to Gibson and Warren (2004), the presence of an intermediate gap helps reactivate the displaced filler earlier in the sentence, thus facilitating its integration at the final gap site. The results showed that only the L2 speakers with naturalistic exposure showed evidence of processing intermediate gaps in a way similar to native speakers. This study provides evidence that naturalistic exposure can facilitate automatic syntactic processing in non-local dependencies, a structure that is argued to be difficult for L2 speakers to process by the SSH (Clahsen & Felser, 2006, 2018).

Using an ERP study, Bowden et al. (2013) examined how native Spanish speakers and two groups of L2 Spanish learners processed Spanish sentences with either semantic or word order errors. The L2 groups consisted of university foreign language learners. The L2 Low group had students with only one year of college Spanish and no immersion experience, with an Intermediate Low proficiency level. The L2 Advanced group had students with 3.5 years of college Spanish plus 1-2 semesters of study abroad immersion, reaching an Advanced Mid proficiency level.

For the semantic violations, the researchers created semantic anomalies by replacing appropriate sentence-final nouns with inappropriate ones. For example, a correct sentence *Tengo muchas millas que correr esta semana* ("I have many miles to run this week") was changed to a semantically anomalous sentence *Tengo muchas millas que correr esta actriz* ("I have many miles to run this actress"). Syntactic violations, specifically word-order violations, were created by reversing the positions of a noun and a verb within a correct sentence. For example, *Tengo que correr muchas millas esta semana* ("I have to run many miles this week") was changed to *Tengo que millas muchas correr esta semana* ("I have to miles many run this week") (Bowden et al., 2013, p. 2498). For semantic violations, all groups showed N400 effects, indicative of similar neurocognitive mechanisms for lexical/semantic processing regardless of L2 proficiency or exposure. For word order violations, the L1 and L2 Advanced groups showed similar LAN and P600 effects, suggesting native-like syntactic processing in this advanced group. The L2 Low group, however, showed only a broad positivity. These findings thus demonstrated that native-like brain processing of syntax can be achieved in university foreign language learners after several years of classroom instruction combined with short-term study abroad immersion. While both classroom instruction and immersion experience appeared to help develop native-like syntactic processing in L2 learners, the study did not fully separate the factors among immersion, types of L2 training as explicit or implicit (see Morgan-Short et al., 2012), and overall proficiency level.

2.4 Summary

In this chapter, I provided a comprehensive overview of the nature L1 and L2 sentence processing. Sentence processing, regardless of in L1 or L2, involves quickly integrating different types of linguistic information in real time. Two prominent features of human sentence processing are incremental processing, where parsers construct partial interpretations as each word unfolds, rather than waiting until the end of a sentence, and predictive processing, where speakers can anticipate upcoming linguistic elements before they appear. L2 speakers appear to show reduced predictive abilities compared to native speakers. Various theoretical models of L1 and L2 sentence processing are discussed in this chapter, including the modular accounts and interactive accounts for L1 processing, and the Shallow Structure Hypothesis (SSH), Declarative/Procedural Model, and Prediction-by-Production Model for L2 processing. A large portion of the chapter was dedicated to the ongoing debates in L2 processing research. One central question is whether L1 and L2 processing are qualitatively similar but quantitatively different, or if there are fundamental qualitative differences. While evidence exists for both views, recent research has moved away from seeking for a binary split. Instead, recent studies tend to adopt a softer view, suggesting that L2 processing may involve a gradual and dynamic process, rather than being categorically different from L1 processing. Among the many factors that modulate L2 processing, this chapter focused on L2 proficiency, years of immersion, and L1-L2 similarity. These variables are central to the experiments in this dissertation and have been shown to play important roles in facilitating native-like L2 processing.

Chapter 3

Filler-Gap Processing in L1 and L2

3.1 Introduction

Syntactic processing is typically straightforward when the order of the linguistic inputs aligns with the based-generated positions of theses elements, such as subject-verb-object (SVO) in English and Mandarin Chinese, or subject-object-verb (SOV) in Japanese and German. In these scenarios, each new word can be effortlessly integrated into the syntactic structure of the sentence. However, a salient property of the syntax in natural languages is the existence of displacement elements which allow a word or phrase to occur in a non-canonical syntactic position. This feature can be illustrated by the relative clause construction (16b), where the *zebra* has been displaced from the object position of *kissed*. Both (16a) and (16b) are about the action of kissing the zebra. In (16a), the relationship between *kissed* and the *zebra* is straightforward to analyze as the direct object *zebra* is located in its canonical position, right after the embedded verb *kissed*. However, in (16b), this relationship is less straightforward during the real-time sentence processing: the *zebra* needs to be interpreted as the direct object of the verb *kissed*, even though it is displaced from the typical post-verbal position. Within the field of sentence processing, the displaced element is usually referred to as a *filler*, and its canonical structural position as a corresponding *gap* (as indicated by the underscore). The relationship between a filler, and its gap has been discussed very early in the sentence processing literature (e.g., Clifton & Frazier, 1989). Such dependencies exist in structures such as relative clauses, topicalization and *wh*-interrogatives across various languages, including for example English, German, Mandarin Chinese, Korean (e.g., Stowe, 1986; Wager & Phillips, 2014; Fiebach at al., 2002; Yang & Liu, 2014; Dong et al., 2021; Kwon, 2008).

(16) a. The hippo kissed the zebra on the nose and then ran far away.

b. The zebra that the hippo kissed ___ on the nose ran far away.

(Hestvik et al., 2007, p. 304)

Proper interpretation of filler-gap constructions requires parsers to recognize the existence of a displaced element, actively look for the gap position, and integrate the filler into a syntactically and semantically appropriate location. This complex process is known as *filler-gap processing*. As discussed in Chapter 2, L1 speakers begin actively searching for potential gaps as soon as they encounter a displaced filler rather than waiting for definitive gap information (e.g., Clifton & Frazier, 1989). The study of filler-gap processing is particularly illuminating for understanding sentence processing similarities and differences between L1 and L2 speakers for two main reasons. First, studying how L2 speakers process filler-gap dependencies offer valuable insights into the linguistic resources used during real-time processing by L1 and L2 speakers, as well as the timing of each step involved in filler-gap resolution. The Shallow Structure Hypothesis (SSH) posits that L2 speakers prioritize semantic and lexical information while underutilizing syntactic cues, especially for complex structures like filler-gap dependencies, and that L2 speakers experience delays in processing (Clahsen & Felser, 2006, 2018). The Trace Reactivation Hypothesis assumes that during the online processing of filler-gap dependencies, parsers postulate syntactic gaps or empty categories (e.g., Bever & McElree, 1988; Nicol & Swinney, 1989; Love & Swinney, 1996). Examining L2 speakers' filler-gap processing would tell us whether such empty categories can be posited by L2 speakers and whether there are delayed or weakened effects. The findings can provide evidence for or against the SSH. Second, non-local dependencies are common but varying in natural languages, allowing for cross-linguistic comparisons. To illustrate, languages like English and Spanish often require syntactic movement. In contrast, *wh*-in-situ languages such as Korean, Mandarin Chinese, and Japanese have been argued to have *wh*-phrases in their base-generated positions without movement (e.g., Li, 1976; Li & Thompson, 1981). This cross-linguistic difference makes these complex structures ideal stimuli for investigating how L1-L2 difference affects the L2 syntactic processing (e.g., White & Juffs, 1998).

This chapter explores the filler-gap processing in both native and second language contexts, with a focus on English and Mandarin Chinese. It begins by introducing the concept of filler-gap dependency and its significance in language processing. The chapter then provides a linguistic background on *wh*-dependencies in English and Mandarin Chinese, including a discussion on whether Mandarin *wh*-dependencies are movement-derived or base-generated based on previous theoretical research. Following this, it delves into the mechanisms of processing filler-gap dependencies, discussing two main hypotheses: the Active Filler Strategy and the Direct Association Hypothesis. A large portion of the chapter is dedicated to examining how English filler-gap dependencies are processed in both L1 and L2, reviewing classic and recent literature using both psycholinguistic and neurolinguistic approaches. The psycholinguistic section covers various effects used in the studies of filler-gap processing, including the *filled-gap effect*, *plausibility mismatch effect*, and debates between *maintenance and retrieval views*. The neurolinguistic section discusses relevant ERP components and recent ERP studies in the field.

3.2 Filler-gap structures in English and Mandarin Chinese: theoretical and empirical evidence

3.2.1 Filler-gap dependencies in English

Filler-gap dependencies can be seen in various types of syntactic structures, including *wh*-questions, embedded *wh*-dependencies, relative clauses, topicalization, cleft sentences, etc., as shown in (17).

(17) a. *Wh*-question: Which book$_i$ did Mary buy _t_i_ at the store?
b. Embedded *wh*-question: I wonder which book$_i$ Mary bought _ t_i _ at the store.

c. Relative clause: The book$_i$ which Mary bought _ t_i _ at the store was interesting.

d. Topicalization: That movie$_i$, I really enjoyed _ t_i _.

e. Cleft sentence: It was the red car$_i$ that John bought _ t_i _.

In Government and Binding Theory (Chomsky, 1981), when a syntactic category is moved, it leaves behind an empty category called a *trace* in the position from which the element was moved. The trace, though phonetically empty, is syntactically active and coindexed with the moved element. By the early 1990s, the copy theory of movement (Chomsky, 1995) replaced the earlier trace theory, proposing that the empty category is actually a *copy*, an unpronounced instance of the moved element itself. For example, in (18a), *which book* has been moved from its base position (after *bought*) to the front of the sentence. The subscript *i* is used to indicate the two copies of the same element.

(18) a. I wonder [which book$_i$ [Mary bought _ e_i _ at the store.]]

b. *I wonder [which book$_i$ [the man [who bought _ e_i _at the store] recommended]].

c. I wonder [which book$_i$ [the man who went to the store recommended _ e_i _]].

Syntactic movement has limitations; it is argued to be governed by *island constraints.* Island constraints define specific syntactic structures, known as *islands*, from which extraction of elements is prohibited (Ross, 1967). Figuratively speaking, elements originating within an island must remain within that domain; they cannot cross the island's boundaries through syntactic operations. Common islands include relative clauses, adjuncts, and complex noun phrases etc. For example, modifying the previous sentence in (18a) to include a relative clause renders it ungrammatical. As in (18b), the relative clause "who bought e_i at the store" constitutes an island. If the *wh*-phrase *which book* is trying to move out of this relative clause to the front of the embedded question, it would violate the island constraint. In English, movement typically cannot cross more than one inflectional/tense phrase (IP/TP) or noun/determiner phrase (NP/DP) boundary at once. Considering the grammatical example "Who did Susan say that Mary met ___?" the movement of *who* is allowed because it moves step-by-step through available specifier positions of a complementizer phrase (CP), never crossing multiple bounding nodes at once. In contrast, sentence (18b) is ungrammatical because there is no intervening CP specifier position to break up this movement into successive steps, and moving *which book* directly from its base position to the front of the embedded question would cross more than one bounding node. Each bounding node is represented with one pair of square brackets. A grammatical alternative appears in (18c), where *which book* is the object of *recommended*, not *bought*, and thus is not trapped within the relative clause island.

Previous research has robustly shown that native speakers respect island constraints during real-time sentence processing (Stowe, 1986; Frazier & Clifton, 1989; Traxler & Pickering, 1996). Using eye-tracking, Traxler and Pickering (1996) examined how native English speakers process filler-verb mismatches in island and non-island conditions, as shown in (19) and (20). For sentences with gaps in islands (20), the results showed no plausibility effect in first fixation times, first-pass reading times, or total reading times at any regions.[13] They also found that parsers reacted to plausibility effects differently in island versus non-island conditions. In non-island conditions (19a) and (19b), participants' first fixations in the verb region (*wrote*) were longer for implausible compared to plausible filler-gap pairs, indicating a plausibility mismatch effect. However, this effect was absent when the verb *wrote* was embedded within a relative clause island as in (20a) and (20b). This suggests that participants adhere to island constraints and do not attempt to fill the gap at any point during the processing of island sentences.

[13] First fixation time refers to the time taken for a participant's gaze to first land on a word region. First-passing reading time refers to the total time spent fixating within a word region before the gaze leaves. Total reading time refers to the total amount of all fixations.

(19) Non-island conditions:

a. Plausible: We like the book that the author **wrote** unceasingly and with great dedication about ___ while waiting for a contract.

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

(20) Island conditions:

c. Plausible: We like the book that the author who **wrote** (*____) unceasingly and with great dedication saw ______ while waiting for a contract.

d. Implausible: We like the city that the author who **wrote** (*___) unceasingly and with great dedication saw _____ while waiting for a contract.

(Traxler & Pickering, 1996, p. 465)

3.2.2 Topic structures in Mandarin Chinese

Unlike the filler-gap dependencies in languages like English, Mandarin Chinese is typically recognized as a *wh*-in-situ language in which *wh*-word remain in their canonical positions in *wh*-questions. Specifically, in (21a), the wh-word *shenme* stays in the object position in Mandarin Chinese, whereas the corresponding *wh*-word *what* would move to the initial position of an interrogative sentence in English. As the canonical word order in Mandarin Chinese is SVO, when a *wh*-word functions as a subject, it likewise remains in its canonical subject position, as shown in (21b).

(21) a. *Suzy* *mai-le* *shenme?*

Suzy buy-Perf[14] what?

'What has Suzy bought?"

b. *Shei* *wenhou-le* *Suzy?*

Who greet-Perf Suzy?

'Who has greeted Suzy?'

Despite being a *wh*-in-situ language, Mandarin Chinese is widely argued to feature *topical prominence*, with a *topic* appearing at the initial position of a sentence, followed by a *comment* that provides additional information about that topic (e.g., Gundel, 1988; Li & Thompson, 1976; Xu, 2006). Consider the examples from Li and Thompson (1976), shown in (22). In (22a), the topic *neichang huo* ("that fire") is associated with the fire brigade's arrival. In (22b), the topic *neike shu* ("that tree") is where the leaves grow. The relation between the topic and its comment is characterized as *aboutness*, meaning that a well-formed topic-comment sentence requires the comment to provide relevant information about the topic (Kuno, 1976; Gundel, 1988; Reinhart, 1981). Examples from Li and Thompson (1976) demonstrate how Mandarin Chinese topic-comment structures satisfy this aboutness requirement, even when the initial topic does not correspond to a syntactic gap within the following clause.

(22) a. *Neichang* *huo* *xingkui* *xiaofangdui* *lai* *de* *kuai.*

that-CL[15] fire fortunately fire-brigade come DE quick

'That fire, fortunately, the fire brigade came quickly.'

(Li & Thompson, 1976, p. 470)

---

[14] Perf is an aspect marker to indicate the completion, equivalent to the prefect aspect in English.

[15] CL represents a classifier, a grammatical marker between demonstratives and nouns.

b. *Neike* *shu* *yezi* *da.*

that-CL tree leaves big.

'That tree, leaves are big.'

(Li & Thompson, 1976, p. 482)

As illustrated in (22), in Mandarin Chinese, topic structures are not moved elements and thus do not need to be co-indexed with an empty category. In other words, some topic structures do not require syntactic gaps for legitimacy; rather, they can be licenced by semantic relationship established in the sentence. Based on the previous discussion, it seems plausible to view such topics as dangling elements, and they should be put outside of the CP as adjuncts in syntax. Mandarin Chinese exhibits various forms of topic structures, but this dissertation focuses on a specific category known as the *gap-type* topic structure (Huang et al. 2009; Dong et al. 2021). This particular type is characterized by the presence of a syntactic gap within the sentence. Gap-type topicalization can be seen in two distinct word orders: OSV and SOV. However, my analysis will primarily concentrate on the former order, where the topicalized element is specifically the direct object of the verb, which has been moved from its original position to the beginning of the sentence. Mandarin Chinese allows topic structures where the topic is assigned a thematic role in the comment through either an overt or covert syntactic dependency. The topic in such structures come from either a *wh*-extraction or a DP-extraction case. A *wh*-extraction sentence may have a surface structure identical to English filler-gap dependency, as shown in (23a). The *wh*-word can also be in-situ, as in (23b). An example of the DP-extraction topic structures in shown in (23c), *shu* ("book") is moved to the left edge of the sentence without carrying an interrogative feature. The DP-extraction topic structure appears to be acceptable but unnatural in English.

(23) a. [Na-ben shu]$_{i,}$ ni du _t_i_ le.

which-CL book you read PERF

'Which book have you read?'

b. Ni du na-ben shu le?

you read which-CL book PERF

'Which book have you read?'

c. Shu$_i$, wo du _t_i_ le.

book I read Perf

'I have read the book.'

Successful topicalization in Chinese requires the topic structure to be established in the previous discourse (Wu, 1999). That is, to undergo successful *wh*-topicalization, the interlocutors in the discourse must have possessed a certain set of entities in the "presupposition background" (Wu, 1999, p. 81). This rule explains why (24b) is felicitous but (24a) is not. (24b) can be felicitously asked in a situation where both the speaker and listener are aware that a list of dishes has already been mentioned in the previous discourse, making it appropriate for the speaker to ask which dish Zhangsan ate among those dishes. However, in (24a), the topicalized *wh*-element *shenme* is not constrained by a particular set as *what* can refer to anything that is not a dish.

(24) a. *Shenme$_i$, Zhangsan chi-le __t_i__?

What Zhangsan eat-Perf

'What did Zhang eat?'

b. [Na-dao cai]$_{i,}$ Zhangsan chi-le _t_i_?

Which-CL dish Zhangsan eat-Perf

'Which dish (is the one that) Zhangsan ate?'

The discussion in Section 3.2.2 helps to explain why Experiment 2 of this dissertation uses *which-phrases* (e.g., "which book") rather than bare *wh*-words (e.g., "who") as stimuli, as these *wh*-phrases are discourse-constrained. Furthermore, the experiment employs gap-type *wh*-extraction topic structures rather than DP-extraction ones, since this sentence type is acceptable in both English and Mandarin Chinese, enabling a more direct cross-linguistic comparison.

3.2.3 Empirical evidence of non-local dependency in Mandarin Chinese

Some studies reporting differences in *wh*-dependency processing between L1 English speakers and L2 speakers whose L1 is Mandarin have attributed these differences to the absence of *wh*-movement, particularly overt *wh*-movement, in Mandarin (e.g., Dallas et al., 2013; Juff, 2005; Covey et al., 2024). However, other research has noted that Mandarin may still involve a non-local dependency (e.g., Huang, 1982). Specifically, in Mandarin Chinese, there exists a non-explicit or covert dependency between the in-situ position and a clause-initial syntactic element.

Recently, several empirical studies have found evidence for covert dependencies in the processing of in-situ *wh*-questions (e.g., Xiang et al., 2014, 2015; Yang et al., 2023; Yang et al., 2010; Kong et al., 2023). Using the speed-accuracy trade-off (SAT) paradigm, Xiang et al. (2014) examined how Mandarin speakers process *wh*-in-situ questions. The study compared *wh*-in-situ questions with declarative sentences of different lengths. Participants were instructed to judge the acceptability of sentences like "Which officials did the city council punish?" (short *wh*-question) and "Which officials did the mayor order the city council to punish?" (long *wh*-question), as well as corresponding declaratives (Xiang et al., 2014, pp. 218–219). By using the SAT technique, researchers could analyze processing speed and accuracy separately. The findings showed that *wh*-questions were processed more slowly than declaratives. Therefore, the authors concluded that processing *wh*-in-situ questions involves constructing a covert long-distance dependency, similar to overt *wh*-movement in languages like English.

In a follow-up study, Xiang et al. (2015) used three methods (sentence completion, self-paced reading, and eye-tracking) to compare *wh*-in-situ questions to declarative sentences across four complexity conditions: Multi-CP, Serial Verb, Adverbial, and Short. In their comprehension experiments, they found evidence of memory interference when processing *wh*-in-situ questions with intervening clause boundaries. This suggested that readers established a connection between the *wh*-phrase and its scope position, even when the *wh*-phrase remains in its original location. Additionally, their sentence completion studies revealed that speakers tended to avoid producing *wh*-in-situ structures, particularly in more complex sentence structures. This avoidance behavior indicated that planning and maintaining the covert long-distance dependency in *wh*-in-situ constructions places a significant burden on working memory during language production.

Yang et al. (2023) recently conducted a study to investigate the processing of *wh*-in-situ questions in Mandarin Chinese, focusing on both bare *wh*-phrases (e.g., "who") and complex *wh*-phrases (e.g., "which classmate"). They conducted two self-paced reading experiments comparing *wh*-questions to declarative sentences with indefinites (for bare *wh*-phrases) and definites (for complex *wh*-phrases). In Experiment 1, bare *wh*-questions were compared to declaratives with indefinites, while Experiment 2 compared complex *wh*-questions to declaratives with definites. Results from both experiments showed that *wh*-questions were processed more slowly than their declarative counterparts, regardless of whether the *wh*-phrase was bare or complex. This study builds on the previous work by Xiang et al. (2014) but addresses some limitations by controlling for (in)definiteness, avoiding sentence-final wrap-up effects, and examining both bare and complex *wh*-phrases. All three studies converge on the conclusion that processing *wh*-in-situ questions in Mandarin Chinese involves constructing a covert dependency, resulting in higher processing costs compared to declarative sentences.

The previous studies all examine the processing of covert *wh*-dependencies in Mandarin Chinese. To the best of my knowledge, there are only three studies that have investigated the processing of overt non-local dependency in Mandarin Chinese. However, these studies did not focus on *wh*-phrases that appear identical to their English counterparts on the surface (Yang & Tao, 2014; Dong et al., 2021; Huang & Kaiser, 2008).

Huang and Kaiser (2008) conducted a self-paced reading experiment to investigate parsers' sensitivity to island constraints in the processing of Mandarin Chinese topic structures. Following Phillips (2006), they manipulated the plausibility of the topic at the parasitic gap, which is either a transitive verb or a coverb.[16] Similar to English, Mandarin Chinese parasitic gaps are sensitive to the island constraints. Specifically, with an adverbial clause, parasitic gaps are acceptable at the object position of a transitive verb. However, they are not acceptable at the object position of a coverb. They prohibit their objects from moving to the initial position of a sentence. The researchers found that Mandarin Chinese readers actively searched for a potential gap after detecting a filler only in transitive conditions, and they were sensitive to the syntactic restrictions created by the coverbs on parasitic gap constructions in Mandarin Chinese. These results showed that Mandarin Chinese topic structures are processed similarly to English filler-gap dependencies, supporting a movement-based analysis of Mandarin Chinese gap-type topic structures.

Building upon Huang and Kaiser's (2008) work, Dong et al. (2021) used the ERP technique to examine the processing of Mandarin Chinese topic structures with or without a complex noun phrase island. They manipulated two independent factors: filler plausibility (whether the filler was a plausible or implausible object of the verb) and islandhood (island vs. non-island conditions). The researchers predicted that if the island constraint was respected, participants would not exhibit a plausibility effect in the island conditions. In the non-island conditions, although they did not observe the expected N400 effect typically associated with processing implausible objects, they found a posterior positivity between 612-796 ms after the first verb in the implausible conditions

[16] The coverb in Mandarin Chinese is a class of verbs having features of both a verb and a preposition (e.g., *tong* "be with").

(e.g., “greet the file”) compared to the plausible ones (e.g., “greeting the teacher”) (Dong et al., 2021, p. 7). This positivity was interpreted as a P600 effect related to an animacy violation, which still suggests that parsers were actively attempting to resolve the dependency during real-time sentence processing. Importantly, the P600 observed in the non-island conditions was absent in the island conditions, indicating that island constraints were indeed respected in Mandarin Chinese. Based on these findings, the researchers concluded that their results are consistent with a movement-based analysis that involves active gap-filling, and that the data support the view that topic structures in Mandarin involve syntactic movement.

In an earlier ERP study, Yang and Tao (2014, p. 91) examined how Mandarin Chinese speakers process grammatical topic structures like “The table, Manager Wang kicked twice" and ungrammatical ones like "The table, Manager Wang kicked Mr. Li," along with a non-topic control condition. Their results showed a Sustained Anterior Negativity (SAN) between the topic noun phrase and the subcategorizing verb, suggesting that the filler was being maintained in working memory. Additionally, at the sentence-final position, grammatical topicalized sentences displayed a posterior P600 compared to the control condition. The researchers interpreted this response as indicative of syntactic dependency establishment. Consequently, the authors claimed that their findings supported the movement account. However, the findings were then found to be unconvincing due to issues with the experimental materials, such as the stimuli being presented without context, and interpretations of the ERP components, such as the interpretation of a positive component occurring at 300-500ms as a P600 (Dong et al., 2021).

Previous studies on L2 filler-gap processing have often attributed the lack of native-like effects in Mandarin Chinese speakers to their first language being a *wh*-in-situ language that lacks overt *wh*-dependencies. However, this assumption overlooks an important feature of Mandarin Chinese: the existence of fronted *wh*-questions in certain contexts. This raises crucial questions regarding the processing of Mandarin Chinese gap-type topic structures: Are fronted *wh*-phrases in

Mandarin Chinese comparable to those in English? Do Mandarin Chinese speakers process fronted *wh*-phrases similarly to how English speakers process non-local dependencies? To investigate this, we need to examine whether a filler-gap dependency is established during the processing of fronted *wh*-phrases in Mandarin Chinese, as shown in (25):

(25) Xiaozhang xiangzhidao nage kehu lingdao tongzhile shixisheng qu zhao__.
Xiaozhang wonder Which-CL client leader inform-Perf intern to find
'Xiaozhang wondered which client the leader informed the intern to go find.'

If native Mandarin Chinese speakers process these overt *wh*-phrases similarly to how native English speakers process filler-gap dependencies in embedded *wh*-questions, it would challenge the assumption that absence of non-local dependency is responsible for non-nativelike processing mechanisms in L1 Mandarin Chinese-L2 English speakers. This finding would necessitate a re-evaluation of some previous research on L2 processing of filler-gap dependencies. Furthermore, Wagers and Phillips (2009) argue that when parsers use the active gap-filling strategy, this suggests that they prepare for the unbounded dependencies and actively connect elements across distances in sentence structures. If this strategy were to be observed in the processing of Mandarin Chinese fronted *wh*-phrases, it would provide indirect evidence for the existence of movement-derived non-local dependencies in Mandarin Chinese, similar to those in English.

3.3 Parsing mechanisms of active gap-filling

The incremental processing of sentences in real time, as discussed in the previous chapter, reveals that sentences are processed as they unfold (e.g., Altmann & Kamide, 1999; Frazier, 1987). Consequently, parsers cannot resolve filler-gap dependencies as soon as a filler is detected, since the gap typically appears later. This is particularly evident in languages like English, where the filler often occurs well before the gap. Thus, two key questions arise regarding the processing of *wh*-dependencies. First, how is the filler-gap dependency ultimately resolved—by the empty category or by the verb? Second, how does the brain handle the filler during the temporal delay between detecting it and finding its gap? Specifically, research has been interested in whether the filler is maintained in working memory or only retrieved when the gap is identified.

3.3.1 Active Filler Strategy

A well-established parsing mechanism for filler-gap processing is the *Active Filler Strategy,* which suggests that once encountering a displaced item, parsers actively search for a possible gap location rather than waiting for conclusive evidence of the gap's actual position (e.g., Clifton & Frazier, 1989). As such, processing errors may occur as parsers attempt to resolve the dependency at each grammatically acceptable position as the sentence unfolds. If later input proves their initial analysis wrong, such as finding another word in what was thought to be the gap position, parsers must revise their initial analysis. The Active Filler Strategy has received substantial empirical support across various languages (e.g., English: Stowe, 1986; Wagers & Phillips, 2014; Japanese: Aoshima et al., 2004; Dutch: Frazier, 1987; German: Felser et al., 2003; Mandarin Chinese: Dong et al., 2021) and through diverse experimental methods (Self-paced reading: Lee, 2004; Pickering & Traxler, 2001; Stowe, 1986; Traxler & Pickering, 1996; Wagers & Phillips, 2009, 2014; Eye-tracking: Chow & Zhou, 2019; Sussman & Sedivy, 2003; ERP: Felser et al., 2003; Fiebach et al., 2002; Garnsey et al., 1989; Hestvik et al., 2007; Kluender & Kutas, 1993; Phillips et al., 2005).

One type of evidence for active gap-filling comes from the parser's immediate sensitivity to the semantic plausibility of the filler as a direct object of a later verb. This is known as the *plausibility effect*. Traxler and Pickering (1996) demonstrated this effect by manipulating the semantic fit between the filler and a potential verb, as in (26). In both sentences, the actual gap occurs after *about*, not after *wrote*. However, the Active Filler Strategy predicts that parsers will attempt to integrate the filler at its first possible gap position, that is, after the verb *wrote*. In (26a), *book* is a plausible object for *wrote*, while *city* in (26b) is not. Using the eye-tracking technique, the researchers measured participants' reading times at the verb *wrote*. They found a significant difference in native English speakers' first fixations at the verb *wrote* between the two conditions. Participants took longer to fixate on the verb *wrote* in sentence (26b), where the filler was semantically implausible, compared to sentence (26a), where the filler was a semantically plausible fit. These results suggest that parsers actively search for gap positions and attempt to integrate the filler at the verb *wrote*, regardless of the semantic association.

(26) 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 & Pickering, 1996, p. 465)

The plausibility effect in active gap-filling has been supported by evidence from ERP studies, particularly through observations of N400, which is typically associated with semantic processing and integration. Larger N400 amplitudes generally indicate greater difficulty with semantic violations (e.g., Kutas & Federmeier, 2011). Garnsey et al. (1989) investigated how participants process sentences containing filler-gap dependencies by measuring their brain activity using ERP technology. The researchers manipulated the semantic plausibility of the filler as a direct object of the verb by creating two conditions: one where the filler was a plausible argument and another where it was implausible, as shown in (27). In sentence (27a), the filler *which customer* is a plausible direct object for the verb *called*, whereas in (27b), *which article* is an implausible object for the same verb. The researchers found a greater N400 amplitude at the verb when the filler was an implausible argument compared to when it was a plausible argument. This immediate brain response to plausibility at the verb suggested that the parser attempted to integrate the filler at the earliest possible gap position.

(27) a. The businessman knew which customer the secretary called __ at home.

b. The businessman knew which article the secretary called __ at home.

(Garnsey et al., 1989, p. 55)

The active gap-filling strategy in sentence processing is also evidenced by the *filled-gap effect*. This effect occurs when parsers predict a gap associated with a preceding filler but encounter another noun phrase already filling that potential gap position. The effect can manifest as a slowdown in reading time, indicating that parsers need to reanalyze their initial interpretation. Stowe's (1986) influential study found this effect using paired sentences such as in (28). In sentence (28b), which contains a *wh*-dependency, readers predicted a gap immediately after the verb *bring*. When they encountered the pronoun *us* instead, they had to reanalyze the previous filler-gap integration, resulting in increased reading time. This reading time slowdown was not observed in the control sentence (28a), where no gap was expected.

(28) a. My brother wanted to know if Ruth will bring us home to Mom at Christmas.

b. My brother wanted to know who Ruth will bring us home to ___ at Christmas.

(Stowe, 1986, p. 234)

It is important to note that Stowe's (1986) study also investigated potential *subject filled-gap effects*. That is, when readers encounter a *wh*-word like *who* at the beginning of an embedded clause, they might predict a gap in the subject position of this clause. If they then encounter a noun *Ruth* in the subject position, this indicates that the potential gap site is already filled. If readers had predicted a gap in this position, finding it occupied by *Ruth* should have led to slower reading times as they needed to revise their initial prediction. However, while Stowe found significant reading time slowdowns at the object position in (28b), the study showed no significant slowdown at the subject position. As Pickering and Barry (1991) pointed out, the interpretation of object filled-gap effects is not straightforward. These effects could be attributed to two distinct processes: either the prediction of a gap position or the direct association of the filler with its licensing verb. This uncertainty requires careful interpretation of studies like Traxler and Pickering's (1996), as we would not know if the reading time slowdown at *wrote* came from the integration of the filler into a post-verbal gap or the unsuccessful direct association with the verb. Also, if N400 is solely associated with semantic mismatch, Garnsey et al.'s (1989) N400 finding seems to suggest that filler-gap dependency is resolved by the verb rather than the gap. A question thus arises: What drives active gap-filling—the existence of an empty category or the thematic argument relationship between the filler and the verb?

Two main accounts have been proposed to explain active gap-filling in sentence processing: the *Trace-Reactivation Hypothesis* and the *Direct Association Hypothesis*. The Trace-Reactivation Hypothesis[17] assumes that parsers postulate empty categories and actively reactivate the antecedent, the moved constitute, at the gap position (e.g., Bever & McElree, 1988; Nicol & Swinney, 1989; Love & Swinney, 1996; Marinis et al., 2005). According to the Trace Reactivation Hypothesis, parsers actively search for an empty category at all syntactically licensed positions as the sentence unfolds. In (29), the filler *which book* is believed to be moved from its base position as the object of *read* to the surface as the specifier of the CP at the clausal boundary. As the sentence unfolds, the filler is assumed to be reactivated first after *say*, and then again after *read*.

[17] The concept of antecedent reactivation at the gap was first demonstrated in the late 1980s, but the explicit term *Trace Reactivation Hypothesis* does not have a widely accepted origin. This concept is referred to as "reactivation" in Nicol & Swinney (1989), "reactivation patterns" in Nicol (1989), and "antecedent reactivation" in Love & Swinney (1996). To the best of my knowledge, the earliest work to explicitly use the term *Trace Reactivation Hypothesis* is Marinis et al. (2005).

(29) [$_{CP}$Which book$_i$ did John say [$_{CP}$ t_i [$_{IP}$ Mary read t_i]]]?

On the contrary, the Direct Association Hypothesis proposes an alternative mechanism for processing sentences containing displaced elements. This hypothesis posits that during sentence processing, parsers establish immediate associations between subcategorizers (verb or preposition) and their arguments (Pickering, 1993; Pickering & Barry, 1991). According to this hypothesis, when a subcategorizer appears in the input, it directly licenses its arguments, including those displaced from their canonical positions. In other words, parsers do not need to assume empty categories or gap sites in the syntactic structure; instead, they simply search for an argument position and directly link the filler with the subcategorizer.

Evidence for the trace-based account comes from findings that filler integration can occur before the verb appears in head-final languages like Japanese (e.g., Nakano et al., 2002). In their study, Nakano et al. (2002) used a cross-modal priming paradigm to examine whether native Japanese speakers reactivate displaced objects at hypothesized preverbal trace positions in long-distance scrambling sentences. They found that participants with higher working memory capacity reactivated the displaced noun phrase before the verb came into play. The timing of such reactivation effects aligns with the trace-based account while challenging the view that filler integration depends on direct association with the subcategorizer.

Additional evidence comes from ERP studies, where different ERP components reflect different ungrammatical brain activities. Hestvik et al. (2007, 2012) compared brain responses between sentences containing filled gap positions and their grammatical counterparts. They presented participants with sentences such as those in (30). In (30a), the parser attempted to link *the zebra* to a gap position after the verb *kissed,* yet they found that position had been already filled by *the camel.* This filled-gap effect elicited an Early Left Anterior Negativity (ELAN), an ERP component typically associated with automatic syntactic parsing at early stages (e.g., Hahne & Friederici, 1999). The occurrence of an ELAN, rather than a N400, indicates the parser's sensitivity to structural information at the gap site, not just semantic cues. In other words, parsers actively predict and search for gap positions based on syntactic structure, rather than relying solely on the semantic relationship between the verb and the argument structure. These findings provide neurophysiological evidence for the psychological reality of an empty category in sentence processing.

(30) a. *The zebra that the hippo kissed the camel on the nose ran far away.

b. The weekend that the hippo kissed the camel on the nose, it was hot.

(Hestvik et al., 2007, p. 304)

In section 3.2.1, I discussed previous empirical evidence showing that parsers are sensitive to island constraints. To recap, island constraints are rules that restrict the extraction of elements from certain structural domains, thus preventing certain types of long-distance movement. The fact that parsers follow these island constraints shows that they are aware of the hierarchical structure of sentence representation. Importantly, parsers do not show any evidence of filler-gap integration within islands (e.g., Traxler & Pickering, 1996; Huang & Kaiser, 2008). This indicates that they are selective and follow structural rules, rather than simply linking fillers to any possible subcategorizers. This selective behavior directly contradicts the predictions of the Direct Association Hypothesis, which would expect fillers to be simply associated with their subcategorizers.

To conclude, previous research appears to align with the psychological existence of an empty category. Thus, the dissertation embraces a parsing model rooted in generative grammar principles, positing that filler-gap dependencies are constructed by movement operations with an empty category being left at the base position.

3.3.2 Maintenance and retrieval view

In the interval between identifying the filler and interpreting it at its corresponding gap, there are three major accounts explaining how the filler is processed in our brain (see Ness & Meltzer-Asscher, 2017 for a review). According to the maintenance view, the filler is stored in short-term memory, allowing for immediate integration when a gap is encountered (e.g., Wanner & Maratsos, 1978; Sprouse et al., 2012). This approach ensures parsers have constant access to the filler. However, continuously maintaining the filler may burden working memory resources. According to the retrieval view, the filler is not actively maintained in a privileged state in working memory. Instead, when parsers identify a potential gap, they must retrieve and reactivate the filler (e.g., McElree, 2001; Lewis & Vasishth, 2005). While this approach eliminates the need for ongoing maintenance and reduces working memory load, it may increase the risk of retrieval difficulties when encountering the gap.

Importantly, the two views—maintenance and retrieval—are not mutually exclusive. Wagers and Phillips (2014, p. 1274) propose a *hybrid maintenance-retrieval model*, in which "coarse-grained" syntactic information about the filler is maintained, while "finer-grained" lexical-semantic information is retrieved at the gap site. They tested this model through three separate self-paced reading experiments. Each experiment tested a specific type of information across varying dependency lengths. In Experiment 1, to test the maintenance of syntactic category information, they compared reading times at filled-gap positions between sentences with NP fillers versus PP fillers. Since only NPs can occupy direct object positions, encountering an overt direct object following an NP filler (a filled-gap scenario) should result in processing difficulty if syntactic information is being maintained. Indeed, the results showed that the syntactic categorical information of the filler (NP filler or PP filler) had an immediate effect on reading times at the gap position across all three dependency lengths. This suggested that syntactic category information was actively maintained in the parsers' working memory regardless of dependency length. In Experiment 2, to test the maintenance of semantic information, they compared plausible versus implausible filler-verb pairs (e.g., "monument...erect" vs. "gods...erect"), where processing difficulty with implausible combinations would suggest maintained semantic information. A plausibility mismatch effect emerged only in short dependencies, while in long dependencies the effect was delayed. This delay suggested that semantic information was not maintained over longer distances but was instead retrieved when needed. In Experiment 3, to test whether preposition information was maintained, they compared matching and mismatching preposition-verb pairs (e.g., "from which…inherit" vs. "to which…inherit"), where processing difficulty with mismatches would suggest that specific preposition information was maintained. Similar to the findings from Experiment 2, a preposition mismatch effect was observed immediately only in short dependencies, but was delayed in long ones, indicating that this finer-grained prepositional information was likely retrieved rather than maintained.

Further support for this hybrid view comes from Felser et al.'s (2003) study of filler-gap dependencies in German using the ERP technique. Their study examined different types of syntactic dependencies, including *wh*-movement, topicalization, and raising constructions. The results revealed a Left Anterior Negativity (LAN) effect at the subject noun phrase for all conditions. The LAN has been associated with working memory load (e.g., Kluender & Kutas, 1993; Felser et al., 2003; Fiebach et al., 2002); thus, the results suggested that working memory processes are engaged in maintaining the filler. At the subcategorizing verb, a P600 effect was observed for *wh*-movement compared to other conditions, indicating higher integration costs for *wh*-dependencies. These findings showed that distinct parsing processes are involved in processing filler-gap dependencies, with memory cost and integration cost functioning separately. According to Felser et al. (2003), processing resources are consumed at various stages of handling the displaced element in sentence processing, including temporary storage in working memory, subsequent retrieval, and final integration with its subcategorizers.

Ness and Meltzer-Asscher (2017) examined whether filler content is actively maintained in working memory during the processing of long-distance dependencies in Hebrew. They compared reading times on a critical NP between the filler and the gap in two structures: object relative clauses and NP-ellipsis sentences. In the NP-ellipsis condition, part of a noun phrase is omitted but can be interpreted from context. For example, in English sentence "Suzy admired the painting by the artist, but not [the painting] by the intern," the bracketed part is omitted. The ellipsis condition served as a control where no expectation of a gap was expected until the ellipsis site is encountered. In contrast, in the object relative clause condition, a filler-gap dependency is established early, requiring the parser to hold the filler in memory before encountering the gap. To test whether the filler is actively maintained in working memory, the researchers introduced interfering NPs that shared features such as animacy and grammatical category with the filler or antecedent. These similar intervening NPs were predicted to cause a *similarity-based interference effect*—that is, if filler content is maintained, then interference should increase processing difficulty before the gap site (Gordon et al., 2001). The results confirmed this prediction: the interfering NPs slow down reading times in the object relative clause condition, and this occurred before the gap. This finding suggested that some information about the displaced element was actively maintained during sentence comprehension, thereby supporting a maintenance or hybrid maintenance/retrieval account.

3.3.3 How does active gap-filling relate to incremental and predictive processing

A particularly interesting question about active gap-filling is whether it aligns with the incremental and predictive processing discussed in the previous chapter. There is a broad consensus that the gap-search process is incremental, as evidenced by the reading time slowdown and the immediate ERP components, such as P600, N400, and LAN at potential gap sites. However, whether it qualifies as predictive processing depends on the syntactic theory adopted (Pickering & Gambi, 2018). To briefly recap using the experimental sentences in Traxler and Pickering (1996), previously shown in (26) and reproduced here as (31) for convenience, the reading time slowdown at *wrote* in (31b) could suggest a prediction of an empty category, yet the same effect could also be interpreted as the filler *city* being linked to *wrote* only when the verb is encountered, without prediction.

(31) 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 & Pickering, 1996, p. 465)

3.4 Processing of English filler-gap dependencies in L1 and L2

This section reviews recent studies on English filler-gap processing that are particularly relevant to this dissertation. The studies selected for discussion are closely related to the experimental manipulations employed in the self-paced reading experiments in Chapter 4 and the ERP experiment in Chapter 5. Specifically, I focus on studies that have examined the effects of filler plausibility and dependency length on processing, as well as those that have investigated the filled-gap effect and reanalysis processes.

3.4.1 Psycholinguistic approaches

Williams et al. (2001, 2006) can be considered among the earliest studies investigating how L2 sentence processing aligns with or differs from native processing using the structure of filler-gap dependencies. The two studies provide evidence for L2 parsers' sensitivity to semantic cues and delay for syntactic cues. Although the two studies did not deliberately test SSH, they were in line with the idea that L2 speakers prioritize semantic information for sentence processing.

Williams et al. (2001) examined how filler plausibility affects the processing of *wh*-questions in English by native speakers and L2 learners with Chinese, German, or Korean as their first language. They were particularly interested in how native speakers and learners handle sentences where initial interpretations might need to be reanalyzed. In a stop-making-sense task, participants read sentences one word at a time and were asked to press a button when they thought the sentence stopped making sense. The experimental sentences are manipulated as in (32). In (32a), the filler *which car* is semantically plausible as a direct object of the first verb in the subsequent input *buy,* while *which friend* cannot be bought, meaning that it is an implausible filler associated with the verb *buy.*

(32) a. Plausible: Which car did the tourist buy the radio for ___?

b. Implausible: Which friend did the tourist buy the radio for ___?

(Williams et al., 2001, p. 536)

The displaced filler, *which car* or *which friend,* was thought to trigger the search for a potential gap. If Active Filler Strategy, as robustly tested in L1 processing, were employed, both native speakers and learners would immediately actively search for a gap. The were two regions of interest of the sentences. The first was the verb *buy*. If participants were sensitive to plausibility information in real time, they would be more likely to indicate that the sentence stopped making sense at this point in the implausible condition. The second region was the direct object after the verb, which consisted of a determiner *the* and a noun. If readers were positing gaps incrementally regardless of plausibility, they would try to fit the *wh*-phrase into the first available spot, regardless of whether it made semantic sense. In both sentences as in (32a) and (32b), readers would initially try to interpret them as *the tourist buy which car* and *the tourist buy which friend.* When they encountered *the radio*, it would cause surprise in both cases because that position was already filled, thus exhibiting an object filled-gap effect. If gap positing were influenced by plausibility, the effect would only appear in the plausible condition. In contrast, in the implausible condition, readers might not attempt to fit *which friend* into that position, and consequently, they would not be surprised by *the radio*.

The findings revealed that at the verb *buy*, both L1 and L2 speakers made more stop-making-sense decisions in the implausible condition as in (32b) compared to the plausible condition as in (32a). This suggests that participants were sensitive to plausibility information at this point. At the direct object position, L1 speakers showed increased reading times at the determiner *the* in the implausible condition and the following noun in the plausible condition, whereas L2 learners only exhibited increased reading times at the noun in the plausible condition. Williams et al. (2001) concluded that both L1 and L2 speakers employ the active gap-filling strategy in processing *wh*-dependencies. However, they found that L1 and L2 speakers differ in both the timing of reanalysis and the linguistic resources used for reanalysis. While the determiner *the* should theoretically serve as a sufficient grammatical cue to trigger reanalysis, only native speakers demonstrated the ability to utilize this cue effectively. The researchers therefore proposed that L2 speakers may be slower in reanalyzing initial parses and that their reanalysis process might be driven by argument competition, where the new noun phrase (e.g., *the radio*) and the filler (e.g., *which friend*) both initially appear to be interpreted as the theme, even though the verb assigns that role to only one argument.

Building on these findings, Williams (2006) conducted two experiments to further examine the role of syntactic cues and working memory in sentence reanalysis. The study revealed that working memory capacity influences both the incrementality of parsing and the ability to use structural cues for reanalysis. The findings suggested that L2 sentence processing may be more working memory intensive, which could explain why L2 speakers struggle to utilize syntactic cues like the determiner *the* as efficiently as L1 speakers during real-time processing.

Previous studies have supported the relationship between working memory and the dependency length between the filler and the gap. According to the Dependency Locality Theory (DLT) (Gibson, 1998, 2000), the number of words between a filler and its corresponding gap directly impacts processing difficulty. As this intervening word count grows, so does the cognitive load on working memory. This occurs because parsers must maintain the filler in their working memory until they encounter its corresponding gap. As discussed in Section 3.3.2, Wagers and Phillips (2014) conducted a series of three experiments with the manipulation of the dependency length and information type, to investigate how memory and control processes interact during the processing of filler-gap dependencies in native speakers. Their findings suggest that syntactic information appears to be actively maintained in working memory, while lexical-semantic details need to be retrieved at the gap.

Wagers and Phillips' (2014) finding that the plausibility effect diminishes with increased dependency length aligns with Dependency Locality Theory (DLT), which predicts that processing becomes harder with more intervening words. However, their results suggest a more nuanced picture than what DLT depicts. The persistence of syntactic information over long distances, contrasted with the loss of semantic information, indicates that not all linguistic information is equally affected by dependency length. The researchers proposed that our brain use two parallel processes when computing filler-gap dependencies. The first process, termed the *leading process*, maintains minimal but crucial information about the filler, primarily its syntactic category. This helps parsers to actively predict potential gap sites and build preliminary structures, even across lengthy dependencies. Then, the *lagging process* comes into play. This second mechanism involves retrieving the more detailed semantic information about the filler from working memory.

Based on Wagers and Phillips' (2014) findings and their hybrid maintenance-retrieval model, several questions arise regarding L2 filler-gap processing. The first question concerns whether L2 learners can maintain basic syntactic category information similar to native speakers during the leading process. The Shallow Structure Hypothesis (Clahsen & Felser, 2006) posits that L2 learners can construct syntactic structures, though less detailed than those of native speakers. If L2 learners can indeed maintain this basic syntactic information, it would be expected to observe active gap-filling effects in L2 processing during the leading process, as this process drives active gap filling according to the dual-process account. However, the Dependency Locality Theory (DLT) predicts that longer dependencies require maintaining filler information over greater distances. Consequently, it can be hypothesized that this active gap-filling effect, triggered by the maintained syntactic information, may be weaker or delayed in long dependencies for L2 learners. The second question concerns whether L2 speakers would retrieve this semantic information more slowly than native speakers. If so, this could result in reduced or delayed plausibility effects in L2 processing. It is also of interest to examine whether the plausibility effect becomes more pronounced as the linear distance between the filler and the gap increases. Thus, the question arises as to whether L2 speakers would show more reduced and delayed plausibility effects in extended filler-gap dependencies.

Chow and Zhou (2019) further investigated the effect of dependency length on active gap-filling. They replicated Wagers and Phillips's (2014) study using an eye-tracking methodology. They manipulated dependency length (Short, +PP for increased linear distance, and +CP for increased structural distance) and the plausibility of filler-verb relationships (plausible vs. implausible) across 48 item sets, doubling the 24 sets used by Wagers and Phillips. Their results revealed evidence of active gap-filling across all dependency lengths, with plausibility effects observed in the critical region for all conditions. Unlike Wagers and Phillips, who found plausibility effects only in short dependencies, Chow and Zhou observed the plausibility effect across all dependency lengths in the critical region. However, the effects were weaker in longer dependencies, partially aligning with Wagers and Phillips' results. They attributed the differences between the two studies to increased statistical power from more item sets and the use of eye-tracking, which allows for more natural reading patterns. Based on their results, Chow and Zhou (2019) proposed that active gap-filling operates regardless of dependency length and suggested that semantic information can be maintained across long dependencies, even though such information may be weakened over distance. They posited that this gradual weakening could be attributed to two factors: the increased cognitive load of maintaining the filler in memory over a longer span, as proposed by the Dependency Locality Theory (Gibson, 1998, 2000), and the potential decay of the filler's mental representation in the parser's memory over time (McElree, 2000).

3.4.2 Neurolinguistic approaches

ERP components

One method used to measure online brain responses to different linguistic information is the ERP technique. To measure ERPs, the electrical activity of the brain in the form of EEG is recorded by electrodes attached to the scalp. These continuous EEG signals enable researchers to examine processing in real time as a stimulus is presented, rather than just focusing on the behavioral performance. Compared to behavioral measures such as the self-paced reading paradigm, ERPs allow investigators to detect sub-conscious cognitive processing with millisecond precision due to their excellent temporal resolution (Luck, 2014).

This subsection examines three ERP components that are particularly relevant to the experiments in the field: N400, P600 and LAN.[18] The N400 is a negative-going voltage that starts

[18] The ERP components are briefly discussed here to provide background for the following review of ERP studies in the field. For a more thorough discussion of the N400, P600, and LAN, see Section 5.2 ERP Components.

around 250 ms and lasts for 200 to 500 ms and generally peaks around 400 ms after the onset of a stimulus (Kutas & Hillyard, 1980). Research has shown that the N400 indexes the difficulty of integrating an input to the existing semantic context (e.g., Kutas & Federmeier, 2000; Kutas & Federmeier, 2011). In other words, it is an indicator showing the difficulty of semantically integrating a word into a sentence, and the more difficult it is to integrate a word into the context, the more negative the N400 is.

The P600 has been associated with syntactic processing. It is a positive-going voltage that peaks around 600ms post-onset of the stimulus (Osterhout & Holcomb, 1992). The P600 is often observed by various syntactic anomalies including phrase structure violations and morpho-syntactic violations (e.g., Osterhout & Holcomb, 1992; Weber-Fox & Neville, 1996). Additionally, the P600 has been associated with syntactic reanalysis and repair (e.g., Friederici et al., 2001). Moreover, complex structures such as filler-gap dependencies and ambiguous structures like garden-path sentences can also trigger a P600 (e.g., Kaan et al., 2000). For example, Kaan et al. (2000) argued that the P600 reflects the underlying cognitive difficulty of general syntactic integration. They reported that the processing of long-distance dependencies, though well-formed, elicited a P600 effect in a study where three types of embedded questions (who, which, and whether) were used. The parsers typically anticipate a gap where "who" and "which" can be integrated, while "whether" does not necessitate integration into any such gap. They found a P600 effect observed in the "who" and "which" conditions at the main verb, but not in the "whether" condition. Their findings support the idea that P600 reflects processing difficulty associated with syntactic integration in general rather than syntactic violations alone.

Another ERP component that may be of interest in this section is the LAN. The LAN is a negative voltage occurring within the 300-500 ms range after the onset of the violation. Although it is predominantly left-lateralized, it can sometimes appear bilaterally (e.g., Pakulak & Neville, 2011; King & Kutas, 1993). The LAN has been associated with the integration of morphosyntactic information in a sentence structure (e.g., Molinaro et al., 2011). In many syntactic violation paradigms, the LAN and P600 can co-occur, often seen as a LAN followed by a P600 (e.g., Bowden et al., 2013). Some researchers view the LAN-P600 complex as reflecting a two-stage process in syntactic processing: early detection (LAN) followed by later reanalysis or repair (P600) (e.g., Friederici, 2002). Additionally, LAN has been argued to index the maintenance effort caused by an unresolved element stored in working memory (e.g., Phillips et al., 2005).

Recent ERP studies in filler-gap processing in L1 and L2

A series of studies have used the ERP technique to investigate the processing of filler-gap dependencies in native and second language speakers. Building on Garnsey et al. (1989), Dallas et al. (2013) employed ERPs to analyze real-time processing of filler-gap dependencies using sentences presented in (14), which are repeated in (33) for convenience, with L1 English speakers and L1 Chinese-L2 English speakers with intermediate or higher English proficiency.

(33) 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, p. 776)

The experimental sentence contains a filler that is either plausible or implausible with the verb. Garnsey et al. (1989) found an N400 effect at the critical verb when the filler was implausible as the direct object of the verb, suggesting that native English speakers had difficulty semantically integrating the filler with the verb. Similarly, Dallas and colleagues (2013) found that native English speakers showed an N400 at the verb when the filler was implausible (which football), whereas L2 group as a whole showed no significant difference between plausible and implausible conditions. It was only when L2 speakers were divided into high and low proficiency groups that a small N400 component was observed among the proficient L2 speakers. These results suggest that as L2 proficiency increases, learners can potentially achieve native-like processing of filler-gap dependencies.

Building on Dallas et al. (2013), Jessen and Felser (2019) examined how L1 and L2 speakers process filler-gap dependencies involving reanalysis. In addition to manipulating plausibility, they introduced reanalysis by including a filled-gap position before the actual gap. Sample stimuli are reproduced in (34), where the filler (*house* vs. *women*) was either plausible or implausible as the direct object of the embedded verb *built*. Notably, Jessen and Felser observed these N400 effects across their entire L2 group, unlike Dallas et al. (2013), who found such effects only in highly proficient L2 learners. At the actual gap position, where reanalysis was expected, only the L2 group exhibited P600 effects for initially plausible fillers, indicating greater difficulty with reanalysis in L2 speakers. Their findings suggested that L1-L2 processing differences emerged during later reanalysis stages rather than during initial filler integration and that L2 speakers experienced greater processing costs during the reanalysis stage.

(34) a. Plausible-at-V:

Bill liked the house that Bob built (__) some ornaments for __ at his workplace.

b. Implausible-at-V:

Bill liked the women that Bob built (__) some ornaments for __ at his workplace.

(Jessen & Felser, 2019, p. 288)

In addition to the plausibility manipulation, some L2 ERP studies employ the filled-gap paradigm to investigate L2 processing mechanisms of filler-gap dependencies. Jessen et al. (2017) used similar participant groups of L1 English speakers and L1-L2 English learners to investigate how they process indirect object *wh*-dependencies. The experimental stimuli consisted of sentence pairs as in (35). Sentence (35a) contains an ungrammatical indirect object *wh*-dependency with a filled gap, while (35b) is a grammatical control sentence without *wh*-movement. This design allowed for the examination of filler storage effects (at *for which*), verb-driven semantic integration effects (at *arranged*), and filled-gap effects (at *for it*). Both groups showed an N400 effect at the verb, suggesting semantic integration of the filler. At the filled gap position, both groups exhibited a positivity being interpreted as a P600 response, though with a more frontal distribution in natives. Interestingly, only L2 speakers showed a sustained LAN following the *wh*-filler, potentially reflecting greater working memory demands.

(35) 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.

(Jessen et al., 2017, p. 1323)

Using similar filled-gap paradigm, Covey et al. (2024) conducted an ERP study to investigate *wh*-dependency processing in English by native speakers and advanced Mandarin Chinese-speaking learners. Their experimental sentences crossed factors of Extraction (no extraction, *wh*-extraction) and Island (non-island, island), as shown in (36).

(36) Non-Island:

a. (No extraction) Jamie wondered if the editor interviewed Dave Campbell with the reporter from the department.

b. (*Wh*-extraction) Jamie wondered who the editor interviewed Dave Campbell with __ from the department.

Island:

c. (No extraction) Jamie wondered if the editor [that interviewed Dave Campbell] kissed the reporter after the meeting.

d. (*Wh*-extraction) Jamie wondered who the editor [that interviewed Dave Campbell] kissed __ after the meeting.

(Covey et al., 2024, p. 11)

The critical regions are *Dave Campbell* in all conditions. The non-island conditions enabled the study of gap prediction in grammatically licensed positions, while the island conditions tested whether participants would avoid predicting gaps in syntactically constrained environments. Native speakers showed an N400 effect at the filled gap in non-island conditions, suggesting that they engaged in gap prediction. In contrast, L2 learners exhibited a P600 effect in the same conditions, indicating integration difficulty rather than gap prediction. Crucially, neither group showed significant ERP effects within islands, demonstrating island sensitivity for both natives and learners. The authors concluded that L2 processing is grammatically guided, as evidenced by island sensitivity. However, the qualitatively different ERP responses (N400 for natives, P600 for learners) in non-island conditions suggest that predictive processing in *wh*-dependency resolution may be limited for learners whose L1 does not have overt *wh*-movement.

Further contributing to this line of research, Dong et al. (2023) examined the effects of proficiency and working memory on filler-gap dependency processing in L2 speakers. Participants included intermediate to highly proficient Chinese-speaking learners of English and native English speakers.[19] For convenience of reference, the experimental stimuli from (7) are reproduced here as (37).

(37) a. Filled-Gap condition:

*The zebra that the hippo kissed the camel on the nose ran far away.

[19] Dong et al. (2023) analyzed the native speakers' data collected by Hestvik et al. (2012), who used the identical experimental design and stimuli.

b. Adjunct condition:

The weekend that the hippo kissed the camel on the nose, it was humid.

c. Object condition:

The zebra said that the hippo kissed the camel on the nose and then ran far away.

d. Trace condition:

The zebra that the hippo kissed on the nose ran far away.

(Dong et al., 2023, p. 9)

Critically, native speakers exhibited a P600 effect at the filled gap site, reflecting syntactic violation detection, whereas L2 learners produced a prefrontal-central positivity that was not modulated by proficiency or working memory capacity. Notably, the findings from Dong et al. (2023) aligned more closely with Covey et al.'s (2024) conclusions about limited L2 predictive processing and supported accounts of L1-L2 differences. Notably, Dong et al.'s findings support Covey et al.'s (2024) claims of limited predictive processing in L2 speakers and contrast with Jessen et al.'s (2017) view that L1 and L2 syntactic processing mechanisms are qualitatively similar.

3.5 Summary

This chapter provided a thorough overview of the processing of filler-gap dependencies processing in both L1 and L2 speakers. The investigation of such structure is significant for understanding how complex syntactic structures are processed between L1 and L2 speakers in real time. Research in filler-gap processing also helps reveal possible cross-linguistic differences in processing strategies and the interaction between syntactic and semantic information during sentence comprehension. As the present dissertation focuses on the processing of such structures in L1 Mandarin Chinese and L1 English, I also reviewed the syntactic properties of these structures in both languages. While English exhibits overt *wh*-movement, creating clear filler-gap dependencies, Mandarin Chinese is typically considered *wh*-in-situ and involve the covert dependency. Although Mandarin Chinese allows for overt topic structures that resemble English filler-gap dependencies, the psychological reality of the filler—whether it is a moved or a dangling element—remains a subject of debate among linguists and psycholinguists. Two main theories have been proposed to explain filler-gap processing: the Active Filler Strategy and the Direct Association Hypothesis. The Active Filler Strategy posits that upon encountering a filler, the parser actively searches for potential gap sites. Instead, the Direct Association Hypothesis suggests that fillers are directly associated with their licensing verbs without positing gaps. The chapter reviews various psycholinguistic and neurolinguistic studies, discussing effects such as filled-gap, plausibility mismatch, and the role of dependency length in processing. Neurolinguistic studies, particularly those using the ERP technique, concentrate on N400, the component related to semantic processing, and P600, the component associated with syntactic processing and reanalysis. These ERP components offer a window into the real-time neural processes underlying sentence processing.

This dissertation focuses on embedded *wh*-dependencies in English for three key reasons. First, they exist in both English and Mandarin. Second, they both require the integration of semantic and syntactic information. Last, their complexity allows us to manipulate various factors, including filler plausibility, dependency length, and reanalysis effects. While the study does not examine L2 speakers' sensitivity to island constraints, as this has been widely tested, it instead concentrates on how L2 speakers use semantic and syntactic information during processing, and the timing/degree differences in their use of these cues compared to L1 speakers. This manipulation helps to test the updated Shallow Structure Hypothesis (Clahsen & Felser, 2018), which suggests that L2 speakers may not lack the ability to use syntactic cues entirely, but rather tend to rely more on semantic information and process linguistic information at different timing compared to L1 speakers.

Chapter 4

The Self-Paced Reading Experiments

4.1 Introduction

The main goal of this dissertation is to investigate how L2 speakers process filler-gap dependencies and whether they can achieve qualitatively similar processing profiles as native speakers with increasing proficiency or years of immersion. This chapter presents two self-paced reading experiments designed to address this question.

Previous research has established that L2 speakers are sensitive to island constraints (e.g., Omaki & Schulz, 2011; Felser et al., 2012; Boxell & Felser, 2017). However, questions remain about how L2 speakers process filler-gap dependencies in the grammatically licit non-island contexts, particularly regarding the use of syntactic and semantic information. The SSH suggests that L2 speakers are capable of using syntactic cues, but they may prefer to use semantic information. The timing and degree to which L1 and L2 speakers use linguistic information (semantic and syntactic) may differ. Therefore, the present study wants to know if syntactic and semantic information is used differently in terms of timing or degree.

While several self-paced reading studies have shown that L2 speakers employ active gap-filling mechanisms (e.g., Williams et al., 2001; Omaki & Schulz, 2011), it remains unclear what drives this active gap-filling process. Is this process driven primarily by syntactic or semantic information? Or is it driven by both types of information? These questions were addressed in Experiment 1. As previously discussed, Wagers and Phillips (2014) found evidence suggesting that syntactic category information is maintained throughout dependency formation and is used to drive active gap-filling, while semantic information may be released from maintenance in long-distance dependencies and has to be retrieved when the actual gap is reached. However, Chow and Zhou (2019) challenged these findings, observing plausibility effects across all dependency lengths in an eye-tracking study, suggesting maintenance of semantic features as well. To address these conflicting results, Experiment 1 manipulated both dependency length and filler plausibility to investigate how L1 and L2 speakers process English *wh*-dependencies in real time. This design allows me to examine the relative contributions of syntactic and semantic information in filler-gap processing across different dependency lengths in both L1 and L2 speakers.

Also at issue is investigating whether the gap-type topic structure in Mandarin Chinese features movement-derived non-local dependency as filler-gap dependencies in English do. This issue was addressed in Experiment 2. Despite empirical evidence supporting the existence of covert *wh*-movement in Mandarin (e.g., Xiang, 2014, 2015; Yang et al., 2023), there is still ongoing debate regarding the existence of overt *wh*-movement in the language. It is important to note that when an overt topic does not necessarily indicate syntactic movement. Instead, such topic structures may be licensed by aboutness, or they may involve base-generated dangling elements that are directly associated with the verb. Given this uncertainty, several studies have attributed the qualitative difference between native English speakers and Mandarin Chinese learners of English to the lack of overt *wh*-movement in their L1 (e.g. Dallas et al., 2013; Dussias & Piñar, 2010; Juff, 2005; Covey et al., 2024). The present investigation aims to test whether such processing differences can indeed be attributed to L1-L2 differences. Experiment 2 used a self-paced reading paradigm to examine how native Mandarin Chinese speakers process gap-type topic structures in real time. Specifically, I aim to investigate 1) whether native Mandarin Chinese speakers actively search for a gap and are sensitive to filler plausibility at the first verb; 2) whether they employ a reanalysis strategy when the initially proposed gap is proven to be filled; and 3) whether they show increased reading times for gap filling when the linear dependency length is longer. This design allows me to investigate the parsing mechanism of gap-type topic structures in Mandarin Chinese and determine whether it is movement-derived as its English counterparts.

Through these experiments, I aim to provide a comprehensive picture of how L2 speakers process *wh*-dependencies in real time, contributing to debates about the nature of L2 sentence processing. I also add to the limited literature about whether Mandarin Chinese features overt *wh*-movement.

4.2 Experiment 1: Filler-Gap processing in L1 and L2 English Speakers

4.2.1 Rationales and predictions

This self-paced reading experiment aims to investigate how L1 English speakers and L2 English speakers with L1 Mandarin Chinese process *wh*-dependencies in English. The experiment adapted the methodology from Experiment 2 of Wagers and Phillips (2014), with several modifications to improve its suitability for L2 speakers and increase the statistical power. Below is a detailed review of the original methodology, followed by an explanation of how it was adapted for the present study.

Wagers and Phillips (2014) manipulated the semantic anomaly effect in a self-paced reading paradigm to measure active dependency formation. Their experiment contrasted sentences in (38). In (38a), (38c) and (38e), interpreting *acorns* as the direct object of *crammed* is semantically plausible. However, in (38b), (38d) and (38e), treating *gods* as the direct object of *erected* creates a semantic anomaly. This design predicted a slowdown in reading times at or shortly after the verb in the implausible condition. The researchers examined this effect across three dependency lengths: short single-clause, long single-clause (+PP), and long bi-clausal (+CP). They hypothesized that if the semantic details of the filler are maintained in working memory, a semantic anomaly effect should be observable regardless of dependency length. In their design, the verb was the critical region for detecting the semantic fit of the filler. To account for potential spillover effects in self-paced reading, they included a four-word direct object region after the verb. The inclusion of spillover region provided ample space to capture any delayed processing effects. The researchers considered a slowdown in reading times to be evidence of active dependency construction if it occurred either immediately at the verb, or within the following four-word direct object region. This slowdown was expected to be more pronounced in the implausible conditions, where the semantic fit between the filler and the verb is poor.

(38) a. Short – Plausible:

The acorns which the squirrels quickly crammed their small puffy cheeks with ___ before scurrying out of the park had fallen from a sickly oak.

b. Short – Implausible:

The cats which the squirrels quickly crammed their small puffy cheeks with ___ before scurrying out of the park were basking in the sun.

c. +PP – Plausible:

The acorns which the squirrels with the bushy black tails quickly crammed their small puffy cheeks with ___ before scurrying out of the park had fallen from a sickly oak.

d. +PP – Implausible:

The cats which the squirrels with the bushy black tails quickly crammed their small puffy cheeks with ___ before scurrying out of the park were basking in the sun.

e. +CP – Plausible:

The acorns which the experienced naturalist concluded that the squirrels quickly crammed their small puffy cheeks with ___ before scurrying out of the park had fallen from a sickly oak.

f. +CP – Implausible:

The cats which the experienced naturalist concluded that the squirrels quickly crammed their small puffy cheeks with ___ before scurrying out of the park were basking in the sun.

(Wagers & Phillips, 2014, p. 1288)

The current experiment builds upon Wagers and Phillips' work with several key modifications. Firstly, I expanded the experimental design to increase statistical power. While the original study used 24 sets of items with 6 conditions (resulting in 4 trials per participant per condition), the present study employed 32 sets of items with 4 conditions, yielding 8 trials per participant per condition. This increase was expected to enhance the study's statistical robustness. Secondly, in contrast to Wagers and Phillips' study, which manipulated three different dependency lengths (Short, +PP, and +CP), the present experiment focused on just two dependency lengths: Short and Long (Linear). This simplification allowed me to isolate the effects of linear distance while keeping structural distance constant. To understand this design choice, it is important to distinguish between linear and structural distance in linguistic dependencies. Linear distance, which refers to the number of intervening words between related elements in a sentence, is closely tied to working memory capacity. According to Gibson (1998, 2000), processing sentences with greater linear distances requires more working memory resources. In filler-gap dependencies, if the filler is maintained in the memory until the gap is posited, longer linear distance would consume more working memory resources and result in processing difficulty. Structural distance, which refers to the number of intervening syntactic nodes in the syntactic tree, is perceived as a measurement of syntactic processing ability independent of general cognitive ability such as working memory (e.g. Hawkins, 1999; O'Grady et al., 2003). In Wagers and Phillips' design, the short condition had 5 intervening words between the filler and gap. The +PP condition increased linear distance by adding a prepositional phrase. The +CP condition introduced a new clause, increasing structural distance while maintaining the same linear distance as the +PP condition (9 intervening words). The present study simplified this design. Specifically, I included only a short distance condition (2 intervening words) and a long-distance condition (7 intervening words). Crucially, I increased only the linear distance without altering the structural distance. As illustrated in Figure 4.1, in the long dependency sentence (tree on the left), the syntactic nodes between the filler and the gap are CP, IP, VP, and NP; in the short dependency sentence (tree on the right), the nodes between the filler and the gap remain the same. This means that the structural distance between the filler and the gap remains constant across conditions. Thus, the added complexity in the long dependency condition is confined within the embedded subject NP, increasing linear but not structural distance between the filler and the gap. This design choice allowed me to isolate the effects of working memory load from syntactic complexity. If, as proposed by Wagers and Phillips (2014), syntactic information is maintained in working memory and consumes cognitive capacity, then extending the linear distance between the filler and the gap would likely cause the syntactic information to decay. As a result, in the long dependency length condition, participants would experience more difficulty integrating the filler into the gap. Lastly, we have adapted our stimuli to better suit L2 participants. Instead of using Wagers and Phillips' experimental materials, I drew from Dallas et al. (2013), whose stimuli were more appropriate for L2 learners. I revised the vocabulary to include more frequently used English words. This adaptation reduced the cognitive load associated with lexical complexity for L2 participants, allowing us to more accurately assess their automatic use of syntactic and semantic information without the interference of vocabulary difficulties.

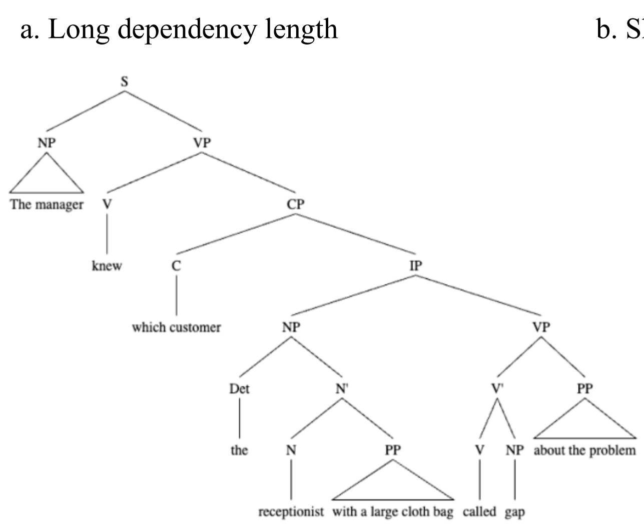


Figure 4.1 Syntactic trees illustrating a) long and b) short dependency length sentences

In this self-paced reading experiment, I employed a 2×2 within-participants design to manipulate two factors: filler plausibility and dependency length. Filler plausibility was varied by creating noun phrases (NPs) that either can or cannot semantically match the gap. Dependency length was manipulated by introducing prepositional phrases (PPs) between the filler and the gap, thereby increasing the linear distance while maintaining the same structural distance. The research questions for this experiment are as follows:

1. How does dependency length (long vs. short) affect the processing of filler-gap dependencies in L1 and L2 English speakers?
2. How does filler plausibility (plausible vs. implausible) affect the processing of filler-gap dependencies in L1 and L2 English speakers?
3. Do L2 proficiency and Years of Immersion (YoI) affect the processing of dependency length and filler plausibility in filler-gap dependency processing for L2 learners?

Both L1 and L2 groups are predicted to make use of the Active Filler Strategy, as reflected by increased reading times at the verb or shortly after. According to the hybrid maintenance/retrieval view, syntactic cues are maintained and trigger active gap filling in L1 speakers, while semantic information is retrieved at the gap. Consequently, L1 speakers are predicted to quickly posit a gap and actively search for it based on the syntactic information associated with the filler-gap dependency. This should result in increased reading times immediately at the verb or shortly after, regardless of the filler's plausibility. The plausibility effect, however, is expected to be observed when the retrieval of semantic information comes into play, probably in the post-critical area.

As for L2 speakers, the processing mechanism may differ. According to the updated Shallow Structure Hypothesis (Clahsen & Felser, 2018), L2 speakers are in fact capable of using syntactic representations during real-time processing, rather than being unable to employ them. However, they tend to "underuse" this type of information (Clahsen & Felser, 2018, p. 695). Regarding this underuse, Clahsen and Felser (2018) argue that L2 speakers face particular difficulty when processing sentences involving complex structures with displaced elements. The *wh*-dependency is such a structure, where parsers need to manipulate abstract elements of grammatical representations—empty categories. They further suggest that grammatical rules are less emphasized in the L2 system compared to non-grammatical factors, and L2 learners might rely more on non-grammatical information than on grammatical rules. Based on the SSH, L2 speakers are predicted to rely more heavily than L1 speakers on surface-level information (e.g., lexical-semantic, pragmatic, probabilistic) to compensate for their less robust ability to construct abstract syntactic representations during real-time resolution of filler-gap dependencies. Also, according to the hybrid maintenance/retrieval model (Wagers & Phillips, 2014), semantic information is retrieved when the gap is reached rather than being stored in memory. For L2 speakers, this means that little syntactic information is maintained in working memory, and they have to reactivate the semantic information at the gap. Therefore, they would face more difficulty at the verb as they initially have little maintained syntactic information left in memory and must retrieve the filler from the gap. Because of this, they would be more sensitive to dependency length, as longer linear distance decays the potential limited syntactic information in memory and makes retrieval more challenging (e.g., Gibson, 1998; Chen et al., 2005).

In summary, the rationales and predictions for the present study are as follows: As maintenance increases working memory load, longer distances would weaken active gap filling at the gap site, resulting in longer reading times when the dependency length is greater. Given L2 speakers' limited working memory capacity, I expect 1) greater difficulty with longer dependencies compared to L1 speakers and 2) potentially more pronounced decay of syntactic information and greater challenges in retrieving semantic information over distance. Additionally, manipulating the semantic fit between the filler and the verb at the gap site enables me to assess when semantic information is accessed during dependency resolution. While L1 speakers do not rely on semantic information for dependency resolution and may show a plausibility effect at the post-verbal regions, L2 speakers are predicted to prioritize semantic information over syntactic information. As a result, they are expected to show a plausibility effect earlier, particularly in shorter dependencies where semantic information can be easily retrieved. Furthermore, previous research has shown that highly proficient L2 speakers can process complex sentences in a manner indistinguishable from native speakers (e.g., Felser & Drummer, 2017; Dallas et al., 2013) and that increased exposure to the target language may result in more automatic grammatical processing (e.g., Steinhauer et al., 2009). Building on these findings, the present study hypothesizes that as L2 proficiency or YoI increase, L2 speakers will process *wh*-dependencies in a manner qualitatively similar to native speakers. This similarity is expected to be reflected in comparable sensitivity to dependency length and plausibility, as well as similar timing of these effects.

4.2.2 Participants

The study included a total of 83 participants who provided informed consent and completed the online study: 28 native English speakers (L1) and 55 Mandarin Chinese learners of English (L2). Recruitment methods differed between the two groups. L1 speakers were recruited from the Integrated System of Participant Research (ISPR) at the University of Ottawa, a platform where students receive course credits for research participation. L2 speakers, on the other hand, were recruited through three channels: the ISPR, word of mouth, and recruitment scripts posted on social media pages. While ISPR participants received course credits, those recruited through the other two channels participated voluntarily. Study advertisements for both groups specified the eligibility criteria that they should have no visual impairments, no history of serious head injury, and no diagnosis of attention-related disorders. All L1 participants reported to be native English speakers and answered "yes" to the following three questions before proceeding to the study: 1) Did you primarily speak English as a child? 2) Is English the language you use most in your life? 3) Do you think and talk to yourself in English? Similarly, all L2 speakers confirmed their native Mandarin speaker status and current residence by answering "yes" to the following questions before proceeding to the study: 1) Did you primarily speak Mandarin Chinese as a child? 2) Do you think and talk to yourself in Mandarin Chinese? 3) Do you reside in an English-speaking country at the time of participating?

After reviewing the demographic and language background questionnaire (Sabourin et al., 2016; see Appendix A), several were excluded based on specific criteria. In the L1 group, three participants were excluded: one due to self-reported visual impairment, another because their native language was not English, and a third for falling outside the 18-35 age range for analysis. In the L2 group, twelve participants were excluded for various reasons: three had not completed the language background questionnaire, four reported a native language other than Mandarin Chinese, one reported a second language other than English, one reported having a visual impairment, and three fell outside the 18-35 age range for analysis. Furthermore, participants with low comprehension accuracy (< 70%) were also excluded based, resulting in the removal of three additional L1 participants and five L2 participants. As such, the final sample consisted of 22 participants in the L1 group, with a mean comprehension accuracy of 88.64%, and 38 participants in the L2 group, with a mean comprehension accuracy of 87.83%. Complete participant information is presented in Table 4.1.

Table 4.1 Participant information for Experiment 1

Group		Age	Self-rated English proficiency /5[a]	Accuracy of measured English Proficiency /100[b]	Age of initial exposure	Age of immersion	Years of immersion
L1	Mean	19.14	5	N/A	0	0	19
(n=22)	*SD*	1.47	0.2	N/A	0	1	2
	Range	18-24	4-5	N/A	0-1	0-4	15-24
L2	Mean	25.97	3.05	64.42	6.39	19.42	6.55
(n=38)	*SD*	3.77	0.57	16.11	2.97	6.41	5.49
	Range	20-35	2-4	22-90	0-12	3-35	0-20

[a] Proficiency levels are self-rated according to this scale: 0 = very low, 1=low, 2=intermediate, 3=advanced, 4=near-native, 5-native.
[b] Proficiency levels are measured by their accuracy of a cloze task that can be found in Appendix B.

To address the third research question examining the effect of L2 speakers' proficiency and YoI on the processing of *wh*-dependencies, L2 participants were categorized into proficiency and immersion groups. Proficiency grouping was based on participants' performance on a cloze task completed before the self-paced reading task and their self-reported scores on English standardized tests. The high-proficiency group comprised 16 Mandarin Chinese learners of English who achieved 65% or higher accuracy on the cloze task and reported standardized test scores not lower than the equivalence of 7 in International English Language Testing System (IELTS). The low-proficiency group included 15 participants who scored below 65% on the cloze task and reported standardized test scores equivalent to IELTS 8 or lower. Seven L2 participants were excluded from proficiency group assignment due to conflicting criteria between their cloze task performance and self-reported test scores. For example, a participant got 88% accuracy on the cloze task; however, his self-reported IELTS score is 5.5. According to International English Language Testing System (IELTS), a score of 5.5 means that the student "has a partial command of the language... handles basic communication in their own fields" ("IELTS band scores," n.d.). Thus, it was not possible to include this participant in either the high or low group for analysis. For immersion groupings, the cutoff point was established at 0.5 standard deviations above the mean immersion duration (M = 6.5 years, SD = 5.65 years), resulting in a threshold of 9 years. This categorization resulted in 13 high-immersion participants (9 or more years) and 25 low-immersion participants (fewer than 9 years). The 0.5 SD above the mean was chosen to capture participants with notably longer immersion experiences while maintaining reasonable sample sizes in both groups. This categorization allows for analysis of the effects of both proficiency and immersion experience on the processing of wh-dependencies in L2 speakers, addressing the study's third research question.

4.2.3 Materials

Thirty-two sets of four conditions each were used in the experiment, in a 2×2 factorial design, in which the filler plausibility (plausible vs. implausible) and dependency length (long vs. short) were manipulated as independent factors. Filler plausibility was manipulated by utilizing noun phrases (NPs) that were either semantically compatible or incompatible with the target verb.[20] Dependency length was manipulated by altering the distance between the filler (e.g., *which customer/which spoon*) and the verb (e.g., *called*). In short conditions, a two-word subject NP separated the filler and verb, while long conditions incorporated an additional five-word prepositional phrase, extending the linear distance to seven words between the filler and the gap. The stimuli were distributed across four lists using a Latin Square design, such that each list contained one condition from each set. Each list also contained 16 filler sentences which were identical in length but varied in syntactic complexity, and some of which also contained a filler-gap dependency (e.g., subject-extracted relative clause). Thus, each participant encountered 32

[20] Filler plausibility is based on frequency, meaning that the semantic fit was determined by whether the filler noun and the verb commonly or uncommonly occur together in natural language usage. Two native English speakers reviewed all items to ensure that the plausible combinations sounded natural and the implausible ones did not.

experimental sentences and 16 filler sentences. To contextualize the experimental sentences and disguise the task as a short story comprehension exercise, each trial began with a background sentence presented in its entirety on the screen. Following the background sentence, the critical experimental sentence was presented word-by-word. To maintain consistency in the position of the critical words across all conditions, a five-word prepositional phrase was strategically placed. In the long conditions, a five-word prepositional phrase followed the background sentence. In the short conditions, this phrase (e.g., *After the last table left*) was added at the beginning of the experimental sentence. This design ensured that the critical regions appeared in the same position across all conditions, allowing for direct comparison of reading times. Among 48 sentences in total, 16 experimental sentences were followed by a yes/no question and 12 filler sentences were followed by a yes/no question to ensure that participants were attending to the stimuli. All yes/no questions were identical for the four conditions in each experimental set so that no matter which list the participant was assigned, the difficulty of the questions was identical. A crucial consideration to accurately measure the plausibility mismatch effect was ensuring that both plausible and implausible fillers fit the broader context of the sentence while being semantically incongruent with the critical verb. That is, both types of fillers were contextually appropriate up until the critical region. This design prevented participants from anticipating any incongruities before reaching the critical region. The filler plausibility match/mismatch, overall naturalness of the sentence, and yes/no questions were validated by three native English speakers. A sample set of stimuli is summarized in Table 4.2. A full set of materials for this experiment can be found in Appendix C.

In this experimental design, the verb serves as the critical region, as it is the initial point where the semantic fit of the filler can be detected. Additionally, the study incorporates measurements of two subsequent words, designated as post-critical region 1 and post-critical region 2, to account for spillover effects. The inclusion of post-critical regions in the analysis is motivated by three considerations. First, according to Wagers and Phillips (2014), the processing of filler-gap dependencies involves two stages: an early stage in which structural expectations are generated based on limited maintained syntactic information, followed by a later stage in which the full filler information is retrieved and reactivated. While Wagers and Phillips (2014, p. 1928) originally referred to this as "two complementary processes", I will use the term *dual-process account* for convenience in the remainder of this discussion.[21] Within this framework, it is plausible that the integration of the filler's semantic information occurs not at the verb itself, but in the post-critical regions that follow. Second, empirical evidence from previous studies supports the examination of these post-critical regions (e.g., Johnson, 2015; Yan & Liu, 2021). For instance, Johnson (2015) found that low-proficiency L2 speakers exhibit filled-gap effects in spillover regions rather than at the critical verb region. This finding underscores the importance of examining areas beyond the verb to capture effects across different proficiency levels in L2 speakers. Last, Wagers and Phillips (2014), in their Experiment 2, acknowledged that in self-paced reading tasks, processing can lag behind the word being read. They argued that a slowdown in reading times in implausible conditions, either at the verb or in following regions, should indicated that readers are actively constructing the filler-gap dependency. In addition to the critical and post-critical regions, the region preceding the verb is measured as the pre-critical region, where no plausibility effect is expected.

[21] The hybrid maintenance/retrieval model and the dual-process account both describe how filler-gap dependencies are processed in real time, but they emphasize different aspects of the parsing mechanism. The hybrid maintenance/retrieval model concerns what types of information are actively maintained or retrieved in working memory, whereas the dual-process account emphasizes when each step of processing occurs.

Table 4.2 Sample experimental stimuli for Experiment 1

Background sentence: The coffee shop was very busy today, and the staff hardly had any time to rest. (After the last table left, …)		
Dependency length	Filler plausibility	Sentence
Short	Plausible	After the last table left, the manager knew *which customer* the receptionist *called* ___ about the problem.
Short	Implausible	After the last table left, the manager knew *which spoon* the receptionist *called* ___ about the problem.
Long	Plausible	the manager knew *which customer* the receptionist with a large cloth bag *called* ___ about the problem.
Long	Implausible	the manager knew *which spoon* the receptionist with a large cloth bag *called* ___ about the problem.
Yes/No question		Was the coffee shop busy today?
Answer		Yes

4.2.4 Procedures

The study was conducted with approval from the University of Ottawa's Research Ethics Board (S-06-21-7042). The entire study was administered online using the Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). Participants received an invitation email containing the experiment link. Upon clicking the link, they were first presented with three screening questions (as mentioned in the Participants section) to confirm their eligibility for either the L1 or L2 group. If participants answered "yes" to all questions, they proceeded to the study; otherwise, they had the option to exit. Participants in both groups were then instructed to complete the consent form and a language background questionnaire (Sabourin et al., 2016; see Appendix A). Following this, L1 speakers proceeded directly to the self-paced reading task. L2 speakers, however, were first asked to complete a cloze task (see Appendix B) and optionally provide their English standardized test scores (from TOEFL, IELTS, or Duolingo, if available) before moving on to the self-paced reading task.

In the self-paced reading task, words were presented in black on a white background. Instructions were displayed on the screen in English. Before the experimental session, participants were presented with five practice trials (varying in structure and length; three with yes/no questions and two without) to familiarize themselves with the task. Each trial began with a fixation point signaling the start of a new sentence. A background sentence was then presented in its entirety on the screen, followed by the 16-word experimental or filler sentence displayed word-by-word at the participant's own pace. The reading time for each word was recorded in milliseconds for each participant. The end of each sentence was marked with a period. In total, each participant encountered 32 experimental stimuli and 16 filler sentences. Of these, 16 experimental sentences and 12 filler sentences were followed by a Yes/No comprehension question to ensure participants were attentive to the task. Participants were instructed to respond by pressing “F” (for yes) or “J” (for no) on the keyboard. Halfway through the self-paced reading task, participants were offered a break of up to 10 minutes. They could either rest during this time or click "continue" to proceed at any point before the 10 minutes elapsed. If participants chose to use the full 10 minutes, the break automatically ended with the display of a fixation cross, signaling the start of the next trial. At the end of the self-paced reading task, a debrief page briefly introduced the research objectives and design of the study. The entire session lasted approximately 1 hour for L2 speakers and 30 minutes for native speakers. The longer duration for L2 speakers was due to the additional cloze task and their anticipated slower reading times.

4.2.5 Data Analysis

Data cleaning was performed using a two-step process: within-participant cleaning followed by between-participant cleaning (see Gosselin & Sabourin, 2024). The first phase employed absolute thresholds, removing reading times below 100 ms and above 2000 ms. This resulted in the exclusion of 3.30% of data points in the L1 group and 1.17% of data points in the L2 group. The relative threshold of ±2.5 standard deviations from each participant's means was not employed. Given the nature of the study, I anticipated legitimate variations in reading times across different experimental conditions and word regions. These variations could be crucial for understanding the plausibility effect and distance effect. Therefore, to avoid removing potential meaning variability in participants' responses, I focused on identifying and removing only the most extreme outliers, which were more likely to be caused by factors unrelated to our experimental manipulations.

The subsequent between-participant screening employed relative thresholds, following the methodology of Aoshima et al. (2004). Reading times above or below 2.5 standard deviations from the mean RT of each experimental condition (short-plausible, short-implausible, long-plausible, long-implausible) at each region (pre-critical, critical, post-critical 1, post-critical 2) were discarded. This step resulted in the removal of 2.86% of L1 data points and 3.25% of L2 data points. Finally, I excluded participants who retained less than 80% of their original trials after these procedures, which led to the removal of two participants from each group, resulting in 20 participants in the L1 group and 36 in the L2 group.

The analysis focused on reading times at four critical positions: pre-critical (Word 12), critical (Word 13), first post-critical (Word 14), and second post-critical regions (Word 15). At each position, two separate ANOVA models were conducted to examine the effects of proficiency and immersion separately. In other words, this design aimed to evaluate whether grouping participants by proficiency or immersion better explained variability in plausibility and dependency length sensitivity. The Proficiency Model included Dependency Length (short, long) and Filler Plausibility (plausible, implausible) as within-subjects factors, and Proficiency Group (L1, L2-High, L2-Low) as a between-subjects factor. The Immersion Model maintained the same within-subjects factors while including Immersion Group (L1, L2-Long, L2-Short) as the between-subjects factor. For significant main effect of Group and significant interactions in both models, follow-up analyses were conducted using pairwise comparisons with Bonferroni correction for multiple comparisons. For each word region, the results were reported with 1) the omnibus ANOVAs for both the Proficiency Model and the Immersion Model, 2) the descriptive mean reading times and standard errors by condition, and 3) follow-up pairwise comparisons.

Although no clear group differences in plausibility sensitivity were observed within the L2 participants, I conducted regression analyses to examine whether Proficiency and Years of Immersion (YoI) influence reading times to make L2 processing more native-like. Specifically, the analyses aimed to determine whether Proficiency and YoI contribute to explaining reading time variance (general reading speed patterns) and whether they modulate plausibility sensitivity in L2 participants. These models used reading time as the dependent variable and included Filler Plausibility, Proficiency (as a continuous variable), and Years of Immersion (YoI, also continuous) as independent variables. I examined these effects at each critical word position using two sequential models. The first model (Model 1) was specified as: Reading Time ~ Filler Plausibility * Proficiency. The second model (Model 2) was specified as: Reading Time ~ Filler Plausibility * Proficiency * YoI. For each word position, I first ran Model 1 to assess the effects of filler plausibility and proficiency on reading times. Following this, I ran Model 2, which included YoI as an additional factor, to determine the unique contribution of immersion experience in explaining variance in reading times, beyond what was already accounted for by proficiency. To evaluate the significance of including YoI, I performed an analysis of variance (ANOVA) comparing Models 1 and 2 for each word position. This allowed me to assess whether the addition of YoI significantly enhanced the model's explanatory power. I also examined the change in R-squared values to quantify the additional variance explained by including YoI in the analysis.

4.2.6 Results

Reading time results for all word regions of interest across all conditions for all proficiency and immersion groups are shown in Figure 4.2.

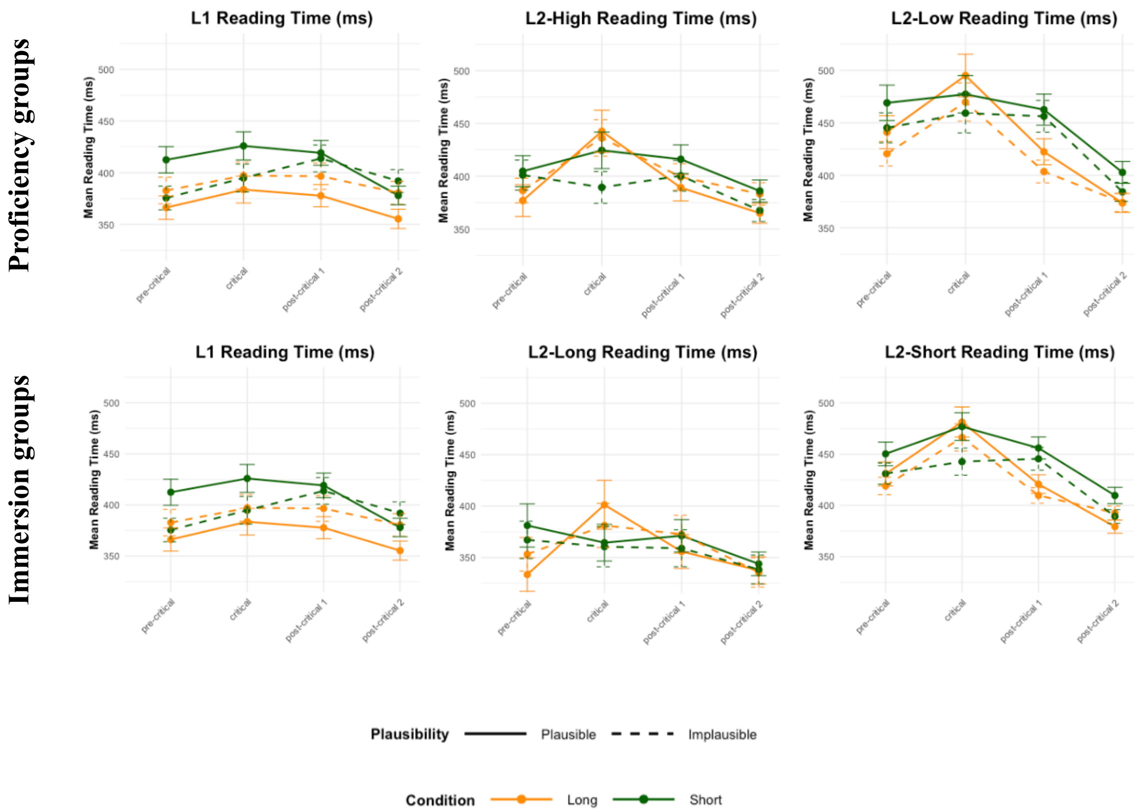


Figure 4.2 Experiment 1 Mean reading times across word regions

Green lines represent short dependency conditions; orange lines represent long dependency conditions. Dashed lines indicate implausible conditions; solid lines indicate plausible conditions.

Pre-critical region (Word 12)

At Word 12, analyses revealed a significant main effect of Dependency Length in both the Proficiency Model ($F(1, 47) = 13.14$, $p < .001$) and the Immersion Model ($F(1, 53) = 14.05$, $p < .001$). In both models, participants showed longer reading times for short-dependency sentences (Proficiency Model: $M = 416$ ms, $SE = 15.4$; Immersion Model: $M = 411$ ms, $SE = 14.6$) compared to long-dependency sentences (Proficiency Model: *M* = 395 ms, *SE* = 14.3; Immersion Model: *M* = 391 ms, *SE* = 13.6). This unexpected increase in reading time for the short dependency in the pre-critical region was likely due to the subject filled-gap effect, which I explored in more detail in the discussion section.

The Immersion Model additionally revealed a significant interaction between Dependency Length and Filler Plausibility ($F(1, 53) = 5.80$, $p = .020$). Follow-up analyses of this interaction showed that in the short dependency condition, implausible items were read significantly faster than plausible items ($t(55) = -2.97$, $p = .004$), while no plausibility effect was observed in the long dependency condition ($t(55) = 0.824$, $p = .413$). Figure 4.2 showed the same pattern. This unexpected finding of slower reading times for plausible fillers can be explained by the relationship between filler animacy and the subject filled-gap effect. Wagers and Pendleton (2016) found that animate fillers are more likely to trigger the subject filled-gap effect, and in the present study, Word 12 happened to be a subject gap position where plausible fillers tended to be animate. Out of 32 experimental sets, 16 had both plausible and implausible fillers being inanimate, 12 had plausible fillers being animate and implausible fillers being inanimate, and 4 had implausible fillers being animate and plausible fillers being inanimate. Consequently, in the short dependency condition, the higher proportion of animate fillers in the plausible condition appeared to cause increased reading times at the subject gap position.

The lack of significant three-way interactions suggested that the two-way interaction between Dependency Length and Filler Plausibility was consistent across different proficiency and immersion groups. Table 4.3 shows the mixed-effects ANOVA results and the follow-up pairwise comparisons for Word 12 reading times.

Table 4.3 ANOVA results and the follow-up pairwise comparisons for Word 12 reading times

1. Omnibus ANOVA results						
	Effect	*F*	*df*	*p*	*Sig.*	η^{2p}
Proficiency Model	Proficiency Group	1.481	2, 47	0.238		0.059
	Dependency Length	13.14	1, 47	< .001	***	0.218
	Filler Plausibility	2.284	1, 47	0.137		0.046
	Proficiency Group × Dependency Length	0.145	2, 47	0.866		0.006
	Proficiency Group × Filler Plausibility	2.096	2, 47	0.134		0.082
	Dependency Length × Filler Plausibility	2.887	1, 47	0.096	.	0.058
Immersion Group	Immersion Group	2.283	2, 53	0.112		0.079
	Dependency Length	14.053	1, 53	< .001	***	0.210
	Filler Plausibility	1.181	1, 53	0.282		0.022
	Immersion Group × Dependency Length	0.424	2, 53	0.657		0.016
	Immersion Group × Filler Plausibility	0.92	2, 53	0.405		0.034
	Dependency Length × Filler Plausibility	5.801	1, 53	0.020	*	0.099

2. Follow-up pairwise Comparisons					
Comparison (Model)	Condition	*t*	*df*	*p*	*Sig.*
Dependency Length (Proficiency)	Long vs. Short	-3.66	49	.001	**
Dependency Length (Immersion)	Long vs. Short	-3.68	55	.001	**
Plausibility at Short Dependency (Immersion)	Implausible vs. Plausible	-2.97	55	.004	**
Plausibility at Long Dependency (Immersion)	Implausible vs. Plausible	0.824	55	.413	

Note: Only significant effects and relevant non-significant effects are reported. Partial eta-squared (η^{2p}) reflects effect size. Significance levels: *** $p < .001$, ** $p < .01$, * $p <. 05$, **.** $p <. 1$ (marginal)

Critical verb region (Word 13)

In the critical verb region, both models showed significant interactions between Group and Dependency Length (Proficiency: $F(2, 47) = 6.53$, $p = .003$; Immersion: $F(2, 53) = 4.33$, $p = .018$). In both models, the L1 group showed a significant trend where the short dependency was associated with longer reading times compared to the long dependency ($t(19) = -2.21$, $p = .039$). In contrast, all L2 groups showed the opposite pattern, with the long dependency condition associated with longer reading times than the short dependency condition, even though only the L2-High group showed a significant difference between long and short conditions ($t(15) = 3.35$, $p = .004$). Thus, at the critical region, the earliest region where the filler-gap integration can be attempted, L2 speakers appeared to be more sensitive to dependency length, while L1 speakers showed a different pattern.

A main effect of Filler Plausibility was observed. This effect was significant in the Proficiency Model ($F(1, 47) = 5.331$, $p = .025$) but only marginally significant in the Immersion Model ($F(1, 53) = 3.343$, $p = .073$). In both models, plausible items elicited longer reading times (Proficiency Model: $M = 438$ ms, $SE = 17.9$; Immersion Model: $M = 433$ ms, $SE = 16.8$) than implausible items (Proficiency Model: $M = 424$ ms, $SE = 16.8$; Immersion Model: $M = 420$ ms, $SE = 16.0$). The direction of the plausibility effect was unexpected, as plausible items elicited longer reading times than implausible items, contrary to typical findings in the literature where processing implausible dependencies usually required more reading times. To better understand the nature of this unexpected, reversed plausibility effect, follow-up pairwise comparisons were conducted at each dependency length. In the Proficiency Model, the analysis revealed a significant plausibility effect only in the short dependency condition ($t(49) = -3.56$, $p < .001$), with plausible items ($M = 442$ ms, $SE = 16.9$) eliciting longer reading times than implausible items ($M = 415$ ms, $SE = 17.1$). No plausibility effect was found in the long dependency condition ($t(49) = -0.159$, $p = .875$). A similar pattern emerged in the Immersion Model, where the plausibility effect was significant only in the short dependency condition ($t(55) = -3.50$, $p < .001$; plausible: $M = 437$ ms, $SE = 15.7$; implausible: $M = 412$ ms, $SE = 16.1$), but not in the long dependency condition ($t(55) = -0.210$, $p = .834$). In addition, as discussed for the pre-critical region, animacy could be a potential factor causing longer reading times at Word 12, and this effect may have spilled over to Word 13. To examine whether animacy was a confounding factor, we conducted an animacy-controlled analysis using only items where both plausible and implausible fillers were inanimate. The results showed that in both the Proficiency Model and Immersion Model, the main effect of Filler Plausibility was no longer significant (Proficiency Model: $F(1, 47) = 2.566$, $p = .116$; Immersion Model: $F(1, 53) = 0.626$, $p = .432$). Furthermore, no interaction between Filler Plausibility and Dependency Length was observed in either model (Proficiency Model: $F(1, 47) = 1.310$, $p = .258$; Immersion Model: $F(1, 53) = 2.435$, $p = .125$). The further investigation building on the reversed plausibility effect in the main analysis may have been driven by the interfering influence of filler animacy rather than the plausibility itself.

The lack of significant three-way interactions in both models suggested that the relationship between Dependency Length and Filler Plausibility remained consistent within each group. Table 4.4 shows the ANOVA results, the follow-up pairwise comparisons and the animacy-controlled ANOVA for Word 13 reading times.

Table 4.4 ANOVA results, the follow-up pairwise comparisons and the animacy-controlled ANOVA for Word 13 reading times

1. Omnibus ANOVA results						
	Effect	*F*	*df*	*p*	*Sig.*	η^{2p}
Proficiency Model	Proficiency Group	1.741	2, 47	0.186		0.069
	Dependency Length	1.73	1, 47	0.195		0.036
	Filler Plausibility	5.331	1, 47	0.025	*	0.102
	Proficiency Group × Dependency Length	6.527	2, 47	0.003	**	0.217
	Proficiency Group × Filler Plausibility	0.277	2, 47	0.760		0.012
	Dependency Length × Filler Plausibility	1.922	1, 47	0.172		0.039
Immersion Model	Immersion Group	2.88	2, 53	0.065	.	0.098
	Dependency Length	0.753	1, 53	0.389		0.014
	Filler Plausibility	3.343	1, 53	0.073	.	0.059
	Immersion Group × Dependency Length	4.33	2, 53	0.018	*	0.14
	Immersion Group × Filler Plausibility	0.881	2, 53	0.420		0.032
	Dependency Length × Filler Plausibility	1.737	1, 53	0.193		0.032

2. Follow-up pairwise Comparisons					
Comparison (Model)	Condition	*t*	*df*	*p*	*Sig.*
Dependency Length – L1 (Proficiency)	Long vs. Short	-2.21	19	.039	*
Dependency Length – L1 (Immersion)	Long vs. Short	-2.21	19	.039	*
Dependency Length – L2-High (Proficiency)	Long vs. Short	3.35	15	.004	**
Filler Plausibility in Short Dependency (Prof.)	Implausible vs. Plausible	-3.56	49	<.001	***
Filler Plausibility in Short Dependency (Im.)	Implausible vs. Plausible	-3.50	55	<.001	***

3. Animacy-controlled ANOVA						
	Effect	*F*	*df*	*p*	*Sig.*	η^{2p}
	Filler Plausibility	2.566	1, 47	0.116		0.052
Proficiency Model	Proficiency Group × Filler Plausibility	0.504	2, 47	0.608		0.021
	Dependency Length × Filler Plausibility	1.310	1, 47	0.258		0.027
	Filler Plausibility	0.626	1, 53	0.432		0.012
Immersion Model	Immersion Group × Filler Plausibility	2.138	2, 53	0.128		0.075
	Dependency Length × Filler Plausibility	2.435	1, 53	0.125		0.044

Note: Only significant effects and relevant non-significant effects are reported. Partial eta-squared (η^{2p}) reflects effect size. Significance levels: *** $p < .001$, ** $p < .01$, * $p <. 05$, **.** $p <. 1$ (marginal)

The first post-critical regions (Word 14)

In the first post-critical region (Word 14), both models revealed a main effect of Dependency Length (Proficiency Model: $F(1,47) = 32.915$, $p < .001$; Immersion Model: $F(1,53) = 15.299$, $p < .001$), with longer reading times in short dependency conditions (Proficiency Model: $M = 428$ ms, $SE = 14.3$; Immersion Model: $M = 422$ ms, $SE = 13.4$) compared to long dependency conditions (Proficiency Model: $M = 399$ ms, $SE = 14.2$; Immersion Model: $M = 396$ ms, $SE = 13.7$). Importantly, there was a significant interaction between Dependency Length and Proficiency Group ($F(2,47) = 3.571$, $p = .036$), suggesting that L1 and L2 speakers processed dependencies differently at this region. Follow-up analyses revealed that the dependency length effect was significant for both L1 speakers ($t(19) = -5.04$, $p < .001$) and L2-Low learners ($t(13) = -4.71$, $p < .001$), but not for L2-High learners ($t(15) = -1.10$, $p = .290$). The Immersion Model showed marginal significance for the interaction between Dependency Length and Immersion Group ($F(2,53) = 2.705$, $p = .076$). Despite this, follow-up analyses were conducted, and the results aligned with those from the Proficiency Model. Both L1 speakers ($t(19) = -5.04$, $p < .001$) and L2-Short immersion learners ($t(23) = -4.09$, $p < .001$) showed significant dependency length effects, while L2-Long immersion learners did not ($t(11) = 0.890$, $p = .890$).

The main effect of Filler Plausibility was not significant in either model (Proficiency Model: $F(1,47) = 0.071$, $p = .791$; Immersion Model: $F(1,53) = 0.083$, $p = .775$). There was only a marginally significant interaction between Filler Plausibility and Dependency Length in the Immersion Model ($F(1,53) = 2.949$, $p = .092$) but not in the Proficiency Model ($F(1,47) = 1.263$, $p = .267$). Further, there was no interaction between Filler Plausibility and Group (Proficiency Model: $F(2,47) = 0.926$, $p = .403$; Immersion Model: $F(2,53) = 1.164$, $p = .320$). The lack of significant three-way interactions in both models suggested that the relationship between Dependency Length and Filler Plausibility remained consistent within each group. Table 4.5 shows the ANOVA results and the follow-up pairwise comparisons for Word 14 reading times.

Table 4.5 ANOVA results and the follow-up pairwise comparisons for Word 14 reading times

1. Omnibus ANOVA results						
	Effect	*F*	*df*	*p*	*Sig.*	η^{2p}
Proficiency Model	Proficiency Group	0.528	2, 47	0.593		0.022
	Dependency Length	32.915	1, 47	< .001	***	0.412
	Filler Plausibility	0.071	1, 47	0.791		0.002
	Proficiency Group × Dependency Length	3.571	2, 47	0.036	*	0.132
	Proficiency Group × Filler Plausibility	0.926	2, 47	0.403		0.038
	Dependency Length × Filler Plausibility	1.263	1, 47	0.267		0.026
Immersion Model	Immersion Group	2.016	2, 53	0.143		0.071
	Dependency Length	15.299	1, 53	< .001	***	0.224
	Filler Plausibility	0.083	1, 53	0.775		0.002
	Immersion Group × Dependency Length	2.705	2, 53	0.076	.	0.093
	Immersion Group × Filler Plausibility	1.164	2, 53	0.32		0.042
	Dependency Length × Filler Plausibility	2.949	1, 53	0.092	.	0.053

2. Follow-up pairwise Comparisons					
Comparison (Model)	Condition	*t*	*df*	*p*	*Sig.*
Dependency Length – L1 (Proficiency)	Long vs. Short	-5.04	19	<.001	***
Dependency Length – L2-High (Proficiency)	Long vs. Short	-1.10	15	.290	
Dependency Length – L2-Low (Proficiency)	Long vs. Short	-4.71	13	<.001	***
Dependency Length – L1 (Immersion)	Long vs. Short	-5.04	19	<.001	***
Dependency Length – L2-Long (Immersion)	Long vs. Short	-0.14	11	.890	
Dependency Length – L2-Short (Immersion)	Long vs. Short	-4.09	23	<.001	***

Note: Only significant effects and relevant non-significant effects are reported. Partial eta-squared (η^{2p}) reflects effect size. Significance levels: *** $p < .001$, ** $p < .01$, * $p <. 05$, **.** $p <. 1$ (marginal)

The second post-critical regions (Word 15)

In the second post-critical region (Word 15), both models showed a significant main effect of Dependency Length (Proficiency Model: $F(1,47) = 11.357$, $p = .002$; Immersion Model: $F(1,53) = 7.741$, $p = .007$), with participants reading more slowly for short dependency items (Proficiency Model: $M = 386$ ms, $SE = 11.1$; Immersion Model: $M = 383$ ms, $SE = 10.4$) compared to long dependency items (Proficiency Model: $M = 373$ ms, $SE = 11.5$; Immersion Model: $M = 370$ ms, $SE = 11.1$).

The main effect of Filler Plausibility was not significant in either model (Proficiency Model: $F(1,47) = 0.867$, $p = .357$; Immersion Model: $F(1,53) = 1.611$, $p = .210$). However, there was a significant interaction between Filler Plausibility and Proficiency Group ($F(2,47) = 4.890$, $p = .012$) and between Filler Plausibility and Immersion Group ($F(2,53) = 3.725$, $p = .031$). Follow-up pairwise analyses revealed that in both Proficiency and Immersion Models, only the L1 group demonstrated a significant plausibility effect ($t(19) = 2.54$, $p = .020$), with longer reading times for implausible fillers. This effect was not significant for either L2 group.

Interestingly, there was a significant interaction between Dependency Length and Filler Plausibility in both analyses (Proficiency Model: $F(1,47) = 6.566$, $p = .014$; Immersion Model: $F(1,53) = 4.891$, $p = .031$), suggesting that the effect of plausibility differed between long and short dependencies. In both Models, follow-up analyses showed that in long dependencies, implausible fillers elicited significantly longer reading times than plausible fillers (Proficiency Model: $t(49) = 2.42$, $p = .019$; Immersion Model: $t(55) = 2.45$, $p = .017$), but this was not significant in short dependencies (Proficiency Model: $t(49) = -1.15$, $p = .255$; Immersion Model: $t(55) = -1.06$, $p = .294$). Given the presence of plausibility effects only in long dependencies, separate two-way ANOVAs for long dependencies were conducted for Word 15 to examine how different groups process plausibility when the dependency became long. Group was taken as the between-subjects factor and Filler Plausibility was taken as the within-subjects factor. In the long dependency condition, both models showed a significant main effect of Filler Plausibility (Proficiency Model: $F(1,47) = 4.802$, $p = .033$; Immersion Model: $F(1,53) = 4.670$, $p = .035$). Neither model showed a significant interaction between Group and Filler Plausibility (Proficiency Model: $F(2,47) = 1.268$, $p = .291$; Immersion Model: $F(2,53) = 0.791$, $p = .459$), suggesting that participants across different groups responded to the plausibility effect similarly. Following the two-way ANOVAs, a closer examination into the group-specific mean reading times revealed different trends between proficiency and immersion groups. In the Proficiency Model, native speakers showed longer reading times for implausible (M = 381 ms, SE = 24.1 ms) than plausible conditions (M = 356 ms, SE = 19.7 ms), reaching marginal significance ($t(19) = 1.97$, $p = .063$). High-proficiency L2 speakers showed a similar trend for implausible (M = 386 ms, SE = 23.3 ms) and plausible conditions (M = 367 ms, SE = 19.2 ms). However, low-proficiency L2 speakers showed identical mean reading times for implausible (M = 374 ms, SE = 17.3 ms) and plausible conditions (M = 374 ms, SE = 18.2 ms). In the Immersion Model, native speakers' mean reading times remained unchanged. L2-Long immersion speakers showed similar reading times for implausible (M = 342 ms, SE = 33.5 ms) and plausible conditions (M = 336 ms, SE = 28.7 ms), while L2-Short immersion speakers showed longer reading times for implausible (M = 394 ms, SE = 12.2 ms) than plausible conditions (M = 381 ms, SE = 11.9 ms).

The three-way interactions between Dependency Length, Filler Plausibility, and Proficiency/Immersion Group were not significant. Table 4.6 shows the ANOVA results, the follow-up pairwise comparisons, and the separate two-way ANOVAs for Word 15 reading times.

Table 4.6 ANOVA results, the follow-up pairwise comparisons and the animacy-controlled ANOVA for Word 15 reading times

1. Omnibus ANOVA results						
	Effect	*F*	*df*	*p*	*Sig.*	η^{2p}
Proficiency Model	Proficiency Group	0.035	2, 47	0.966		0.001
	Dependency Length	11.36	1, 47	0.002	**	0.195
	Filler Plausibility	0.867	1, 47	0.357		0.018
	Proficiency Group × Dependency Length	2.014	2, 47	0.145		0.079
	Proficiency Group × Filler Plausibility	4.890	2, 47	0.012	*	0.172
	Dependency Length × Filler Plausibility	6.566	1, 47	0.014	*	0.123
Immersion Model	Immersion Group	1.835	2, 53	0.170		0.065
	Dependency Length	7.741	1, 53	0.007	**	0.127
	Filler Plausibility	1.611	1, 53	0.210		0.029
	Immersion Group × Dependency Length	0.846	2, 53	0.435		0.031
	Immersion Group × Filler Plausibility	3.725	2, 53	0.031	*	0.123
	Dependency Length × Filler Plausibility	4.891	1, 53	0.031	*	0.028

2. Follow-up pairwise Comparisons					
Comparison (Model)	Condition	*t*	*df*	*p*	*Sig.*
Plausibility Effect – L1 (Proficiency)	Implausible vs. Plausible	2.54	19	0.020	*
Plausibility Effect – L2-High (Proficiency)	Implausible vs. Plausible	0.14	15	0.891	
Plausibility Effect – L2-Low (Proficiency)	Implausible vs. Plausible	-1.94	13	0.074	.
Plausibility Effect – L1 (Immersion)	Implausible vs. Plausible	2.54	19	0.020	*
Plausibility Effect – L2-Long (Immersion)	Implausible vs. Plausible	-0.01	11	0.991	
Plausibility Effect – L2-Short (Immersion)	Implausible vs. Plausible	-0.96	23	0.347	
Plausibility Effect in Long Dependency (Proficiency)	Implausible vs. Plausible	2.42	49	0.019	*
Plausibility Effect in Short Dependency (Proficiency)	Implausible vs. Plausible	-1.15	49	0.255	
Plausibility Effect in Long Dependency (Immersion)	Implausible vs. Plausible	2.45	55	0.017	*
Plausibility Effect in Short Dependency (Immersion)	Implausible vs. Plausible	-1.06	55	0.294	

3. Two-way ANOVAs in the long dependency condition						
	Effect	*F*	*df*	*p*	*Sig.*	η^{2p}
Proficiency Model	Proficiency Group	0.044	2, 47	0.957		0.002
	Filler Plausibility	4.8	1, 47	0.033	*	0.008
	Proficiency Group × Filler Plausibility	1.27	2, 47	0.291		0.004

Immersion Model	Immersion Group	1.37	2, 53	0.263		0.046
	Filler Plausibility	4.67	1, 53	0.035	*	0.007
	Immersion Group × Filler Plausibility	0.79	2, 53	0.459		0.002

Note: Only significant effects and relevant non-significant effects are reported. Partial eta-squared (η^{2p}) reflects effect size. Significance levels: *** $p < .001$, ** $p < .01$, * $p <. 05$, **.** $p <. 1$ (marginal)

Regression analyses in the long dependency condition

To address the third question, I ran a series of linear regressions in the L2 group with proficiency and YoI as continuous variables. As the previous ANOVAs showed, the expected plausibility effect was observed only in the long dependency condition, as the plausibility effect might be masked by the potential subject filled-gap effect in the short conditions. The linear regression was only done in the long condition. The analyses consistently showed that including YoI significantly improved model fit across all word regions. In the pre-critical region (Word 12), Model 2, which included YoI, explained significantly more variance than Model 1 ($R^2 = 0.081$ vs. $R^2 = 0.009$; $F(4, 544) = 10.66, p < .001$). This pattern was consistent across the critical region (Word 13: $R^2 = 0.069$ vs. $R^2 = 0.011$; $F(4, 549) = 8.593, p < .001$), the first post-critical region (Word 14: $R^2 = 0.045$ vs. $R^2 = 0.007$; $F(4, 554) = 5.568, p < .001$), and the second post-critical region (Word 15: $R^2 = 0.058$ vs. $R^2 = 0.006$; $F(4, 550) = 7.620, p < .001$). In brief, the inclusion of YoI in Model 2 resulted in substantial increases in explained variance, ranging from 3.8% in the first post-critical region to 7.2% in the pre-critical region. This consistent improvement suggested that YoI played a significant role in explaining reading time variability during filler-gap processing, beyond what is accounted for by proficiency alone. Figures 4.3 illustrates how Proficiency and YoI respectively correlate with reading times. The plots highlight the general trend of reduced reading times with increasing proficiency and immersion. Figure 4.4 compares the R-square values of Model 1 (Proficiency only) and Model 2 (with YoI added) across word regions. Complete regression results for the long dependency condition are provided in the Appendix D.

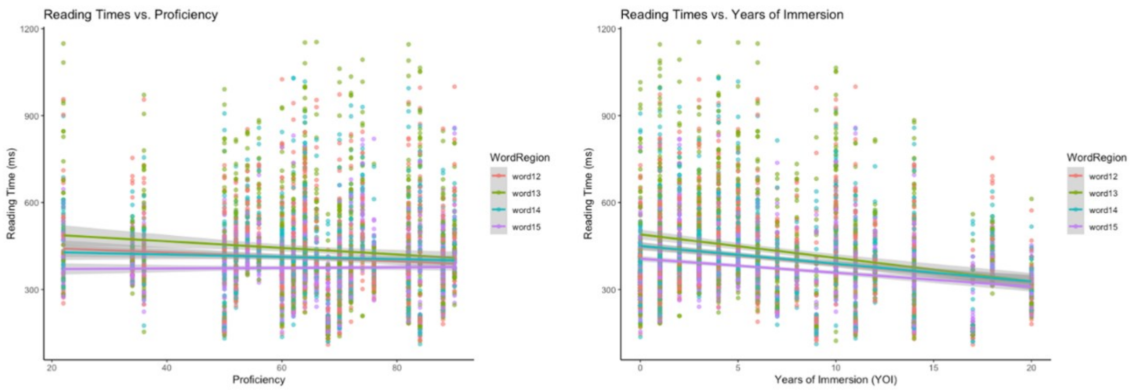


Figure 4.3 Correlations of Proficiency and Years of Immersion (YoI) with reading times

Figure 4.4 R-squared comparison between Model 1 and Model 2 across word regions

However, it is important to note that while the inclusion of YoI improved the model's ability to predict overall reading times, neither proficiency nor YoI showed specific effects on plausibility processing. The significant improvement in model fit appeared to reflect YoI's contribution to explaining general reading speed patterns, rather than sensitivity to plausibility. The interaction between filler plausibility and proficiency was not significant across regions, and similarly, the interaction between filler plausibility and YoI showed no consistent pattern. The three-way interaction (Plausible × Proficiency × YoI) was also not significant across regions, suggesting that neither proficiency nor immersion, alone or in combination, clearly modulated sensitivity to plausibility, despite their contribution to explaining general reading time patterns. Figures 4.5 illustrates the interactions of Filler Plausibility with Proficiency and YoI, respectively. These plots demonstrate that neither predictor significantly modulates sensitivity to plausibility.

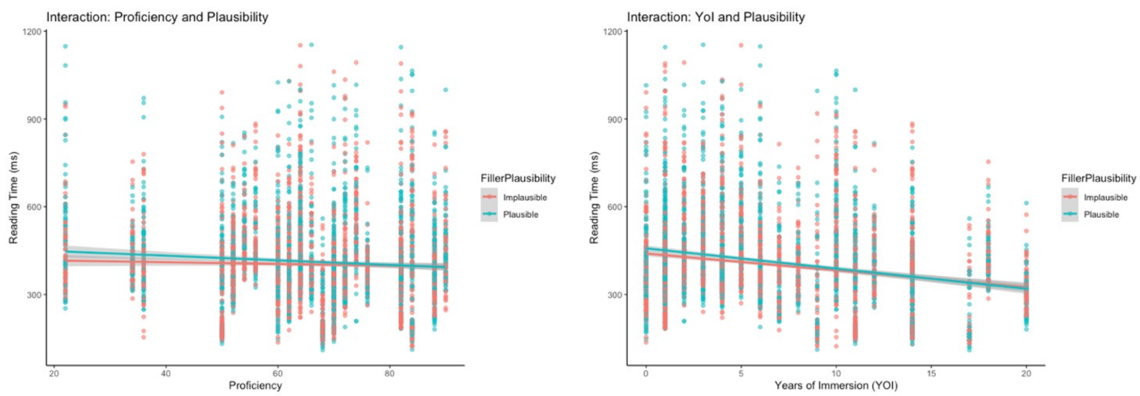


Figure 4.5 Interaction effects of Proficiency and YoI with Filler Plausibility

4.2.7 Discussion

The research objective of Experiment 1 was to examine whether L1 and L2 speakers use similar mechanisms to process wh-dependencies with the manipulation of dependency length and filler plausibility, and whether L2 speakers' parsing mechanisms would be modulated by proficiency and YoI.

Active Filler Strategy

According to Figure 4.2, the results were suggestive of a potential active gap-filling strategy, as evidenced by increased reading time at the verb region, which occurred before the actual gap, particularly in the long dependencies. Although the increased reading times at the verb alone could not definitively suggest active gap filling due to the lack of a non-gap control condition, when considered together with the delayed plausibility effect at the second post-critical region, these results provided evidence supporting the Active Filler Strategy, consistent with previous studies (e.g., William et al., 2001; Wagers & Phillips, 2014). According to the mean reading time graphs in Figure 4.2, L1 speakers showed a slight increase in reading time from the pre-critical to the critical region in both conditions, while L2 speakers demonstrated a more pronounced increase, particularly in the long condition. The steeper slope for L2 speakers between these regions suggested that they had greater sensitivity to dependency length during active gap-filling. A plausible interpretation is that when parsers encountered a verb, which signaled the approaching of a syntactic gap, they started to actively resolve the dependency. However, this integration process seemed more demanding for L2 speakers, particularly when the distance between the filler and the verb was long, which led to more pronounced increases in reading time from the pre-critical to the critical word.

Despite the increased reading time at the critical verb region, the statistical model did not show a main effect of filler plausibility with implausible fillers being processed with longer time, nor significant interactions between group and filler plausibility at the critical word. These findings are compatible with Wagers and Phillips's (2014) dual-process account. Under this framework, the observed reading time elevation at the critical region was likely driven by maintained syntactic information – such as the postulation of a syntactic gap or an empty category linked to the filler. In contrast, the plausibility of the filler-verb combination may not yet have been retrieved at this stage and thus would not influence processing until later in the sentence. In addition, the results at the critical verb were consistent with the updated Shallow Structure Hypothesis (Clahsen & Felser, 2018), which suggests that L2 speakers are in fact capable of using syntactic representations during real-time processing but less effectively. The two accounts could explain why L2 speakers showed greater difficulty at the verb region: syntactic information was applied first for active gap-filling, but L2 speakers tended to have limited syntactic information maintained in memory to trigger filler integration.

Dependency length effect

Among the three research questions I intended to address in Experiment 1, the first research question focused on whether dependency length (long vs. short) affected the processing of filler-gap dependencies in L1 and L2 English speakers. Contrary to traditional theories, which predict increased processing difficulty for longer dependencies (Gibson, 1998, 2000), the present study revealed an unexpected pattern: participants, at all four regions, spent more time reading in short dependency conditions compared to long dependency conditions. This effect was particularly pronounced in the pre-critical and critical regions but persisted to the two post-critical regions as well. This finding might be explained by the subject filled-gap effect, a phenomenon where readers experience processing difficulty when encountering an unexpected noun phrase in the subject position that could potentially be the gap site for a preceding filler (e.g. Stowe, 1986; Lee, 2004). In my experimental design, the pre-critical word in the short dependence sentences occupied the subject gap position, whereas the pre-critical word in the long dependency sentences did not. See Table 4.2 again for clarification. For example, in the Short-Plausible condition, after the filler *which customer*, there are two possible gap positions: the subject gap position filled by *the receptionist* and the object gap position occupied by an empty element after *called*. The observed reading time slowdown at the precritical region in the short conditions could reflect parsers' surprise when encountering the filled subject gap *the receptionist*, resulting in increased processing difficulty.

Table 4.2 Sample experimental stimuli for Experiment 1

Background sentence: The coffee shop was very busy today, and the staff hardly had any time to rest. (After the last table left, …)		
Dependency length	Filler plausibility	Sentence
Short	Plausible	After the last table left, the manager knew *which customer* the receptionist *called* ___ about the problem.
Short	Implausible	After the last table left, the manager knew *which spoon* the receptionist *called* ___ about the problem.
Long	Plausible	the manager knew *which customer* the receptionist with a large cloth bag *called* ___ about the problem.
Long	Implausible	the manager knew *which spoon* the receptionist with a large cloth bag *called* ___ about the problem.
Yes/No question		Was the coffee shop busy today?
Answer		Yes

This unexpected longer reading times due to the subject filled-gap effect was not observed in Stowe (1986), and the subject gap was not explicitly measured in Wagers and Phillips (2014). Stowe (1986) reported filled-gap effects only in object noun phrase positions, not in subject gap positions. The present study closely followed the design of Wagers and Phillips (2014), which manipulated dependency length by inserting a prepositional phrase between the filler and the verb, as illustrated in (38) and reproduced here as (39) for convenience. Based on the findings of the present study, in the short-condition sentences in (39), readers might have initially anticipated a subject gap after *which*. This expectation aligns with the Active Filler Strategy (Frazier & Clifton, 1989), which posits that the parser seeks to associate a filler with the earliest possible gap position. When encountering a noun phrase that fills the subject position, it is reasonable to expect a reading time slowdown. For example, after encountering *the acorns which*, parsers might predict a structure like "The acorns which ___ fell from the tree...," where the gap is in the subject position. Upon finding that the subject gap is occupied by a noun phrase (e.g., *the squirrel*), a reading time slowdown would likely occur at *the squirrel*. This slowdown may also spill over to subsequent regions. However, no such effect was reported in Experiment 2 of Wagers and Phillips (2014). A plausible explanation is that, in their stimuli, an adverb (*quickly*) was inserted between the relative clause subject (*the squirrel*) and the verb. This adverb may have acted as a buffer, mitigating the processing difficulty typically associated with the filled subject gap. The present study, in contrast, lacked such a buffer, and this may explain the observed slowdown in reading times starting from the pre-critical region for the short dependency conditions.

(39) a. Short – Plausible:

The acorns which the squirrels quickly crammed their small puffy cheeks with ___ before scurrying out of the park had fallen from a sickly oak.

b. Short – Implausible:

The cats which the squirrels quickly crammed their small puffy cheeks with ___ before scurrying out of the park were basking in the sun.

c. +PP – Plausible:

The acorns which the squirrels with the bushy black tails quickly crammed their small puffy cheeks with ___ before scurrying out of the park had fallen from a sickly oak.

d. +PP - Implausible:

The cats which the squirrels with the bushy black tails quickly crammed their small puffy cheeks with ___ before scurrying out of the park were basking in the sun.

e. +CP – Plausible:

The acorns which the experienced naturalist concluded that the squirrels quickly crammed their small puffy cheeks with ___ before scurrying out of the park had fallen from a sickly oak.

f. +CP – Implausible:

The cats which the experienced naturalist concluded that the squirrels quickly crammed their small puffy cheeks with ___ before scurrying out of the park were basking in the sun.

(Wagers & Phillips, 2014, p. 1288)

Although it was not the primary research question, the unexpected reading time pattern found in this experiment may suggest that both L1 and L2 readers engaged in predictive processing during filler-gap dependency resolution. Specifically, the longer reading times in the short dependency conditions could potentially be explained by the subject filled-gap effect. This interpretation aligns with other studies on filler-gap sentence processing that emphasize the role of prediction (e.g., Johnson 2015, Yan & Liu, 2021; Lee, 2004). For instance, Johnson (2015) examined how Korean learners of English process *wh*-dependencies, focusing on the subject filled-gap effect that happens before the occurrence of the verb. The study found a notable subject filled-gap effect when the potential subject gap immediately followed the filler, as the design of the sentences of the short dependency condition in my study, and this effect was observed in both native English speakers and Korean learners of English. This finding suggested that both L1 and L2 speakers predictively processed subject gaps. Similarly, Yan & Liu (2021) reported subject filled-gap effects in Mandarin learners of English when the filler and the subject gap were adjacent.

While the subject filled-gap effect is a plausible explanation for the findings in the present study, it remains speculative without direct evidence. However, if this explanation holds, it would provide further evidence for the idea that L2 speakers can predictively resolve *wh*-dependencies using syntactic information, rather than relying solely on lexical-semantic cues from the verb. Furthermore, the possibility that L2 speakers engage in syntax-driven predictive processing suggests that the cognitive mechanisms underlying predictive processing may be fundamentally similar across native and non-native speakers, as proposed by Kaan (2014). If confirmed, these findings would cast doubt on the common misconception of the Shallow Structure Hypothesis (Clahsen & Felser, 2006), which claims that L2 learners are unable to utilize abstract syntactic information in real-time language processing.

An intriguing finding regarding dependency length was that while L1 and both groups of L2 speakers showed elevated reading times at the subject gap filled by a noun phrase in the short dependency length conditions, different patterns emerged when encountering the critical verb. In the critical verb region, the L1 group showed significantly longer reading times in the short dependency condition compared to the long dependency condition. In contrast, L2 speakers demonstrated longer reading times for long dependencies compared to short ones, with this pattern being most pronounced among highly proficient L2 speaker. This observation can be explained by the following two theories. First, the findings can be interpreted by Wagers and Phillips' (2014) dual-process account, which suggests that filler-gap processing involves two processes: a leading process that relies on maintained syntactic cues, and a lagging process that addresses the filler's semantic information. The increased reading time observed at the critical word region likely stems from the parser's use of stored syntactic information, while semantic details remain unaddressed. As the distance between the filler and the gap increases, the strength of the syntactic cues that drive active gap-filling diminishes in L2 as linear distance between the filler and the gap require working memory resources (Gibson, 1998, 2000). Consequently, L2 speakers may experience greater challenges when confronted with long dependency structures, resulting in increased reading times. This interpretation is consistent with the updated Shallow Structure Hypothesis (Clahsen & Felser, 2018), which suggests that L2 speakers are capable of using syntactic representations during real-time processing, but they do so less effectively than L1 speakers.

However, this account alone does not explain why the dependency length effect was more pronounced in high-proficiency L2 speakers than in low-proficiency L2 speakers. If lower-proficiency learners maintain even less syntactic information in memory, it would be reasonable to expect them to show greater, not lesser, difficulty with long dependencies. The observed pattern instead suggests that lower-proficiency learners may not be fully engaging in structure-driven gap-filling at all, making them less sensitive to the dependency length.

To further account for the findings about the dependency length effect, a second explanation can be drawn from the Direct Association Hypothesis. This hypothesis highlights the role of lexical information—particularly verb subcategorization—in processing filler-gap dependencies. It is important to note that the occurrence of the critical verb marks the first point at which the direct association between the verb and the filler becomes available. A possible explanation for the less pronounced dependency length effect observed in low-proficiency L2 speakers is that verb subcategorization information facilitates the integration of the filler when the verb is encountered. For low-proficiency L2 speakers, establishing a filler-gap dependency may still be highly lexically driven, triggered upon reaching the subcategorizing verb and recognizing its argument structure. The filler-verb association, once formed at the verb, may support the establishment of the dependency and ease processing with low-proficiency L2 speakers. As a result, the increased linear distance in long dependencies imposes less additional processing cost on low-proficiency L2 speakers compared to high-proficiency L2 speakers.

In summary, L2 speakers' greater sensitivity to dependency length at the critical verb likely results from two factors: weaker maintained syntactic information, making active gap filling more challenging, and greater reliance on verb-subcategorization, requiring additional cognitive resources.

Plausibility effect

The second research question of the current experiment examined whether filler plausibility (plausible vs. implausible) affected the processing of filler-gap dependencies in L1 and L2 English speakers. The discussion of the filler plausibility effect focuses on two aspects: 1) the unexpected reversed plausibility effect, where plausible fillers were processed more slowly than implausible ones at the pre-critical and critical regions, and 2) the typical plausibility effect, characterized by longer reading times for implausible fillers.

At both the pre-critical and critical regions, an unexpected pattern emerged: longer reading times for plausible fillers compared to implausible ones. This reversed plausibility effect appeared primarily in the short dependency conditions, supported by the following findings. First, follow-up pairwise comparisons revealed that the significant reversed plausibility effect occurred only in short dependency conditions, while long dependency conditions showed no reading time differences. Second, trend lines in Figure 4.2 showed that in short dependency conditions, plausible fillers were read more slowly than implausible ones. Third, when animacy was controlled by only considering inanimate fillers in both plausible and implausible conditions, this unexpected, reversed plausibility effect disappeared. These findings suggested that the reversed plausibility effect likely reflected a stronger subject filled-gap effect triggered by animate fillers in the plausible conditions under short dependency conditions. This interpretation is in line with Wagers and Pendleton's (2016) study, which investigated how filler animacy affects the processing of filler-gap dependencies. They conducted two self-paced reading experiments using a filled-gap design. In both experiments, they manipulated two factors: the animacy of the filler (animate vs. inanimate) and the filler category (NP vs. PP). They found a significant filled-gap effect, reflected by slower reading times, at the relative clause subject position, but only when the filler was animate. Wagers & Pendleton (2016) concluded that the filled-gap effect in the subject position was stronger for animate fillers than for inanimate ones. Reflecting on the current experiment, out of 32 experimental sets, 16 sets had both plausible and implausible fillers being inanimate, which likely would not cause differing levels of subject filled-gap effects. However, in the remaining sets, 12 had plausible fillers being animate and implausible fillers being inanimate, while 4 had the reversed pattern. This imbalance may have led participants to experience greater surprise in the plausible conditions where animate fillers were more likely to be considered as the subject of the clause, leading to longer reading times for plausible fillers.

The expected plausibility effect, characterized by processing difficulty when the filler was a semantically implausible direct argument, was found in the second post-critical region. The overall analyses revealed significant interactions between Filler Plausibility and both Proficiency and Immersion Groups. Follow-up pairwise analyses showed that only L1 English speakers showed a significant plausibility effect, with longer reading times for implausible fillers. The initial Plausibility × Group interaction might lead to a premature conclusion that L2 speakers lacked sensitivity to plausibility and thus would suggest qualitative L1-L2 difference in processing. However, the presence of Dependency Length × Filler Plausibility interaction prompted further analyses and prevented us from concluding qualitative L1-L2 differences. Follow this interaction, pairwise analyses showed that implausible fillers led to significantly longer reading times only in long dependencies, while no such effect was found in short dependencies. When analyzing long dependencies separately, all groups showed similar sensitivity to plausibility, as evidenced by the significant main effect of Filler Plausibility and the absence of interaction between Group and Filler Plausibility.

In the present study, both L1 and L2 speakers exhibited reading time slowdown at the critical word, suggesting the use of the Active Filler Strategy. The observation of the filler plausibility effect only in the second post-critical region, even among native speakers, suggested that initial gap-filling was primarily driven by syntactic information, with semantic integration occurring after initial syntactic parsing. Because of the main effect of Filler Plausibility, along with the absence of Filler Plausibility and Group interaction, it is plausible that both native and second language speakers were capable of applying syntactic cues to resolve filler-gap dependencies. This pattern was consistent with the study of Traxler and Pickering (1996), who found that readers initially postulated gaps at the earliest possible position, regardless of plausibility, and only later showed sensitivity to semantic fit. The delayed plausibility effect also supported the dual-process account proposed by Wagers and Phillips (2014). Furthermore, it provided evidence for the Delayed Integration Hypothesis (Wagers & Phillips, 2009), which suggests that while initial syntactic dependencies are formed rapidly, semantic integration may be delayed. Taken together, these findings contributed to a broader understanding of sentence processing, lending support to the "syntax-first" models, which posit a distinction between initial structure building and subsequent semantic interpretation (e.g., Friederici, 2002, p. 78). Importantly, the finding that semantic integration occurred only after initial syntactic parsing challenged alternative theories of sentence processing. Specifically, it cast doubt on the interactive account or constraint-based model, which suggests that multiple sources of information, including both syntactic and semantic cues, are processed in parallel from the beginning of sentence parsing (e.g., Trueswell et al., 1994; Clifton, 2003). The clear delay in semantic integration observed in our study was inconsistent with the simultaneous processing of syntactic and semantic information proposed by these models.

As discussed in Chapter 2, Phillips and Ehrenhofer (2015) identified a crucial gap in research supporting or opposing the SSH. They claimed that most studies focused on whether L2 learners are sensitive to structural cues used by native speakers during parsing, but they did not adequately address whether L2 learners prioritize semantic information over structural cues. The current study addressed this gap. The findings revealed that L2 learners were sensitive to both structural and semantic cues: potential subject filled-gap and active gap-filling effects (structural) were observed at the critical region, while plausibility effects (semantic) emerged later in the second post-critical region. This delay in semantic integration suggested that L2 learners did not simply prioritize semantic information over structural cues, as being predicted by the Shallow Structure Hypothesis. Instead, the findings suggested that in L2 processing of filler-gap dependencies, structural information was processed rapidly, similar to L1 speakers, while semantic integration occurred at a later stage.

The effect of L2 proficiency and years of immersion

The third research question investigated whether L2 proficiency and years of immersion (YoI) modulated the processing of dependency length and filler plausibility in filler-gap dependency processing for L2 learners.

On dependency length

Regarding the dependency length, at the critical verb region, both L2 proficiency groups and L2 immersion groups showed longer reading times for long dependencies. Higher proficiency L2 speakers exhibited greater sensitivity to dependency length, as evidenced by the significant effect in the L2-High group. Lower proficiency L2 speakers showed a similar trend, but the effect was not statistically significant. Neither immersion group reached statistical significance. Interestingly, L1 speakers showed a contrasting pattern, suggesting shorter reading times for long dependencies. As discussed previously, active gap-filling appears to be driven by the syntactic information maintained in the working memory. The contrasting pattern in L1 speakers may be explained by the relatively strong syntactic information maintained, even in longer dependency conditions. While filler integration tended to result in longer reading times at the critical verb region compared to the pre-critical region in both short and long conditions, resolving the filler-gap at the verb likely imposed a lower processing load for L1 speakers. Consequently, in the short conditions, the potential subject filled-gap effect may still have a residual influence, leading to the situation where short conditions still had longer reading times in L1 speakers.

Compared to the L1 group, who consistently showed longer reading times at the critical region for short dependency conditions, L2 speakers exhibited a contrasting pattern, with longer reading times associated with long dependency conditions. This raised a question: what factor contributed to the reversal observed in L2 speakers at the critical region and compensated for the potential subject filled-gap effect? Cunnings (2017a, 2017b) argued that while L2 learners can process syntactic structures with similar complexity to L1 speakers, they differ in their ability to access and use previously processed linguistic information. To be specific, when L2 speakers retrieve relevant information from memory, they are more vulnerable to interference from multiple potential antecedents present in memory. This may be particularly evident when the filler and the verb are separated by longer distances. In this case, L2 speakers appear to struggle more with the fading of filler information in working memory, leading to increased reading time when they need to retrieve and integrate filler information at the verb, as predicted by the distance effect (e.g., Gibson, 1998).

Interestingly, the follow-up pairwise comparisons revealed that only the L2-High proficiency group showed a significant effect, with longer reading times for long dependencies compared to short ones. This elevated reading times for long dependencies observed in L2-High proficiency speakers at the critical verb region did not immediately dissipate, and it persisted into Word 14. A possible explanation is that L2 speakers with higher proficiency may allocate greater cognitive resources to processing filler-gap dependencies than low proficiency L2 speakers, who are less capable of constructing such a complex structure. The more effort they invest in maintaining and integrating relevant linguistic information, the more sensitive they become to factors like the distance between filler and verb. However, this interpretation needs further empirical evidence showing relationship between L2 proficiency and cognitive resource allocation during syntactic processing.

In summary, all L2 groups demonstrated longer reading times for long dependencies at the critical verb region, in contrast to L1 speakers who showed shorter reading times for long dependencies. Higher proficiency appeared to be associated with greater sensitivity to dependency length.

On filler plausibility

The expected plausibility effect, characterized by processing difficulty for semantically implausible fillers, was observed only in the second post-critical region. The omnibus ANOVA analyses showed that only L1 speakers demonstrated a significant plausibility effect, exhibiting longer reading times for implausible fillers. However, subsequent analyses of the long dependency condition revealed that all groups, regardless of L2 proficiency of years of immersion (YoI), showed similar sensitivity to filler plausibility. The results were consistent with the Continuity Hypothesis (Hopp, 2010; McDonald, 2006). The similarity in plausibility effects between L2 speakers and L1 speakers suggested that L2 learners may employ the same universal parsing strategies as native speakers.

The mean reading time results from the two-way ANOVAs suggest a potential gradient effect of proficiency on plausibility sensitivity. High-proficiency L2 speakers, in parallel with L1 speakers, showed a trend toward longer reading times for implausible filler integration compared to plausible filler integration. In contrast, low proficiency L2 speakers showed identical reading times for plausible and implausible conditions in the second post critical region. This finding was consistent with the concept of the gradient difference rather than the categorical distinction in L2 syntactic processing in the updated SSH (Clahsen & Felser, 2018). The current experiment suggested that this gradient effect may extend beyond syntactic processing to include semantic processing. However, given the absence of a significant interaction between Filler Plausibility and Group, this observation should be interpreted merely as a descriptive trend, with proficiency being a potential factor in more native-like processing rather than as empirical evidence. Regarding the effect of immersion on the plausibility effect, all immersion groups exhibited a similar pattern to L1 speakers, with implausible fillers being processed with longer reading times than plausible fillers.

The absence of any trend of an immersion effect on plausibility sensitivity was somewhat surprising, given that previous research has suggested that immersion can lead to more native-like processing (e.g., Pliatsikas & Marinis, 2013; Bowden et al., 2013). This may be explained in several ways. First, general language exposure during immersion may improve overall processing speed, but the development of sensitivity to semantic anomalies in filler-gap dependencies may require more targeted parsing strategies. This suggested that while immersion may enhance general language processing, it might not automatically lead to the acquisition of specialized skills needed for complex syntactic structures like filler-gap dependencies. Additionally, the quality, rather than just the quantity, of language exposure during immersion may play an important role (Dussias & Sagarra, 2007). While the present study measured YoI, it did not account for the nature of linguistic interactions during this period. Consequently, participants, despite long periods of immersion, may not have had sufficient exposure to the specific types of complex sentences that would encourage native-like processing of filler-gap dependencies.

The regression analyses consistently showed that including YoI significantly improved model fit across all word regions, leading to notable increases in explained variance (R^2). This increase suggested that YoI played a significant role in explaining reading time variability during filler-gap processing, beyond what was accounted for by proficiency alone. However, interactions between filler plausibility and either YoI or proficiency were not statistically significant, nor was the three-way interaction (Plausible × Proficiency × YoI) across regions. These results suggested that while proficiency and YoI contributed to general reading times during filler-gap processing, neither directly modulated sensitivity to filler plausibility.

In conclusion, the current experiment suggested that L2 speakers, regardless of proficiency level or years of immersion, showed native-like sensitivity to filler plausibility in the second post-critical region. Neither factor directly correlated with filler plausibility processing. Although high-proficiency L2 speakers showed a trend toward more native-like sensitivity to semantically implausible filler integration, this effect remained descriptive rather than statistically significant. Years of immersion did not appear to influence sensitivity to semantic anomalies in filler-gap processing.

4.3 Experiment 2: Filler-Gap processing in L1 Mandarin Chinese

4.3.1 Rationales and predictions

The present experiment aims to investigate the gap-type topic structure in Mandarin Chinese from a processing perspective. Note that Mandarin Chinese topic structures include diverse types (see Huang et al., 2009 for a review). The gap-type topic structures examined in the present experiment specifically refer to those involving on overt gap and having the OSV word order, mirroring the English-like filler-gap structures. The motivation of this research objective is that in many previous studies comparing L1 and L2 sentence processing, the detected parsing differences between native English speakers and Mandarin Chinese learners of English have been attributed to the lack of overt *wh*-movement in the participants' L1 (e.g., Dallas et al., 2013; Dussias & Piñar, 2010; Juff, 2005; Covey et al., 2024). However, it remains a question whether Mandarin Chinese topic structures involve movement like English. While Mandarin Chinese is widely known as a *wh*-in-situ language and several early studies suggest that Mandarin Chinese does not involve A'-movement as English does (e.g., Li & Thompson, 1981; Ning, 1993), some recent studies found evidence for the construction of a filler-gap dependency in real-time processing (overt filler-gap dependency: Yang & Liu, 2014; Dong et al., 2021; Huang & Kaiser, 2008; covert filler-gap: e.g., Xiang, 2014, 2015; Yang et al., 2023).

Specifically, the aims of this experiment are twofold. First, this study aims to investigate whether gap-type topic structures involve movement by examining the presence of the active gap-filling strategies in processing filler-gap dependencies in Mandarin Chinese. Wagers and Phillips (2009, p. 398) suggest that the active gap-filling strategy can be seen as two features of *wh*-dependencies: "unboundedness and indirect evidence of the tail," as this parsing strategy suggests that parsers are prepared to handle the unbounded nature of dependencies and the necessity to connect empty elements across potentially long distances in sentence structures. If Mandarin speakers engage in active gap-filling during real-time sentence processing, this would provide evidence for a movement-based account. In contrast, if they only attempt to establish a referential relationship between a topic and its co-referent without actively positing a gap, the data would be more consistent with a base-generation account. Therefore, one approach to address this aim is to investigate whether native Mandarin Chinese speakers actively search for a gap before the unambiguous gap is confirmed. The second aim is to determine the driving forces behind the active gap-filling strategy in gap-type topic structures, specifically examining whether semantic information, syntactic information, or both guide this process, and to identify when different sources of information become available during processing. According to Wagers and Phillips (2014) and findings from Experiment 1, syntactic information of the filler is maintained and is responsible for triggering the construction of the filler-gap dependency, while semantic information comes into play later as it is not maintained in working memory and has to be retrieved later on. Taken together, the findings are expected to illuminate the mechanisms of processing overt gap-type topic structures in Mandarin Chinese and would explain whether the previous lack of native-like processing in L1 Mandarin Chinese-L2 English speakers comes from the absence of such dependency in Mandarin Chinese.

The Mandarin Chinese experimental sentences appear to have a surface structure similar to those in Experiment 1. The sentences contain a fronted element, such as "After the morning meeting ended, everyone wondered which client the supervisor had told the intern to contact." The key question is whether there is a movement-derived filler-gap dependency involved or the fronted element *which client* simply serves as a base-generated topic followed by a related comment, as proposed by Li and Thompson (1981).

The present self-paced reading experiment employed a 2×2 within-participants design to manipulate two factors: filler plausibility and dependency length. Filler plausibility was varied by selecting NPs that were either plausible or implausible as direct objects of the first verb. While this first verb position presented a potential gap site, it was not the ultimate gap position. Importantly, all sentences were globally plausible because in all conditions the filler was always a legitimate argument of the final verb, which introduced the actual gap. Dependency length was manipulated by introducing PPs between the filler and the gap, thereby increasing the linear distance while maintaining the same structural distance. These intervening phrases were attached to the existing structure without adding significant structural complexity. The core syntactic nodes that needed to be crossed remained the same in both short and long conditions. The specific research questions are as follows:

1. Is there a plausibility effect found at the first verb being encountered?
2. Is reanalysis required when an initially posited gap turns out to be filled by another noun phrase, and is this reanalysis more difficult when the initial integration is plausible compared to when it is implausible?
3. How does the dependency length between the filler and the possible gaps affect the processing of filler-gap dependencies?

Two competing accounts, movement-derived and base-generated, can explain the processing of Mandarin Chinese topic structures, each making distinct predictions that address the three research questions. Regarding the plausibility effect, the base-generation account predicts that parsers would not actively search for gaps but rather employ a bottom-up strategy. Under this account, parsers would wait until encountering the final verb to establish a referential dependency. Therefore, no plausibility effect should be detected before the final verb. The movement account, in contrast, predicts longer reading times for implausible versus plausible fillers at the first verb or shortly thereafter. Following the findings from Experiment 1, if Mandarin Chinese topic structures involve movement-derived dependencies like English, I expect elevated reading times at the first verb compared to the pre-verbal condition, indicating the integration difficulty. If increased reading times at the first verb are observed without plausibility effects, this would support a syntactic-first account, suggesting that active gap-filling is primarily driven by syntactic information. However, if both increased reading times and plausibility effects are detected at the first verb, this would indicate that both syntactic and semantic information guide the active search for potential gaps.

For the second research question concerning the reanalysis effect, the two accounts again make different predictions. Under the base-generation account, no plausibility effects should emerge at the first verb, as parsers would not attempt premature integration of the *wh*-phrase following a bottom-up processing mechanism. Therefore, parsers would simply process the first verb with its following object as a local relationship, and thus no reanalysis effect is expected. Since all of the tested sentences are globally grammatical, no plausibility effect is expected at this position either. The movement account, in contrast, predicts longer reading times for plausible versus implausible fillers at the direct object noun phrase regions, as initially plausible integration would require stronger reanalysis.

Regarding the dependency length effect, the base-generated and movement-derived accounts again make contrasting predictions. Under the base-generation account, the linear distance between the *wh*-phrase and potential gap positions should not affect processing, as no filler-gap dependency needs maintenance in working memory. In contrast, under the movement account, maintaining a filler's syntactic information across longer distances consumes more working memory capacity, making gap-filling more costly (Gibson, 1998, 2000). This increased cost in long dependency conditions would likely be reflected in slower reading times at any potential gap positions in long conditions compared to short conditions. Moreover, the plausibility effect or the reanalysis effect,[22]if detected, should be modulated by dependency length considering the absence

[22] The plausibility effect refers to longer reading times for implausible conditions than for plausible ones at the first verb or shortly after. The reanalysis effect refers to longer reading times for plausible-at-verb than for implausible-at-verb conditions at the filled gap position or shortly after, when the parser revises its initial interpretation. Reanalysis tends to be more effortful when the filler is initially plausible as the verb’s direct object, compared to when the filler-verb pairing is initially implausible.

of plausibility effect in the long dependency in the study of Wagers and Phillips (2014). In short dependencies, I expect robust plausibility effects at or shortly after the first verb or robust reanalysis effect at/shortly after the noun phrase as the filler's semantic features are readily retrievable. In long dependencies, plausibility effects or reanalysis effects might be attenuated due to increased difficulty in retrieving certain information. This interaction would provide evidence for movement-derived processing, as it reflects both the cost of maintaining moved elements and the impact of working memory constraints on filler-gap integration.

In addition to these specific predictions, if Mandarin Chinese topic structures involve movement-derived dependencies, native speakers' reading time patterns would be expected to closely mirror those observed for native English speakers in Experiment 1, as both experiments manipulated filler plausibility and dependency length.

4.3.2 Participant

Eighty-two native Mandarin Chinese speakers participated in this study, recruited through word of mouth on a voluntary basis. Prior to participation, they confirmed that they: 1) were native Mandarin Chinese speakers, 2) thought and talked to themselves in Mandarin Chinese, 3) currently resided in Mainland China, and 4) had no history of language or reading difficulties, head injury, or attention-related disorders. These criteria were verified before participants provided informed consent and began the online study.

Following a review of the demographic and background questionnaire (Sabourin et al., 2016), four participants were excluded: one for not providing age information, two for reporting visual impairments, and one whose parents were native Korean speakers. All the remaining 78 participants reported native-level proficiency in Mandarin Chinese. While some participants reported knowledge of additional languages (English, German, Japanese, French, and Spanish), none exceeded an advanced proficiency level in these languages. Furthermore, following previous studies in the field (Huang & Kaiser, 2008; Ng & Wicha, 2014), participants with lower than 75%

overall accuracy on the comprehension questions were excluded, resulting in the removal of fifteen additional participants. The final sample consisted of 63 participants (39 females, 24 males) ranging in age from 21 to 35 years (M = 27.97 years, SD = 3.81 years). The mean comprehension accuracy across all sentences was 84%, with experimental sentences showing a mean accuracy of 72%. These accuracy rates demonstrate that participants were attentive during the reading task.

4.3.3 Materials

Twenty-four sets of experimental sentences were constructed for this experiment. As illustrated in Table 4.7, each set comprised four conditions, based on a 2 × 2 design crossing the factors Filler Plausibility (plausible vs. implausible) and Dependency Length (long vs. short). Filler Plausibility was manipulated by utilizing *wh*-phrases (e.g., *which customer/which file*) that were either semantically compatible or incompatible with the first verb (e.g., *called*), where initial filler integration was expected to occur. Dependency Length was manipulated by altering the distance between the fronted *wh*-phrase (e.g., *which customer/which file*) and the first verb (e.g., *called*). In short conditions, only a one-word subject noun phrase separated the filler and first verb. In long conditions, an additional four-word prepositional phrase was inserted, creating a five-word distance between the filler and the first verb. At the actual gap licensor position, where the sentence becomes globally plausible, the distance between the fronted *which*-phrase and the verb varied by condition: four words in short dependency sentences and eight words in long dependency sentences. While this position could show dependency length effects, no interaction between plausibility and dependency length was expected here.

Each sentence was divided into 12 regions, with each region consisting of a word rather than an individual character. For example, in the sentence *Zài huìyì jiéshù hòu, dàjiā xiǎngzhīdào nǎge kèhù lǐngdǎo tōngzhīle shíxíshēng qù zhǎo* ("When the meeting finished, everyone wondered which customer the leader notified the intern to find"), *jiéshù* ("to finish") was segmented as one region because it is a single word composed of two characters (*jié* and *shù*). This word-based segmentation approach allowed for analyzing reading times based on meaningful linguistic units in Mandarin Chinese. These 24 sets of experimental items were distributed into four lists using a Latin Square design, such that each list contained one condition from each set. An additional set of 24 fillers were created and added to each list. These filler sentences did not contain topic structures.

The critical regions for analysis were: 1) Region 9, the first verb where temporary filler integration was expected to occur, 2) Region 10, the noun phrase that signals the need for reanalysis, and 3) Region 12, the final verb serving as the gap licensor to introduce the actual gap. The first verb (Region 9) serves as the critical region, as it is the initial point where filler integration is expected to happen. The study incorporates measurements of the subsequent three regions. These regions are included not only to capture spillover effects but also to observe other processing effects. Specifically, the noun phrase at Region 10 forces readers to abandon their initial analysis of the wh-phrase as the verb's direct object. Reanalysis difficulty should be reflected in reversed plausibility effects, with sentences containing initially plausible fillers causing greater processing difficulty than those containing implausible ones at the direct object NP and/or shortly after, namely at Region 11. Then, at Region 12, where the final verb introduces the actual gap position, dependency length effects was expected. In addition to the critical and post-critical regions, the region preceding the verb is measured as the pre-critical region, where no plausibility effect is expected. I expect processing difficulty to be indexed by longer reading times.

Table 4.7 Sample experimental stimuli for Experiment 2

Short dependency, Plausible filler:											
1	2	3	4	5	6	7	8	9	10	11	12
在	会议	结束	后	大家	想知道	哪个客户	领导	通知了	实习生	去	找
Zài	huìyì	jiéshù	hòu	dàjiā	xiǎngzhīdào	nǎge kèhù	lǐngdǎo	tōngzhī-le	shíxíshēng	qù	zhǎo
at	*meeting*	*finish*	*after*	*everyone*	*wondered*	*which customer*	*leader*	*notify-PERF*	*intern*	*to*	*find*
'When the meeting finished, everyone wondered which customer the leader notified the intern to find'											
Short dependency, Implausible filler:											
1	2	3	4	5	6	7	8	9	10	11	12
在	会议	结束	后	大家	想知道	哪个文件	领导	通知了	实习生	去	找
Zài	huìyì	jiéshù	hòu	dàjiā	xiǎngzhīdào	nǎge wénjiàn	lǐngdǎo	tōngzhī-le	shíxíshēng	qù	zhǎo
at	*meeting*	*finish*	*after*	*everyone*	*wondered*	*which file*	*leader*	*notify-PERF*	*intern*	*to*	*find*
'When the meeting finished, everyone wondered which file the leader notified the intern to find'											
Long dependency, Plausible filler:											
1	2	3	4	5	6	7	8	9	10	11	12
大家	想知道	哪个客户	领导	在	会议	结束	后	通知了	实习生	去	找
dàjiā	xiǎngzhīdào	nǎge kèhù	lǐngdǎo	zài	chénhuì	jiéshù	hòu	tōngzhī-le	shíxíshēng	qù	zhǎo
everyone	*wondered*	*which customer*	*leader*	*at*	*meeting*	*finish*	*after*	*notify-PERF*	*intern*	*to*	*find*
'Everyone wondered which customer the leader notified the intern to find when the meeting finished'											
Long dependency, Implausible filler:											
1	2	3	4	5	6	7	8	9	10	11	12
大家	想知道	哪个文件	领导	在	会议	结束	后	通知了	实习生	去	找
dàjiā	xiǎngzhīdào	nǎge wénjiàn	lǐngdǎo	zài	chénhuì	jiéshù	hòu	tōngzhī-le	shíxíshēng	qù	zhǎo
everyone	*wondered*	*which file*	*leader*	*at*	*meeting*	*finish*	*after*	*notify-PERF*	*intern*	*to*	*find*
'Everyone wondered which file the leader notified the intern to find when the meeting finished'											

4.3.4 Procedures

The study received ethical approval from the University of Ottawa's Research Ethics Board (S-06-21-7042) and was conducted entirely online through the Gorilla Experiment Builder platform (Anwyl-Irvine et al., 2020). Participants received an invitation email containing the experiment link. Upon clicking the link, they were first presented with a set of four screening criteria to confirm their eligibility as native Mandarin Chinese speakers. They were required to meet all the following conditions: (1) they self-reported Mandarin Chinese as their native language; (2) they use Mandarin Chinese during internal thought and self-talk; (3) they must currently reside in Mainland China; and (4) they should not have any history of language or reading impairments, brain injuries, or Attention Deficit Hyperactivity Disorder (ADHD). Participants who reported to meet all these criteria proceeded to the experiment; others were notified their ineligibility and had the option to give up on the task. Qualified participants completed an informed consent form and a language background questionnaire (Sabourin et al., 2016; see Appendix C) before proceeding to the self-paced reading task.

The experiment utilized a self-paced reading paradigm where words appeared in black text on a white background. After receiving the instructions displayed in Mandarin Chinese, participants completed five practice trials to familiarize themselves with the procedure. The experimental structure consisted of trials beginning with a fixation point, followed by a 12-word sentence presented one word at a time. Participants controlled the reading pace, with reading times (RTs) recorded in milliseconds for each word. A period appearing on a separate screen marked sentence endings. Each participant read 48 sentences in total (24 experimental and 24 fillers). All sentences were followed by a Yes/No comprehension question to ensure participants were attentive to the task. Participants were instructed to respond by pressing F (for yes) or J (for no) on the keyboard. Participants were offered an optional 10-minute break at the midpoint of the self-paced reading task. They could choose to take a break or click "continue" to proceed. The session concluded with a debrief explaining the study's objectives and design. The entire procedure lasts approximately 25 minutes.

4.3.5 Data Analysis

The analysis included five consecutive regions (Regions 8-12), which included three theoretically critical regions: Region 9 (the verb where temporary filler integration was expected), Region 10 (the noun phrase triggering reanalysis), and Region 12 (the final verb serving as the actual gap licensor position). Although Regions 8 and 11 were not critical for testing our hypotheses, they were included to establish the baseline and capture potential spillover effects.

Data cleaning was conducted in two stages. First, following Xiang et al. (2015), extreme values were removed based on the absolute threshold: RTs exceeding 4000ms were excluded. Additionally, a lower threshold of 100ms was set to eliminate unreasonably fast responses, though no observations in any region fell below our established lower threshold of 100ms. This initial cleaning resulted in the removal of 19 data points out of 7560 total observations (0.25%). In the critical region specifically, 5 out of 1512 observations (0.03%) exceeded the upper threshold of 4000ms and were removed. Second, following Yang et al. (2023), outliers were identified and removed based on relative thresholds within each experimental condition. Specifically, RTs above or below 3 standard deviations from the mean RT of each experimental condition (short-plausible, short-implausible, long-plausible, long-implausible) at each region (Region 8, Region 9, Region 10, Region 11, Region 12) were discarded. This procedure resulted in the removal of an additional 157 data points from the remaining 7541 observations (2.08%). The proportion of data removed was relatively consistent across regions: Region 8 (1.99%), Region 9 (2.19%), Region 10 (2.32%), Region 11 (1.92%), and Region 12 (1.99%).

To address the typically right-skewed distribution of RTs, all remaining valid observations were transformed using the natural logarithm (ln), consistent with previous studies investigating covert or overt dependencies in Mandarin Chinese (Xiang et al., 2015; Yang et al., 2023). The resulting log-transformed RTs were analyzed using repeated-measures ANOVAs with the ez package in R (R Core Team, 2024). The within-subject factors were Filler Plausibility (plausible vs. implausible) and Dependency Length (long vs. short). Effect sizes were reported as partial eta-squared (pes). For all significant and marginally significant effects, follow-up pairwise comparisons were conducted using paired t-tests. For significant interactions between Filler Plausibility and Dependency Length, the effect of each factor was analyzed separately at each level of the other factor. The analyses were conducted separately for each region of interest (Regions 8-12).

As discussed in Experiment 1, previous research has demonstrated that filled-gap effects in subject positions are more pronounced with animate fillers than inanimate fillers in filler-gap dependencies (Wagers & Pendleton, 2016). This animacy effect was assumed to be relevant to an unexpected finding in Experiment 1, where a plausibility effect was observed at the critical verb with plausible filler integration being processed more slowly than its implausible counterparts. When an animacy-controlled analysis was conducted using only inanimate fillers for both plausible and implausible conditions in Experiment 1, this unexpected plausibility effect disappeared. Based on this finding from Experiment 1, an animacy-controlled analysis of Experiment 2 data was performed using 16 out of 24 experimental sets. In these 16 sets, the animacy was carefully balanced: 5 sets had animated fillers in the plausible condition and inanimate in the implausible condition, 5 had inanimate fillers in the plausible condition and animate in the implausible condition, and 6 had inanimate fillers for both conditions. However, the results showed no qualitative differences between the overall analysis and the animacy-controlled analysis. Therefore, the following sections present figures from both data sets but report only the statistical data from the analysis that includes all experimental sets.

4.3.6 Results

Figures 4.6 and 4.7 illustrated the mean log RTs across regions for all experimental items (Figure 4.6) and animacy-controlled items only (Figure 4.7). At the pre-critical region (Region 8), the animacy-controlled analysis (Figure 4.7) showed a reduced difference between short plausible and short implausible conditions compared to the overall analysis (Figure 4.6). This reduction was consistent with previous findings that animate fillers are more likely to trigger subject filled-gap effects (Wagers & Pendleton, 2016), as the animacy-controlled items contained animacy-balanced fillers. Both figures demonstrated that sentences with short dependencies (green lines) generally elicited longer RTs compared to long dependencies (orange lines), particularly pronounced in Regions 8-11. This unexpected reversed dependency length effect gradually diminished by Region 12. This pattern paralleled the findings in Experiment 1 with English materials, likely caused by the subject filled-gap effect in the pre-critical region of short conditions, with this effect persisting across subsequent regions.

For all conditions, reading times increased from Region 8 to Region 9, suggesting that parsers experienced additional processing load when encountering the verb. At Region 9 (the first verb), the patterns differed between the two analyses. In the overall analysis (Figure 4.6), short dependency conditions showed longer RTs for plausible than implausible fillers, possibly due to spillover from the higher RTs for plausible fillers in the pre-critical region. In contrast, the animacy-controlled analysis (Figure 4.7) revealed longer RTs for implausible than plausible fillers in short dependencies, aligning with the prediction that parsers would show increased reading times for implausible fillers. At Region 11, plausible fillers (solid lines) led to longer reading times than implausible ones (dashed lines) in short dependency conditions, indicating a reanalysis effect. At the actual gap licensor position (Region 12), the differences between dependency length and plausibility conditions reduced.

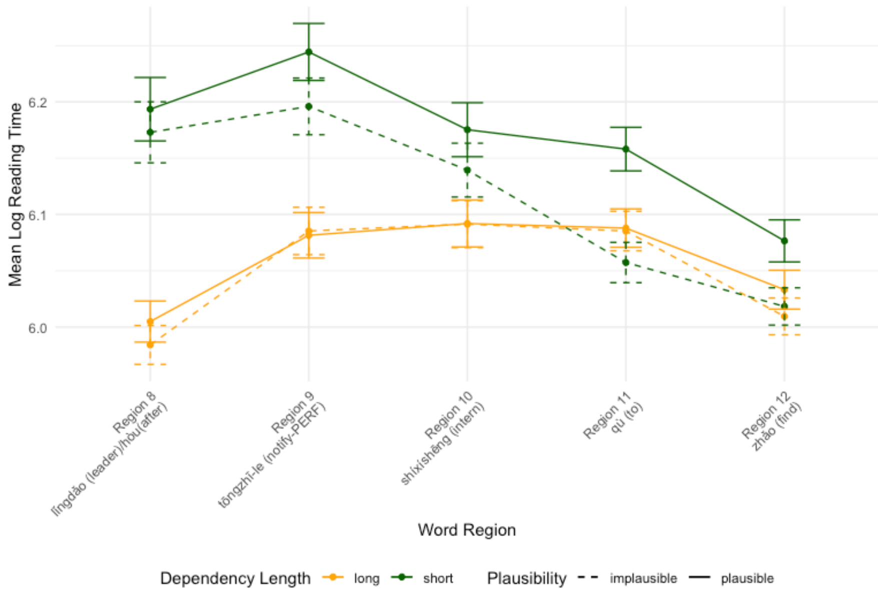


Figure 4.6 Experiment 2 Mean Log reading times of all experimental sets across word regions

Green lines represent short dependency conditions; orange lines represent long dependency conditions. Dashed lines indicate implausible conditions; solid lines indicate plausible conditions.

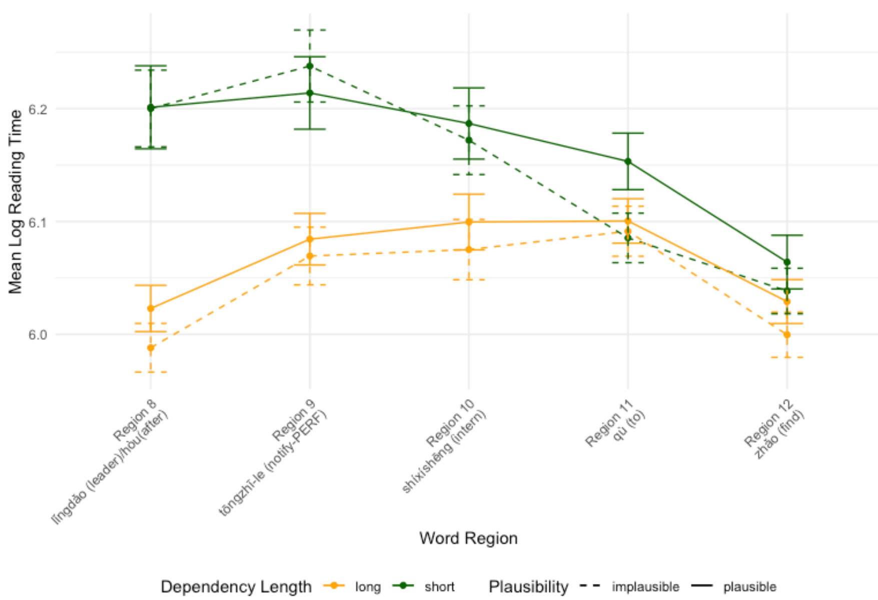


Figure 4.7 Experiment 2 Mean Log reading times of animacy-controlled experimental sets across word regions

Green lines represent short dependency conditions; orange lines represent long dependency conditions. Dashed lines indicate implausible conditions; solid lines indicate plausible conditions.

Table 4.8 presents the omnibus ANOVA results and follow-up analyses for each word region. Unlike Experiment 1, which examined complex group-specific interactions, Experiment 2 focused on the main effects of dependency length and filler plausibility, as well as their interaction.

In the pre-critical region (Region 8), analysis revealed a robust main effect of Dependency Length ($F(1, 62) = 50.84$, $p < .001$), with longer reading times in short dependency conditions ($M = 6.19$, $SE = 0.035$) compared to long dependency conditions ($M = 6.00$, $SE = 0.023$). Neither the main effect of Filler Plausibility ($F(1, 62) = 0.70$, $p = .407$) nor the interaction ($F(1, 62) = 0.10$, $p = .750$) reached significance.

In the critical first verb region (Region 9), a significant effect of Dependency Length was observed ($F(1, 62) = 42.60$, $p < .001$), with slower reading times found in short dependency conditions ($M = 6.23$, $SE = 0.031$) compared to long dependency conditions ($M = 6.09$, $SE = 0.027$). No significant effect of Filler Plausibility was found ($F(1, 62) = 1.63$, $p = .207$), and the interaction between Dependency Length and Filler Plausibility was not significant ($F(1, 62) = 0.84$, $p = .363$).

In the noun phrase region (Region 10) where the reanalysis effect was expected, a significant main effect of Dependency Length was found ($F(1, 62) = 13.48$, $p < .001$). Reading times were slower in the short dependency condition ($M = 6.16$, $SE = 0.029$) compared to the long dependency condition ($M = 6.10$, $SE = 0.026$). A marginally significant effect of Filler Plausibility emerged ($F(1, 62) = 3.69$, $p = .059$), with the plausible-at-the-verb condition showing a trend toward longer reading times ($M = 6.14$, $SE = 0.028$) compared to the implausible-at-the verb condition ($M = 6.12$, $SE = 0.028$). The interaction remained non-significant ($F(1, 62) = 0.61$, $p = .439$).

In Region 11, analyses revealed a marginal effect of Dependency Length ($F(1, 62) = 3.39$, $p = .070$), but significant effects of Filler Plausibility ($F(1, 62) = 21.01$, $p < .001$) and their interaction ($F(1, 62) = 9.87$, $p = .003$). Simple effects analyses showed that the reversed plausibility effect was significant only in short dependency conditions ($t(62) = -4.93$, $p < .001$), with plausible fillers leading to longer reading times than implausible ones. This pattern aligned with the predicted reanalysis difficulty being reflected in reversed plausibility effects, where sentences containing initially plausible fillers were expected to cause greater processing difficulty at the direct object NP and/or shortly after.

At the actual gap licensor position (Region 12), a significant main effect of Filler Plausibility was observed ($F(1, 62) = 10.57$, $p = .002$). Pairwise comparisons revealed that plausible fillers elicited significantly longer reading times ($M = 6.06$, $SE = 0.022$) compared to implausible ones ($M = 6.02$, $SE = 0.020$; $t(62) = -3.25$, $p = .002$). A marginal main effect of Dependency Length was found ($F(1, 62) = 3.49$, $p = .067$). The interaction between these two factors was not significant ($F(1, 62) = 1.66$, $p = .203$).

Table 4.8 Experiment 2 Omnibus ANOVA results and follow-up analyses across critical regions (Regions 8-12)

Analysis Type	Effect/Comparison	*F/t*	*p*	*Sig.*	η^{2p}
Region 8 (Pre-critical Region)					
ANOVA	Dependency Length	$F(1,62) = 50.84$	< .001	***	0.451
	Filler Plausibility	$F(1,62) = 0.70$	.407		0.011
	Interaction	$F(1,62) = 0.10$	.750		0.002
Follow-up Tests	Long vs. Short	$t(62) = -7.13$	< .001		
Region 9 (First Verb)					
ANOVA	Dependency Length	$F(1,62) = 42.60$	< .001	***	0.407
	Filler Plausibility	$F(1,62) = 1.63$	.207		0.026
	Interaction	$F(1,62) = 0.84$	.363		0.013
Follow-up Tests	Long vs. Short	$t(62) = -7.08$	< .001	***	
Region 10 (Noun Phrase)					
ANOVA	Dependency Length	$F(1,62) = 13.48$	< .001	***	0.179
	Filler Plausibility	$F(1,62) = 3.69$	.059	.	0.056
	Interaction	$F(1,62) = 0.61$	.439		0.010
Follow-up Tests	Long vs. Short	$t(62) = -3.67$	< .001	***	
	Implausible vs. Plausible	$t(62) = -1.92$	.059	.	

Region 11					
ANOVA	Dependency Length	$F(1,62) = 3.39$	.070	.	0.052
	Filler Plausibility	$F(1,62) = 21.01$	< .001	***	0.253
	Interaction	$F(1,62) = 9.87$	.003	**	0.137
Follow-up Tests	Long vs. Short	$t(62) = -1.84$	.070	.	
	Implausible vs. Plausible (Overall)	$t(62) = -4.58$	< .001	***	
	Implausible vs. Plausible (Long)	$t(62) = -0.36$	.719		
	Implausible vs. Plausible (Short)	$t(62) = -4.93$	< .001	***	
Region 12 (Gap Licensor)					
ANOVA	Dependency Length	$F(1,62) = 3.49$	.067	.	0.053
	Filler Plausibility	$F(1,62) = 10.57$	.002	**	0.146
	Interaction	$F(1,62) = 1.66$	.203		0.026
Follow-up Tests	Long vs. Short	$t(62) = -1.87$	.067	.	
	Implausible vs. Plausible	$t(62) = -3.25$	.002	**	

Note: Partial eta-squared (η^{2p}) reflects effect size. Significance levels: *** $p < .001$, ** $p < .01$, * $p <. 05$, . $p <. 1$ (marginal)

4.3.7 Discussion

The present study investigated the real-time processing of topic structures in Mandarin Chinese, examining whether these constructions involve movement-derived or base-generated dependencies. The results provide several important insights into the nature of filler-gap processing in Mandarin Chinese.

The first research question examined whether there is a plausibility effect at the first verb. Previous research has considered the plausibility effect as an indicator of active gap-filling strategy (William et al., 2001; Wagers & Phillips, 2014). As mentioned earlier, Wagers and Phillips (2009) propose that when parsers employ active gap-filling, it provides evidence for two key syntactic features: unbounded dependencies and abstract traces. This is because active gap-filling presupposes that the parser has identified the fronted constituent as having undergone movement and posits an associated empty category (or trace) in the base position, thereby initiating an active search to link the filler with its gap. Although no plausibility effect was found at the first verb position, two pieces of evidence suggested that Mandarin Chinese speakers engaged in active gap searching: (1) increased reading times at the first verb compared to pre-verbal conditions, and (2) a reanalysis effect at the actual gap licensor region. However, it is important to note that increased reading times at the verb alone cannot definitively demonstrate active gap filling due to the lack of the direct comparison with a non-gap condition. These increased reading times might simply reflect the greater processing demands of verbs compared to other word categories, rather than indicating active dependency resolution. Therefore, these results must be interpreted in conjunction with the reanalysis effect. The following paragraphs discussed both the elevated reading times and the reanalysis effect in detail.

According to Figure 4.6 and Figure 4.7, there was obvious reading time slowdown in the region of the first verb, which was identical to what was shown in Experiment 1. Admittedly, there could be two accounts that explain this reading time slowdown at the first verb: the Direct Association Hypothesis (Pickering, 1991, 1993; Pickering & Barry, 1991) and the Trace Reactivation Hypothesis (e.g., Bever & McElree, 1988; Nicol & Swinney, 1989; Love & Swinney, 1996; Marinis et al., 2005). The Direct Association Hypothesis is a bottom-up approach where the verb's properties trigger the integration, whereas the Trace Reactivation Hypothesis is more top-down in that the stored filler drives active gap search. The findings in the present study are more compatible with the Trace Reactivation Hypothesis, because if *wh*-phrases were directly linked to the verb's subcategorization properties, we would also expect a reading time slowdown at the final verb *find,* where the verb must find its direct object and therefore incur integration costs. The lack of reading time slowdown at the actual gap licensor (Region 12) could be attributed to: 1) the filler information not being fully released after the first verb, and 2) the maintained information that drove the active gap-filling having faded. It is important to note that at the first verb position, there was no plausibility effect or interaction effect between filler plausibility and dependency length, indicating that the elevated reading time at the first verb position was likely not driven by semantic information but rather by syntactic information. However, the absence of plausibility effect at the first verb position cannot be used to rebut active gap filling, as effects in self-paced reading paradigms often lag and spill over to following regions. For example, Wagers and Phillips (2014) showed that a slowdown in reading time in implausible conditions, either at the verb or in following regions, indicated that readers were actively constructing the filler-gap dependency. Similarly, in my Experiment 1, the plausibility effect where implausible fillers were processed slower than plausible fillers were found in the second post-critical region. Therefore, the reading time slowdown at the first verb *notify*, but not at the final verb *find*, is compatible with a movement analysis characterized by an active gap-filling strategy.

For the research questions concerning the reanalysis effect, the results showed a crucial pattern that supported reanalysis processes and thus endorsed active gap-filling. While no immediate plausibility effect was observed at the first verb (Region 9), the effect began to emerge at the noun phrase region (Region 10), where a marginally significant effect of Filler Plausibility was found, with plausible-at-the-verb fillers showing a trend toward longer reading times. This effect then reached statistical significance at Region 11, with a significant main effect of Filler Plausibility and the interaction between Dependency Length and Plausibility. Importantly, the direction of this effect showed that initially plausible filler-verb pairs led to greater processing difficulty than implausible ones, particularly in short dependency conditions. This reanalysis effect also spilled over to Region 12. The reversed plausibility effect provided evidence for reanalysis: when readers encountered Region 10 (the noun phrase), they tended to abandon their initial filler-gap integration. The presence of this reanalysis effect suggested that semantic integration had occurred earlier, even though no explicit plausibility effect was detected at the verb region. This was likely because reanalysis costs—which reflect the difficulty of revising an initial semantic commitment—can only arise if readers have previously integrated the filler's meaning with the verb. The greater reanalysis cost for initially plausible filler-gap integration suggested that readers had committed more strongly to the initial misanalysis when the filler was semantically compatible with the first verb. This pattern aligned with the findings in English filler-gap processing where a reanalysis effect was observed in the form of a P600 component when initial parsing commitments required reanalysis (Jessen & Felser, 2019).

Interestingly, the reanalysis effect was only detected in the short conditions but not in the long conditions at Region 11. This pattern can be explained through the dual aspects of filler-gap processing: syntactic and semantic. In short dependencies, both syntactic and semantic features of the filler tend to remain highly accessible, leading to stronger initial integration at the first verb and consequently stronger reanalysis costs when this integration must be abandoned. In long dependencies, however, the results aligned with Wagers and Phillips's (2014) interpretation that the lexicosemantic features of the filler become less accessible when actively building long-distance dependencies. Importantly, this does not mean that fillers' semantic features are completely forgotten in long dependencies. Instead, as Wagers and Phillips (2014, p. 1292) noted, these features become "specifically inaccessible or ineffective for active dependency formation."

The third research question, examining the dependency length effect, revealed unexpected results that mirrored those found in Experiment 1. Contrary to the initial prediction based on the movement account, significant longer reading times were consistently observed in short dependency conditions across multiple regions from Region 8 to Region 10. This pattern may be explained by the subject filled-gap effect: in short conditions, the pre-critical region contained a subject that filled a potential gap position, causing residual processing difficulty that spilled over into subsequent regions, similar to what was observed in Experiment 1. If the subject filled-gap effect was indeed responsible for the unexpected longer reading times in short dependencies, longer reading times would be expected in Region 4 of long dependency sentences, as this was where the subject filled-gap effect would occur. This hypothesis was confirmed, as reflected by the longer mean reading times in the long dependency condition compared to the short dependency condition at Region 4 (see Figure 4.8). If the subject filled-gap effect holds, it may provide evidence for movement-based processing rather than base-generation, as it suggests that Mandarin Chinese speakers behave similarly to English speakers: they actively predict and attempt to fill the subject gap, the place where the verb has not yet appeared. When the subject gap position is unexpectedly filled, processing difficulty arises and spills over.

The plausibility effect or the reanalysis effect, if detected, was predicted to be modulated by dependency length. This prediction was based on Wagers and Phillips's (2014) observation that plausibility effects disappeared when dependencies became longer. The results were consistent with this prediction. Region 11 showed a significant interaction between Dependency Length and Filler Plausibility, with the reanalysis effect appearing only in short dependency conditions. It was not until Region 12 when a main effect of Filler Plausibility without interaction with Dependency length was found. In short dependencies, robust reanalysis effects suggest that the semantic features of the filler remain readily accessible, whereas in long dependencies, the delayed reanalysis effects indicate that access to these features is reduced over distance.

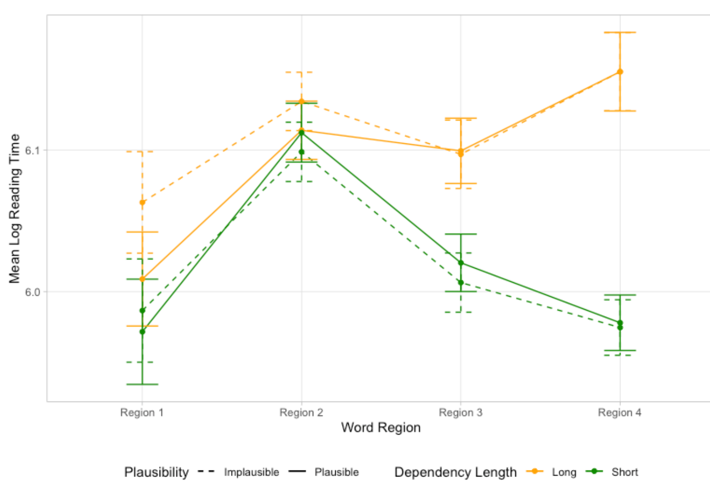


Figure 4.8 Experiment 2 Mean Log reading times of all experimental sets from Region 1 to Region 4

Green lines represent short dependency conditions; orange lines represent long dependency conditions. Dashed lines indicate implausible conditions; solid lines indicate plausible conditions.

In summary, the processing patterns found in Experiment 2 regarding Mandarin Chinese topic structures showed evidence of active dependency formation similar to English *wh*-dependencies. Three key findings supported this conclusion. First, the reading time slowdown at the first verb indicated active gap-filling strategies, suggesting that readers were actively searching for a gap position rather than waiting for bottom-up verb information. Second, there was an evident reanalysis effect showing that even before the actual gap position, parsers engage in filler integration semantically. Finally, the potential subject filled-gap effects demonstrated that Mandarin speakers, like English speakers, predictively posited gaps even before encountering verbs. Therefore, the present experiment suggested that, although Mandarin Chinese is typically considered a *wh*-in-situ language, Mandarin Chinese speakers employ processing mechanisms for gap-type topic structures similar to those English speakers use for *wh*-dependencies.

4.4 General Discussion

The two experiments included in this chapter provided evidence about the nature of filler-gap processing across English and Mandarin Chinese in L1 and L2 contexts. Experiment 1 investigated how L1 and L2 English speakers process *wh*-dependencies in real time, while Experiment 2 examined whether native Mandarin Chinese speakers process overt topic structures in ways similarly to how native English speakers process English *wh*-dependencies. The findings offered insights into three aspects of sentence processing: (1) the sequential timing of using syntactic and semantic information in the process of active gap-filling, (2) the similar parsing mechanisms of active gap-filling across English and Mandarin Chinese, and (3) the nature of L2 processing mechanisms being qualitative similar.

Active gap-filling process

Both experiments provided evidence for Friederici's (2002, p. 78) "syntax-first" model and Wagers and Phillips's (2014) dual-process account. In Experiment 1, both L1 and L2 speakers showed immediate sensitivity to syntactic information, as evidenced by elevated reading times at the critical verb region and increased reading times at the filled subject gap position, while plausibility effects emerged later in the second post-critical region. Similarly, in Experiment 2, Mandarin Chinese speakers showed greater processing difficulty at the first verb position, with the reanalysis effect being detected shortly after. This non-simultaneous involvement of syntactic and semantic processing challenged interactive or constraint-based models that propose simultaneous processing of multiple information sources (e.g., Trueswell et al., 1994; Clifton, 2003). Instead, the findings suggested a processing profile where active gap-filling was driven by syntactic information, with semantic integration occurring at a later stage.

Similar processing mechanisms across English and Mandarin Chinese

Similarity in processing mechanisms across English, a *wh*-fronted language, and Mandarin Chinese, a *wh*-in-situ language, was found. Despite Mandarin Chinese being a *wh*-in-situ language, Experiment 2 revealed that native Mandarin Chinese speakers process topic structures using active gap-filling strategies similar to those employed by English speakers processing *wh*-dependencies. This was evidenced by: (1) immediate reading time slowdown at the first potential gap site, (2) sensitivity to plausibility, which was reflected by plausibility effects or reanalysis effects when initial parsing decision had to be revised, and (3) potential subject filled-gap effects in both languages. These shared processing mechanisms had important implications for theories of L2 processing. To be specific, previous studies have attributed L2 processing difficulties to the absence of *wh*-movement in learners' L1 when their L1 is a *wh*-in-situ language (e.g., Dallas et al., 2013; Dussias & Piñar, 2010). However, the results suggested that such processing difficulties cannot stem from L1-L2 differences, as Mandarin Chinese speakers showed similar active gap-filling strategies in their L1.

The nature of L2 processing mechanisms being qualitative similar

The findings of Experiment 1 provided evidence against qualitative difference between L1 and L2 processing. In Experiment 1, L2 learners demonstrated sensitivity to syntactic information, even though their processing appeared to be more effortful in several ways: they showed greater sensitivity to dependency length, delayed sensitivity to semantic anomalies, and plausibility effects modulated by proficiency. These patterns suggested that L2 processing difficulties may not stem from an inability to construct detailed syntactic representations, but rather from reduced efficiency in maintaining and retrieving multiple sources of information. This interpretation aligned with both the RAGE hypothesis (Grüter et al., 2013, 2021) and Continuity Hypothesis (Hopp, 2010; McDonald, 2006). As for sensitivity to filler plausibility, both L1 and L2 speakers showed similar processing mechanisms, regardless of L2 proficiency or immersion experience. Although there was a descriptive trend suggesting that high-proficiency L2 learners processed filler plausibility more like native speakers, this pattern did not reach significance and thus warranted further investigation.

4.5 Conclusion

This chapter explored how parsers process filler-gap dependencies in both English and Mandarin Chinese, with both L1 and L2 speakers. The findings were threefold. First, syntactic and semantic information were integrated at different time points rather than simultaneously, with syntactic information being processed first in both L1 and L2 speakers and in both English and Mandarin Chinese. Moreover, the findings from these two experiments suggested that native Mandarin Chinese speakers process gap-type topic structures in a similar way to how English speakers process *wh*-dependencies. Therefore, previous qualitative differences in processing mechanisms between L2 Mandarin speakers and L1 native English speakers should not be attributed to differences between the two languages. Lastly, while L2 speakers showed greater processing difficulty with long-distance dependencies compared to L1 speakers, they demonstrated native-like filler-gap integration profiles, as evidenced the reading time slowdown at the critical verb during initial integration attempts and the sensitivity to filler semantic cues shortly after. These findings aligned more with the view that L1 and L2 speakers share qualitatively similar processing mechanisms.

Chapter 5

The ERP Experiment

5.1 Introduction

This chapter presents an ERP study addressing the central question of this dissertation: how L2 speakers process filler-gap dependencies and whether they employ parsing mechanisms similar to those of L1 speakers. The present study specifically examines whether L1 and L2 speakers show similar brain responses at three critical regions during *wh*-dependency processing: (1) the initial verb where filler plausibility is tested, (2) the filled-gap position where reanalysis is required, and (3) the final licensing element at the actual gap site, thereby investigating how syntactic and semantic information contribute to successful dependency resolution in both groups.

In the previous chapter using self-paced reading paradigms, we found that although L2 speakers seemed to process *wh*-dependencies more effortfully and slowly than native speakers, their underlying parsing mechanisms became increasingly native-like as their language proficiency improved. Specifically, in both L1 and L2 speakers, syntactic cues seemed to be employed to drive active gap filling, and sensitivity to semantic information was found across the groups, with L2 high-proficiency L2 speakers showing a descriptive trend of being more native-like. This finding aligns with the Convergence Hypothesis, which states that the mechanisms underlying L2 processing are dynamic and converge more toward L1 patterns as language proficiency increases (Steinhauer et al. 2009). However, as behavioral responses are assumed to lag behind neurocognitive processing, it is premature to draw the conclusion that L2 processing relies on the same mechanisms as L1. It is possible that while proficient L2 speakers may achieve native-like behavioral performance, they might recruit different neurocognitive strategies to accomplish similar processing outcomes (e.g. Dong et al., 2023; Covey et al., 2024). Even if L2 speakers use qualitatively similar mechanisms to L1, their processing may be less automatic and more effortful.

This could be reflected in different ERP response amplitudes (Jesson et al., 2019; Rossi et al., 2006; Batterink & Neville, 2013) and different distributions of ERP patterns (Jessen et al., 2017; Andersson et al., 2019). Neuroimaging evidence has indeed shown both intensified activation in specific brain regions (Rüschemeyer et al., 2006) and broader recruitment of brain areas (Wartenburger et al., 2003) during L2 processing. Additionally, L2 speakers may exhibit delayed use of grammatical information during sentence processing, as proposed in the updated SSH (Clahsen & Felser, 2018). This delay has been demonstrated in filler-gap resolution through eye-tracking studies (Boxell & Felser, 2017; Felser et al., 2012).

Compared to behavioral measures, ERPs can reflect real-time electrophysiological brain activity of cognitive processes. Their millisecond-level temporal resolution enables us to detect processing mechanisms that are difficult to observe with purely behavioral measures. Furthermore, the well-documented set of language-related ERP components in native speakers provides crucial reference points for evaluating L2 processing mechanisms. The present study thus uses ERPs to investigate whether the similar behavioral outcomes observed between L1 and L2 speakers reflect genuinely similar underlying neural profiles or L2 speakers achieve comparable performance through different neurocognitive mechanisms or with different ERP amplitude, distribution or timing.

5.2 ERP Components

The present study focuses primarily on three ERP components widely associated with sentence processing: N400, P600, LAN. The N400 is a central-parietal, negative-going wave occurring in the 300-500 ms range and peaking around 400 ms post-stimulus (Kutas & Hillyard, 1980). This neural signature typically reflects semantic integration difficulty and semantic expectation violation (Kutas & Federmeier, 2000; Kutas & Federmeier, 2011; Luck, 2014). When the brain encounters words that are semantically unexpected or difficult to integrate into the context, the N400 amplitude increases. Beyond its well-known association with semantic anomaly detection, recent research has shown that N400 is also linked to lexical prediction (Michel, 2014; Covey et al., 2024; Federmeier, 2007; Van Petten & Lukas; 2012). When words align with lexical-semantic expectations, they elicit smaller N400 responses compared to unexpected items. For example, in a study investigating filled-gap effects and island constraints in the processing of wh-dependencies, Covey and colleagues (2024) found that native speakers exhibited larger N400 responses at the filled gap position compared to conditions where no gap was predicted. They attributed this N400 effect to predictive processing in native speakers, suggesting that native speakers engage in gap prediction. Their findings aligned with Michel (2014), who compared brain responses to gaps appearing in unexpected positions (illicit island conditions) versus predicted positions (licit non-island conditions). The results showed that a larger N400 amplitude occurred at unexpected gaps within illicit islands compared to licit non-island conditions. These findings collectively indicate that the N400 not only reflects semantic integration but also serves as an index of the brain's predictive mechanisms during both semantic and syntactic processing. However, it remains unclear whether L2 speakers employ similar predictive mechanisms when processing filler-gap dependencies. Recently, Covey and colleagues (2024) provided the only evidence in this domain, and they found P600 rather than N400 responses in L2 speakers at gap prediction positions. They attributed this to qualitative differences between L1 and L2 processing, suggesting that while L1 speakers engage in prediction, L2 speakers experience syntactic integration difficulty.

With regard to syntactic processing, a commonly elicited component is the P600. It is a relatively large positive-going wave emerging within 500-900 ms and typically peaking at approximately 600 ms (Osterhout & Holcomb, 1992). The P600 is associated with the detection of syntactic violations (e.g., Osterhout & Holcomb, 1992; Weber-Fox & Neville, 1996), the processing of complex syntactic structures such as dependency resolution (e.g., Kaan et al., 2000; Kaan & Swaab, 2003), and the need for syntactic reanalysis (e.g., Friederici et al., 2001; Gouvea et al., 2010). Specifically in ERP literature on *wh*-dependencies, the P600 is shown to reflect filler integration (e.g., Kaan et al., 2000; Phillips et al., 2005; Jessen & Felser, 2019). For example, Kaan et al. (2000) found that the amplitude of the P600 component was greater when parsers processed verbs embedded in *wh*-clause structures, which index a potential gap to be filled, compared to when these same verbs appeared in whether-clauses, where no gap is required. This larger P600 was attributed to increased processing costs caused by integrating an antecedent *wh*-filler stored in working memory into its gap. Beyond syntactic processing difficulties, the P600 component has been shown to index various semantic anomalies, including animacy constraint violations (e.g., Kuperberg et al., 2003), semantically impossible role assignments (e.g., Kim & Osterhout, 2005), and strong semantic incongruities that generate both N400 and P600 responses (e.g., DeLong & Kutas, 2020). Some recent studies also suggest that the P600 reflects a more general revision mechanism in language processing, responding to both structural and semantic complexities that require reanalysis, even if there is no clear violations (Van Petten & Luka 2012; Brouwer et al. 2012; Brouwer & Crocker 2017). Notably, studies of L2 processing reveal that language proficiency plays a crucial role in the occurrence of P600 effects, with only advanced L2 learners showing P600 responses. Even when present, these P600 responses may not be canonical, often displaying more broadly distributed patterns (e.g., Mickan & Lemhöfer, 2020; for further discussion, see van Hell, 2023; Caffarra et al., 2015).

Another relevant ERP component is the LAN. It is a negative-going wave appearing in the time-window of 300-500 ms. Although occurring in the same temporal window and showing similar negative voltage shifts as the N400, the LAN's distribution is distinctly left-lateralized anterior or bilateral (e.g., Pakulak & Neville, 2010, 2011; King & Kutas, 1993). The LAN has been associated with automatic early detection of morphosyntactic violations, such as number agreement violations, gender agreement violations, and tense-marking violations (see Molinaro et al., 2011 for a review). Additionally, it is linked to increased working memory load, particularly in filler-gap processing, where the LAN reflects processing costs associated with maintaining a filler in working memory (e.g., Kluender & Kutas, 1993; Felser et al., 2003; Fiebach et al., 2002; Phillips et al., 2005). The LAN is frequently followed by a P600 component. Although the exact nature of the LAN's role in parsing mechanisms remains inconclusive, it is generally considered a marker of native-like automaticity (e.g., Gunter et al., 2000). However, even among native speakers, the LAN response shows considerable individual variation and is not uniformly present. Consequently, the absence of a LAN does not necessarily indicate non-native-like processing (Antonicelli & Rastelli, 2023). While some ERP studies have observed LAN-P600 patterns in L2 learners during word order violations (Bowden et al., 2013; Morgan-Short et al., 2012) or word category violations (Steinhauer et al., 2006), no study has revealed this complex in filler-gap processing.

5.3 Experiment 3: Filler-Gap processing in L1 and L2 English Speakers using ERPs

5.3.1 Rationales and predictions

The present chapter addresses three main research objectives. First, we examine whether L2 speakers, like native speakers, demonstrate sensitivity to filler plausibility upon encountering a potential gap position. Second, we investigate whether L2 speakers exhibit neural responses similar to native speakers during syntactic reanalysis when the initial gap position turns to be incorrect. Third, by manipulating the gap licensor type (preposition vs. verb), we examine whether participants show different ERP responses depending on the type of gap licensor, and whether L2 speakers show different neuroimaging responses from L1 speakers.

To address these objectives, this study extends Jessen & Felser's (2019) work by incorporating a novel variable of gap licensor type in addition to the filler plausibility manipulation. According to a corpus analysis by Atkinson and Omaki (2023), prepositional object (PO) gaps (gaps licensed by prepositions) occur far less frequently than direct object (DO) gaps (gaps licensed by transitive verbs requiring a direct object). This raises the question of why PO gaps are less probable. If this is due to syntactic integration difficulty, prepositions may elicit a P600 effect. If it stems from a violation of lexical expectations, an N400 response may occur. Alternatively, if the difficulty arises from maintaining the filler in working memory, ERP components related to working memory may be observed.

Following Jessen and Felser, we manipulated filler plausibility at the first verb to examine whether L1 and L2 speakers engage in active gap-filling and exhibit sensitivity to filler plausibility. Garnsey et al. (1989) first found that L1 speakers immediately integrate fillers with verbs, evidenced by an N400 effect for implausible fillers. Building on this, Dallas et al. (2013) observed this effect only in L1 speakers and highly proficient L2 learners. Notably, their L2 participants (L1 Chinese) showed native-like N400 effects only in simpler constructions without gaps (in *whether* complements), suggesting syntactic complexity may constrain L2 semantic processing. More recently, Jessen and Felser (2019) found N400 effects in both L1 and L2 (L1 German) groups, though with different scalp distributions and timing - left-frontal and longer-lasting (300-650 ms) in L1 versus right-frontal and shorter (400-550 ms) in L2 speakers. Jessen and Felser suggested that these different scalp distributions actually indicated similar processing patterns. Based on their findings, we hypothesized that both L1 and L2 speakers would show N400 effects at the first verb when encountering implausible fillers, though the L2 group might show different scalp distributions or amplitude strength compared to the L1 group.

Following Jessen and Felser's design, we also filled the potential gap after the first verb with a NP, creating a filled-gap position requiring reanalysis. As previously discussed, Hestvik et al. (2007, 2012) found that encountering an NP in a predicted gap position elicits an anterior negativity (ELAN/LAN) followed by a P600 in native English speakers, suggesting syntactic category violation and subsequent reanalysis. L2 research, however, has produced mixed findings. It is worth noting that, in the present study, both plausible and implausible conditions contained a filled-gap position. Our primary interest, following Jessen and Felser, was in how the plausibility of the initial filler-gap integration affects the reanalysis process at this filled-gap position. Jessen and Felser (2019) considered both the noun phrase (filled-gap position) and the subsequent preposition (actual gap position) as the disambiguating regions. Notably, at the filled-gap position, neither L1 nor L2 speakers showed any significant ERP effects. The significant effects emerged at the preposition region, where L2 learners showed larger P600 responses to plausible fillers while native speakers showed no effects. The researchers attributed this P600 to a reanalysis effect, specifically a reversed plausibility effect, with plausible integrations being more costly to reanalyze. In addition, some studies suggest qualitatively similar processing between native and non-native speakers. For instance, Jessen et al. (2017) found P600 effects for both native English speakers and German learners of English when processing filled indirect object gaps, though the L2 positivity was more globally distributed and of higher amplitude than the natives' frontal P600. However, more studies indicate fundamental differences between L1 and L2 processing. Similarly, Covey et al. (2024) found that while native speakers showed N400 effects at filled gap positions, reflecting predictive processing, Mandarin Chinese learners of English instead produced P600 effects, suggesting they experienced integration difficulty rather than prediction failure. More recently, Dong et al. (2023) found that Chinese learners of English showed a qualitatively different ERP component (prefrontal-central positivity) compared to natives' canonical P600. According to Leckey and Federmeier (2020), the P600 may actually comprise a collection of related but distinct brain responses. Specifically, the canonical posterior-distributed P600 appears associated with syntactic violations requiring repair, whereas grammatical but complex syntactic structures lead to more frontally distributed P600 (Friederici, Hahne & Saddy, 2002) or broadly distributed P600 (Hagoort & Brown, 2000). Based on these findings, we predicted that native English speakers would show components associated with syntactic reanalysis (P600) or prediction (N400) at the Noun Phrase. For L2 learners, we predicted a similar but more pronounced effect for plausible conditions, probably a larger P600. This is because, if semantic information were prioritized during incremental processing, the plausible integration might be more firmly established during the initial semantic processing stage, which could make it harder to reanalyze when follow-up syntactic cues come into play.

A thorough review examining ERP studies of L2 sentence processing published between 2019-2020 revealed that late L2 learners' language proficiency is a key predictor of their ERP responses (Antonicelli & Rastelli, 2023). Based on this perspective and the findings from the previous chapter, we anticipated that L2 learners' proficiency levels would affect their brain responses, with higher proficiency participants showing more native-like ERP patterns in terms of component latency, amplitude, and topographical distribution. Four specific research questions are investigated in the present experiment.

1. Do L1 and L2 speakers show similar brain responses (e.g., N400) at the first verb when encountering implausible versus plausible fillers?
2. Do L1 and L2 speakers show similar brain responses (e.g., P600, N400) at the noun phrase after the first verb when they need to reanalyze an initially plausible filler-gap dependency?
3. Do L1 and L2 speakers show different brain responses (e.g., P600, LAN) when processing filler-gap dependencies licensed by prepositions versus transitive verbs?
4. How does L2 proficiency level modulate ERP responses compared to L1 speakers across the three critical regions?

5.3.2 Participants

The final analysis of the study included 11 native English speakers (L1) (8 females, mean age 21.09, range 18-27 years) and 20 Mandarin Chinese learners of English (L2) (12 females, mean age 23.20, range 19-32 years). Participants were recruited through two channels: (1) the Integrated System of Participant Research (ISPR) at the University of Ottawa, where students received course credits for participation, and (2) voluntary participation in response to recruitment advertisements posted on campus. On both ISPR and recruitment posters, it was specified that participants were to have no visual impairments, no history of serious head injury, and no diagnosed attention-related disorders. All participants were right-handed and had normal or corrected-to-normal vision. Table 5.1 provides a comprehensive summary of participant demographics for both groups.

Two additional L2 participants were tested but subsequently excluded from analysis. One participant scored only 65% for all sentences and 50% for experimental sentences, substantially below the L2 group mean of 87.40% and 81.88%, respectively, suggesting either insufficient attention during the experiment or difficulty comprehending the experimental stimuli. The other participant did not complete the cloze task following the ERP study. The L2 group's English proficiency was assessed using a cloze proficiency task administered after the EEG recording. Based on the median score of the cloze task, L2 participants were divided into high-proficiency (n=9, mean score = 71/100) and low-proficiency (n=11, mean score = 42/100) subgroups to examine potential proficiency effects on neural responses.

Table 5.1 Participant information for Experiment 3

Group		Age	Self-rated English proficiency /5[a]	Accuracy of measured English Proficiency /100[b]	Age of initial exposure	Age of immersion	Year of immersion
L1	Mean	21.09	5	N/A	0.27	0.55	20.55
(n=11)	*SD*	2.88	0	N/A	0.90	1.21	2.84
	Range	18-27	5-5	N/A	0-3	0-3	18-24
L2	Mean	23.20	2.95	54.90	7	17.50	5.70
(n=20)	*SD*	2.95	0.69	17.22	2.29	3.52	2.98
	Range	19-32	2-4	30-82	3-12	12-27	1-12

[a] Proficiency levels are self-rated according to this scale: 0 = very low, 1=low, 2=intermediate, 3=advanced, 4=near-native, 5-native.

[b] Proficiency levels are measured by their accuracy of a cloze task that can be found in Appendix B.

5.3.3 Materials

Twenty-four sets of experimental sentences were designed as exemplified in Table 5.2. The complete list can be found in Appendix F. Each set consisted of four target sentences. All sentences contained a *wh*-clause introduced by the complementizer *which* followed by a noun phrase, such as *which article* in Sentences A and C and *which woman* in Sentences B and D. The independent variables were filler plausibility (plausible vs. implausible) and gap licensor type (transitive verb vs. preposition). Filler plausibility was manipulated by utilizing NPs that were either semantically compatible or incompatible with the first verb, which introduced a potential gap site. The gap licensor referred to the specific word that creates the actual gap of the *wh*-filler. The two factors were not crossed in the design. For filler plausibility, Sentences A and C represented the plausible condition, while Sentences B and D represented the implausible condition. For gap licensor type, Sentences A and B contained transitive verbs, whereas Sentences C and D contained prepositions as gap licensors. We chose not to cross these factors because ERP studies require a large number of trials per condition for accurate measurement, as ERP signals have small effect size (Luck, 2014). By designing the stimuli this way, each participant experienced 12 trials for the filler plausibility condition and 12 trials for the gap licensor condition, which doubled the number of trials per condition compared to a fully crossed design. The target sentences were distributed across four separate lists using a Latin Square design, such that each list contains one sentence from each experimental set. All experimental sentences were globally plausible. This was important as most previous studies employed ultimately ungrammatical sentences (e.g., Dallas et al., 2013; Jessen et al., 2017; Dong et al., 2023), making it difficult to determine whether ERP components were elicited by grammatical violations or complex sentence construction. To the best of my knowledge, only two ERP studies investigating L2 filler-gap processing have employed globally grammatical sentences (Jessen & Felser, 2019; Covey et al., 2024). Each list also included 5 practice items and 24 filler sentences that were globally plausible and varied in structure. The items in each list were pseudo-randomized with limits on repetition, such that no more than two experimental sentences from the same condition and no more than three fillers appeared in sequence. All sentences were followed by a yes/no comprehension question to disguise the experiment as a language comprehension task and to measure participant attention. All sentences were reviewed and revised by two native English speakers in the ERP linguistics lab to ensure they sounded natural to native speakers and that the two independent variables of the experimental stimuli were appropriately manipulated.

Table 5.2 Sample experimental stimuli for Experiment 3

Type	Filler Plausibility	Gap Licensor	Sentence
A	Plausible	VT	Lucy identifies which article the author **wrote** a column to describe ___ in the newspaper.
B	Implausible	VT	Lucy identifies which woman the author **wrote** a column to describe ___ in the newspaper.
C	Plausible	PREP	Lucy identifies which article the author **wrote** a column about ___ in the newspaper.
D	Implausible	PREP	Lucy identifies which woman the author **wrote** a column about ___ in the newspaper.
Yes/No question			Did the author write a column?
Answer			Yes

There were three critical regions. The first critical region was at the First Verb in the *wh*-clause, where filler integration was first attempted (e.g., *which article/which woman... wrote*). If participants utilized semantic information during active gap filling, their sensitivity to filler plausibility should be detected at this region. The second critical region was the Noun Phrase following the First Verb, where the expected gap was filled, and the initial filler-gap integration was reanalyzed. At this region, we anticipated a reversed plausibility effect. Specifically, the reversed plausibility effect could occur when previously plausible integrations are harder to reanalyze because these integrations would be more robust. Alternatively, the reversed plausibility effect could arise from the collapse of expectations, as an initially plausible integration strengthens the expectation of a syntactic gap. In Jessen and Felser's study, they found no effects at the noun phrase region but observed an enhanced P600 for plausible conditions at the following preposition (their actual gap site) in the L2 group. However, as their preposition served as both the reanalysis trigger and the actual gap site, we would not know if the P600 indeed reflected reanalysis difficulty or if it was merely an effect of filler-gap integration. Therefore, in the current study, we constrained the reanalysis region to the noun phrase position, which preceded the actual gap, allowing us to isolate reanalysis effects from gap integration effects. The third critical region was at the actual Gap Licensor, which was either a transitive verb (Vt) or a preposition (Prep). While participants had previously encountered both plausible and implausible filler integrations at the first verb, the two plausibility conditions were evenly balanced. Therefore, any effects observed at this actual gap position should primarily reflect the structural and semantic properties of the gap licensors themselves (Vt vs. Prep), independent of earlier plausibility manipulations.

5.3.4 Procedure

The study was conducted with the approval of the University of Ottawa's Research Ethics Board (file #: S-10-23-9693). All participants completed the informed consent and language background questionnaire. After fitting the ERP cap and facial electrodes, participants were guided to a sound-attenuated room where they sat in front of a computer monitor and used a keyboard with only two active buttons: green for "Yes" and red for "No." The computer screen was 20 inches (31.5 x 38 cm) and positioned approximately 100 cm from the participant. Words were presented in black font on a white screen using Presentation version 17.2 (Neurobehavioral Systems). Before starting the task, participants were instructed to relax their facial muscles and read the sentences silently for comprehension. They were informed that a practice session would be provided to familiarize them with the task. All participants cooperated to silence their electronic devices before beginning.

When ready, participants self-initiated the task. Sentences were presented one word at a time using the Rapid Serial Visual Presentation (RSVP) paradigm. Each sentence began with a fixation to indicate the start of a new sentence. For presentation rate, previous studies in the field of L2 filler-gap processing have used varying durations for word presentation (ranging from 350 ms to 550 ms) and intervals (ranging from 300 ms to 500 ms) (Dallas et al., 2013; Jessen et al., 2017; Jessen & Felser, 2019; Covey et al., 2024). We set each word to be presented for 500 ms with an interval of 400 ms, as both L1 and L2 speakers reported this pace felt natural. After each sentence, a yes/no comprehension question appeared entirely at once on the screen, and participants were instructed to respond using the green button for "Yes" and red for "No." These comprehension questions were designed to maintain participants' attention rather than evaluate their understanding of filler-gap dependencies. The EEG session took approximately 30 minutes to complete. Following the EEG recording, L2 speakers were allotted up to 30 minutes for the cloze proficiency task. The entire session lasted approximately 90 minutes for L1 speakers and 120 minutes for L2 speakers, including electrode cap fitting and removal.

5.3.5 EEG recording and pre-processing

Continuous EEG was recorded using an elastic cap (Compumedics Quik-Cap Neuroscan) with 64 channels arranged according to the 10-20 electrode layout system. Vertical electrooculogram (VEOG) was monitored using electrodes placed above and below the left eye (vEOGu and vEOGl), and horizontal electrooculogram (HEOG) was monitored from electrodes positioned on both temples (hEOGr and hEOGl). The online EEG was recorded against reference electrode located between Cz and CPz on the midline of the cap, and then re-referenced offline to the average of the left and right mastoid electrodes (M1 and M2). Before beginning the task, impedances were checked to be below 10kΩ. Participants were shown their real-time EEG signals associated with facial muscle movements, including eyeblinks, frowning, and jaw tension, and were instructed to relax and minimize unnecessary muscle movements. The recording was performed using Compumedics Neuroscan Scan 4.5 software, sampling at 1000 Hz. The signal was processed through a SynAmps amplifier with a 200 Hz low-pass filter, while no online high-pass filtering was applied.

EEG data cleaning was performed using Brain Vision Analyzer Version 2.2.0 (Brain Products GmbH, Herrsching, Germany). After re-referencing, the data were filtered offline using a 0.01 Hz high-pass filter and a 30 Hz low-pass filter. To isolate and correct artifacts from vertical and horizontal eye movements, an Independent Components Analysis (ICA) algorithm (Infomax) was applied to the EEG data (see Jessen et al., 2017). For segmentation, the data were segmented into intervals from -200 ms to 1000 ms, time-locked to the onset of the three critical regions. These segments underwent baseline correction using the 200 ms pre-stimulus period. Artifact rejection was performed with amplitude thresholds set at ±100 μV. Non-functioning electrode signals were interpolated from surrounding electrodes prior to averaging (Covey et al., 2024). When calculating the ERPs for each experimental condition per participant, all trials were included in the averaging process, regardless of performance on the post-trial comprehension questions.

Forty-three electrodes were divided into nine regions of interest (ROIs) for analysis following Gosselin and Sabourin (2021). As illustrated in Figure 5.1, the nine ROIs were: the frontal-left region (FL: F5, F3, F1, FC5, FC3, FC1), the frontal-midline region (FM: Fz, FCz), the frontal-right region (FR: F2, F4, F6, FC2, FC4, FC6), the central-left region (CL: C5, C3, C1, CP5, CP3, CP1), the central-midline region (CM: Cz, CPz), the central-right region (CR: C2, C4, C6, CP2, CP4, CP6), the posterior-left region (LP: P5, P3, P1, PO5, PO3, O1), the posterior-midline region (PM: Pz, POz, Oz), and the posterior-right region (PR: P2, P4, P6, PO4, PO6, O2).

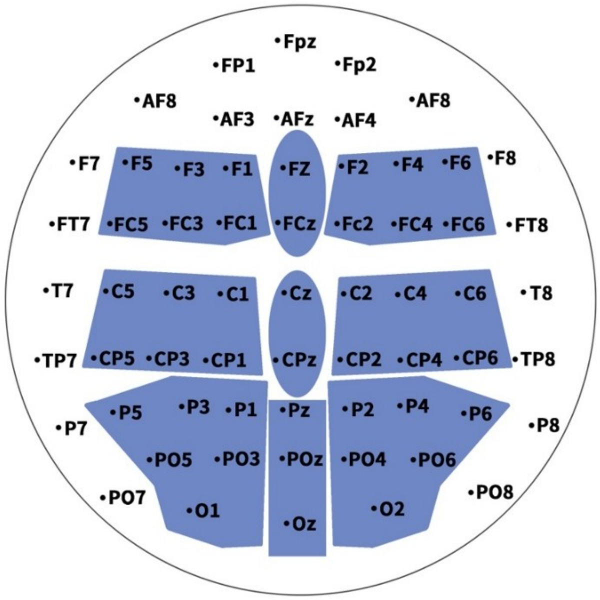


Figure 5.1 Electrode groupings by ROI

5.3.6 Analysis

At each of the three critical regions, the time-windows for statistical analysis were first chosen based on visual inspection of the canonical time windows associated with the relevant ERP components. For example, if a pronounced N400 was observed within the canonical 300-500 ms time-window for either group, this 300-500 ms time-window was selected for analysis. In cases where visual inspection revealed potential but ambiguous components, a sliding window analysis was employed in each group. In other words, separate statistical tests for each 100 ms time-window were performed to precisely identify the onset and duration of the potential significant effects. Mauchly's test was applied to examine the sphericity assumption, and Greenhouse-Geisser corrections were implemented when violations were detected ($p < .05$). When interaction effects (Condition × ROI) reached significance in two consecutive 100 ms intervals, we proceeded with analyses encompassing the broader time-window.

The mean amplitude was the dependent variable and was analyzed using ANOVAs for each critical region. The preliminary three-way ANOVA included two within-subject factors: Condition (plausible vs. implausible for First Verb and Noun Phrase; Vt vs. Prep for Gap Licensor) and ROI (9 regions), and one between-subject factor: Group (L1 vs. L2-High vs. L2-Low). Follow up group-independent two-way ANOVAs incorporated Condition and ROI as within-subject factors. Mauchly's test was also conducted for ROI factor and any interactions involving ROI to examine the sphericity assumption, and Greenhouse-Geisser corrections were applied when violations were detected (p <.05). ROI-specific one-way ANOVAs with Condition as the within-subject factor were conducted in the scenarios where a significant Condition × ROI interactions or a significant main effect of Condition was found in the two-way ANOVA. Additionally, when visual inspection revealed component-like effects (e.g., P600, N400) in certain regions, even without significant overall effects, the one-way ANOVAs were performed to avoid Type II errors. For significant three-way ROI × Condition × Group interactions, which indicates the topographical distribution of the condition effect differs between groups, we first conducted separate Condition × ROI ANOVAs for each group, followed by one-way ANOVAs detecting the ROI specific main effect of Condition. If no significant main or interaction effects were found in separate L2-High and L2-Low analyses, the two groups were combined into a single L2 group to increase statistical power. This analysis approach was consistently applied across all three critical regions. All analyses were conducted using RStudio (Version 4.4.1).

5.4 Results

5.4.1 Behavioral results

For comprehension accuracy to end-of-trial yes/no questions, the L1 group achieved a mean of 93.18% (range: 89.58% - 97.92%) for all sentences (filler and experimental) and 89.77% (range: 83.33% - 95.83%) for experimental sentences only. The L2 group showed a mean accuracy of 87.40% (range: 79.17% - 95.83%) for all sentences and 81.88% (range: 66.67% -

95.83%) for experimental sentences. The high accuracy rate across groups indicate that participants were actively engaging in the tasks for comprehension. No trials were excluded from the ERP analysis based on behavioral performance.

5.4.2 ERP results

First Verb

The ERP analyses were conducted separately for three critical regions. The only effects to be discussed include those involving Condition, Group, and any interactions involving Condition. We first focused on the First Verb region, where parsers were predicted to attempt to resolve the dependency for the first time. Two time-windows were analyzed: 300-500 ms and 500-600 ms. The 300-500 ms window was chosen based on the observation of a more negative-going wave for implausible fillers in L1 speakers in the canonical N400 time-window, as shown in Figure 5.2. While no individual group (L1 vs. L2 High vs. L2 Low) showed significant effects in any 100 ms time-window, combining the two L2 groups revealed a significant Condition × ROI interaction in the 500-600 ms time-window ($F(8,152) = 2.446$, $p = .016$, corrected $p = 0.063$, $\eta p^2 = 0.114$). We selected this time-window to investigate the potential delayed N400-like response in L2 speakers. Table 5.3 summarizes the ANOVA results for the First Verb Region in both time-windows for analysis.

Table 5.3 ANOVA results for the First Verb region

300-500 ms (L1 vs. L2-Combined)						
	Effect	*F*	*df*	*p*	*Sig.*	η^{2p}
Three-way omnibus ANOVA	Condition	4.84	1, 29	0.036	*	0.099
One-way ROI-specific ANOVA in L1	Condition in CM	5.435	1, 10	.042	*	0.352
One-way ROI-specific ANOVA in L2-Combined	Condition in PR	3.304	1, 19	.085	.	0.148

500-600 ms (L1 vs. L2-High vs. L2-Low)						
	Effect	*F*	*df*	*p*	*Sig.*	η^{2p}
Three-way omnibus ANOVA	Condition × ROI	2.873	8, 224	.036	*	0.093
One-way ROI-specific ANOVA in L2-High	Condition in PR	5.51	1, 8	.047	*	0.408

Note: Only significant effects and relevant non-significant effects are reported. Partial eta-squared (η^{2p}) reflects effect size. Significance levels: *** $p < .001$, ** $p < .01$, * $p <. 05$, **.** $p <. 1$ (marginal)

For the 300-500 ms time-window, a preliminary three-way mixed ANOVA with Group (L1, L2-High, L2-Low), Condition (Plausible, Implausible), and ROI (9 levels) was conducted. All main effects and interactions of interest were non-significant. Follow-up two-way ANOVAs (Condition × ROI) for each group separately revealed no significant effects. After combining the L2-High and L2-Low groups into a single L2 group, the three-way mixed ANOVA revealed a significant main effect of Condition ($F(1, 29) = 4.84$, $p = .036$, $\eta p^2 = .099$). The mean amplitudes showed that the L1 group exhibited a larger negativity for implausible compared to plausible conditions (-1.203 µV vs. -0.157 µV, difference = -1.046 µV), while the L2-Combined group showed a smaller difference between conditions (1.470 µV vs. 1.797 µV, difference = -0.327 µV). All remaining effects failed to reach significance. For both groups, the two-way ANOVA analyses revealed only a significant main effect of ROI. Given the significant main effect of Condition in the three-way ANOVA and the presence of clear N400-like patterns in the L1 group's waveforms, I conducted ROI-specific analyses to gauge where the Condition effect occurs. For the L1 group, a significant effect of plausibility was found at the CM region ($F(1,10) = 5.435$, $p = .042$, $\eta G^2 = 0.352$), with implausible fillers eliciting a more negative response (-2.894 µV) than plausible fillers (-0.810 µV, difference = 2.084 µV), consistent with the observed central negativity distribution. It matched the canonical N400 topography and latency. No other regions showed significant effects. For the L2-Combined group, analyses revealed a marginally significant effect at the PR region ($F(1,19) = 3.304$, $p = .085$, $\eta p^2 = 0.148$), with plausible fillers eliciting a more positive response (3.867 µV) than implausible fillers (2.927 µV, difference = 0.939 µV), with a broader posterior positivity distribution. No effects were found in other regions. Figures 5.2 and 5.3 show the grand average ERPs at each ROI and the topographic maps of the two conditions in the 300-500 ms time-window for L1 and L2 groups, respectively. ERP waveforms are plotted across nine scalp ROIs that are arranged from left to right and from anterior to posterior across the scalp. These include FL (frontal-left), FM (frontal-midline), FR (frontal-right), CL (central-left), CM (central-midline), CR (central-right), PL (posterior-left), PM (posterior-midline), and PR (posterior-right). The ERP waveforms are plotted from -200 ms to 1000 ms relative to the onset of the critical word.

L1:

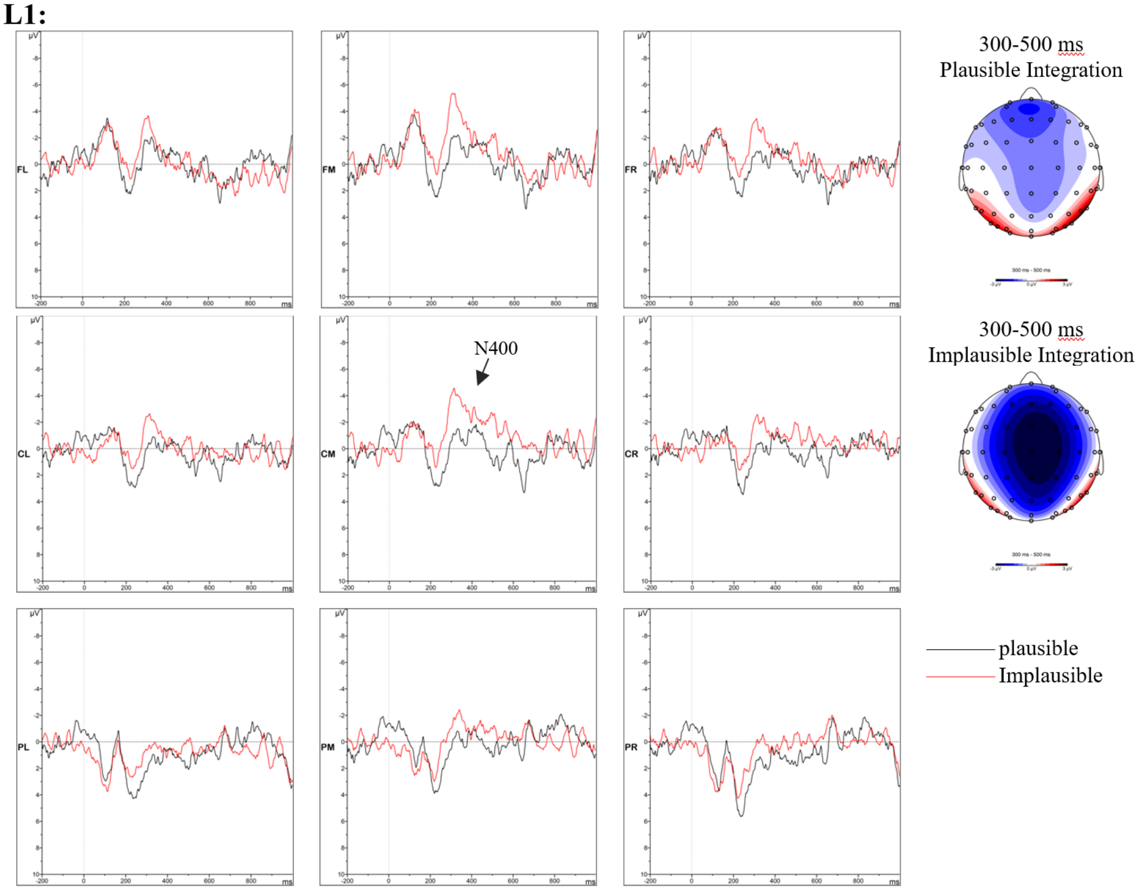


Figure 5.2 L1 group's grand average ERPs at the First Verb region for plausible and implausible conditions for the nine ROIs

L2-Combined:

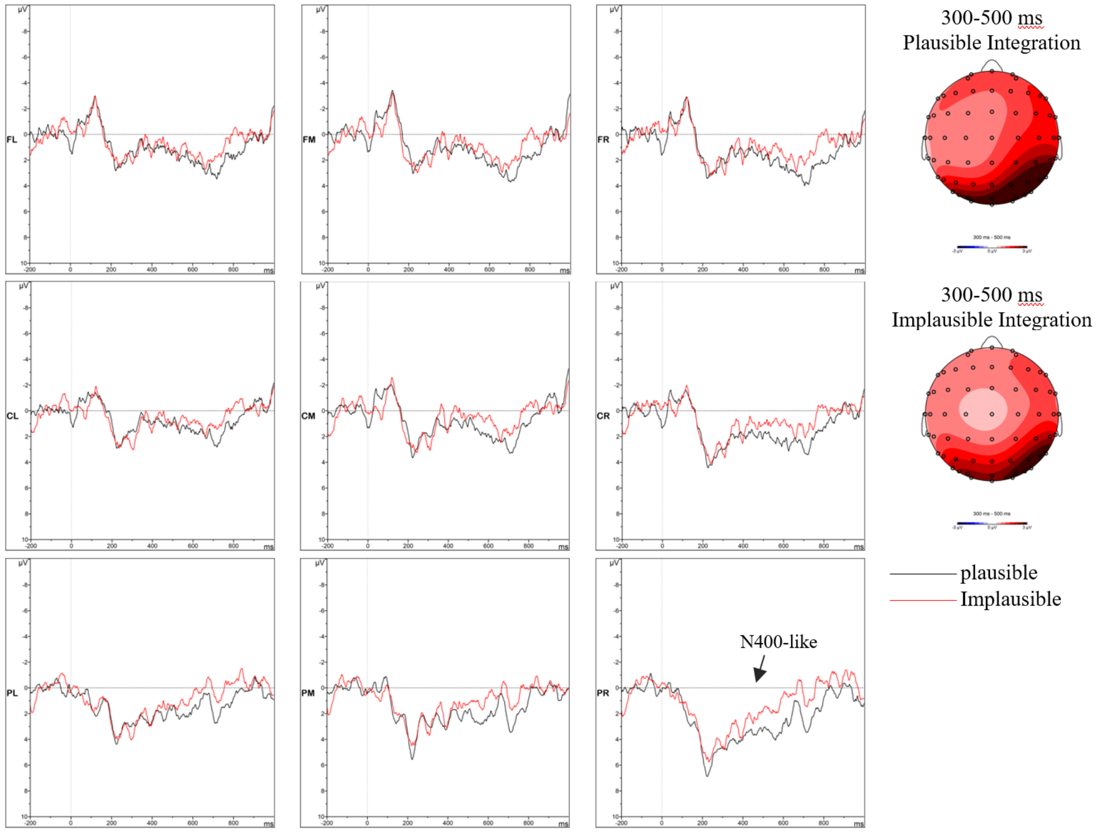


Figure 5.3 L2-Combined group's grand average ERPs at the First Verb region for plausible and implausible conditions for the nine ROIs

For the 500-600 ms time-window, a preliminary three-way mixed ANOVA with three different proficiency groups (L1, L2-High, L2-Low) revealed a significant Condition × ROI interaction (F(8, 224) = 2.873, p = .005, corrected p = .036, ηp^2 = 0.093). Group effect and interactions were non-significant. Mean amplitudes showed more negative responses for implausible conditions across all groups (L1: -0.329 μV vs. 0.556 μV; L2-High: 1.59 μV vs. 2.23 μV; L2-Low: 0.658 μV vs. 1.85 μV). Following the significant interaction (Condition × ROI), I conducted two-way ANOVAs for each group. However, none of the groups showed significant main effects or interactions. Although these analyses revealed no significant effects, visual inspection of the L2-High group's waveforms showed a clear negative-going deflection for implausible conditions, particularly prominent in the PR region, where a marginal N400 effect had been observed earlier in the 300-500 ms window. I then conducted separate one-way ANOVAs examining the Condition effect in each ROI for each group. In the L2-High group, a significant Condition effect was observed in the PR region ($F(1,8) = 5.51$, $p = .047$, $\eta p^2 = 0.408$), with implausible conditions eliciting more negative-going waveforms (1.60 μV) compared to plausible conditions (3.31 μV). This posterior-right distribution was consistent with the pattern observed in the 300-500 ms window, where L2 speakers also showed a marginally significant Condition effect in the PR region. Notably, the effect became more robust for the L2-High group in this later time-window. No other ROIs showed significant Condition effects. No Condition effect was detected in any ROI for the L1 and the L2 Low groups. Figure 5.4 shows the L2-High group's grand average ERPs at the first verb region for plausible and implausible conditions across nine ROIs, and Figure 5.5 shows the same ERPs for L2-Low. The L1 group's grand average ERPs are presented in Figure 5.2.

L2-High:

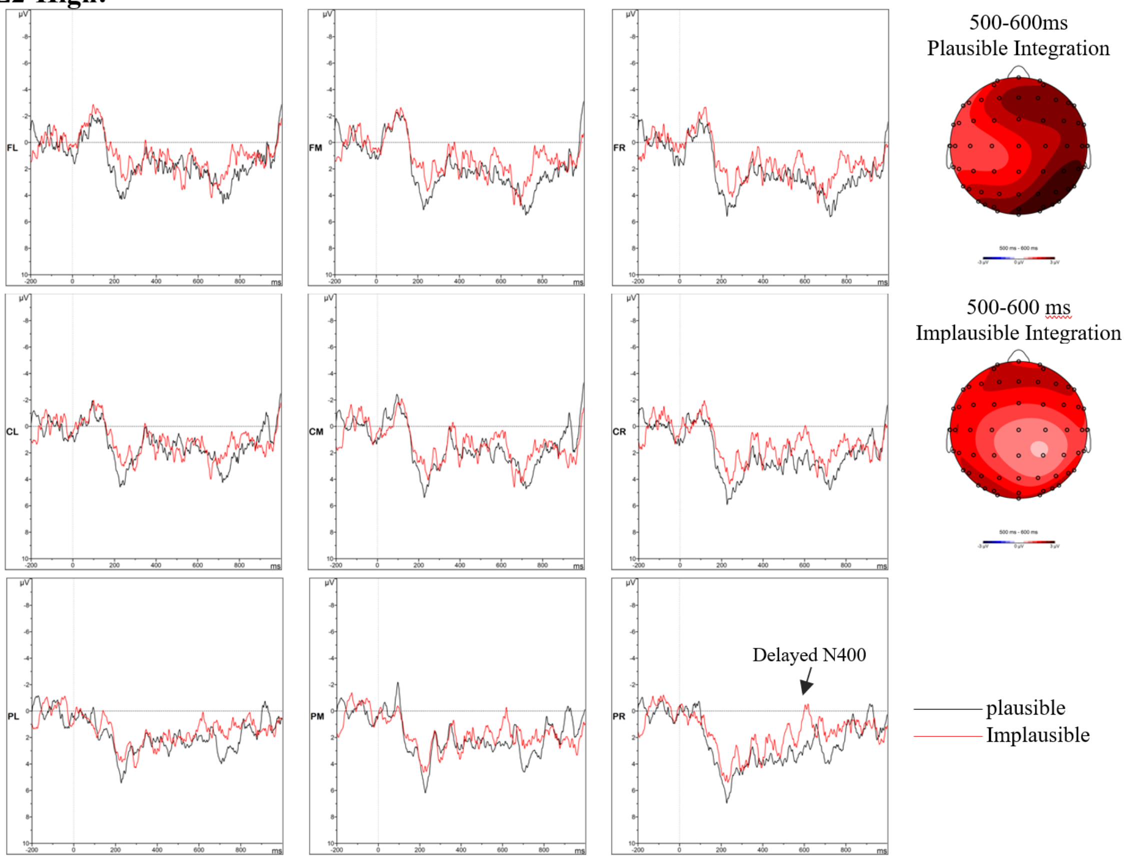


Figure 5.4 L2-High group's grand average ERPs at the First Verb region for plausible and implausible conditions across nine ROIs

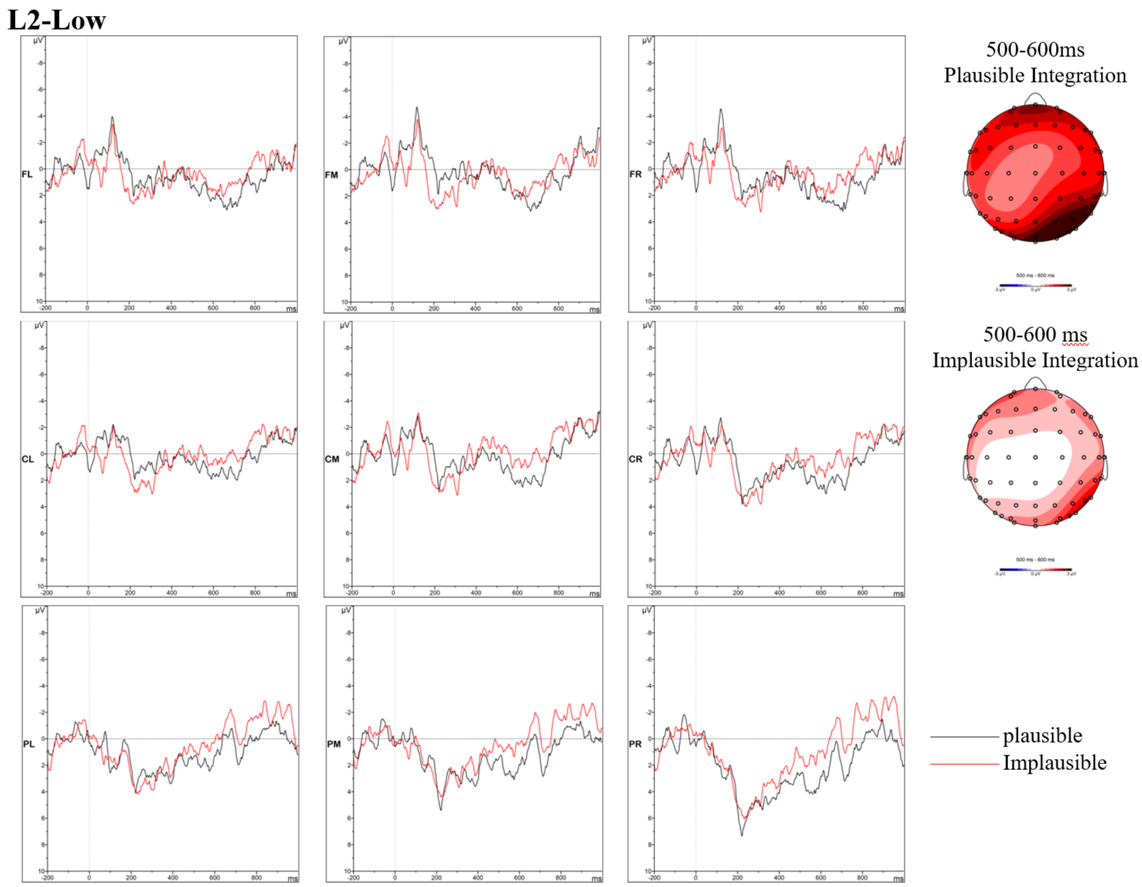


Figure 5.5 L2-Low group's grand average ERPs at the First Verb region for plausible and implausible conditions across nine ROIs

Filled-Gap Noun Phrase

The second critical region was at the Noun Phrase region after the first verb, where the expectation violation or the syntactic reanalysis was expected. The sliding window analyses using 100 ms intervals failed to show any significant effects of Condition or its interactions in any group. Visual inspection of the waveform patterns revealed seemingly larger negative-going waves for plausible conditions than implausible ones across all groups in the 300-500 ms time-window. Also, in the L2-Low group, there appeared to be more positive-going waves for plausible versus implausible conditions. Based on these observations and previous literature on reanalysis and filled-gap effects (e.g., Jessen & Felser, 2019; Covey et al., 2024), we selected two time-windows for analysis: 300-500 ms to examine potential N400/LAN effects, and 550-750 ms to investigate possible P600 effects. Table 5.4 summarizes the significant effects and the observed trend, which may only be seen in the wave patterns and descriptively rather than statistically proved.

Table 5.4 ANOVA results for the Noun Phrase region

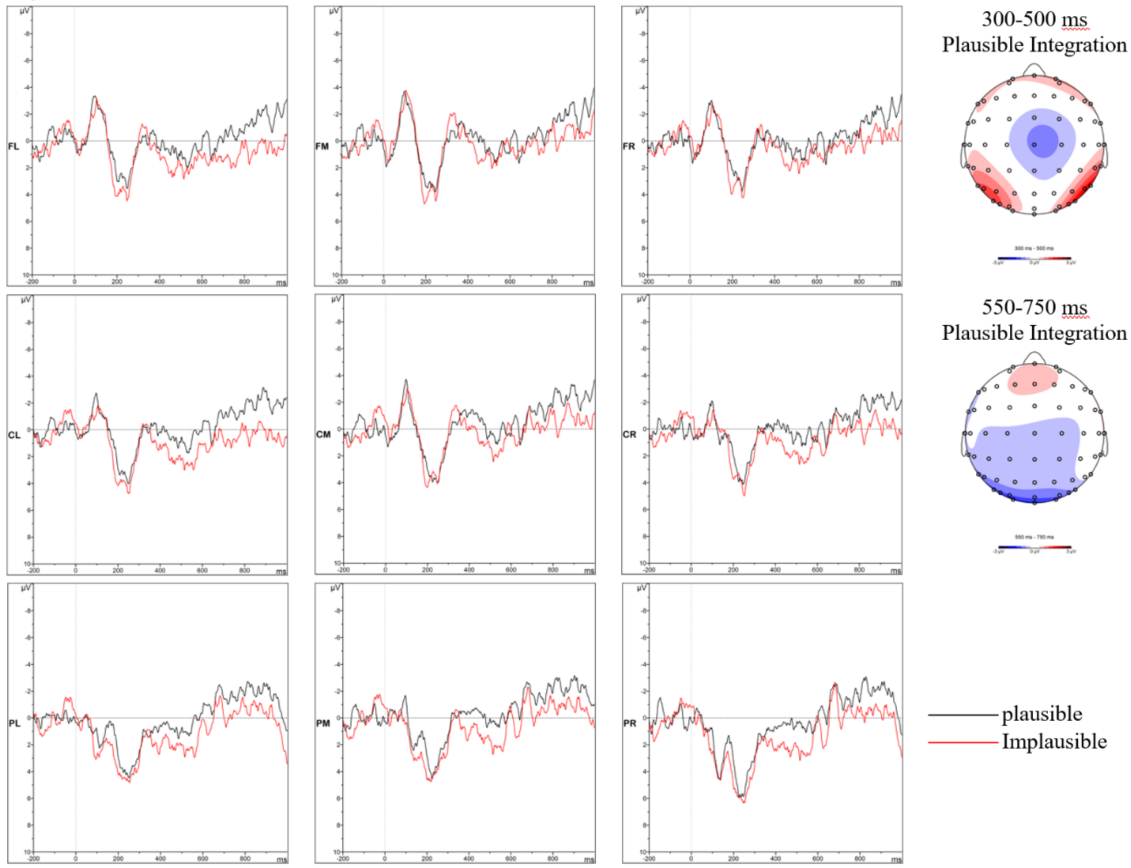
300-500 ms (L1 vs. L2-High vs. L2-Low)						
	Effect	*F*	*df*	*p*	*Sig.*	η^{2p}
One-way ROI-specific ANOVA in L1	Condition in CM *(implausible -1.42 µV vs. plausible 0.564 µV)*	2.848	1, 10	0.122	ns	0.222
One-way ROI-specific ANOVA in L2-High	Condition in FM	4.201	1, 8	.075	**.**	0.344
One-way ROI-specific ANOVA in L2-Low	Condition in FM *(implausible 0.632 µV vs. plausible 1.28 µV)*	0.212	1, 10	.655	ns	0.021
550-750 ms (L1 vs. L2-High vs. L2-Low)						
	Effect	*F*	*df*	*p*	*Sig.*	η^{2p}
One-way ROI-specific ANOVA in L2-Low	Condition in FL	3.443	1, 10	.093	**.**	0.256
One-way ROI-specific ANOVA in L2-Low	Condition in FM	4.418	1, 10	.062	**.**	0.306

Note: Only significant effects and relevant non-significant effects are reported. Partial eta-squared (η^{2p}) reflects effect size. Significance levels: *** $p < .001$, ** $p < .01$, * $p <. 05$, **.** $p <. 1$ (marginal). "ns" indicates a non-significant trend observed in the ERP wave patterns.

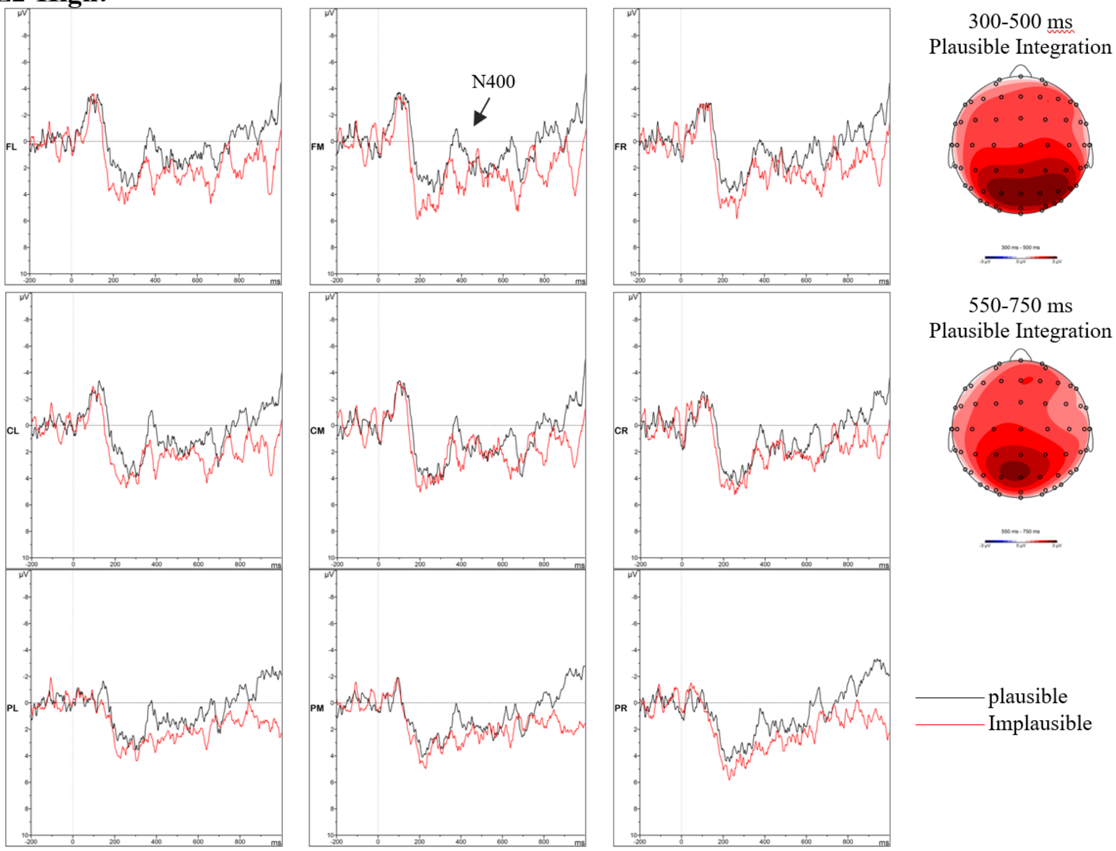
In the 300-500 ms window, the three-way ANOVA revealed no significant effects. After L2-High and L2-Low were combined into a single L2 group, the results remained non-significant. Follow-up separate two-way ANOVAs (Condition × ROI) for each group revealed no evidence of an effect of Condition. However, given the clear negative-going waveforms for plausible conditions across groups and the different topographic distributions between groups (central negativity for L1 speakers and posterior positivity for L2 speakers), separate one-way ANOVAs examining the condition effect at each ROI within each group were conducted. Though not statistically significant, L1 and low-proficient L2 speakers showed more negative amplitudes for plausible than for implausible conditions across all ROIs. Only the L2-High group showed a marginally significant condition effect at the frontal-middle region (FM: $F(1,8) = 4.201$, $p = .075$, $\eta p^2 = 0.344$), with more negative amplitudes for plausible compared to implausible conditions (1.24 μV vs. 2.35 μV). Figure 5.6 showed the grand average ERPs at the Noun Phrase region across all groups (L1, L2 High, and L2 Low).

Analysis of the 550-750 ms time-window using three-way mixed ANOVA yielded no significant effects. The mean amplitudes showed that L1 and L2-High groups exhibited more negative-going waves for plausible than implausible conditions (L1: -0.452 μV vs. 0.583 μV; L2-High: 1.21 μV vs. 2.17 μV), while L2-Low showed the opposite pattern (2.06 μV vs. 1.59 μV). No other main effects or interactions reached significance. In the follow-up separate two-way ANOVAs (Condition × ROI), no significant main effects or interactions were found in any group. Given that visual inspection revealed component-like effects, P600, in the frontal areas in the L2-Low group, we conducted ROI-specific one-way ANOVAs to explore these distinct wave patterns. The analyses revealed marginally significant condition effects in the L2-Low group at FL ($F(1,10) = 3.443$, $p = .093$, $\eta p^2 = 0.256$) and FM ($F(1,10) = 4.418$, $p = .062$, $\eta p^2 = 0.306$), showing more positive amplitudes for plausible than implausible conditions (FL: 3.49 μV vs. 2.23 μV; FM: 3.55 μV vs. 1.99 μV). The L1 and L2-High groups showed descriptively more negative amplitudes for plausible than implausible conditions across ROIs, though no region reached significance.

L1:


FL
FM
FR
CL
CM
CR
PL
PM
PR
300-500 ms
Plausible Integration
550-750 ms
Plausible Integration
plausible
Implausible

L2-High:


FL
FM
FR
CL
CM
CR
PL
PM
PR
N400
300-500 ms
Plausible Integration
550-750 ms
Plausible Integration
plausible
Implausible

L2-Low:

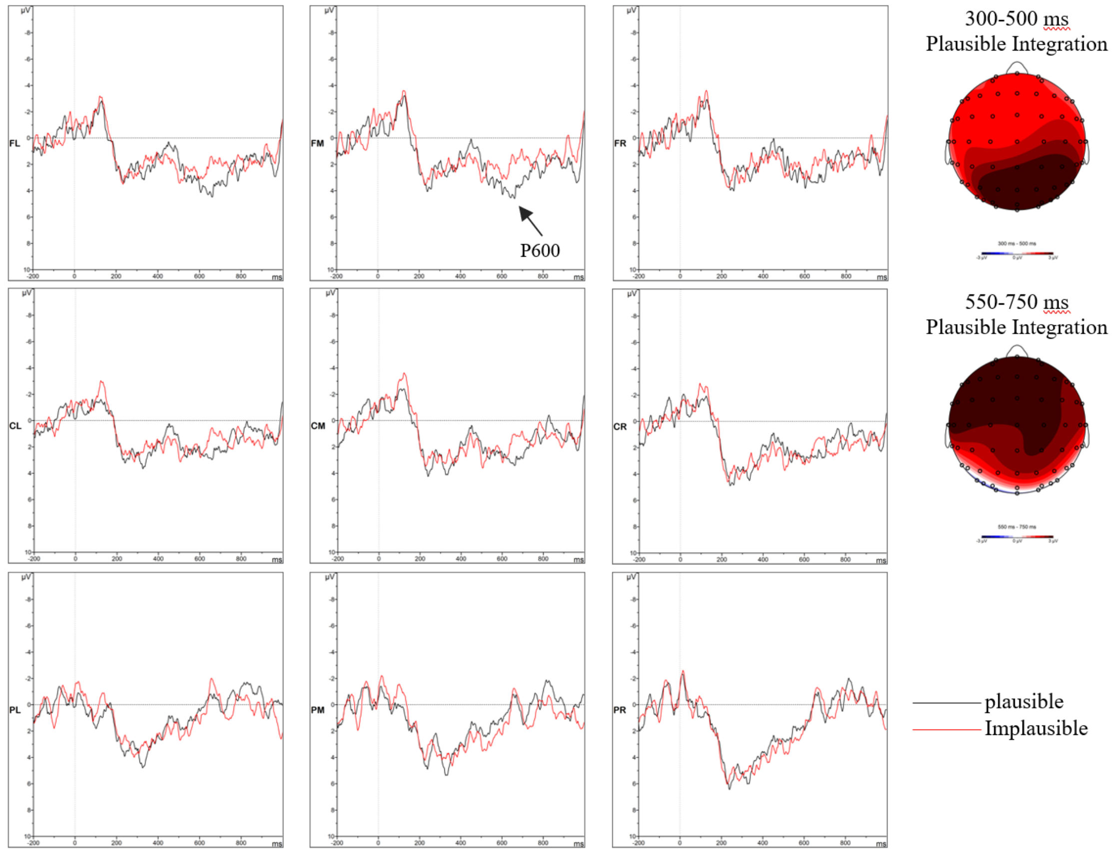


Figure 5.6 Grand average ERPs at the Noun Phrase region for all groups (L1, L2-High, and L2-Low)

Gap Licensor

The third critical region was at the Gap Licensor region, where the real syntactic gap was introduced. Three time-windows were selected for analysis. The early time-window (250-320 ms) was selected based on visual inspection and previous literature showing that function words elicit an N280 component different from verbs at left anterior sites (Luck, 2014). Both the 300-500ms and 500-900 ms time-windows were selected based on visual inspection of the waveforms and the consistent condition effects and interactions observed in the 100 ms timeline analyses. These two time-windows were also consistent with the canonical time periods for LAN/N400 and P600 components. Table 5.5 summarizes the ANOVA results for the Gap Licensor region.

Table 5.5 ANOVA results for the Gap Licensor region

250-320 ms (L1 vs. L2-High vs. L2-Low)						
	Effect	*F*	*df*	*p*	*Sig.*	η^{2p}
Three-way omnibus ANOVA	Condition	6.277	1, 28	.018	*	0.183
	Group × ROI	1.860	16, 224	.025	*	0.117
	Condition × ROI	14.144	8, 224	<.001	***	0.336
Two-way ANOVA in L1	Condition × ROI	3.933	8, 80	.025	*	0.012
Two-way ANOVA in L2-High	Condition × ROI	3.147	8, 64	.048	*	0.019
Two-way ANOVA in L2-Low	Condition	6.720	1, 10	.027	*	0.124
	Condition × ROI	8.830	8, 80	.002	**	0.083
One-way ROI-specific ANOVA in L1	Condition in FL	6.381	1, 10	.030	*	0.389
	Condition in FM	10.257	1, 10	.009	**	0.506
	Condition in FR	8.634	1, 10	.015	*	0.463
One-way ROI-specific ANOVA in L2-High	Condition in FL	5.562	1, 8	.046	*	0.058
One-way ROI-specific ANOVA in L2-Low	Condition in FL	16.167	1, 10	.002	**	0.618
	Condition in FM	9.375	1, 10	.012	*	0.484
	Condition in FR	13.317	1, 10	.004	**	0.571
	Condition in CL	9.343	1, 10	.012	*	0.483
	Condition in CM	5.664	1, 10	.039	*	0.362
	Condition in CR	5.125	1, 10	.047	*	0.339
300-500 ms (L1 vs. L2-High vs. L2-Low)						
	Effect	*F*	*df*	*p*	*Sig.*	η^{2p}
Three-way omnibus ANOVA	Condition	6.634	1, 28	.016	*	0.192
	Group × ROI	2.891	16, 224	.039	*	0.171
	Condition × ROI	9.671	8, 224	<.001	***	0.257
Two-way ANOVA in L1	Condition	4.647	1, 10	.057	.	0.317
Two-way ANOVA in L2-High	Condition × ROI	3.494	8, 64	.040	*	0.304
Two-way ANOVA in L2-Low	Condition	3.508	1, 10	.091	.	0.264
	Condition × ROI	5.131	8, 80	.021	*	0.339
One-way ROI-specific ANOVA in L1	Condition in FL	8.237	1, 10	.017	*	0.452
	Condition in FM	12.263	1, 10	.006	**	0.551
	Condition in FR	8.899	1, 10	.014	*	0.471

One-way ROI-specific ANOVA in L2-Low	Condition in FL	11.388	1, 10	.007	**	0.532
	Condition in FM	7.466	1, 10	.021	*	0.427
	Condition in FR	10.615	1, 10	.009	**	0.515
500-900 ms (L1 vs. L2-High vs. L2-Low)						
	Effect	*F*	*df*	*p*	*Sig.*	η^{2p}
Three-way omnibus ANOVA	Condition	11.405	1, 28	.002	**	0.289
	Condition × ROI	16.135	8, 224	<.001	***	0.366
Two-way ANOVA in L1	Condition	15.905	1, 10	.003	**	0.614
	Condition × ROI	3.098	8, 80	.079	.	0.236
Two-way ANOVA in L2-High	Condition × ROI	4.237	8, 64	.021	*	0.346
Two-way ANOVA in L2-Low	Condition	4.185	1, 10	.068	.	0.295
	Condition × ROI	10.440	8, 80	<.001	***	0.511
One-way ROI-specific ANOVA in L1	Condition in FL	18.662	1, 10	.002	**	0.651
	Condition in FM	29.741	1, 10	<.001	***	0.748
	Condition in FR	20.052	1, 10	.001	**	0.667
	Condition in CL	16.407	1, 10	.002	**	0.621
	Condition in CM	12.799	1, 10	.005	**	0.561
	Condition in CR	13.095	1, 10	.005		0.567
	Condition in PL	5.631	1, 10	.039		0.360
	Condition in PM	5.379	1, 10	.043		0.350
One-way ROI-specific ANOVA in L2-Low	Condition in FL	9.882	1, 10	.010		0.497
	Condition in FM	5.974	1, 10	.035		0.374
	Condition in FR	7.894	1, 10	.018		0.441
	Condition in CL	5.928	1, 10	.035		0.372
	Condition in CR	5.367	1, 10	.043		0.349

Note: Only significant effects and relevant non-significant effects are reported. Partial eta-squared (η^{2p}) reflects effect size. Significance levels: *** $p < .001$, ** $p < .01$, * $p <. 05$, **.** $p <. 1$ (marginal).

In the 250-320 ms time-window, the three-way ANOVA revealed a significant main effect of Condition ($F(1,28) = 6.277$, $p = .018$, $\eta p^2 = 0.183$) and a significant Condition × ROI interaction ($F(8,224) = 14.144$, $p < .001$, corrected $p < .001$, $\eta p^2 = 0.336$). While a Group × ROI interaction was also observed ($F(16,224) = 1.860$, $p = .025$, $\eta p^2 = 0.117$), no significant main effect of Group or interactions involving Condition were found. The two-way ANOVAs (Condition × ROI) were then conducted for each group. For the L1 group, this analysis revealed a significant Condition ×

ROI interaction ($F(8,80) = 3.933$, $p < .001$, corrected $p = .025$, $\eta p^2 = 0.012$). The L2-High group also showed a significant Condition × ROI interaction ($F(8,64) = 3.147$, $p = .005$, corrected $p = .048$, $\eta p^2 = 0.019$). The L2-Low group demonstrated a significant main effect of Condition ($F(1,10) = 6.720$, $p = .027$, $\eta p^2 = 0.124$) and a significant Condition × ROI interaction ($F(8,80) = 8.830$, $p < .001$, corrected $p = .002$, $\eta p^2 = 0.083$). Given significant Condition × ROI interactions in all groups, additional one-way ANOVAs were conducted at each ROI to pinpoint which regions the condition effects were. Follow-up one-way ANOVAs revealed different patterns across groups. The L1 group showed significant condition effects in frontal regions: FL ($F(1,10) = 6.381$, $p = .030$, $\eta p^2 = 0.389$), FM ($F(1,10) = 10.257$, $p = .009$, $\eta p^2 = 0.506$), and FR ($F(1,10) = 8.634$, $p = .015$, $\eta p^2 = 0.463$). The L2-High group showed a significant condition effect only in FL ($F(1,8) = 5.562$, $p = .046$, $\eta p^2 = 0.058$). The L2-Low group exhibited the most widespread effects, with significant condition differences in frontal and central regions: FL ($F(1,10) = 16.167$, $p = .002$, $\eta p^2 = 0.618$), FM ($F(1,10) = 9.375$, $p = .012$, $\eta p^2 = 0.484$), FR ($F(1,10) = 13.317$, $p = .004$, $\eta p^2 = 0.571$), CL ($F(1,10) = 9.343$, $p = .012$, $\eta p^2 = 0.483$), CM ($F(1,10) = 5.665$, $p = .039$, $\eta p^2 = 0.362$), and CR ($F(1,10) = 5.125$, $p = .047$, $\eta p^2 = 0.339$).

For the 300-500 ms time-window, the three-way analysis revealed a significant main effect of Condition ($F(1,28) = 6.634$, $p = .016$, $\eta p^2 = 0.192$) and a significant Condition × ROI interaction ($F(8,224) = 9.671$, $p < .001$, corrected $p = < .001$, $\eta p^2 = 0.257$). A significant Group × ROI interaction was also observed ($F(16,224) = 2.891$, $p < .001$, corrected $p = .039$, $\eta p^2 = 0.171$), though this effect was not of primary interest. The three-way interaction of Group × Condition × ROI was not significant ($F(16,224) = 1.063$, $p = .392$, $\eta p^2 = 0.071$). After the two-way ANOVAs for each group, the L1 group showed a marginally significant main effect of Condition ($F(1,10) = 4.647$, $p = .057$, $\eta p^2 = 0.317$), while Condition × ROI interaction ($F(8,80) = 1.645$, $p = .125$, corrected $p = .200$, $\eta p^2 = 0.141$) did not reached significance. The L2-High group demonstrated only a significant Condition × ROI interaction ($F(8,64) = 3.494$, $p = .002$, corrected $p = .040$, $\eta p^2 = 0.304$), with no significant main effect of Condition ($F(1,8) = 0.313$, $p = .591$, $\eta p^2 = 0.038$). The L2-Low group showed a significant Condition × ROI interaction ($F(8,80) = 5.131$, $p < .001$, corrected $p = .021$, $\eta p^2 = 0.339$). The main effect of Condition approached significance but not reached significance ($F(1,10) = 3.508$, $p = .091$, $\eta p^2 = 0.260$). For L1 speakers, one-way ANOVAs revealed significant effects of Condition in three frontal regions: FL ($F(1,10) = 8.237$, $p = .017$, $\eta p^2 = 0.452$), FM ($F(1,10) = 12.263$, $p = .006$, $\eta p^2 = 0.551$), and FR ($F(1,10) = 8.899$, $p = .014$, $\eta p^2 = 0.471$). No significant effects of Condition were found in any of the ROIs in the L2-High group. In the L2-Low Group, significant effects of Condition were found in three frontal-central regions: FL ($F(1,10) = 11.388$, $p = .007$, $\eta p^2 = 0.532$), FM ($F(1,10) = 7.466$, $p = .021$, $\eta p^2 = 0.427$), and FR ($F(1,10) = 10.615$, $p = .009$, $\eta p^2 = 0.515$).

For the 500-900 ms time-window, the preliminary three-way ANOVA analysis revealed a significant main effect of Condition ($F(1,28) = 11.405$, $p = .002$, $\eta p^2 = 0.289$). A significant Condition × ROI interaction was also found ($F(8,224) = 16.135$, $p < .001$, corrected $p < .001$, $\eta p^2 = 0.366$). Neither the main effect of Group nor any interactions involving Group reached significance. After the two-way ANOVAs for each group in the 500-900 ms time-window, L1 group showed a significant main effect of Condition ($F(1,10) = 15.905$, $p = .003$, $\eta p^2 = 0.614$) and a significant Condition × ROI interaction ($F(8,80) = 3.098$, $p = .004$, corrected $p = .079$, $\eta p^2 = 0.236$). The L2-High group showed a significant Condition × ROI interaction ($F(8,64) = 4.237$, $p < .001$, corrected $p = .021$, $\eta p^2 = 0.346$). The L2-Low group showed a marginally significant Condition effect ($F(1,10) = 4.185$, $p = .068$, $\eta p^2 = 0.295$) and a significant Condition × ROI interaction ($F(8,80) = 10.440$, $p < .001$, corrected $p < .001$, $\eta p^2 = 0.511$). One-way ANOVAs revealed robust Condition effects in L1 group across multiple regions: FL ($F(1,10) = 18.662$, $p = .002$, $\eta p^2 = 0.651$), FM ($F(1,10) = 29.741$, $p < .001$, $\eta p^2 = 0.748$), FR ($F(1,10) = 20.052$, $p = .001$, $\eta p^2 = 0.667$), CL ($F(1,10) = 16.407$, $p = .002$, $\eta p^2 = 0.621$), CM ($F(1,10) = 12.799$, $p = .005$, $\eta p^2 = 0.561$), CR ($F(1,10) = 13.095$, $p = .005$, $\eta p^2 = 0.567$), PL ($F(1,10) = 5.631$, $p = .039$, $\eta p^2 = 0.360$), and PM ($F(1,10) = 5.379$, $p = .043$, $\eta p^2 = 0.350$). In the L2-Low group, significant effects emerged in FL ($F(1,10) = 9.882$, $p = .010$, $\eta p^2 = 0.497$), FM ($F(1,10) = 5.974$, $p = .035$, $\eta p^2 = 0.374$), FR ($F(1,10) = 7.894$, $p = .018$, $\eta p^2 = 0.441$), CL ($F(1,10) = 5.928$, $p = .035$, $\eta p^2 = 0.372$), and CR ($F(1,10) = 5.367$, $p = .043$, $\eta p^2 = 0.349$). No significant effects were found in the L2-High group.

Figures 5.7-5.9 present the grand average ERPs and topographic maps at the Gap Licensor region for the L1, L2-High, and L2-Low groups, respectively. Each figure displays the ERP waveforms across all ROIs and the corresponding topographic distributions across the three time-windows (250-320 ms, 300-500 ms, and 500-900 ms).

L1:

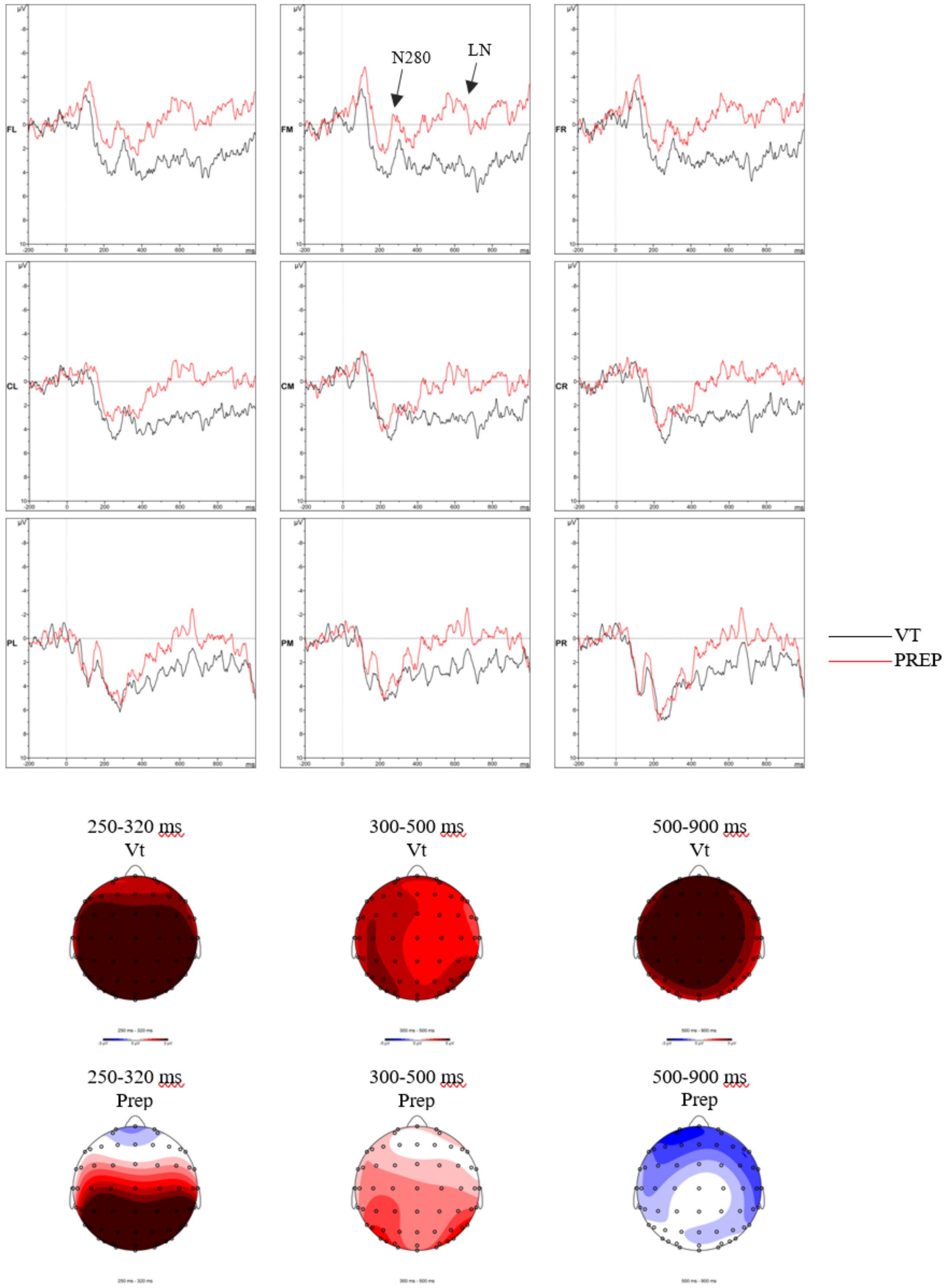


Figure 5.7 L1 group's the grand average ERPs and topographic maps of different time-windows at the Gap Licensor region

L2-High:

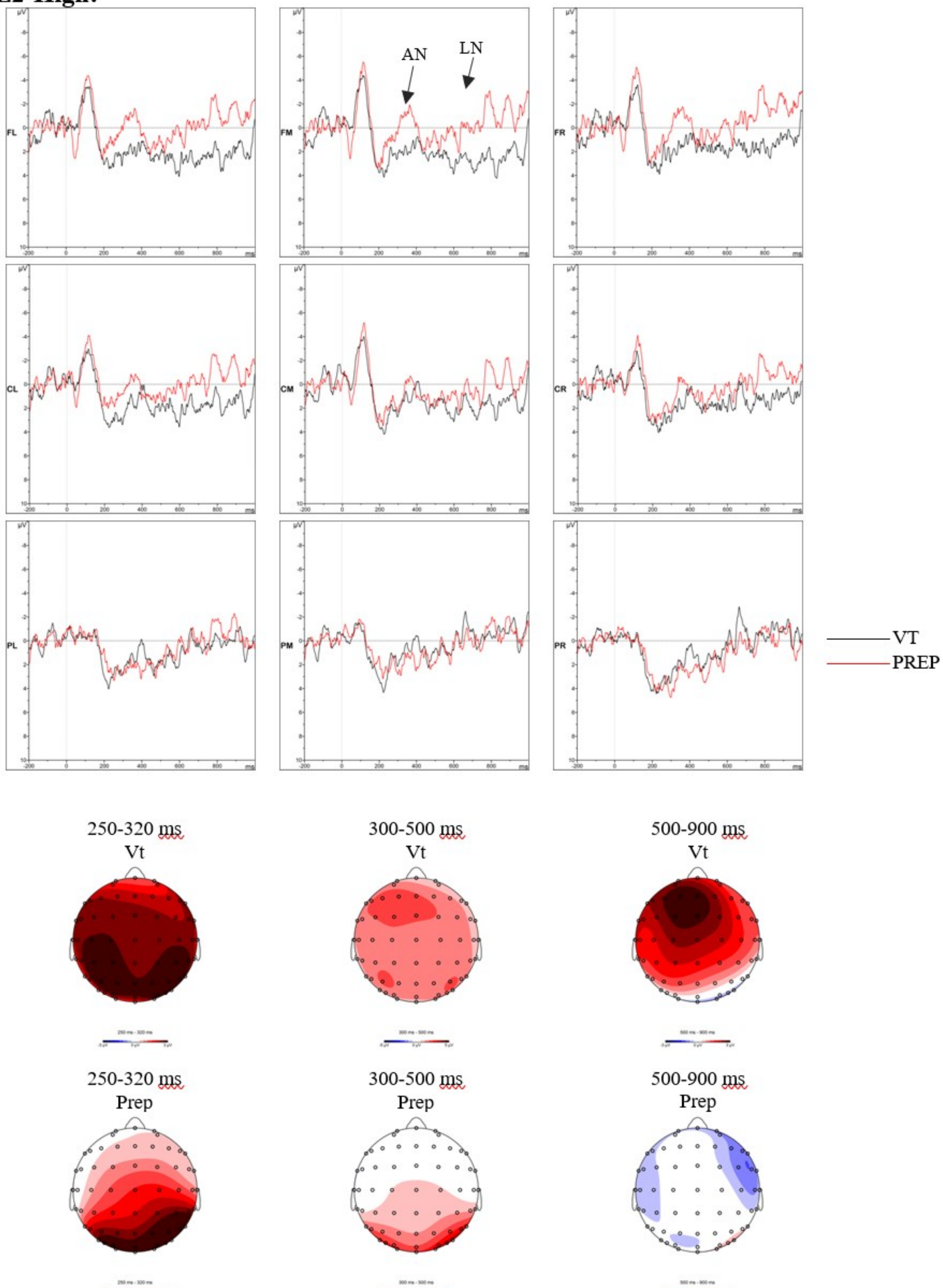


Figure 5.8 L2-High group's the grand average ERPs and topographic maps of different time-windows at the Gap Licensor region

L2-Low:

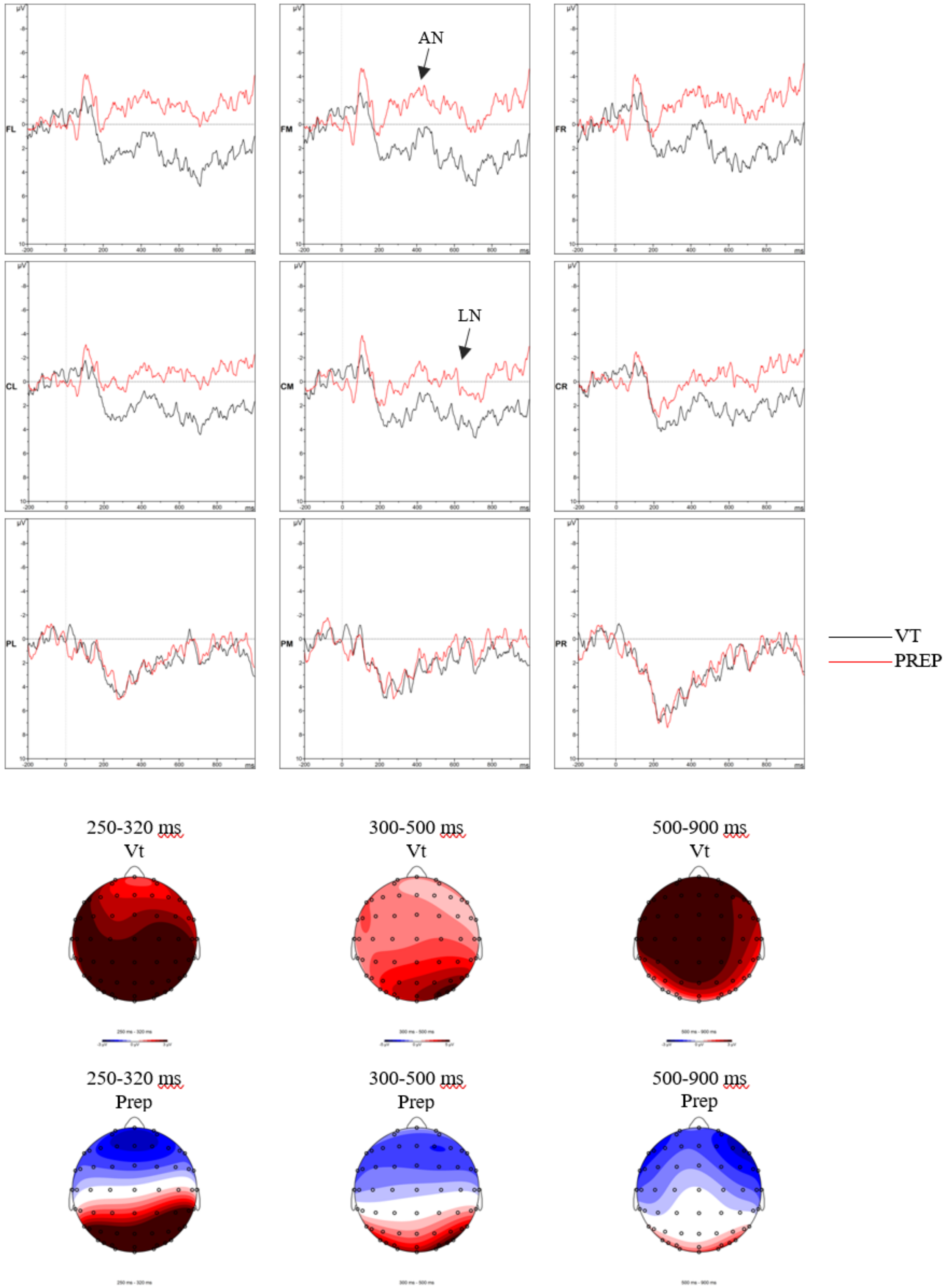


Figure 5.9 L2-Low group's the grand average ERPs and topographic maps of different time-windows at the Gap Licensor region

5.5 Discussion

This current ERP study examined the neural mechanisms underlying filler-gap dependency processing in both native English speakers and L2 learners to address the central question of this dissertation about whether L2 speakers employ parsing mechanisms similar to or different from those used by native speakers. Two key indicators of L1-L2 processing difference would be: a main effect of Group, which would indicate overall processing differences between L1 and L2 speakers, and a significant Condition × Group interaction, which would suggest that L1 and L2 speakers respond differently to the experimental manipulation. The results of the current study revealed that, across all three critical regions (First Verb, Noun Phrase, Gap Licensor), neither of these effects reached significance. While we did find some the Group × ROI interaction at the Gap Licensor region for the N280 time-window, this significance would only suggest differences in the topographic distribution of neural activity rather than fundamental differences in how the groups process the experimental manipulation. The discussion of the results found in each critical region is as follow.

First Verb

At this region, we predicted that both L1 and L2 speakers would show N400 effects at the first verb when encountering implausible fillers, though L2 speakers might exhibit different patterns in terms of timing, scalp distribution, or amplitude strength. The results aligned with this prediction.

Given a significant main effect of Condition in the 300-500 ms window and no significant Group or Condition × Group interactions, our results suggested that both groups demonstrated sensitivity to plausibility, as indexed by the N400. However, their sensitivity to plausibility manifested with distinct temporal and topographical characteristics. Specifically, L1 speakers showed an N400 in the canonical central-midline region, with implausible fillers eliciting a more negative response, while L2 speakers showed only a marginally significant effect in the posterior-right region. Given its marginal significance, we interpreted this as an N400-like effect. In the later window (500-600 ms), we observed a significant Condition × ROI interaction, with all groups (L1, L2-High, L2-Low) showing more negative-going waves for implausible conditions. Specifically, L2-High group showed a significant N400 effect in the posterior-right region in the 500-600 ms time-window. Importantly, while the L2-High group's response in the posterior-right region was only marginal during the 300-500 ms window, it strengthened to become a significant N400 effect in this later window.

In line with previous studies (e.g., Jessen & Felser, 2019), both L1 and L2 speakers showed an N400 at the first potential gap, which was interpreted to support active gap-filling. However, for L2 speakers, particular in the L2-High group, we observed a temporal progression in their response: from an N400-like effect in the early window (300-500 ms) to a significant N400 effect during the later window (500-600 ms). This aligns with previous findings of attenuated and delayed but qualitatively similar semantic processing in L2 speakers (e.g., Williams et al., 2001; Jessen & Felser, 2019; Dallas et al., 2013). As discussed in Chapter 3, based on the trace-reactivation hypothesis, the presence of plausibility sensitivity at the First Verb region suggests predictive processing is taking place, since this region precedes the actual syntactic gap. The fact that participants attempt dependency resolution at this point indicates they are actively predicting potential gap sites. The attenuated N400-like effect observed in the 300-500 ms time-window is consistent with the RAGE Hypothesis (Grüter & Rohde, 2013, 2021), which posits that L2 speakers exhibit diminished predictive processing. In our case, although L2 speakers could predict a forthcoming gap after the verb, the prediction was less effective, resulting in reduced sensitivity to semantic information. Additionally, considering the parsing mechanism specifically under active gap filling, according to Wagers and Phillips's dual-process account (Wagers & Phillips, 2014), active gap-filling would require maintained syntactic cues to drive the process, followed by retrieval of semantic information to complete the dependency resolution. Kaan's (2014) review discusses a range of individual factors, such as frequency information, accuracy of word-related knowledge, and motivation, that contribute to variability in predictive processing in native speakers; these individual factors similarly influence L2 processing. Therefore, L2 speakers' attenuated N400 effect in the 300-500 ms region could be attributed to their reduced ability to generate expectations or their relatively limited vocabulary capacity, which hampers their ability to retrieve semantic information of the filler during the prediction process.

The delayed N400 effect we observed in the 500-600 ms time-window reflects a broader pattern in L2 processing research, where non-native speakers exhibit slower sentence processing and delayed effects compared to native speakers (e.g., Dussias & Pinar, 2010; Williams, 2001; Fernandez et al., 2020). This processing speed difference is well-illustrated in Fernandez and colleagues' (2020) eye-tracking study, which examined real-time comprehension of *wh*-dependencies. They tracked participants' ability to anticipate patient/theme while listening to *wh*-questions such as "Who did the wolf attack?" at varying speech rates. Their three groups for comparison include the young native English speakers (aged 18-24), older native speakers (aged 40-75), and German-L1 English learners. Results revealed clear processing speed differences: while young native speakers could predict up to 5.5 syllables per second and older natives up to 4.5 syllables per second, L2 speakers only showed predictive processing at the slowest rate of 3.5 syllables per second. This demonstrates that when actively constructing filler-gap dependencies, L2 speakers require more processing time than native speakers to generate predictions.

As the delayed N400 was only significant for the L2-High group, proficiency appeared to be a factor in making L2 processing more native-like. This proficiency effect is consistent with Dallas et al.'s (2013) findings, where only L2 speakers with high proficiency yielded N400 at the verb for implausible integration. However, the difference between the present study and Dallas and colleagues' study is that they found this sensitivity to plausibility only in simpler, no-gap sentences for high-proficiency L2 speakers. The effect did not reach significance in their more complex filler-gap sentences. While these results highlight the importance of proficiency, we must approach the relationship between proficiency and prediction cautiously. The differences in prediction abilities between L1 and L2 speakers may not simply be attributed to general language proficiency but could also reflect differences in how words are represented in their mental lexicon. It is possible that our high-proficiency participants had more expansive vocabulary and reading skills that made their lexical representations more robust, thus facilitating the retrieval of semantic information.

Our findings can be interpreted through Pickering and Gambi's (2018) prediction-by-production model, which proposes two distinct prediction mechanisms. The attenuated N400-like effect in L2 speakers during the 300-500 ms window likely reflects their initial reliance on prediction-by-association, an automatic but less precise mechanism. Then the significant N400 effects only in high-proficiency L2 speakers during the later 500-600 ms window suggests their eventual engagement with prediction-by-production, a more resource-intensive mechanism that enables precise predictions. This finding may suggest that while L2 speakers can utilize predictive mechanisms, they require more time and cognitive resources than L1 speakers. Further evidence comes from the topographic distribution. Similar to Jessen and Felser (2019), we found different topographic distributions between groups, with L2 effects being more posterior. This pattern potentially reflects additional processing costs in integrating filler-gap dependencies in L2 speakers.

Thus far, it appears that both L1 and L2 speakers engage in active gap filling, as evidenced by the main effect of Condition and absence of Group effects or interactions. In other words, both groups seem to predict a syntactic gap shortly after encountering the verb. However, the efficiency of using semantic information at this position differs between groups. Specifically, semantic information appears to be effectively involved in predictive processing for L1 speakers but less effective and delayed in L2 speakers. This efficiency of prediction appears to improve with higher proficiency, probably because high-proficiency speakers have better vocabulary capacity or more cognitive resources for the prediction-by-production stage than L2 low speakers.

The attenuated and delayed N400 in L2 speakers undermines previous research that claimed an optimistic view about L2 speakers' ability to use semantic information. While accounts in the field have recognized L2 speakers' less robust and less immediate use of syntactic information, they appear to argue that lexical semantic information was not problematic for L2 speakers, and that L2 speakers may prioritize semantic non-grammatical information over grammatical cues during language processing (Clahsen & Felser, 2006, 2018; Hopp, 2022). Even as the updated SSH acknowledged L2 speakers' capacity to utilize grammatical information, it maintained that L2 speakers can be more strongly guided by semantic information. However, the findings of the present study suggest that even for semantic information, L2 speakers show less automaticity, at least regarding the deployment of semantic information in the prediction processing of filler-gap dependencies. While native speakers exhibited immediate sensitivity to semantic constraints at the first verb, L2 speakers showed an attenuated and delayed response, indicating reduced automaticity in processing semantic information.

Noun Phrase

This region was the first potential gap where participants met strong evidence showing that reanalysis would be required. In the discussion about the predictions, we mentioned two possible indicators of reanalysis. On the one hand, the reanalysis could be reflected by a reversed plausibility effect that occurs when the initial incremental plausible integration at the gap is impeded, requiring re-evaluation of the syntactic integration, indexed by a P600. Note that we use "incremental" here because the gap has already been reached here and thus the syntactic reanalysis at this region could only provide evidence for incremental processing. On the other hand, the reanalysis could be reflected by a reversed plausibility effect that arises from the collapse of a previously established expectation for a syntactic gap. In this case, participants would be surprised at the contextually unexpected word. This could be indexed by an N400, as less predictability would result in larger N400 amplitude (see Federmeier, 2007 for a review). Our findings suggested an N400 effect only in the L2-High group and a P600 effect only in the L2-Low group in the one-way ANOVAs conducted separately in each region for each group.

The broadly distributed negative-going wave for plausible conditions in the L2-High group could be interpreted as an N400 component, suggesting that semantic information was employed during predictive processing. With plausible match of filler and the verb, L2-High speakers had stronger prediction for an actual gap and experienced greater expectation collapse, eliciting N400. In contrast, in implausible conditions, the semantic mismatch between the filler and verb might have already signaled that a gap was unlikely, resulting in weaker predictions and thus no significant N400. This is consistent with the finding at First Verb: even though the sensitivity to semantic mismatch was attenuated and delayed compared to native speakers, this semantic information was indeed involved, making the prediction of the gap larger for the plausible condition, leading to more surprise at the actual gap. Also, the broadly distributed effect reflects the less localized and less automated processing nature typical of L2 speakers.

However, despite descriptively larger negative amplitudes for plausible conditions in L1 speakers, the expected significant N400 effect was not found. This pattern aligns with Jessen & Felser (2019), who also found no effects at the object noun position in both L1 and L2 participants, despite their showing robust N400 effects at the initial verb. However, our findings differ from Covey et al. (2024), who found clear N400 effects at filled gap positions in L1 speakers. This discrepancy across studies could be explained by two key factors. First, the experimental designs differed in the manipulation of the condition. While Covey and colleagues (2024) compared filled-gap versus non-filled-gap conditions, creating a strong contrast in prediction, both the current study and Jessen & Felser (2019) manipulated plausibility at the filled gap position, resulting in more subtle effects based on reversed plausibility. This difference in design might have led to weaker prediction differences in our study and Jessen & Felser's study. Second, and perhaps more crucially, the absence of significant N400 effects at the noun phrase in L1 speakers might be explained by their ability to recover from expectation collapse immediately at the determiner rather than the noun. According to Williams et al. (2001), L1 speakers showed increased reading times at the determiner *the* in the implausible condition and also increased reading times at the following noun in the plausible condition, whereas L2 learners only showed increased reading times at the noun in the plausible condition. This suggested that L1 speakers could begin reanalysis as early as the determiner *the* before reaching the noun. By the time L1 speakers reached our critical region, Noun Phrase, they had already initiated and possibly completed their reanalysis process due to prediction failure. In contrast, L2-High speakers, who showed weaker and delayed N400 effects at the First Verb, exhibited N400-like effects at the noun phrase.

The L2-Low group showed no N400 effects at either the First Verb or the filled-gap Noun Phrase position. This suggests that low-proficient participants may not form strong predictions at the First Verb, as indicated by the absence of N400 effects there. Without the initial predictions, they consequently would not experience prediction violations at the filled gap position. In other words, the absence of N400 at the filled gap position in the L2-Low group may be attributed to their lack of initial prediction formation at the verb. While both L1 and L2-Low speakers showed statistically null N400 effects at the filled gap position, the underlying mechanisms appear to be different. For native speakers, the null effect likely results from rapid reanalysis following their expectation violation. In contrast, L2-Low speakers might not engage in strong predictive processing during filler-gap dependency resolution, resulting in no observable effects at either position. Therefore, the observed pattern of N400 effects at the First Verb and the Noun Phrase region across all groups appears to suggest that L2 speakers predictive processing may gradually approximate native-like patterns as L2 learner' proficiency increase.

In the later time-window (550-750 ms), our prediction was based on the prevailing hypothesis that L2 speakers rely on or prefer to use semantic information (e.g., Clahsen & Felser, 2018). If semantic information were prioritized during incremental processing in L2 speakers, the P600 would be larger for plausible conditions, as L2 speakers could first use semantic information to establish the filler-gap dependency in the plausible condition, making reanalysis more difficult. Alternatively, if L2 speakers did not prioritize semantic information but could use syntactic information, they would identify the syntactic gap and attempt to resolve the dependency regardless of semantic congruency, showing no difference between plausible and implausible conditions. As discussed earlier, P600 is often associated with difficulty in syntactic integration or reanalysis (Kaan et al., 2000). While L2-High speakers showed larger N400 for plausible conditions at the filled gap, suggesting engagement in prediction-based processing, the frontally distributed P600-like in L2-Low speakers likely indicated syntactic integration difficulty rather than unsuccessful prediction. This syntactic integration difficulty appearing only in plausible conditions may suggest that L2-Low speakers attempt integration only when supported by semantic information. Although marginal, we tentatively interpreted this effect as P600 for two reasons. First, previous studies in filler-gap processing have reported similar processing difficulty in L2 speakers. Williams et al. (2001) found that L2 speakers, in a stop-making-sense judgement task, showed reanalysis effects only at the noun, not at the determiner, suggesting they need substantial lexical-semantic information to trigger reanalysis. Similarly, Covey and colleagues (2024) observed P600 effects rather than N400 effects at filled gap positions, indicating integration difficulty rather than expectation failure. Second, the topographic distribution of this P600 aligns with established patterns of syntactic processing difficulty. Specifically, frontally distributed P600 has been associated with the reanalysis during the processing of grammatical but complex syntactic structures (Friederici, Hahne & Saddy, 2002). For example, Dong et al. observed a prefrontal-central positivity in L2 learners at the filled gap site. Importantly, the lack of reversed plausibility effect (as indexed by larger P600 for plausible conditions) in L1 and L2-High groups cannot be interpreted as an absence of syntactic reanalysis. Rather, this finding suggests that at the filled-gap position, both groups likely may have performed syntactic reanalysis regardless of whether the semantic information was plausible or implausible. In other words, their ability to posit syntactic gaps and resolve filler-gap dependencies seems to function independently of the sentence's initial plausibility.

The findings at the Noun Phrase region provided evidence for the Trace-Reactivation Hypothesis. In other words, filler-gap dependencies are driven, at least in part, by the abstract syntactic trace. If filler-gap dependencies were entirely verb-driven, the parser would complete the dependency only at the verb. In such a case, L2-High speakers would not exhibit an N400 reflecting prediction collapse, as no predictions would have been made. Similarly, L2-Low speakers would not show a P600 at the filled gap position, as they would not engage in syntactic reanalysis.

Gap Licensor

At this region, we manipulated the type of gap licensor (transitive verb vs. preposition) to shed light on why PO gaps are found to be less preferred than DO gaps. It is important to note that the gap licensor position occurs before the actual syntactic gap; thus, the interpretation of the results at this position should be associated to predictive processing (see Pickering and Gambi, 2018). The results showed that all groups showed more positive-going waveforms for verbs across all time-windows. The consistent Condition and Condition × ROI interactions across time-windows suggested that both L1 and L2 speakers distinguish between verb and preposition licensors. This is not surprising as different linguistic information can be captured by distinct ERP components in the brain. Friederici (2002, 2004) has suggested that language processing occurs in sequential model, with syntactic processing preceding semantic processing. Their model proposes that morphosyntactic operations happen first (marked by ELAN brain response), followed by categorical evaluation (marked by LAN), then semantic processing (marked by N400), and finally syntactic reanalysis (marked by P600). Each stage corresponds to specific electrical patterns in the brain. In addition, Group × ROI interactions in early time-windows (250-320 ms, 300-500 ms) but not the latter time-window (500-900 ms) may suggest that L1 and L2 speakers initially recruited different brain regions for processing, though they achieve similar outcomes. However, these findings might merely reflect that all groups processed verbs and prepositions differently, not necessarily from the same gap-filling mechanism. To better understand whether these effects reflect similar gap-filling mechanisms or merely lexical category recognition, we will examine the specific ERP patterns for each group separately.

N280 is an ERP component characterized by a negative-going deflection peaking at around 280 ms after the onset of the stimuli, and it has been suggested that this component is triggered by words with a primarily syntactic role, such as function words (e.g., prepositions in this study), but does not appear in content words (e.g., verbs in this study) (Luck, 2014). To the best of my knowledge, no previous studies in filler-gap processing have reported this component. In the early 250-320 time-window in the present study, the more negative-going waves for the preposition gap licensor appeared to demonstrate that all groups process function words in a similar manner. The timing of the 250-320 window might occur immediately following the early morphosyntactic processing stage described in the sequential model (Friederici, 2002, 2004). This early morphosyntactic processing stage is typically indexed by ELAN (Hahne & Friederici, 1999). It is possible that the N280 effect is associated with immediate syntactic demands for prepositions compared to verbs. Specifically, when parsers encounter a preposition, the brain has no morphological variations to process, and the parser must immediately commit to the processing of a noun phrase. In contrast, verbs can introduce richer morphosyntactic information, such as tense, number, person, and aspect, which allows for greater flexibility and enables the parser to postpone certain decisions. An alternative explanation of the N280 is that it reflects increased processing difficulty when a prepositional gap licensor is interpreted as introducing an adjunct rather than an argument. Schütze and Gibson (1999, p. 417) found that when reading sentences such as "The company lawyers considered employee demands for a month/for a raise but they did not act," readers preferred prepositional phrases that serve as NP arguments (e.g., *for a raise*) over those that act as VP modifiers (e.g., *for a month*). The researchers suggested that this was because *for a raise* was an argument that completed the meaning of *demands* rather than just an adjunct that added optional time information, as in *for a month*. A similar preference may have influenced the ERP responses in the present study. In the sentence "Lucy identifies which article the author wrote a column to describe/about ___ in the newspaper," participants may have initially interpreted *about* as introducing an adjunct (e.g., "about an hour"), increasing processing demands and resulting in an N280. In contrast, *describe* unambiguously licenses a direct object gap as its argument, allowing for a smoother syntactic integration. Although all groups recognized the prepositions, the wave patterns reflected a clearer 280 ms peak in L1 groups only, and the topographical distributions differed. L1 speakers showed significant condition effects across frontal regions (FL, FM, FR). Similarly, the L2-High group exhibited a comparable but more focused effect, limited to the left frontal region (FL). In contrast, the L2-Low group showed widespread activation across frontal and central regions (FL, FM, FR, CL, CM, CR), presumably due to their need for greater cognitive effort and broader neural recruitment for preposition processing. It is important to note that despite these differences, the lack of significant Group × Condition interactions in the ANOVA suggests that all groups successfully processed the syntactic information, albeit with varying neural efficiency.

In the 300-500 ms time-window, across all groups, there was a significant difference in brain responses to verbs versus prepositions as gap licensors, with more negative amplitudes observed for prepositions than for verbs. This effect showed a frontal distribution (FL, FM, FR) in both L1 and L2-Low groups. Although no specific ROI reached significance for the L2-High group, their waveform patterns similarly showed more negative-going deflections for prepositions compared to verbs in frontal areas (FL, FM, FR). The anterior negativity observed in the present experiment is unlikely to be a LAN for two main reasons. First, LAN effects are typically left-lateralized, whereas in the present investigation, the distribution extends across the FL, FM, and FR regions in all groups. Second, LAN is often seen as an indicator of native-like processing, and previous research has rarely reported LAN in non-native speakers, even those with high L2 proficiency (for reviews, see Caffarra et al., 2015; Antonicelli & Rastelli, 2023). However, the present study showed that this anterior negativity was present across all groups, as evidenced by a main effect of Condition in the omnibus ANOVA. Therefore, the anterior negativity is unlikely to be a LAN.

Previous studies have not reached a consensus on the nature of this frontally distributed anterior negativity; this effect has been labeled with different names: anterior negativity (Friederici, 2002; Dallas et al., 2013), bilateral LAN (Hestvik et al., 2012; King & Kutas, 1995), or sustained anterior negativity (SAN) (Phillips, 2005). For example, Hestvik et al. (2012) compared native speakers' ERPs at the object fill-gap position in sentences (40a) and (40b). In participants with high working memory capacity, the researchers found a 100-200 ms early left anterior negativity (ELAN) at *the camel* in (40a). This ELAN persisted bilaterally from 200 ms and peaked at 500 ms. The researchers interpreted this subsequent bilateral anterior negativity as a bilateral LAN. In contrast, participants with low working memory capacity showed a delayed anterior negativity that appeared at 400 ms. The ELAN/LAN pattern, along with the following P600 at the object filled-gap position *the camel,* were taken as the indicators of word category violation, as the parser was not expecting a noun phrase at the gap reserved for the displaced filler. It is important to note that their experimental design also highlighted working memory differences between conditions. To illustrate, sentence (40a) required participants to maintain *the zebra* in memory until encountering the gap at *the camel*, while sentence (40b) contained no such incomplete dependency.

(40) a. *The zebra that the hippo kissed the camel on the nose ran far away.

b. The weekend that the hippo kissed the camel on the nose he ran far away.

(Hestvik et al., 2012, p. 428)

In addition to being interpreted as a bilateral LAN related to word category violation, the anterior negativity has also been interpreted as a component related to working memory load. King & Kutas (1995) compared how native English speakers processed two types of sentences containing relative clauses. In (41a), the word *reporter* is the subject of the relative clause, while in (41b) it is the object of the relative clause. The results showed that when *reporter* is the object in the relative clause, as in (41b), participants showed greater bilateral anterior negativity at *the senator* compared to the same word region in (41a). The researchers interpreted this anterior negativity as a LAN reflecting increased working memory costs caused by maintaining a filler in working memory. Similar anterior negativity effects were also observed by Dallas and colleagues (2013), who found significantly more negativity in the frontal regions when a noun acted as a filler that requiring storage in working memory compared to when it appeared as a direct object in its original post-verbal position.[23]

(41) a. The reporter who harshly attacked the senator admitted the error.

b. The reporter who the senator harshly attacked admitted the error.

Consistent with previous research linking anterior negativity to working memory load, the anterior negativity (AN) observed in the present study may also reflect different working memory

[23] The two nouns occurred in different sentence positions. One appeared early in the sentence while the other appeared later. The author suggested that this positioning might also explain the observed effects.

demands during filler maintenance and retrieval. Verbs provide both syntactic and semantic cues that can facilitate filler retrieval from working memory, potentially reducing working memory load. In contrast, prepositions offer primarily syntactic cues, making filler retrieval more demanding, and thus the parsers need to continue engaging working memory to maintain the filler until it can be fully integrated. This interpretation aligns with Wagers and Phillips' dual-process account of filler-gap processing (Wagers & Phillips, 2014). When computing sentences like “Lucy identifies which article the author wrote a column to describe/about ____”, the stronger AN for *about* likely reflects increased demands on the lagging process. In other words, at the gap licensor position, the parsers first reactivate the maintained syntactic information about the filler to predict the gap. Yet in the subsequent lagging process, they find that there is no semantic information to be used for filler retrieval. This may potentially result in higher working memory costs or continuing filler maintenance, possibly manifested as a greater AN component compared to transitive verb condition.

In the 500-900 ms time-window, all groups processed transitive verbs and prepositions differently in the Gap Licensor region. This pattern was evident in waveforms, where prepositions elicited greater negativity than transitive verbs across all groups. This unexpected Late Negativity has been found, though not often, in several previous studies of sentence processing (Covey et al., 2024; Díaz et al., 2016; Fromont et al., 2020). Covey and colleagues (2024) observed a late negativity (LN) in native speakers at the filled gap position (noun phrase) in both island and non-island conditions. They attributed this effect to the difficulty in thematic role assignment, specifically when the filled gap position required integration of two competing arguments—displaced *wh*-phrases versus filled NPs. Although they examined a different region than the present study (filled-gap noun phrase versus preposition gap licensor), their thematic role assignment account could explain our observed Late Negativity. Following their interpretation, LN reflects thematic role assignment challenges. In the preposition condition, the displaced *wh*-phrase (e.g., which article) and the preposition *about* create a dependency that typically requires thematic role assignment. However, compared to transitive verbs, the absence of semantic cues from the preposition prevents it from assigning a thematic role to the filler *which column*, thus leads to thematic role assignment difficulty. Notably, Covey et al. (2024) reported that participants with superior attentional control (indicated by better Stroop scores) exhibited larger LN. They proposed that enhanced attentional control facilitates predictive processing, resulting in more pronounced thematic assignment conflicts and consequently larger negativity. However, our findings did not reveal this relationship between predictive ability and LN. Given our earlier observation that L2-Low speakers showed reduced predictive processing capacity in the filled-gap noun phrase section, Covey and colleagues' hypothesis would predict smaller late negativity in this group compared to L1 and L2-High groups. Our results did not support this prediction. Our findings aligned more closely with Fromont et al. (2020), who linked LN to increased working memory demands for maintaining, integrating, and reanalyzing linguistic information. Their study demonstrated that working memory capacity modulated ERP responses in both L1 and L2 speakers, with higher working memory scores associated with LN. In the present study, both L1 and low-proficient L2 speakers showed broadly distributed LN, whereas high-proficient speakers showed no LN effect in any specific ROI despite their high proficiency. This pattern was consistent with the one observed in the 300-500 ms time-window, where LAN may also suggest working memory load associated with filler maintenance and retrieval. Therefore, it seems apt that the effects in both 300-500 and 500-900 ms time-windows manifest similar working memory demands.

5.6 Conclusion

The current study aimed to explore the neural mechanisms underlying filler-gap processing in both L1 and L2 speakers. The absence of Group × Condition interactions across all critical regions challenges theories proposing fundamental differences between L1 and L2 processing and provides evidence for qualitatively similar processing mechanisms. At the First Verb, all participants showed sensitivity to plausibility, with effects varying in topographical distribution. L2 speakers appeared to show the N400 in a less efficient and delayed manner. This pattern potentially indicated L2 speakers' involvement of semantic information in predictive processing. At the Noun Phrase region, while omnibus analyses revealed no significant main effects or interactions, interesting patterns emerged in waveform and ROI-specific analyses. The L2-High group exhibited a marginally significant N400 effect for plausible conditions, which aligned with effects seen for prediction violation in previous studies (e.g., Covey et al., 2024). In contrast, the L2-Low group showed a marginally significant frontally-distributed P600-like effect, consistent with difficulties in syntactic reanalysis observed in previous studies (e.g., Jessen et al., 2017). At the Gap Licensor region, all groups showed similar sensitivity to the distinction between verbs and prepositions across multiple time-windows. This was manifested through the N280 component, which indicates quick recognition of function words versus content words (Luck, 2014), as well as Anterior Negativity and Late Negativity effects, which reflect working memory load for maintaining and retrieving a filler in working memory. The findings of the present study suggested that while the overall parsing mechanism of gap-filling appears qualitatively similar across L1 and L2 speakers, individual differences in proficiency could influence how readers respond to prediction. Specifically, higher proficiency learners might engage in predictive processing and show sensitivity to prediction violations through N400 effects, whereas lower proficiency learners might involve in incremental processing only and show syntactic reanalysis difficulty through P600 effects. However, this potential relationship between proficiency and ERP patterns (N400 versus P600) requires further investigation due to the statistical power of the current study, as effects in ROI-specific analyses reached only marginal significance.

Chapter 6

Conclusion

6.1 Overview of the study and summary of findings

The goal of this dissertation was to examine how L2 speakers process sentences in real time and whether they employ qualitatively similar sentence processing mechanisms to native speakers. This goal was achieved by investigating L1 and L2 speakers' parsing mechanisms of filler-gap dependencies, a non-local construction requiring parsers to build a relationship between a filler and its corresponding gap, and by adopting both self-paced reading and ERP methodologies. The self-paced reading provided behavioral evidence through the measurement of parsers' reading times, while ERP provided neural evidence through more precise measurement of our brains' immediate responses to linguistic inputs.

The study addressed four key research questions. The first explored whether the online processing mechanisms of filler-gap dependencies are qualitatively similar between L1 and L2 speakers. To investigate, I used the manipulation of filler plausibility and dependency length for the self-paced reading experiments, and filler plausibility and gap licensor for the ERP study. The second research question asked how L1 and L2 speakers utilize syntactic and semantic information to resolve ongoing dependencies. Specifically, do L2 speakers heavily rely on lexical-semantic information in filler-gap processing, or are they able to map sentences using syntactic cues, namely the empty category trace in the present study? If they are able to employ syntactic information as native speakers do, would there be timing and degree differences, for example, being slower or less sensitive to the gap? The third investigated to what extent proficiency and language immersion can explain L1 and L2 processing differences. If L2 speakers show different performances in processing filler-gap dependencies, would being more proficient or being immersed in their second language environment make them more native-like? The fourth question, though not falling into the category of L2 processing, addressed a new and less researched field: whether Mandarin Chinese topic structures involve similar processing mechanisms to English filler-gap dependencies. This was aimed to evaluate many previous studies that attributed the different processing profiles to L1-L2 differences between Mandarin Chinese and English (e.g., Dallas et al., 2013; Dussias & Piñar, 2010; Juff, 2005; Covey et al., 2024).

To investigate these questions, three experiments were conducted in the dissertation. Overall, findings provided evidence for qualitatively similar parsing mechanisms between L1 and L2 speakers and showed that relatively ineffective or delayed effects could be compensated for by higher proficiency. The first experiment used a self-paced reading paradigm to examine how L1 and L2 speakers process *wh*-dependencies in real time with the manipulation of both dependency length and filler plausibility. Both groups appeared to employ the Active Filler Strategy (e.g., Clifton & Frazier, 1989), evidenced by their reading time slowdown at the earliest acceptable gap licensor, compared to the pre-critical word, and their sensitivity to plausibility at the second post-critical region. Although the slowdown in reading time at the critical verb provided only indirect evidence for active gap filling due to the absence of a direct comparison with a non-gap sentence, the subsequent plausibility effect indicated that parsers attempted to integrate the filler and experienced surprise upon detecting a semantic anomaly. Together, these findings suggested that the immediate elevation in reading times at the critical region reflected syntactic integration difficulty, while the delayed plausibility effects in the post-critical region indicated semantic integration difficulty. This pattern aligns with the garden-path model that both groups process syntactic information before semantic information, with syntactic cues driving initial gap-filling, and semantic integration occurring at a later stage (Frazier, 1987). This is also consistent with the dual-process account (Wagers & Phillips, 2014), the syntactic-first model (Friederici, 2002), while casting doubt on the holistic perspective of the interactive account or constraint-based model, which argues that multiple sources of information are processed simultaneously (e.g., Trueswell et al., 1994; Clifton et al., 2003; MacDonald et al., 1994). The study also found that higher proficiency L2 speakers showed descriptively greater sensitivity to dependency length. Interestingly, while years of immersion affected overall processing speed, it did not significantly modulate sensitivity to semantic plausibility across different groups.

The second experiment, also using self-paced reading, investigated whether gap-type topic structures in Mandarin Chinese involve movement-derived non-local dependencies similar to English *wh*-dependencies. This experiment explored whether native Mandarin Chinese speakers actively search for gaps, employ reanalysis strategies, and show sensitivity to dependency length. Although not directly comparing the two participant groups, this experiment addressed whether previous processing differences between native English speakers and Mandarin Chinese learners of English stem from L1-L2 differences. In the present study, native Mandarin Chinese speakers processed topic structures in ways remarkably similar to how English speakers process *wh*-dependencies. The observed reading time slowdown at the first verb, without a corresponding slowdown at the final verb, suggested that Mandarin speakers actively searched for a gap and attempted to resolve the dependency at the first possible gap position. Second, Mandarin Chinese speakers also demonstrated reanalysis effects when their initially plausible construction of the dependency had to be reanalyzed. Third, both Experiments 1 and 2 showed unexpected elevated reading times at the pre-critical position in the short conditions, possibly due to a subject-filled gap effect. This complicated the measurement of dependency length, but may show that in both languages, parsers are able to predict a subject gap before the verb.

The third experiment utilized ERP methodology to examine the neural underpinnings of L1 and L2 filler-gap processing and the role of proficiency in L2 processing. There were three critical regions: the initial verb where filler-gap integration was first attempted, the filled-gap position requiring reanalysis, and the final licensing element at the actual gap site. This neurophysiological approach provided insights into whether behaviorally similar processing profiles between L1 and proficient L2 speakers reflect genuinely similar underlying neural mechanisms. Across all critical regions, the absence of significant Group × Condition interactions suggested qualitatively similar processing mechanisms between L1 and L2 speakers. The parsing differences that were found do not appear to represent fundamental differences but are more likely explained by delayed timing and an attenuation of degree. At the First Verb position, both the L1 and L2 groups demonstrated sensitivity to plausibility through N400 effects, though L2 speakers showed this effect in a delayed and less efficient manner. Specifically, L2 speakers overall demonstrated a marginally significant N400 effect in the posterior-right region during the canonical 300-500 ms time window, and this effect became significant in the same region during the later 500-600 ms time window for high-proficient L2 speakers. At the filled-gap noun phrase position, proficiency levels may have influenced processing strategies: high-proficiency L2 learners showed marginally significant N400 effects indicating prediction violations, while low-proficiency learners exhibited marginally significant P600 effects suggesting syntactic reanalysis difficulties. Finally, at the gap licensor position, all groups similarly distinguished between verbs and prepositions, as shown through N280, Anterior Negativity, and Late Negativity components, indicating working memory demands for maintaining and retrieving fillers. These findings revealed that while L1 and L2 speakers employ fundamentally similar parsing mechanisms, L2 proficiency could potentially facilitate their predictive processing, though these proficiency-related effects require further investigation due to marginal statistical significance for N400 and P600 at the filled-gap noun phrase.

6.2 General discussion

The General Discussion addresses three main topics in the dissertation: *wh*-dependency resolution, parsing mechanisms in L1 versus L2 processing, as well as incremental and predictive processing.

Wh-dependency resolution

The Active Filler Strategy posits that upon encountering a displaced item, parsers actively search for potential gap locations rather than waiting for unambiguous evidence about the actual gap location (Clifton & Frazier, 1989). The findings in the present study provided several pieces of evidence supporting this parsing mechanism. The most direct evidence came from the N400 effect at the first verb from both L1 and L2 speakers. Although L2 speakers' N400 effect was less pronounced and delayed, this showed sensitivity to plausibility at the first verb position. In the behavioral self-paced reading experiment, both L1 and L2 readers showed increased reading times at the first verb, along with the plausibility effect that emerged shortly after. These findings together aligned with the active searching strategy. Previous studies have shown that self-paced reading tasks tend to show lagging effects that spill over to post-critical regions (e.g., Wagers & Phillips, 2014; Johnson, 2015); therefore, as long as the plausibility effect appears at the first potential gap position or shortly after, it should be considered evidence for active gap filling. Further, the self-paced reading experiment revealed an unexpected but informative potential subject filled-gap effect in short dependency conditions. Both L1 and L2 readers showed increased reading times when encountering a noun phrase in the subject position that could potentially be a gap site. This effect, if indeed caused by the filled subject, would align with Johnson (2015) and Yan & Liu's (2021) findings of subject filled-gap effects, suggesting predictive processing of subject gaps in both L1 and L2 speakers. Future studies could target this effect, and if supported, this evidence would be particularly compelling for Active Filler Strategy because it would show parsers actively predicting gap positions before encountering the verb without the engagement of a verb.

The increased reading time at the critical verb, along with the plausibility effect occurring at the second post-critical region, aligned well with the dual-process account proposed by Wagers and Phillips (2014). If we interpret reading time slowdowns as indicators of integration difficulty, our results suggest that parsers attempted to integrate the filler at the verb position regardless of plausibility, followed by a later stage of semantic evaluation. This temporal sequence supports the proposed distinction between a leading syntactic process and a lagging semantic process. One might argue that the increased reading times at the verb position simply reflect that verbs are generally more difficult to process. However, this interpretation lacks strong empirical support. While previous studies have demonstrated that adjuncts are more costly to process than arguments (Schütze & Gibson, 1999), and the ERP study in the present dissertation showed that functional words might elicit larger N280 components than content words, to the best of my knowledge, there has been no established evidence suggesting that verbs are inherently more difficult to process than other word categories. Importantly, even when readers had already experienced processing difficulty from a subject filled-gap effect at the pre-verbal position, they still showed noticeable reading time increases at the verb. Therefore, the reading time patterns more likely reflect integration-related processing rather than general verb processing difficulty. The timing of these effects—with integration difficulty preceding plausibility effects—seems to provide further support for the dual-process account (Wagers & Phillips, 2014). This sequential processing mechanism is consistent with both the garden path model (Frazier, 1987) and the "syntax-first" model (Friederici, 2002, p. 78). All these perspectives fall within Fodor's (1983) broader framework of modular account, which assumes that modules responsible for different linguistic information operate separately and in a sequential order.

The Active Filler Strategy was also evident in the processing of Mandarin Chinese topic structures. At the actual gap region, a reversed plausibility effect was observed. Although no plausibility effect was found at the first verb, initially plausible fillers appeared to form more robust integrations, making the reanalysis more effortful and thus resulting in longer reading times at the actual gap. The fact that readers performed reanalysis eventually suggests they must have attempted integration earlier than the actual gap position. If readers were not actively predicting gaps, the plausibility of the filler would not have affected reanalysis difficulty. Notably, this effect appeared only in short dependencies, aligning with Wagers and Phillips's (2014) observation that sensitivity to plausibility mismatch vanished for sentences with longer linear distance. Importantly, the processing patterns observed in the self-paced reading of Mandarin Chinese topic structures showed remarkable similarities to how native English speakers process English *wh*-dependencies, including unexpected longer reading times for short conditions than for long ones, increased reading times at the verb, and delayed plausibility-related effects that spilled over to later regions. These findings suggest that the gap-type topic structures in Mandarin Chinese might be similarly movement-derived, which calls into question previous studies that have attributed L1-L2 processing differences in filler-gap dependencies to the lack of syntactic dependency in Mandarin Chinese (e.g., Dallas et al., 2013; Dussias et al., 2015; Juff, 2005; Covey et al., 2024).

L1 versus L2 processing

Our findings showed no salient evidence for qualitative differences between L1 and L2 processing and thus more aligned with the qualitative similarity between L1 and L2 processing. The most direct evidence came from the ERP experiment, where no Group × Condition interaction was found across all critical regions. L2 speakers demonstrated a tendency toward native-like processing.[24] Specifically, at the first verb, L2-High speakers showed N400 effects that progressed from an attenuated response in the early window (300-500 ms) to a significant effect in the later window (500-600 ms), suggesting that higher proficiency could enable more native-like semantic sensitivity, albeit with some delay. At the gap licensor position, all groups exhibited N280 and working memory-related Anterior Negativity and Late Negativity. Controversy might arise from the findings at the Noun Phrase region, where L2-High speakers showed N400 effects indicating prediction-based processing similar to L1 speakers in a recent study (Covey et al., 2024), while L2-

[24] L2 speakers reported a mean proficiency of 2.95/5, with an average age of exposure at 7 years and 5.7 years of immersion. L2 participants were divided into high-proficiency (n = 9, mean = 71/100) and low-proficiency (n = 11, mean = 42/100) subgroups based the median of the cloze task scores.

Low speakers exhibited P600 effects suggesting different processing strategies. One might argue against qualitative similarity based on these different ERP components. However, it is important to note that claims of qualitative differences should be based on evidence that individual factors cannot account for processing differences. In a recent filler-gap study, Dong et al. (2023) found pre-frontal central positivity at the filled-gap position, which they interpreted as the indicator of qualitative differences between L1 and L2 processing, as proficiency and working memory (WM) had no effect on L2 ERP amplitude. However, the present ERP study presents a different case. The L2-high group showed an N400 for plausible conditions, similar to the pattern observed in native speakers in Covey et al. (2024). Moreover, although L1 speakers did not show a significant N400 effect, they exhibited more negative-going waveforms for plausible conditions, as evident in both wave patterns and descriptive data. Therefore, while L2-low speakers showed a P600, this may not indicate qualitatively different processing from L1 and L2-high speakers, but rather suggests that L2-low speakers may not yet be capable of engaging in predictive processing and rely on incremental processing, leading to greater reanalysis difficulty when their initial plausible integration must be revised.

The self-paced reading experiment revealed patterns supporting qualitatively similar but potentially gradient processing mechanisms between L1 and L2 speakers. For plausibility, initial analyses showed categorical L1-L2 differences with significant Plausibility × Group interactions. However, due to the presence of a significant Plausibility × Dependency interaction, follow-up analyses were conducted on long dependencies only, and the results revealed a main effect of Plausibility across groups. In addition, higher-proficiency L2 speakers showed a trend toward native-like sensitivity to implausibility, while lower-proficiency speakers showed no such trend. This aligns with theories proposing gradient rather than categorical differences in L2 processing (Clahsen & Felser, 2018). This could lend support to the view that L2 processing can become more native-like with increased proficiency (e.g., Steinhauer et al., 2014). While L2 speakers, particularly those with higher proficiency, showed increased reading times for longer dependencies, L1 speakers showed an opposite pattern. This might be consistent with the resource deficit account, which suggests that L2 speakers are resource constraint and can potentially achieve more native-like processing when they have sufficient resources (Hopp, 2010; McDonald, 2006).

Prediction in L2 filler-gap processing

The discussion on predictive processing primarily draws from the ERP study results, as ERP methods provide better temporal resolution than behavioral methods. The investigation of filler-gap dependencies provides a good linguistic resource for investigating predictive processing, as assuming the existence of syntactic gaps, any position preceding the actual would be a meaningful position where we can uncover parsers' predictive processing mechanisms. In our study, the N400 effect observed at the filled-gap noun phrase for initial plausible integration cannot be attributed to a semantic violation, as the preceding filler-verb integration remains semantically plausible. Instead, the N400 effect found in the L2-High group at the filled-gap position suggests these speakers were actively predicting a gap after encountering the verb and experienced a prediction violation. Notably, while L2-High speakers showed an N400, L2-Low speakers exhibited a P600. This pattern might suggest that L2 speakers' ability to engage in predictive processing could become more native-like as their language proficiency develops. This interpretation aligns with Phillips and Ehrenhofer's (2015) proposal that limited predictive processing abilities in L2 might hinder learning. This perspective accounts for the relationship between low proficiency and limited predictive processing.

Two hypotheses can be generated from the ERP study in the present thesis. First, semantic congruency seems be a prerequisite for prediction. In our case, both L1 and L2-High groups showed sensitivity to filler plausibility at the first verb, before the syntactic gap. In the 300-500 ms window, L1 speakers showed a trend toward more negative amplitude for plausible conditions, while L2-

High speakers showed marginally significant N400 effects. This suggests prediction may occur only when both syntactic and semantic information have been validated. In other words, if there is implausible integration, parsers would not make predictions. Second, L2-Low speakers' frontally-distributed P600 response has been associated with syntactic reanalysis (Friederici, Hahne & Saddy, 2002). This suggests that L2 speakers, though not yet capable of prediction, at least engaged in incremental processing. Under incremental processing, these speakers process incoming words sequentially, leading to increased reanalysis demands when the appearance of a noun phrase conflicts with previously resolved dependency. In summary, it is plausible to suggest that L2 speakers could gradually transition from incremental to predictive processing with increasing proficiency, and this proficiency might include better vocabulary knowledge, larger available resources, or reduced time constraints.

6.3 Limitations and future research

One major limitation came from the unexpected longer reading times at the pre-critical and critical positions for sentences with short dependencies compared to those with long dependencies in the self-paced reading experiments, making the interpretation of dependency length effects unclear. Future studies should carefully control the pre-critical positions, making them identical in both animacy and syntactic structure.

Another limitation was that the study did not manipulate plausibility-for-First Verb at the Gap Licensor position in the ERP experiment. To illustrate, while all sentences were ultimately grammatical, the study could have compared cases where fillers were either plausible or implausible with the first verb to examine how this initial integration affects processing at the Gap Licensor position. Previous research by Jessen and Felser (2019) has demonstrated that at the actual gap position, L2 learners, unlike L1 speakers, exhibited a P600 reanalysis effect when processing fillers that could have been plausible direct objects, suggesting that L2 speakers might have more difficulty recovering from their initial analysis. Future studies could address this limitation by systematically manipulating plausibility at both positions. This design could reveal how initial plausible versus implausible integration affects subsequent gap processing and might provide a more complete picture of reanalysis processes in both L1 and L2 filler-gap dependency resolution.

Further, while proficiency was measured through a cloze test, immersion was assessed only through questionnaires. Although we had provided instructions and clarification for offline testing, some aspects of immersion remained subjective. For example, participants varied in whether they considered their international school experience as immersion. Future studies measuring immersion should establish more objective standards, such as years of formal education in L2-immersed higher education institutions or years in full-time employment requiring L2 as the sole language, while also controlling for exposure to L1 media and interactions.

Last, future research could investigate the relationship between semantic plausibility and syntactic prediction in filler-gap processing. Studies might compare how L1 and L2 speakers process two types of structures: syntactically well-formed but semantically implausible filler-gap dependencies versus non-filler-gap constructions with similar semantic implausibility. If syntactic gap prediction occurs independently of semantic plausibility, we might observe prediction-related effects at potential gap sites independent of semantic licensing. This comparison could help clarify whether syntactic cues alone might trigger predictive processing in both L1 and L2 speakers, or whether semantic congruency is indispensable for prediction, as our current findings suggest.

References

Abrahamsson, N. (2012). Age of onset and nativelike L2 ultimate attainment of morphosyntactic and phonetic intuition. *Studies in Second Language Acquisition, 34*(2), 187–214. https://doi.org/10.1017/S0272263112000022

Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychologica, 128*(3), 466–478. https://doi.org/10.1016/j.actpsy.2008.03.014

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Altmann, G. T. M., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science, 33*(4), 583–609. https://doi.org/10.1111/j.1551-6709.2009.01022.x

Andersson, A., Sayehli, S., & Gullberg, M. (2019). Language background affects online word order processing in a second language but not offline. *Bilingualism: Language and Cognition, 22*(4), 802–825. https://doi.org/10.1017/S1366728918000573

Antonicelli, G., & Rastelli, S. (2023). Event-related potentials in the study of L2 sentence processing: A scoping review of the decade 2010–2020. *Language Acquisition, 30*(2), 163–200. https://doi.org/10.1080/10489223.2022.2141633

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*(2), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Aoshima, S. (2003). *The grammar and parsing of wh-dependencies* [Doctoral dissertation, University of Maryland]. ProQuest Dissertations & Theses.

Aoshima, S., Phillips, C., & Weinberg, A. (2004). Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language, 51*(1), 23–54. https://doi.org/10.1016/j.jml.2004.03.001

Aoun, J., & Li, Y. A. (1993). Wh-elements in situ: Syntax or LF? *Linguistic Inquiry, 24*(2), 199–238.

Atkinson, E., & Omaki, A. (2023). Adaptation of Gap Predictions in Filler-Gap Dependency Processing during Reading. *Languages, 8*(4), 285-. https://doi.org/10.3390/languages8040285

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 3–76). Cambridge University Press.

Batterink, L., & Neville, H. (2013). Implicit and explicit second language training recruit common neural mechanisms for syntactic processing. *Journal of Cognitive Neuroscience, 25*(6), 936–951. https://doi.org/10.1162/jocn_a_00354

Berghoff, R. (2022). L2 processing of filler-gap dependencies: Attenuated effects of naturalistic L2 exposure in a multilingual setting. *Second Language Research, 38*(2), 373–393. https://doi.org/10.1177/0267658320945757

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and language development* (pp. 279–362). John Wiley & Sons.

Bever, T. G., & McElree, B. (1988). Empty categories access their antecedents during comprehension. *Linguistic Inquiry, 19*(1), 35–43.

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. M. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 41–68). Cambridge University Press.

Bley-Vroman, R. (1990). The logical problem of second language learning. *Linguistic Analysis, 20*(1), 3–49.

Bowden, H. W., Steinhauer, K., Sanz, C., & Ullman, M. T. (2013). Native-like brain processing of syntax can be attained by university foreign language learners. *Neuropsychologia, 51*(13), 2492–2511. https://doi.org/10.1016/j.neuropsychologia.2013.09.004

Boxell, O., & Felser, C. (2017). Sensitivity to parasitic gaps inside subject islands in native and non-native sentence processing. *Bilingualism: Language and Cognition, 20*(3), 494–511. https://doi.org/10.1017/S1366728915000942

Boxell, O., Felser, C., & Cunnings, I. (2017). Antecedent contained deletions in native and non-native sentence processing. *Linguistic Approaches to Bilingualism, 7*(5), 554–582. https://doi.org/10.1075/lab.15006.box

Brothers, T., Hoversten, L. J., & Traxler, M. J. (2021). Bilinguals on the garden path: Individual differences in syntactic ambiguity resolution. *Bilingualism: Language and Cognition, 24*(3), 612–627. https://doi.org/10.1017/S1366728920000711

Brouwer, H., & Crocker, M. W. (2017). On the proper treatment of the N400 and P600 in language comprehension. *Frontiers in Psychology, 8*, Article 1327. https://doi.org/10.3389/fpsyg.2017.01327

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research, 1446*, 127–143. https://doi.org/10.1016/j.brainres.2012.01.055

Caffarra, S., Molinaro, N., Davidson, D., & Carreiras, M. (2015). Second language syntactic processing revealed through event-related potentials: An empirical review. *Neuroscience and Biobehavioral Reviews, 51*, 31–47. https://doi.org/10.1016/j.neubiorev.2015.01.010

Canales Viquez, A. J. (2012). *Online processing of wh-dependencies in English by native speakers of Spanish* [Doctoral dissertation]. ProQuest Dissertations & Theses.

Casey, B., Galvan, A., & Hare, T. A. (2005). Changes in cerebral functional organization during cognitive development. *Current Opinion in Neurobiology, 15*(2), 239–244. https://doi.org/10.1016/j.conb.2005.03.012

Cheng, L. L.-S. (1991). *On the typology of wh-questions* [Doctoral dissertation, Massachusetts Institute of Technology]. ProQuest Dissertations & Theses.

Chomsky, N. (1981). *Lectures on government and binding.* Foris.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use.* Praeger.

Chomsky, N. (1995). *The minimalist program.* MIT Press.

Chow, W.-Y., & Zhou, Y. (2019). Eye-tracking evidence for active gap-filling regardless of dependency length. *Quarterly Journal of Experimental Psychology, 72*(6), 1297–1307. https://doi.org/10.1177/1747021818804988

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics, 27*(1), 3–42. https://doi.org/10.1017/S0142716406060024

Clahsen, H., & Felser, C. (2018). Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition, 40*(3), 693–706. https://doi.org/10.1017/S0272263117000250

Clifton, C., & Frazier, L. (1989). Comprehending sentences with long-distance dependencies. In G. N. Carlson & M. K. Tanenhaus (Eds.), *Linguistic structure in language processing* (pp. 273–317). Springer. https://doi.org/10.1007/978-94-009-2729-2_8

Clifton, C., Traxler, M. J., Taha Mohamed, M., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language, 49*(3), 317–334. https://doi.org/10.1016/S0749-596X(03)00070-6

Corps, R. E., Liao, M., & Pickering, M. J. (2023). Evidence for two stages of prediction in non-native speakers: A visual-world eye-tracking study. *Bilingualism: Language and Cognition, 26*(1), 231–243. https://doi.org/10.1017/S1366728922000499

Covey, L., Fiorentino, R., & Gabriele, A. (2024). Island sensitivity in L2 learners: Evidence from acceptability judgments and event-related potentials. *Second Language Research, 40*(1), 19–50. https://doi.org/10.1177/02676583221116039

Cunnings, I. (2017a). Interference in native and non-native sentence processing. *Bilingualism: Language and Cognition, 20*(4), 712–721. https://doi.org/10.1017/S1366728916001243

Cunnings, I. (2017b). Parsing and working memory in bilingual sentence. *Bilingualism: Language and Cognition, 20*(4), 659–678. https://doi.org/10.1017/S1366728916000675

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. https://doi.org/10.1111/lang.12026

DeKeyser, R. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning, 62*(s2), 189–200. https://doi.org/10.1111/j.1467-9922.2012.00712.x

Dekydtspotter, L., Schwartz, B. D., & Sprouse, R. A. (2006). The comparative fallacy in L2 processing research. In *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference* (pp. 33–40). Cascadilla Press.

DeLong, K. A., & Kutas, M. (2020). Comprehending surprising sentences: Sensitivity of post-N400 positivities to contextual congruity and semantic relatedness. *Language, Cognition and Neuroscience, 35*(8), 1044–1063. https://doi.org/10.1080/23273798.2019.1708960

Díaz, B., Erdocia, K., de Menezes, R. F., Mueller, J. L., Sebastián-Gallés, N., & Laka, I. (2016). Electrophysiological correlates of second-language syntactic processes are related to native and second language distance regardless of age of acquisition. *Frontiers in Psychology, 7*, Article 133. https://doi.org/10.3389/fpsyg.2016.00133

Dong, Z. R. (2014). *Processing of tense morphology and filler-gap dependencies by Chinese second language speakers of English* [Doctoral dissertation]. ProQuest Dissertations & Theses.

Dong, Z., Rhodes, R., & Hestvik, A. (2021). Active gap filling and island constraint in processing the Mandarin "gap-type" topic structure. *Frontiers in Communication, 6*. https://doi.org/10.3389/fcomm.2021.650659

Dong, Z. R., Han, C., Hestvik, A., & Hermon, G. (2023). L2 processing of filled gaps: Non-native brain activity not modulated by proficiency and working memory. *Linguistic Approaches to Bilingualism, 13*(6), 767–800. https://doi.org/10.1075/lab.20058.don

Dowens, M. G., Guo, T., Guo, J., Barber, H., & Carreiras, M. (2011). Gender and number processing in Chinese learners of Spanish: Evidence from event-related potentials. *Neuropsychologia, 49*(6), 1651–1659. https://doi.org/10.1016/j.neuropsychologia.2011.02.034

Dowens, M. G., Vergara, M., Barber, H. A., & Carreiras, M. (2010). Morphosyntactic processing in late second language learners. *Journal of Cognitive Neuroscience, 22*(8), 1870–1887. https://doi.org/10.1162/jocn.2009.21304

Dussias, P. E. (2003). Syntactic ambiguity resolution in L2 learners: Some effects of bilinguality on L1 and L2 processing strategies. *Studies in Second Language Acquisition, 25*(4), 529–557. https://doi.org/10.1017/S0272263103000238

Dussias, P. E., & Piñar, P. (2010). Effects of reading span and plausibility in the reanalysis of wh-gaps by Chinese–English second language speakers. *Second Language Research, 26*(4), 443–472. https://doi.org/10.1177/0267658310373326

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish–English bilinguals. *Bilingualism: Language and Cognition, 10*(1), 101–116. https://doi.org/10.1017/S1366728906002847

Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in*

Second Language Acquisition, 35(2), 353–387. https://doi.org/10.1017/S0272263112000915
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*(3), 345–354. https://doi.org/10.1016/S0093-934X(03)00135-4
Felser, C., Cunnings, I., Batterham, C., & Clahsen, H. (2012). The timing of island effects in nonnative sentence processing. *Studies in Second Language Acquisition, 34*(1), 67–98. https://doi.org/10.1017/S0272263111000507
Felser, C., & Drummer, J.-D. (2017). Sensitivity to crossover constraints during native and non-native pronoun resolution. *Journal of Psycholinguistic Research, 46*(3), 771–789. https://doi.org/10.1007/s10936-016-9465-8
Felser, C., & Roberts, L. (2007). Processing wh-dependencies in a second language: A cross-modal priming study. *Second Language Research, 23*(1), 9–36. https://doi.org/10.1177/0267658307071600
Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics, 24*(3), 453–489. https://doi.org/10.1017/S0142716403000237
Fernandez, L. B., Engelhardt, P. E., Patarroyo, A. G., & Allen, S. E. (2020). Effects of speech rate on anticipatory eye movements in the visual world paradigm: Evidence from aging, native, and non-native language processing. *Quarterly Journal of Experimental Psychology, 73*(12), 2348–2361. https://doi.org/10.1177/1747021820948019
Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*(1), 11–15. https://doi.org/10.1111/1467-8721.00158
Ferreira, F., & Patson, N. D. (2007). The "good enough" approach to language comprehension. *Language and Linguistics Compass, 1*(1–2), 71–83. https://doi.org/10.1111/j.1749-818X.2007.00007.x
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*(2), 250–272. https://doi.org/10.1016/S0749-596X(02)00004-9
Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology.* MIT Press.
Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Bilingualism: Language and Cognition, 14*(3), 379–399. https://doi.org/10.1017/S136672891000012X
Foucart, A., Martin, C. D., Moreno, E. M., Costa, A., & Greene, R. L. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(5), 1461–1469. https://doi.org/10.1037/a0036756
Foucart, A., Ruiz-Tada, E., & Costa, A. (2016). Anticipation processes in L2 speech comprehension: Evidence from ERPs and lexical recognition tasks. *Bilingualism: Language and Cognition, 19*(1), 213–219. https://doi.org/10.1017/S1366728915000486
Frazier, L. (1987). Syntactic Processing: Evidence from Dutch. *Natural Language and Linguistic Theory*, *5*(4), 519–559. https://doi.org/10.1007/BF00138988
Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, *4*(2), 93–126. https://doi.org/10.1080/01690968908406359
Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*(2), 178–210. https://doi.org/10.1016/0010-0285(82)90008-1

Frenck-Mestre, C. (2002). An on-line look at sentence processing in the second language. In *Advances in Psychology* (Vol. 134, pp. 217–236). Elsevier Science & Technology. https://doi.org/10.1016/S0166-4115(02)80012-7

Friederici, A. D., Mecklinger, A., Spencer, K. M., Steinhauer, K., & Donchin, E. (2001). Syntactic parsing preferences and their online revisions: A spatio-temporal analysis of event-related brain potentials. *Cognitive Brain Research, 11*(3), 305–323. https://doi.org/10.1016/S0926-6410(00)00062-3

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(2), 78–84. https://doi.org/10.1016/S1364-6613(00)01839-8

Friederici, A. D., Hahne, A., & Saddy, D. (2002). Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. *Journal of Psycholinguistic Research, 31*(1), 45–63. https://doi.org/10.1023/A:1014376204525

Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences, 99*(1), 529–534. https://doi.org/10.1073/pnas.012611199

Friederici, A. D. (2004). Event-related brain potential studies in language. *Current Neurology and Neuroscience Reports, 4*(6), 466–470. https://doi.org/10.1007/s11910-004-0070-7

Fromont, L. A., Royle, P., & Steinhauer, K. (2020a). Growing random forests reveals that exposure and proficiency best account for individual variability in L2 (and L1) brain potentials for syntax and semantics. *Brain and Language, 204*, Article 104770. https://doi.org/10.1016/j.bandl.2020.104770

Fromont, L. A., Steinhauer, K., Royle, P., & Molinaro, N. (2020b). Verbing nouns and nouning verbs: Using a balanced design provides ERP evidence against "syntax-first" approaches to sentence processing. *PloS One, 15*(3), e0229169. https://doi.org/10.1371/journal.pone.0229169

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. https://doi.org/10.1007/BF01069046

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*(1), 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain* (pp. 95–126). MIT Press. https://doi.org/10.7551/mitpress/3654.001.0001

Gosselin, L., & Sabourin, L. (2024). Language-specific cognitive flexibility is related to code-switching habits and interactional context; domain-general cognitive flexibility is not. *Journal of Cognitive Psychology.* https://doi.org/10.1080/20445911.2024.2365463

Grüter, T., Lau, E., & Ling, W. (2020). How classifiers facilitate predictive processing in L1 and L2 Chinese: The role of semantic and grammatical cues. *Language, Cognition and Neuroscience, 35*(2), 221–234. https://doi.org/10.1080/23273798.2019.1648840

Grüter, T., & Rohde, H. (2013). L2 processing is affected by RAGE: Evidence from reference resolution. Abstract from *12th Conference on Generative Approaches to Second Language Acquisition (GASLA)*, United States.

Grüter, T., & Rohde, H. (2021). Limits on expectation-based processing: Use of grammatical aspect for co-reference in L2. *Applied Psycholinguistics, 42*(1), 51–75. https://doi.org/10.1017/S0142716420000582

Gundel, J. K. (1988). *Universals of topic-comment structure.* In M. Hammond, E. A. Moravcsik, & J. R. Wirth (Eds.), *Studies in syntactic typology* (pp. 209–239). Amsterdam: John Benjamins.

Gunter, T. C., Friederici, A. D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience, 12*(4), 556–568. https://doi.org/10.1162/089892900562336

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory Interference During Language Processing. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 27*(6), 1411–1423. https://doi.org/10.1037/0278-7393.27.6.1411

Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes, 25*(2), 149–188. https://doi.org/10.1080/01690960902965951

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*(11), 1531–1549. https://doi.org/10.1016/S0028-3932(00)00053-1

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience, 11*(2), 194–205. https://doi.org/10.1162/089892999563328

Hahne, A., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition, 4*(2), 123–141. https://doi.org/10.1017/S1366728901000232

Hawkins, J. A. (1999). Processing complexity and filler-gap dependencies across grammars. *Language, 75*(2), 244–285. https://doi.org/10.2307/417261

Hawkins, R., & Yuet-hung Chan, C. (1997). The partial availability of universal grammar in second language acquisition: The "failed functional features hypothesis." *Second Language Research, 13*(3), 187–226. https://doi.org/10.1191/026765897671476153

Hernandez, A., Li, P., & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends in Cognitive Sciences, 9*(5), 220–225. https://doi.org/10.1016/j.tics.2005.03.003

Hernandez, A. E., & Li, P. (2007). Age of acquisition: Its neural and computational mechanisms. *Psychological Bulletin, 133*(4), 638–650. https://doi.org/10.1037/0033-2909.133.4.638

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*(6), 425–438. https://doi.org/10.1007/s10936-011-9197-8

Hestvik, A., Maxfield, N., Schwartz, R. G., & Shafer, V. (2007). Brain responses to filled gaps. *Brain and Language, 100*(3), 301–316. https://doi.org/10.1016/j.bandl.2006.07.007

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua, 120*(4), 901–931. https://doi.org/10.1016/j.lingua.2009.06.004

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research, 29*(1), 33–56. https://doi.org/10.1177/0267658312461803

Hopp, H. (2015). Semantics and morphosyntax in predictive L2 sentence processing. *International Review of Applied Linguistics in Language Teaching (IRAL), 53*(3), 277–306. https://doi.org/10.1515/iral-2015-0014

Hopp, H. (2017). Cross-linguistic lexical and syntactic co-activation in L2 sentence processing. *Linguistic Approaches to Bilingualism, 7*(1), 96–130. https://doi.org/10.1075/lab.14027.hop

Hopp, H. (2022). Second language sentence processing. *Annual Review of Linguistics, 8*(1), 235–256. https://doi.org/10.1146/annurev-linguistics-030821-054113

Huettig, F., Audring, J., & Jackendoff, R. (2022). A parallel architecture perspective on pre-activation and prediction in language processing. *Cognition, 224*, 105050. https://doi.org/10.1016/j.cognition.2022.105050

Huang, C.-T. J. (1982). *Logical relations in Chinese and the theory of grammar* [Doctoral dissertation, Massachusetts Institute of Technology]. ProQuest Dissertations & Theses.

Huang, C. T. J., Li, A., & Li, Y. F. (2009). *Syntax of Chinese.* Cambridge University Press. https://doi.org/10.1017/CBO9781139166935

Huang, Y. C., & Kaiser, E. (2008). Investigating filler-gap dependencies in Chinese topicalization. In K. M. Chan & H. Kang (Eds.), *Proceedings of the 20th North American Conference on Chinese Linguistics* (Vol. 2, pp. 927–941). The Ohio State University.

Indefrey, P. (2006). A meta-analysis of hemodynamic studies on first and second language processing: Which suggested differences can we trust and what do they mean? *Language Learning, 56*(s1), 279–304. https://doi.org/10.1111/j.1467-9922.2006.00365.x

Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language, 98*, 1–11. https://doi.org/10.1016/j.jml.2017.09.002

Janacsek, K., Fiser, J., & Nemeth, D. (2012). The best time to acquire new skills: Age-related differences in implicit sequence learning across the human lifespan. *Developmental Science, 15*(4), 496–505. https://doi.org/10.1111/j.1467-7687.2012.01150.x

Jessen, A., & Felser, C. (2019). Reanalysing object gaps during non-native sentence processing: Evidence from ERPs. *Second Language Research, 35*(2), 285–300. https://doi.org/10.1177/0267658317753030

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. https://doi.org/10.1007/s10936-017-9496-9

Johnson, A. M. (2015). *Individual differences in predictive processing: Evidence from subject filled-gap effects in native and nonnative speakers of English* [Doctoral dissertation]. ProQuest Dissertations & Theses.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*(1), 60–99. https://doi.org/10.1016/0010-0285(89)90003-0

Juffs, A. (2005). The influence of first language on the processing of wh-movement in English as a second language. *Second Language Research, 21*(2), 121–151. https://doi.org/10.1191/0267658305sr255oa

Juffs, A., & Harrington, M. (1996). Garden path sentences and error data in second language sentence processing. *Language Learning, 46*(2), 283–323. https://doi.org/10.1111/j.1467-1770.1996.tb01237.x

Jung, S. (2017). *The effects of age of immersion and working memory on second language processing of island constraints: An eye-movement study* [Doctoral dissertation]. ProQuest Dissertations & Theses.

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism, 4*(2), 257–282. https://doi.org/10.1075/lab.4.2.05kaa

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15*(2), 159–201. https://doi.org/10.1080/016909600386084

Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience, 15*(1), 98–110. https://doi.org/10.1162/089892903321107855

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research, 32*(1), 37–55. https://doi.org/10.1023/A:1021933015362

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52*(2), 205–225. https://doi.org/10.1016/j.jml.2004.10.002

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. https://doi.org/10.1162/jocn.1995.7.3.376

Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience, 5*(2), 196–214. https://doi.org/10.1162/jocn.1993.5.2.196

Kong, D., & Hsu, Y.-Y. (2023). Processing covert dependency: An eye-tracking study of scope interpretations of embedded wh-questions in Mandarin. *PLOS ONE, 18*(5), e0285873. https://doi.org/10.1371/journal.pone.0285873

Kuno, S. (1976). *Subject, theme, and the speaker's empathy: A reexamination of relativization phenomena*. In C. N. Li (Ed.), Subject and topic (pp. 417–444). New York: Academic Press.

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Brain Research. Cognitive Brain Research, 17*(1), 117–129. https://doi.org/10.1016/S0926-6410(03)00086-7

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463–470. https://doi.org/10.1016/S1364-6613(00)01560-6

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component. *Annual Review of Psychology, 62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Kwon, N. (2008). *Processing of syntactic and anaphoric gap-filler dependencies in Korean: Evidence from self -paced reading time, ERP and eye-tracking experiments*. ProQuest Dissertations & Theses.

Leckey, M., & Federmeier, K. D. (2020). The P3b and P600(s): Positive contributions to language comprehension. *Psychophysiology, 57*(7), e13351. https://doi.org/10.1111/psyp.13351

Lee, M. W. (2004). Another look at the role of empty categories in sentence processing (and grammar). *Journal of Psycholinguistic Research, 33*(1), 51–73. https://doi.org/10.1023/B:JOPR.0000010514.50468.30

Lenneberg, E. H. (1967). *Biological foundations of language.* Wiley.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29*(3), 375–419. https://doi.org/10.1207/s15516709cog0000_25

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. https://doi.org/10.1016/j.jml.2010.07.003

Li, C. N. (1976). Subject and topic: A new typology of language. In C. N. Li (Ed.), *Subject and topic* (pp. 458–489). Academic Press.

Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar.* University of California Press.

Love, T., & Swinney, D. (1996). Coreference processing and levels of analysis in object-relative constructions: Demonstration of antecedent reactivation with the cross-modal priming paradigm. *Journal of Psycholinguistic Research, 25*(1), 5–24. https://doi.org/10.1007/BF01708418

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). MIT Press.

MacDonald, M. C., & MacWhinney, B. (1999). Distributional information in language comprehension, production, and acquisition: Three puzzles and a moral. In B. MacWhinney (Ed.), *The emergence of language* (1st ed., pp. 177–196). Psychology Press. https://doi.org/10.4324/9781410602367-6

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*(4), 676–703. https://doi.org/10.1037/0033-295X.101.4.676

MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 49–67). Oxford University Press.

MacWhinney, B. (2013). The logic of the unified model. In *The Routledge handbook of second language acquisition* (pp. 211–227). Routledge. https://doi.org/10.4324/9780203808184

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition, 27*(1), 53–78. https://doi.org/10.1017/S0272263105050035

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*(4), 574–588. https://doi.org/10.1016/j.jml.2013.08.001

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language, 55*(3), 381–401. https://doi.org/10.1016/j.jml.2006.06.006

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research, 29*(2), 111–123. https://doi.org/10.1023/A:1005184709695

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(3), 817–835. https://doi.org/10.1037/0278-7393.27.3.817

McRae, K., Matsuki, K., & van Gompel, R. P. G. (2013). Constraint-based models of sentence processing. In R. P. G. van Gompel (Ed.), *Sentence processing* (1st ed., pp. 51–77). Psychology Press. https://doi.org/10.4324/9780203488454-3

Michel, D. (2014). *Individual cognitive measures and working memory accounts of syntactic island phenomena* [Doctoral dissertation]. ProQuest Dissertations & Theses.

Mickan, A., & Lemhöfer, K. (2020). Tracking syntactic conflict between languages over the course of L2 acquisition: A cross-sectional ERP study. *Journal of Cognitive Neuroscience, 32*(5), 822–846. https://doi.org/10.1162/jocn_a_01528

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. https://doi.org/10.1017/S1366728914000881

Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex, 47*(8), 908–930. https://doi.org/10.1016/j.cortex.2011.02.019

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2007). An ERP investigation of second language processing: Effects of proficiency and explicit and implicit training. *Journal of Cognitive Neuroscience, Supplement*, 164.

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation

patterns. *Journal of Cognitive Neuroscience, 24*(4), 933–947. https://doi.org/10.1162/jocn_a_00119
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. https://doi.org/10.1023/A:1021260920232
Ness, T., & Meltzer-Asscher, A. (2017). Working memory in the processing of long-distance dependencies: Interference and filler maintenance. *Journal of Psycholinguistic Research, 46*(6), 1353–1365. https://doi.org/10.1007/s10936-017-9499-6
Ng, S., & Wicha, N. Y. Y. (2014). Processing gap-filler dependencies in Chinese: What does it tell us about semantic processing? *Journal of Memory and Language, 74*, 16–35. https://doi.org/10.1016/j.jml.2014.04.002
Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research, 18*(1), 5–19. https://doi.org/10.1007/BF01069043
Ning, C. (1993). *The overt syntax of relativization and topicalization in Chinese* [Doctoral dissertation]. ProQuest Dissertations & Theses.
Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., & Bunting, M. F. (2014). Clearing the garden-path: Improving sentence processing through cognitive control training. *Language and Cognitive Processes, 29*(2), 186–217. https://doi.org/10.1080/01690965.2012.758297
O'Grady, W., Lee, M., & Choo, M. (2003). A subject-object asymmetry in the acquisition of relative clauses in Korean as a second language. *Studies in Second Language Acquisition, 25*(4), 433–448. https://doi.org/10.1017/S0272263103000172
Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience, 17*(8), 1212–1228. https://doi.org/10.1162/0898929055002436
Omaki, A. (2010). *Commitment and flexibility in the developing parser* [Doctoral dissertation]. ProQuest Dissertations & Theses.
Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island constraints in second-language sentence processing. *Studies in Second Language Acquisition, 33*(4), 563–588. https://doi.org/10.1017/S0272263111000313
Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*(6), 785–806. https://doi.org/10.1016/0749-596X(92)90039-Z
Pakulak, E., & Neville, H. J. (2010). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience, 22*(12), 2728–2744. https://doi.org/10.1162/jocn.2009.21329
Pakulak, E., & Neville, H. J. (2011). Maturational constraints on the recruitment of early processes for syntactic processing. *Journal of Cognitive Neuroscience, 23*(11), 2752–2765. https://doi.org/10.1162/jocn.2010.21586
Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition, 25*(4), 501–528. https://doi.org/10.1017/s0272263103000214
Penfield, W., & Roberts, L. (1959). *Speech and brain mechanisms.* Princeton University Press.
Phillips, C. (2006). The real-time status of island phenomena. *Language, 82*(4), 795–823. https://doi.org/10.1353/lan.2006.0217
Phillips, C., & Ehrenhofer, L. (2015). The role of language processing in language acquisition. *Linguistic Approaches to Bilingualism, 5*(4), 409–453. https://doi.org/10.1075/lab.5.4.01phi

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. https://doi.org/10.1016/j.cogbrainres.2004.09.012
Pickering, M. (1993). Direct association and sentence processing: A reply. *Language and Cognitive Processes, 8*(2), 163–196. https://doi.org/10.1080/01690969308406953
Pickering, M., & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes, 6*(3), 229–259. https://doi.org/10.1080/01690969108406944
Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin, 144*(10), 1002–1044. https://doi.org/10.1037/bul0000158
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. https://doi.org/10.1037/0278-7393.27.6.1401
Pliatsikas, C., & Marinis, T. (2013). Processing empty categories in a second language: When naturalistic exposure fills the (intermediate) gap. *Bilingualism: Language and Cognition, 16*(1), 167–182. https://doi.org/10.1017/S136672891200017X
R Core Team. (2024). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/
Reinhart, T. (1981). Pragmatics and linguistics: *An analysis of sentence topics. Philosophica,* 27, 53–94.
Roberts, L., Marinis, T., Felser, C., & Clahsen, H. (2007). Antecedent Priming at Trace Positions in Children's Sentence Processing. *Journal of Psycholinguistic Research*, *36*(2), 175–188. https://doi.org/10.1007/s10936-006-9038-3
Rodriguez, G. A. (2008). *Second language sentence processing: Is it fundamentally different?* [Doctoral dissertation]. ProQuest Dissertations & Theses.
Ross, J. R. (1976). *Constraints on variables in syntax* [Doctoral dissertation]. Massachusetts Institute of Technology.
Rossi, S., Gugler, M. F., Friederici, A. D., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience, 18*(12), 2030–2048.
Rüschemeyer, S.-A., Zysset, S., & Friederici, A. D. (2006). Native and non-native reading of sentences: An fMRI experiment. *NeuroImage, 31*(1), 354–365. https://doi.org/10.1016/j.neuroimage.2005.11.047
Sabourin, L. (2006). Does the shallow structures proposal account for qualitative differences in first and second language processing? *Applied Psycholinguistics, 27*(1), 81–84. https://doi.org/10.1017/S0142716406060140
Sabourin, L., Brien, C., & Burkholder, M. (2014). The effect of age of L2 acquisition on the organization of the bilingual lexicon: Evidence from masked priming. *Bilingualism: Language and Cognition, 17*(3), 542–555. https://doi.org/10.1017/S1366728913000643
Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research, 24*(3), 397–430. https://doi.org/10.1177/0267658308090186
Sabourin, L., Leclerc, J.-C., Lapierre, M., Burkholder, M., & Brien, C. (2016). The language background questionnaires in L2 research: Teasing apart the variables. In L. Hracs (Ed.), *Proceedings of the 2016 Annual Conference of the Canadian Linguistics Association.* Canadian Linguistics Association.
Schlenter, J. (2023). Prediction in bilingual sentence processing: How prediction differs in a later learned language from a first language. *Bilingualism: Language and Cognition, 26*(2), 253–267. https://doi.org/10.1017/S1366728922000736

Schütze, C. T., & Gibson, E. (1999). Argumenthood and English prepositional phrase attachment. *Journal of Memory and Language, 40*(3), 409–431. https://doi.org/10.1006/jmla.1998.2619
Song, K., Chang, H., & Wang, Y. (2022). Processing of Chinese base-generated-topic sentences by L1-Korean speakers: An eye-tracking study. *Brain Sciences, 12*(11), Article 1573. https://doi.org/10.3390/brainsci12111573
Sprouse, J., Wagers, M., & Phillips, C. (2012). A test of the relation between working-memory capacity and syntactic island effects. *Language, 88*(1), 82–123. https://doi.org/10.1353/lan.2012.0004
Staub, A., & Clifton, C. (2006). Syntactic prediction in language comprehension: Evidence from either...or. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(2), 425–436. https://doi.org/10.1037/0278-7393.32.2.425
Steinhauer, K. (2014). Event-related potentials (ERPs) in second language research: A brief introduction to the technique, a selected review, and an invitation to reconsider critical periods in L2. *Applied Linguistics, 35*(4), 393–417. https://doi.org/10.1093/applin/amu028
Steinhauer, K., White, E., Cornell, S., Genesee, F., & White, L. (2006). The neural dynamics of second language acquisition: Evidence from event-related potentials. *Journal of Cognitive Neuroscience, 25*(Suppl.), 99. https://doi.org/10.1177/0267658308098995
Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research, 25*(1), 13–41. https://doi.org/10.1177/0267658308098995
Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, *1*(3), 227–245. https://doi.org/10.1080/01690968608407062
Sussman, R. S., & Sedivy, J. (2003). The time-course of processing syntactic dependencies: Evidence from eye movements. *Language and Cognitive Processes, 18*(2), 143–163. https://doi.org/10.1080/01690960143000498
Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition, 17*(2), 277–293. https://doi.org/10.1017/S1366728913000370
Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition, 27*(2), 173–204. https://doi.org/10.1017/S0272263105050102
Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules.* MIT Press.
Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language, 35*(3), 454–475. https://doi.org/10.1006/jmla.1996.0025
Trenkic, D., Mirkovic, J., & Altmann, G. T. M. (2014). Real-time grammar processing by native and non-native speakers: Constructions unique to the second language. *Bilingualism: Language and Cognition, 17*(2), 237–257. https://doi.org/10.1017/S1366728913000321
Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition, 73*(2), 89–134. https://doi.org/10.1016/S0010-0277(99)00032-3
Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33*(3), 285–318. https://doi.org/10.1006/jmla.1994.1014
Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition, 4*(2), 105–122. https://doi.org/10.1017/S1366728901000220
Ullman, M. T. (2006). The declarative/procedural model and the shallow structure hypothesis. *Applied Psycholinguistics, 27*(1), 97–105. https://doi.org/10.1017/S014271640606019X

Ullman, M. T., & Sanz, C. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In *Mind and context in adult second language acquisition* (pp. 141–178). Georgetown University Press.

van Bergen, G., & Flecken, M. (2017). Putting things in new places: Linguistic experience modulates the predictive power of placement verb semantics. *Journal of Memory and Language, 92*, 26–42. https://doi.org/10.1016/j.jml.2016.05.003

van Gompel, R. P. G., Pickering, M. J., & Gaskell, M. G. (2012). Syntactic parsing. In *The Oxford handbook of psycholinguistics.* Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198568971.013.0017

van Hell, J. G. (2023). The neurocognitive underpinnings of second language processing: Knowledge gains from the past and future outlook. *Language Learning, 73*(S1), 95–138. https://doi.org/10.1111/lang.12601

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*(2), 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015

Wagers, M. W., & Pendleton, E. (2016). Structuring expectation: Licensing animacy in relative clause comprehension. In K.-M. Kim et al. (Eds.), *Proceedings of the 33rd West Coast Conference on Formal Linguistics* (pp. 29–46). Cascadilla Proceedings Project.

Wagers, M. W., & Phillips, C. (2009). Multiple dependencies and the role of the grammar in real-time comprehension. *Journal of Linguistics, 45*(2), 395–433. https://doi.org/10.1017/S0022226709005726

Wagers, M. W., & Phillips, C. (2014). Going the distance: Memory and control processes in active dependency construction. *Quarterly Journal of Experimental Psychology, 67*(7), 1274–1304. https://doi.org/10.1080/17470218.2013.858363

Wanner, E., & Maratsos, M. (1978). An ATN approach to comprehension. In M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 119–161). MIT Press.

Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., & Perani, D. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron, 37*(1), 159–170. https://doi.org/10.1016/S0896-6273(02)01150-9

Weber-Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience, 8*(3), 231–256.

Weber-Fox, C. M., & Neville, H. J. (1999). Functional neural subsystems are differentially affected by delays in second language immersion: ERP and behavioral evidence in bilinguals. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 23–38). Lawrence Erlbaum Associates.

White, L., Juffs, A., Martohardjono, G., Flynn, S., & O'Neil, W. (1998). Constraints on wh-movement in two different contexts of nonnative language acquisition: Competence and processing. In *The generative study of second language acquisition* (1st ed., pp. 111–129). Psychology Press. https://doi.org/10.4324/9781315806471-9

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience, 16*(7), 1272–1288. https://doi.org/10.1162/0898929041920487

Williams, J. N. (2006). Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition, 9*(1), 71–88. https://doi.org/10.1017/S1366728905002385

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. https://doi.org/10.1017/S0142716401004030

Wolpert, M., Zhang, H., Baum, S., & Steinhauer, K. (2024). Native and non-native parsing of adjective placement–An ERP study of Mandarin and English sentence processing. *Brain and Language, 254*, Article 105427. https://doi.org/10.1016/j.bandl.2024.105427

Wu, J. (1999). *Syntax and semantics of quantification in Chinese* [Doctoral dissertation]. ProQuest Dissertations & Theses.

Xiang, M., Dillon, B., Wagers, M., Liu, F., & Guo, T. (2014). Processing covert dependencies: An SAT study on Mandarin wh-in-situ questions. *Journal of East Asian Linguistics, 23*(2), 207–232. https://doi.org/10.1007/s10831-013-9115-1

Xiang, M., Wang, S., & Cui, Y. (2015). Constructing covert dependencies—The case of Mandarin wh-in-situ dependency. *Journal of Memory and Language, 84*, 139–166. https://doi.org/10.1016/j.jml.2015.05.006

Xu, L. (2006). Topicalization in Asian languages. In M. Everaert & H. van Riemsdijk (Eds.), *The Blackwell companion to syntax* (Vol. 5, pp. 137–174). Blackwell Publishing. https://doi.org/10.1002/9780470996591.ch73

Yan, H., & Liu, J. (2021). Subject filled-gap effects: A self-paced reading study with Mandarin learners of English. *Journal of Psycholinguistic Research, 50*(5), 1049–1064. https://doi.org/10.1007/s10936-021-09781-4

Yang, Y., & Liu, T. (2014). Neural mechanisms of trace in Chinese topicalized constructions. *Social Sciences in China, 35*(1), 86–111. https://doi.org/10.1080/02529203.2013.875661

Yang, Y., Pablos, L., & Cheng, L. L.-S. (2023). The processing mechanisms of Mandarin wh-questions. *Journal of Chinese Linguistics, 51*(1), 147–171. https://doi.org/10.1353/jcl.2023.0009

Yang, C. L., Perfetti, C. A., & Liu, Y. (2010). Sentence integration processes: An ERP study of Chinese sentence comprehension with relative clauses. *Brain and Language, 112*(2), 85–100. https://doi.org/10.1016/j.bandl.2009.10.005

Zawiszewski, A., Gutiérrez, E., Fernández, B., & Laka, I. (2011). Language distance and non-native syntactic processing: Evidence from event-related potentials. *Bilingualism: Language and Cognition, 14*(3), 400–411. https://doi.org/10.1017/S1366728910000350

Appendices

Appendix A

Language Background Questionnaire

Language Background Questionnaire - Short Version

Participant information – *To be filled out by the researcher*		
Project code: ____________	Today's date: ____________	Participant code: ______________

1. Biographical information

Month and year of birth: ____________________ Current Age: ________ Gender: _____________

Your native language(s)? ____
Mother's native language(s)? ____
Father's native language(s)? ____
Other caregivers' native language(s)? ____

Are you left- or right-handed? ____
Do you have any known visual impairments (including colour-blindness)? ____________________
Do you have any known hearing impairments? ____
Have you ever had a serious head injury? ____

What is your current profession? ____

Place of birth (city, country): ____

Please list all the places where you have lived, when you lived there, and for how long. Please list these in chronological order. ____
____
____

2. Basic language information

For an example, see page 3. Please ask if you have any questions.

In the **first row**, please list all the spoken or signed languages of which you have ANY current or previous knowledge in order of **DOMINANCE**, i.e. how comfortable you are using them.

In the **second row**, please write the age at which you were **first exposed** to this language, i.e. when it first became present in your environment.

In the **third row**, please indicate whether you were ever **IMMMERSED** in this language, i.e. if you received significant exposure to this language in either a community/home setting **or** a school (greater than 20%). If so, please also indicate the age or grade at which this began.

Dominance	Most ⟷ Least				
Language:	1.	2.	3.	4.	5.
Age/grade of 1st exposure:					
Immersed: Age/grade:	YES / NO ______	YES / NO ______	YES / NO ______	YES / NO ______	YES / NO ______

Language Background Questionnaire - Short Version

For each of the above-listed languages, please provide your current and highest ever attained level of proficiency. This consists of your **general** proficiency. If you feel that you are between categories, simply indicate where along the scale you feel your proficiency is best represented.

Language:	Current level of general proficiency					
	very low	low	intermediate	advanced	near-native	native
________	Highest level of proficiency ever attained					
	very low	low	intermediate	advanced	near-native	native

Language:	Current level of general proficiency					
	very low	low	intermediate	advanced	near-native	native
________	Highest level of proficiency ever attained					
	very low	low	intermediate	advanced	near-native	native

Language:	Current level of general proficiency					
	very low	low	intermediate	advanced	near-native	native
________	Highest level of proficiency ever attained					
	very low	low	intermediate	advanced	near-native	native

Language:	Current level of general proficiency					
	very low	low	intermediate	advanced	near-native	native
________	Highest level of proficiency ever attained					
	very low	low	intermediate	advanced	near-native	native

Language:	Current level of general proficiency					
	very low	low	intermediate	advanced	near-native	native
________	Highest level of proficiency ever attained					
	very low	low	intermediate	advanced	near-native	native

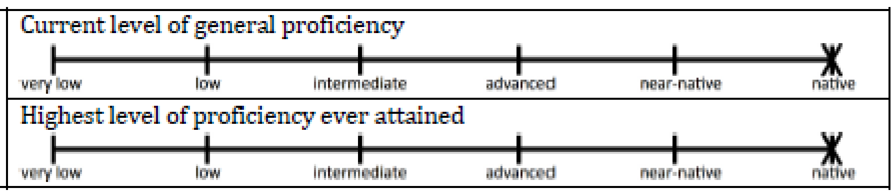
Language Background Questionnaire - Short Version

EXAMPLE 1: Julia's mother is a native speaker of Spanish, and her father is a native speaker of English. However, Julia's mother passed away when she was four years old, and Julia therefore did not get much exposure to Spanish after that age. Julia spoke Spanish natively until the age of four (highest level of proficiency ever attained) and considers her current level of proficiency to be advanced. Her current and highest ever level of proficiency in English is native. Furthermore, Julia was in French immersion all throughout high school. When she graduated, her level of proficiency was advanced, but she now feels that it is closer to intermediate. She also took a beginner-level course in Japanese when she was in her first year of university.

Dominance:	Most ⟵			⟶	Least
Language:	1. English	2. Spanish	3. French	4. Japanese	5.
Age/grade of 1st exposure:	Birth (age=0)	Birth (age=0)	First grade (age=6)	First year univ. (age=18)	
Immersion: Age/grade:	**<u>YES</u>** / NO Age: Birth	**<u>YES</u>** / NO Age: Birth	**<u>YES</u>** / NO Age: 12	YES / **<u>NO</u>** Age: _______	YES / NO Age: _______

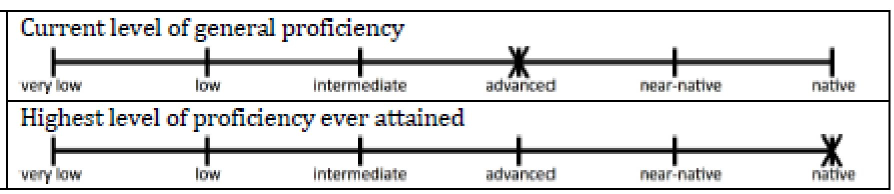
Language:

<u>English</u>



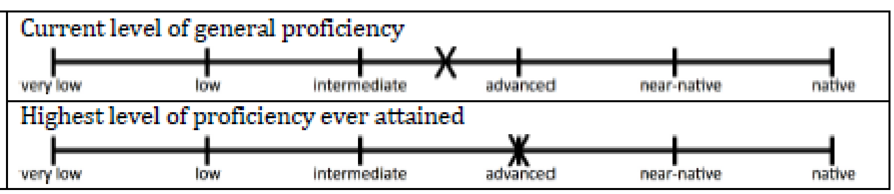
Language:

<u>Spanish</u>



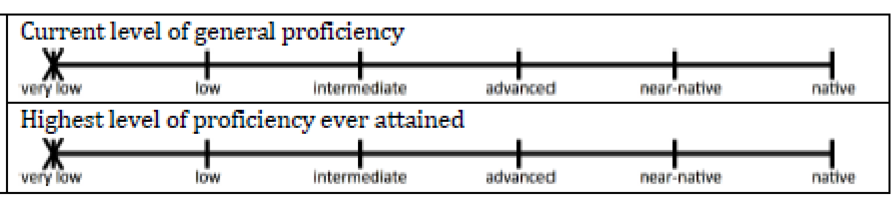
Language:

<u>French</u>



Language:

<u>Japanese</u>



Appendix B

Cloze task

Cloze Task

Man is the only living creature that can make and use tools. He is the most teachable of living beings, earning the name of Homo sapiens. (1) ______ ever restless brain has used the (2) ______ and the wisdom of his ancestors (3) ______ improve his way of life. Since (4) ______ is able to walk and run (5) ______ his feet, his hands have always (6) ______ free to carry and to use (7) . Man's hands have served him well (8) his life on earth. His development, (9) ______ can be divided into three major (10) ______, is marked by several different ways (11) ______ life.

Up to 10,000 years ago, (12) ______ human beings lived by hunting and (13) ______. They also picked berries and fruits, (14) ______ dug for various edible roots. Most (15) ______, the men were the hunters, and (16) ______ women acted as food gatherers. Since (17) ______ women were busy with the children, (18) ______ men handled the tools. In a (19) ______ hand, a dead branch became a (20) to knock down fruit or to (21) for tasty roots. Sometimes, an animal (22) ______ served as a club, and a (23) ______ piece of stone, fitting comfortably into (24) ______ hand, could be used to break (25) ______ or to throw at an animal. (26) ______ stone was chipped against another until (27) ______ had a sharp edge. The primitive (28) ______ who first thought of putting a (29) ______ stone at the end of a (30) ______ made a brilliant discovery: he (31) ______ joined two things to make a (32) ______ useful tool, the spear. Flint, found (33) ______ many rocks, became a common cutting (34) ______ in the Paleolithic period of man's (35) ______. Since no wood or bone tools (36) ______ survived, we know of this man (37) ______ his stone implements, with which he (38) ______ kill animals, cut up the meat, (39) ______ scrape the skins, as well as (40) pictures on the walls of the (41) where he lived during the winter.

(42) ______ the warmer seasons, man wandered on (43) ______ steppes of Europe without a fixed (44) ______, always foraging for food. Perhaps the (45) ______ carried nuts and berries in shells (46) ______ skins or even in light, woven (47) ______. Wherever they camped, the primitive people (48) ______ fires by striking flint for sparks (49) ______ using dried seeds, moss, and rotten (50) for tinder. With fires that he kindled himself, man could keep wild animals away and could cook those that he killed, as well as provide warmth and light for himself.

Appendix C
Experiment 1 Experimental Stimuli

Set	Distance	Plausibility	filler-gap	yes/no questions	answer
1	long	plausible	The coffee shop was very busy today, and the staff hardly had any time to rest. After the last table left, **the manager knew which customer the receptionist with a large cloth bag called about the problem.**	Was the coffee shop busy today?	Yes
1	long	implausible	The coffee shop was very busy today, and the staff hardly had any time to rest. After the last table left, **the manager knew which spoon the receptionist with a large cloth bag called about the problem.**	Was the coffee shop busy today?	Yes
1	short	plausible	The coffee shop was very busy today, and the staff hardly had any time to rest. **After the last table left, the manager knew which customer the receptionist called about the problem.**	Was the coffee shop busy today?	Yes
1	short	implausible	The coffee shop was very busy today, and the staff hardly had any time to rest. **After the last table left, the manager knew which spoon the receptionist called about the problem.**	Was the coffee shop busy today?	Yes
2	long	plausible	The jewelry store was busy with people looking for gifts for special occasions. As the evening settled in, **the jeweler checked which necklace the model with the long blonde hair wore for her anniversary.**	Were people looking for gifts at the jewelry store?	Yes
2	long	implausible	The jewelry store was busy with people looking for gifts for special occasions. As the evening settled in, **the jeweler checked which banker the model with the long blonde hair wore for her anniversary.**	Were people looking for gifts at the jewelry store?	Yes
2	short	plausible	The jewelry store was busy with people looking for gifts for special occasions. **As the evening settled in, the jeweler checked which necklace the model wore for her anniversary.**	Were people looking for gifts at the jewelry store?	Yes
2	short	implausible	The jewelry store was busy with people looking for gifts for special occasions. **As the evening settled in, the jeweler checked which banker the model wore for her anniversary.**	Were people looking for gifts at the jewelry store?	Yes

3	long	plausible	Despite having a busy schedule, the editor still kept an eye on all staff. As the new week began, **the editor knew which manuscript the author with a relatively good reputation revised over the weekend.**	Did the editor have a busy schedule?	Yes
3	long	implausible	Despite having a busy schedule, the editor still kept an eye on all staff. As the new week began, **the editor knew which professor the author with a relatively good reputation revised over the weekend.**	Did the editor have a busy schedule?	Yes
3	short	plausible	Despite having a busy schedule, the editor still kept an eye on all staff. **As the new week began, the editor knew which manuscript the author revised over the weekend.**	Did the editor have a busy schedule?	Yes
3	short	implausible	Despite having a busy schedule, the editor still kept an eye on all staff. **As the new week began, the editor knew which professor the author revised over the weekend.**	Did the editor have a busy schedule?	Yes
4	long	plausible	The stadium was filled with excited fans waiting for the game to start. In the midst of this, **the umpire asked which player the coach with the pair of glasses threatened before the game.**	Were the fans excited for the game to start?	Yes
4	long	implausible	The stadium was filled with excited fans waiting for the game to start. In the midst of this, **the umpire asked which football the coach with the pair of glasses threatened before the game.**	Were the fans excited for the game to start?	Yes
4	short	plausible	The stadium was filled with excited fans waiting for the game to start. **In the midst of this, the umpire asked which player the coach threatened before the game.**	Were the fans excited for the game to start?	Yes
4	short	implausible	The stadium was filled with excited fans waiting for the game to start. **In the midst of this, the umpire asked which football the coach threatened before the game.**	Were the fans excited for the game to start?	Yes
5	long	plausible	After the surgery, the hospital was noisy with patients talking to each other. By paying attention to everyone, **the doctor noticed which nurse the patient in the blue hospital gown complimented after the surgery.**	Was the hospital quiet with patients resting?	No

5	long	implausible	After the surgery, the hospital was noisy with patients talking to each other. By paying attention to everyone, **the doctor noticed which bed the patient in the blue hospital gown complimented after the surgery.**	Was the hospital quiet with patients resting?	No
5	short	plausible	After the surgery, the hospital was noisy with patients talking to each other. **By paying attention to everyone, the doctor noticed which nurse the patient complimented after the surgery.**	Was the hospital quiet with patients resting?	No
5	short	implausible	After the surgery, the hospital was noisy with patients talking to each other. **By paying attention to everyone, the doctor noticed which bed the patient complimented after the surgery.**	Was the hospital quiet with patients resting?	No
6	long	plausible	As the semester went on, the school's grades improved a lot. After thinking about this thoroughly, **the dean asked which teachers the class with a surprisingly big progress admired most that fall.**	Did the school's grades get worse?	No
6	long	implausible	As the semester went on, the school's grades improved a lot. After thinking about this thoroughly, **the dean asked which tests the class with a surprisingly big progress admired most that fall.**	Did the school's grades get worse?	No
6	short	plausible	As the semester went on, the school's grades improved a lot. **After thinking about this thoroughly, the dean asked which teachers the class admired most that fall.**	Did the school's grades get worse?	No
6	short	implausible	As the semester went on, the school's grades improved a lot. **After thinking about this thoroughly, the dean asked which tests the class admired most that fall.**	Did the school's grades get worse?	No
7	long	plausible	The property management office was often busy dealing with many issues. Looking back at recent incidents, **the realtor forgot which agent the tenants in the striking yellow sweaters employed late last month.**	Was the office often busy?	Yes
7	long	implausible	The property management office was often busy dealing with many issues. Looking back at recent incidents, **the realtor forgot which meeting the tenants in the striking yellow sweaters employed late last month.**	Was the office often busy?	Yes

7	short	plausible	The property management office was often busy dealing with many issues. **Looking back at recent incidents, the realtor forgot which agent the tenants employed late last month.**	Was the office often busy?	Yes
7	short	implausible	The property management office was often busy dealing with many issues. **Looking back at recent incidents, the realtor forgot which meeting the tenants employed late last month.**	Was the office often busy?	Yes
8	long	plausible	The courtroom was very tense as the important trial went on. During the short break time, **the lawyer established which witness the police in their bullet proof vests confronted about the crime.**	Was the trial important?	Yes
8	long	implausible	The courtroom was very tense as the important trial went on. During the short break time, **the lawyer established which car the police in their bullet proof vests confronted about the crime.**	Was the trial important?	Yes
8	short	plausible	The courtroom was very tense as the important trial went on. **During the short break time, the lawyer established which witness the police confronted about the crime.**	Was the trial important?	Yes
8	short	implausible	The courtroom was very tense as the important trial went on. **During the short break time, the lawyer established which car the police confronted about the crime.**	Was the trial important?	Yes
9	long	plausible	In the busy hospital, nurses and doctors were talking about patient care. In one of these talks, **the nurses asked which surgeon the doctor with the short white beard contacted about the procedure.**	Was the hospital empty that day?	No
9	long	implausible	In the busy hospital, nurses and doctors were talking about patient care. In one of these talks, **the nurses asked which briefcase the doctor with the short white beard contacted about the procedure.**	Was the hospital empty that day?	No
9	short	plausible	In the busy hospital, nurses and doctors were talking about patient care. **In one of these talks, the nurses asked which surgeon the doctor contacted about the procedure.**	Was the hospital empty that day?	No
9	short	implausible	In the busy hospital, nurses and doctors were talking about patient care. **In one of these talks, the nurses asked which briefcase the doctor contacted about the procedure.**	Was the hospital empty that day?	No

10	long	plausible	The kitchen in the fancy restaurant was full of chefs making tasty food. Amidst all the busy preparation, **the chef decided which carrot the trainee of the renowned gourmet restaurant roasted as an appetizer.**	Were the chefs doing nothing in the kitchen?	No
10	long	implausible	The kitchen in the fancy restaurant was full of chefs making tasty food. Amidst all the busy preparation, **the chef decided which waiter the trainee of the renowned gourmet restaurant roasted as an appetizer.**	Were the chefs doing nothing in the kitchen?	No
10	short	plausible	The kitchen in the fancy restaurant was full of chefs making tasty food. **Amidst all the busy preparation, the chef decided which carrot the trainee roasted as an appetizer.**	Were the chefs doing nothing in the kitchen?	No
10	short	implausible	The kitchen in the fancy restaurant was full of chefs making tasty food. **Amidst all the busy preparation, the chef decided which waiter the trainee roasted as an appetizer.**	Were the chefs doing nothing in the kitchen?	No
11	long	plausible	The meeting room was filled with discussions about past successes and future plans. After reviewing all the records, **the company announced which conference the professor from the prestigious mathematics institute attended in the past.**	Were past successes discussed in the meeting?	Yes
11	long	implausible	The meeting room was filled with discussions about past successes and future plans. After reviewing all the records, **the company announced which coffee the professor from the prestigious mathematics institute attended in the past.**	Were past successes discussed in the meeting?	Yes
11	short	plausible	The meeting room was filled with discussions about past successes and future plans. **After reviewing all the records, the company announced which conference the professor attended in the past.**	Were past successes discussed in the meeting?	Yes
11	short	implausible	The meeting room was filled with discussions about past successes and future plans. **After reviewing all the records, the company announced which coffee the professor attended in the past.**	Were past successes discussed in the meeting?	Yes
12	long	plausible	The local government was working on new plans to improve the community. During a very important meeting, **the senator overlooked which initiative the governor from a small rural village proposed in the spring.**	Was the local government not working today?	No

12	long	implausible	The local government was working on new plans to improve the community. During a very important meeting, **the senator overlooked which tree the governor from a small rural village proposed in the spring.**	Was the local government not working today?	No
12	short	plausible	The local government was working on new plans to improve the community. **During a very important meeting, the senator overlooked which initiative the governor proposed in the spring.**	Was the local government not working today?	No
12	short	implausible	The local government was working on new plans to improve the community. **During a very important meeting, the senator overlooked which tree the governor proposed in the spring.**	Was the local government not working today?	No
13	long	plausible	In the book, the night was silent as the thief went through a rich neighborhood. In the most thrilling chapter, **the thief decided which earrings in the most luxurious house he would steal for his wife.**	Did the story take place at night?	Yes
13	long	implausible	In the book, the night was silent as the thief went through a rich neighborhood. In the most thrilling chapter, **the thief decided which basement in the most luxurious house he would steal for his wife.**	Did the story take place at night?	Yes
13	short	plausible	In the book, the night was silent as the thief went through a rich neighborhood. **In the most thrilling chapter, the thief decided which earrings he would steal for his wife.**	Did the story take place at night?	Yes
13	short	implausible	In the book, the night was silent as the thief went through a rich neighborhood. **In the most thrilling chapter, the thief decided which basement he would steal for his wife.**	Did the story take place at night?	Yes
14	long	plausible	Planning the wedding was very tricky, but they planned everything well. In the midst of this, **the bride knew which caterer the groom with the full thick beard hired for the wedding.**	Were they going to cancel the wedding?	No
14	long	implausible	Planning the wedding was very tricky, but they planned everything well. In the midst of this, **the bride knew which drink the groom with the full thick beard hired for the wedding.**	Were they going to cancel the wedding?	No
14	short	plausible	Planning the wedding was very tricky, but they planned everything well. **In the midst of this, the bride knew which caterer the groom hired for the wedding.**	Were they going to cancel the wedding?	No

14	short	implausible	Planning the wedding was very tricky, but they planned everything well. **In the midst of this, the bride knew which drink the groom hired for the wedding.**	Were they going to cancel the wedding?	No
15	long	plausible	The house smelled like fresh pastries when the family got together to celebrate. During the happy celebration time, **the girl noted which pastry her mother with bright read curly hair baked for her birthday.**	Was it a classmate reunion?	No
15	long	implausible	The house smelled like fresh pastries when the family got together to celebrate. During the happy celebration time, **the girl noted which song her mother with bright read curly hair baked for her birthday.**	Was it a classmate reunion?	No
15	short	plausible	The house smelled like fresh pastries when the family got together to celebrate. **During the happy celebration time, the girl noted which pastry her mother baked for her birthday.**	Was it a classmate reunion?	No
15	short	implausible	The house smelled like fresh pastries when the family got together to celebrate. **During the happy celebration time, the girl noted which song her mother baked for her birthday.**	Was it a classmate reunion?	No
16	long	plausible	The library was a quiet place for reading and had special times for storytelling. During a fun storytelling time, **the students knew which book the librarian with the soft gentle voice read to the class.**	Was the library noisy for group talks?	No
16	long	implausible	The library was a quiet place for reading and had special times for storytelling. During a fun storytelling time, **the students knew which ladder the librarian with the soft gentle voice read to the class.**	Was the library noisy for group talks?	No
16	short	plausible	The library was a quiet place for reading and had special times for storytelling. **During a fun storytelling time, the students knew which book the librarian read to the class.**	Was the library noisy for group talks?	No
16	short	implausible	The library was a quiet place for reading and had special times for storytelling. **During a fun storytelling time, the students knew which ladder the librarian read to the class.**	Was the library noisy for group talks?	No

17	long	plausible	The kitchen in the restaurant was very busy as lunchtime got closer. Before it got really busy, **the restaurant decided which vegetables the cook with the bright blue apron chopped for the salad.**	Was the kitchen getting busy for dinner?	No
17	long	implausible	The kitchen in the restaurant was very busy as lunchtime got closer. Before it got really busy, **the restaurant decided which spoons the cook with the bright blue apron chopped for the salad.**	Was the kitchen getting busy for dinner?	No
17	short	plausible	The kitchen in the restaurant was very busy as lunchtime got closer. **Before it got really busy, the restaurant decided which vegetables the cook chopped for the salad.**	Was the kitchen getting busy for dinner?	No
17	short	implausible	The kitchen in the restaurant was very busy as lunchtime got closer. **Before it got really busy, the restaurant decided which spoons the cook chopped for the salad.**	Was the kitchen getting busy for dinner?	No
18	long	plausible	The garage was messy with tools while the mechanic was fixing a car. Right after a sudden accident, **the mechanic examined which finger the driver with the short white beard cut during the accident.**	Was the mechanic fixing a car?	Yes
18	long	implausible	The garage was messy with tools while the mechanic was fixing a car. Right after a sudden accident, **the mechanic examined which hammer the driver with the short white beard cut during the accident.**	Was the mechanic fixing a car?	Yes
18	short	plausible	The garage was messy with tools while the mechanic was fixing a car. **Right after a sudden accident, the mechanic examined which finger the driver cut during the accident.**	Was the mechanic fixing a car?	Yes
18	short	implausible	The garage was messy with tools while the mechanic was fixing a car. **Right after a sudden accident, the mechanic examined which hammer the driver cut during the accident.**	Was the mechanic fixing a car?	Yes
19	long	plausible	The security team was extremely watchful because of recent break-ins in the neighborhood. During the course of investigation, **the guard verified which person the robber with the long messy hair intimidated during the burglary.**	Did the neighborhood have crimes recently?	Yes

19	long	implausible	The security team was extremely watchful because of recent break-ins in the neighborhood. During the course of investigation, **the guard verified which table the robber with the long messy hair intimidated during the burglary.**	Did the neighborhood have crimes recently?	Yes
19	short	plausible	The security team was extremely watchful because of recent break-ins in the neighborhood. **During the course of investigation, the guard verified which person the robber intimidated during the burglary.**	Did the neighborhood have crimes recently?	Yes
19	short	implausible	The security team was extremely watchful because of recent break-ins in the neighborhood. **During the course of investigation, the guard verified which table the robber intimidated during the burglary.**	Did the neighborhood have crimes recently?	Yes
20	long	plausible	The stadium was packed as the home team played an important match. At the game's halfway point, **the coach saw which leg the athlete from the prestigious football club bruised during the game.**	Was the match important to the home team?	Yes
20	long	implausible	The stadium was packed as the home team played an important match. At the game's halfway point, **the coach saw which shoe the athlete from the prestigious football club bruised during the game.**	Was the match important to the home team?	Yes
20	short	plausible	The stadium was packed as the home team played an important match. **At the game's halfway point, the coach saw which leg the athlete bruised during the game.**	Was the match important to the home team?	Yes
20	short	implausible	The stadium was packed as the home team played an important match. **At the game's halfway point, the coach saw which shoe the athlete bruised during the game.**	Was the match important to the home team?	Yes
21	long	plausible	The zoo was crowded with excited visitors going on tours led by the zookeeper. To the surprise of everyone, **the zookeeper saw which animal the lion with the bushy yellow tail attacked during the tour.**	Were there only a few visitors at the zoo?	No
21	long	implausible	The zoo was crowded with excited visitors going on tours led by the zookeeper. To the surprise of everyone, **the zookeeper saw which river the lion with the bushy yellow tail attacked during the tour.**	Were there only a few visitors at the zoo?	No

21	short	plausible	The zoo was crowded with excited visitors going on tours led by the zookeeper. **To the surprise of everyone, the zookeeper saw which animal the lion attacked during the tour.**	Were there only a few visitors at the zoo?	No
21	short	implausible	The zoo was crowded with excited visitors going on tours led by the zookeeper. **To the surprise of everyone, the zookeeper saw which river the lion attacked during the tour.**	Were there only a few visitors at the zoo?	No
22	long	plausible	The dessert station at the fancy restaurant was full of sweet treats. During the preparation for dinner, **the baker decided which chocolate in the luxury gourmet restaurant she would melt for the dessert.**	Was dinner service being prepared?	Yes
22	long	implausible	The dessert station at the fancy restaurant was full of sweet treats. During the preparation for dinner, **the baker decided which cookbook in the luxury gourmet restaurant she would melt for the dessert.**	Was dinner service being prepared?	Yes
22	short	plausible	The dessert station at the fancy restaurant was full of sweet treats. **During the preparation for dinner, the baker decided which chocolate she would melt for the dessert.**	Was dinner service being prepared?	Yes
22	short	implausible	The dessert station at the fancy restaurant was full of sweet treats. **During the preparation for dinner, the baker decided which cookbook she would melt for the dessert.**	Was dinner service being prepared?	Yes
23	long	plausible	The national park was alive with wildlife and visitors enjoying the natural beauty. On one of the trails, **the campers saw which deer the ranger with the grey working coverall fed in the park.**	Did visitors enjoy the national park?	Yes
23	long	implausible	The national park was alive with wildlife and visitors enjoying the natural beauty. On one of the trails, **the campers saw which tent the ranger with the grey working coverall fed in the park.**	Did visitors enjoy the national park?	Yes
23	short	plausible	The national park was alive with wildlife and visitors enjoying the natural beauty. **On one of the trails, the campers saw which deer the ranger fed in the park.**	Did visitors enjoy the national park?	Yes

23	short	implausible	The national park was alive with wildlife and visitors enjoying the natural beauty. **On one of the trails, the campers saw which tent the ranger fed in the park.**	Did visitors enjoy the national park?	Yes
24	long	plausible	The community garden was a special project for people who loved gardening. Throughout the entire planting season, **the gardener specified which tree the assistant with very rough dry hands planted in her garden.**	Was the project for people who disliked gardening?	No
24	long	implausible	The community garden was a special project for people who loved gardening. Throughout the entire planting season, **the gardener specified which cat the assistant with very rough dry hands planted in her garden.**	Was the project for people who disliked gardening?	No
24	short	plausible	The community garden was a special project for people who loved gardening. **Throughout the entire planting season, the gardener specified which tree the assistant planted in her garden.**	Was the project for people who disliked gardening?	No
24	short	implausible	The community garden was a special project for people who loved gardening. **Throughout the entire planting season, the gardener specified which cat the assistant planted in her garden.**	Was the project for people who disliked gardening?	No
25	long	plausible	On the ship, the crew was busy with their duties as they sailed across the open sea. When the crew was hungry, **the sailor decided which filet the cook with a mastery of spices broiled for the crew.**	Did the crew sail across a lake?	No
25	long	implausible	On the ship, the crew was busy with their duties as they sailed across the open sea. When the crew was hungry, **the sailor decided which plate the cook with a mastery of spices broiled for the crew.**	Did the crew sail across a lake?	No
25	short	plausible	On the ship, the crew was busy with their duties as they sailed across the open sea. **When the crew was hungry, the sailor decided which filet the cook broiled for the crew.**	Did the crew sail across a lake?	No
25	short	implausible	On the ship, the crew was busy with their duties as they sailed across the open sea. **When the crew was hungry, the sailor decided which plate the cook broiled for the crew.**	Did the crew sail across a lake?	No

26	long	plausible	The church was peaceful, with sunlight coming through the big glass on a sunny morning. Just before the ceremony began, **the priest saw which flowers the bride under her soft smooth veil smelled at the church.**	Was the church filled with sunlight?	Yes
26	long	implausible	The church was peaceful, with sunlight coming through the big glass on a sunny morning. Just before the ceremony began, **the priest saw which poem the bride under her soft smooth veil smelled at the church.**	Was the church filled with sunlight?	Yes
26	short	plausible	The church was peaceful, with sunlight coming through the big glass on a sunny morning. **Just before the ceremony began, the priest saw which flowers the bride smelled at the church.**	Was the church filled with sunlight?	Yes
26	short	implausible	The church was peaceful, with sunlight coming through the big glass on a sunny morning. **Just before the ceremony began, the priest saw which poem the bride smelled at the church.**	Was the church filled with sunlight?	Yes
27	long	plausible	The travel agency was known for its great tours and helpful staff. Before finalizing their travel plans, **the tourists asked which guide from the reputable local agency they should email for the trip.**	Were the tourists planning a trip?	Yes
27	long	implausible	The travel agency was known for its great tours and helpful staff. Before finalizing their travel plans, **the tourists asked which route from the reputable local agency they should email for the trip.**	Were the tourists planning a trip?	Yes
27	short	plausible	The travel agency was known for its great tours and helpful staff. **Before finalizing their travel plans, the tourists asked which guide they should email for the trip.**	Were the tourists planning a trip?	Yes
27	short	implausible	The travel agency was known for its great tours and helpful staff. **Before finalizing their travel plans, the tourists asked which route they should email for the trip.**	Were the tourists planning a trip?	Yes
28	long	plausible	The art school was buzy with students getting ready for the festival. In the crowded school lobby, **the assistant misunderstood which coffee from the specialty dessert cafe she should brew for the meeting.**	Were students preparing for a festival?	Yes

28	long	implausible	The art school was buzy with students getting ready for the festival. In the crowded school lobby, **the assistant misunderstood which donuts from the specialty dessert cafe she should brew for the meeting.**	Were students preparing for a festival?	Yes
28	short	plausible	The art school was buzy with students getting ready for the festival. **In the crowded school lobby, the assistant misunderstood which coffee she should brew for the meeting.**	Were students preparing for a festival?	Yes
28	short	implausible	The art school was buzy with students getting ready for the festival. **In the crowded school lobby, the assistant misunderstood which donuts she should brew for the meeting.**	Were students preparing for a festival?	Yes
29	long	plausible	The lounge was a favorite spot for artists to relax and share stories after their performances. On one quite late evening, **the bartender wondered which cigar the musician with the black stud earrings smoked in the bathroom.**	Was the lounge primarily used for formal performances?	No
29	long	implausible	The lounge was a favorite spot for artists to relax and share stories after their performances. On one quite late evening, **the bartender wondered which guitar the musician with the black stud earrings smoked in the bathroom.**	Was the lounge primarily used for formal performances?	No
29	short	plausible	The lounge was a favorite spot for artists to relax and share stories after their performances. **On one quite late evening, the bartender wondered which cigar the musician smoked in the bathroom.**	Was the lounge primarily used for formal performances?	No
29	short	implausible	The lounge was a favorite spot for artists to relax and share stories after their performances. **On one quite late evening, the bartender wondered which guitar the musician smoked in the bathroom.**	Was the lounge primarily used for formal performances?	No
30	long	plausible	Before starting the cafeteria rounds, the staff members went over the health and safety standards. During the busy lunchtime session, **the supervisor checked which beverages the students from the linguistics research lab drank in the cafeteria.**	Did the staff skip reviewing the health and safty stardards?	No

30	long	implausible	Before starting the cafeteria rounds, the staff members went over the health and safety standards. During the busy lunchtime session, **the supervisor checked which cashier the students from the linguistics research lab drank in the cafeteria.**	Did the staff skip reviewing the health and safty stardards?	No
30	short	plausible	Before starting the cafeteria rounds, the staff members went over the health and safety standards. **During the busy lunchtime session, the supervisor checked which beverages the students drank in the cafeteria.**	Did the staff skip reviewing the health and safty stardards?	No
30	short	implausible	Before starting the cafeteria rounds, the staff members went over the health and safety standards. **During the busy lunchtime session, the supervisor checked which cashier the students drank in the cafeteria.**	Did the staff skip reviewing the health and safty stardards?	No
31	long	plausible	In the university library, students were focused on their studies, flipping through books quietly. During the routine inspection rounds, **the librarian checked which article the student with the spiky blond hair scattered in the library.**	Did students in the university library study quietly?	Yes
31	long	implausible	In the university library, students were focused on their studies, flipping through books quietly. During the routine inspection rounds, **the librarian checked which room the student with the spiky blond hair scattered in the library.**	Did students in the university library study quietly?	Yes
31	short	plausible	In the university library, students were focused on their studies, flipping through books quietly. **During the routine inspection rounds, the librarian checked which article the student scattered in the library.**	Did students in the university library study quietly?	Yes
31	short	implausible	In the university library, students were focused on their studies, flipping through books quietly. **During the routine inspection rounds, the librarian checked which room the student scattered in the library.**	Did students in the university library study quietly?	Yes
32	long	plausible	The housekeeper worked on arranging next week's tasks for all the employees in the mansion. When crafting the weekly agenda, **the housekeeper forgot which carpet the maid with very skillful dexterous hands cleaned over the weekend.**	Was the housekeeper focused on summarizing the previous week?	No

32	long	implausible	The housekeeper worked on arranging next week's tasks for all the employees in the mansion. When crafting the weekly agenda, **the housekeeper forgot which contract the maid with very skillful dexterous hands cleaned over the weekend.**	Was the housekeeper focused on summarizing the previous week?	No
32	short	plausible	The housekeeper worked on arranging next week's tasks for all the employees in the mansion. **When crafting the weekly agenda, the housekeeper forgot which carpet the maid cleaned over the weekend.**	Was the housekeeper focused on summarizing the previous week?	No
32	short	implausible	The housekeeper worked on arranging next week's tasks for all the employees in the mansion. **When crafting the weekly agenda, the housekeeper forgot which contract the maid cleaned over the weekend.**	Was the housekeeper focused on summarizing the previous week?	No

Appendix D

Experiment 1 Linear Regression Results in the Long Dependency Condition

Region	Model	Predictor	Estimate	*Std. Error*	*t*	*P*	*R²*	Overall p-value
Word 12 (Pre-critical)	1	Intercept	401.74	39.47	10.179	<0.001***	0.009	0.186
		Filler Plausibility Plausible	82.96	56.14	1.478	0.140		
		Proficiency	-0.06	0.59	-0.104	0.917		
		Plausible × Proficiency	-1.25	0.84	-1.490	0.137		
	2	Intercept	474.03	72.38	6.549	<0.001***	0.081	<0.001***
		Filler Plausibility Plausible	82.68	103.13	0.802	0.423		
		Proficiency	-0.61	1.04	-0.586	0.558		
		YOI	-7.58	6.85	-1.107	0.269		
		Plausible × Proficiency	-0.962	1.486	-0.647	0.518		
		Plausible × YOI	0.859	9.699	0.089	0.929		
		Proficiency × YOI	0.031	0.099	0.309	0.757		
		Plausible × Proficiency × YOI	-0.058	0.140	-0.414	0.679		
Word 13 (Critical)	1	Intercept	472.94	53.33	8.869	<0.001***	0.011	0.106
		Filler Plausibility Plausible	96.79	75.55	1.281	0.201		
		Proficiency	-0.52	0.79	-0.656	0.512		
		Plausible × Proficiency	-1.23	1.13	-1.095	0.274		
	2	Intercept	616.65	99.24	6.214	<0.001***	0.069	<0.001***
		Filler Plausibility Plausible	113.90	141.21	0.807	0.420		
		Proficiency	-1.83	1.43	-1.281	0.201		
		YOI	-15.23	9.41	-1.620	0.106		
		Plausible × Proficiency	-1.448	2.033	-0.712	0.477		
		Plausible × YOI	-1.234	13.392	-0.092	0.927		
		Proficiency × YOI	0.098	0.135	0.725	0.469		
		Plausible × Proficiency × YOI	0.013	0.193	0.068	0.946		

Word 14 (Post-critical 1)	1	Intercept	365.89	35.85	10.206	<0.001***	0.007	0.290
		Filler Plausibility Plausible	92.70	50.62	1.831	0.068		
		Proficiency	0.48	0.53	0.907	0.365		
		Plausible × Proficiency	-1.39	0.75	-1.847	0.065		
	2	Intercept	445.41	67.43	6.605	<0.001***	0.045	<0.001***
		Filler Plausibility Plausible	25.78	95.30	0.271	0.787		
		Proficiency	-0.31	0.97	-0.322	0.748		
		YOI	-8.61	6.41	-1.344	0.180		
		Plausible × Proficiency	-0.273	1.374	-0.198	0.843		
		Plausible × YOI	7.727	9.046	0.854	0.393		
		Proficiency × YOI	0.069	0.092	0.747	0.455		
		Plausible × Proficiency × YOI	-0.136	0.130	-1.044	0.297		
Word 15 (Post-critical 2)	1	Intercept	331.95	28.16	11.786	<0.001***	0.006	0.347
		Filler Plausibility Plausible	37.57	39.93	0.941	0.347		
		Proficiency	0.65	0.42	1.544	0.123		
		Plausible × Proficiency	-0.71	0.60	-1.188	0.235		
	2	Intercept	352.70	52.50	6.718	<0.001***	0.058	<0.001***
		Filler Plausibility Plausible	11.85	74.64	0.159	0.874		
		Proficiency	0.79	0.76	1.045	0.296		
		YOI	-1.87	4.99	-0.376	0.707		
		Plausible × Proficiency	-0.426	1.076	-0.396	0.692		
		Plausible × YOI	2.815	7.070	0.398	0.691		
		Proficiency × YOI	-0.044	0.072	-0.617	0.538		
		Plausible × Proficiency × YOI	-0.026	0.102	-0.255	0.799		

Note: . $p<0.1$; * $p<0.05$; ** $p<0.01$; *** $p<0.001$
R^2 values are Multiple R-squared for each model.
Model 2 includes YOI (Years of Immersion) as an additional predictor.

Appendix E

Experiment 2 Mandarin Chinese Experimental Stimuli

<table>
<tr><th>set</th><th>dependency length</th><th>plausibility</th><th>Regions (1-13)</th><th>question</th><th>answer</th></tr>
<tr><td>1</td><td>short</td><td>plausible</td><td>在|晨会|结束|后,|小张|想知道|哪个客户|领导|通知了|实习生|去|找|。|</td><td>实习生会去找一个客户吗？</td><td>是</td></tr>
<tr><td>1</td><td>short</td><td>implausible</td><td>在|晨会|结束|后,|小张|想知道|哪个文件|领导|通知了|实习生|去|找|。|</td><td>实习生会去找一个客户吗？</td><td>否</td></tr>
<tr><td>1</td><td>long</td><td>plausible</td><td>小张|想知道|哪个客户|领导|在|晨会|结束|后|通知了|实习生|去|找|。|</td><td>实习生会去找一个客户吗？</td><td>是</td></tr>
<tr><td>1</td><td>long</td><td>implausible</td><td>小张|想知道|哪个文件|领导|在|晨会|结束|后|通知了|实习生|去|找|。|</td><td>实习生会去找一个客户吗？</td><td>否</td></tr>
<tr><td>2</td><td>short</td><td>plausible</td><td>在|研讨会|结束|后，|护士|想知道|哪位患者|医院|通知了|医生|去|检查|。|</td><td>医生会去检查一个处方吗？</td><td>否</td></tr>
<tr><td>2</td><td>short</td><td>implausible</td><td>在|研讨会|结束|后，|护士|想知道|哪个处方|医院|通知了|医生|去|检查|。|</td><td>医生会去检查一个处方吗？</td><td>是</td></tr>
<tr><td>2</td><td>long</td><td>plausible</td><td>护士|想知道|哪位患者|医院|在|研讨会|结束|后|通知了|医生|去|检查|。|</td><td>医生会去检查一个处方吗？</td><td>否</td></tr>
<tr><td>2</td><td>long</td><td>implausible</td><td>护士|想知道|哪个处方|医院|在|研讨会|结束|后|通知了|医生|去|检查|。|</td><td>医生会去检查一个处方吗？</td><td>是</td></tr>
<tr><td>3</td><td>short</td><td>plausible</td><td>在|病房|巡查|后，|患者|不清楚|哪个护士|医生|联络了|助理|去|安排|。|</td><td>助理会去给患者安排一个护士吗？</td><td>是</td></tr>
<tr><td>3</td><td>short</td><td>implausible</td><td>在|病房|巡查|后，|患者|不清楚|哪个药物|医生|联络了|助理|去|安排|。|</td><td>助理会去给患者安排一个护士吗？</td><td>否</td></tr>
<tr><td>3</td><td>long</td><td>plausible</td><td>患者|不清楚|哪个护士|医生|在|病房|巡查|后|联络了|助理|去|安排|。|</td><td>助理会去给患者安排一个护士吗？</td><td>是</td></tr>
<tr><td>3</td><td>long</td><td>implausible</td><td>患者|不清楚|哪个药物|医生|在|病房|巡查|后|联络了|助理|去|安排|。|</td><td>助理会去给患者安排一个护士吗？</td><td>否</td></tr>
<tr><td>4</td><td>short</td><td>plausible</td><td>在|年终|大会|前，|小王|不确定|哪个设计师|经理|委托了|评审团|去|选择|。|</td><td>评审团会去选择一个项目吗？</td><td>否</td></tr>
<tr><td>4</td><td>short</td><td>implausible</td><td>在|年终|大会|前，|小王|不确定|哪个项目|经理|委托了|评审团|去|选择|。|</td><td>评审团会去选择一个项目吗？</td><td>是</td></tr>
<tr><td>4</td><td>long</td><td>plausible</td><td>员工|不确定|哪个设计师|经理|在|年终|大会|上|委托了|评审团|去|选择|。|</td><td>评审团会去选择一个项目吗？</td><td>否</td></tr>
<tr><td>4</td><td>long</td><td>implausible</td><td>员工|不确定|哪个项目|经理|在|年终|大会|上|委托了|评审团|去|选择|。|</td><td>评审团会去选择一个项目吗？</td><td>是</td></tr>
</table>

5	short	plausible	在|飞机|起飞|前，|乘客|想了解|哪位机长|航空公司|指派了|调度员|去|联络|。|	调度员会去联络一位机长吗？	是
5	short	implausible	在|飞机|起飞|前，|乘客|想了解|哪个机场|航空公司|指派了|调度员|去|联络|。|	调度员会去联络一位机长吗？	否
5	long	plausible	乘客|想了解|哪位机长|航空公司|在|飞机|起飞|前|指派了|调度员|去|联络|。|	调度员会去联络一位机长吗？	是
5	long	implausible	乘客|想了解|哪个机场|航空公司|在|飞机|起飞|前|指派了|调度员|去|联络|。|	调度员会去联络一位机长吗？	否
6	short	plausible	在|紧急|会议|后，|侦探|想查明|哪个人|警局|派遣了|便衣警察|去|调查|。|	便衣警察会去调查一条线索吗？	否
6	short	implausible	在|紧急|会议|后，|侦探|想查明|哪个线索|警局|派遣了|便衣警察|去|调查|。|	便衣警察会去调查一条线索吗？	是
6	long	plausible	侦探|想查明|哪个人|警局|在|紧急|会议|后|派遣了|便衣警察|去|调查|。|	便衣警察会去调查一条线索吗？	否
6	long	implausible	侦探|想查明|哪个线索|警局|在|紧急|会议|后|派遣了|便衣警察|去|调查|。|	便衣警察会去调查一条线索吗？	是
7	short	plausible	在|展览|结束|后，|人们|在探讨|哪位艺术家|评审团|邀请了|专家|去|点评|。|	专家会去点评一位艺术家吗？	是
7	short	implausible	在|展览|结束|后，|人们|在探讨|哪幅画|评审团|邀请了|专家|去|点评|。|	专家会去点评一位艺术家吗？	否
7	long	plausible	人们|在探讨|哪位艺术家|评审团|在|展览|结束|后|邀请了|专家|去|点评|。|	专家会去点评一位艺术家吗？	是
7	long	implausible	人们|在探讨|哪幅画|评审团|在|展览|结束|后|邀请了|专家|去|点评|。|	专家会去点评一位艺术家吗？	否
8	short	plausible	在|期末|考试|前，|教授|要确定|哪个学生|教务处|联系了|辅导员|去|找|。|	辅导员会去找一份材料吗？	否
8	short	implausible	在|期末|考试|前，|教授|要确定|哪份材料|教务处|联系了|辅导员|去|找|。|	辅导员会去找一份材料吗？	是
8	long	plausible	教授|要确定|哪个学生|教务处|在|期末|考试|前|联系了|辅导员|去|找|。|	辅导员会去找一份材料吗？	否
8	long	implausible	教授|要确定|哪份材料|教务处|在|期末|考试|前|联系了|辅导员|去|找|。|	辅导员会去找一份材料吗？	是
9	short	plausible	在|新学期|开始|时，|学生|不知道|哪个讲座|学校|要举办|活动|去|宣传|。|	学校要举办活动去宣传一个讲座吗？	是

9	short	implausible	在|新学期|开始|时，|学生|不知道|哪位教授|学校|要举办|活动|去|宣传|。|	学校要举办活动去宣传一个讲座吗？	否
9	long	plausible	学生|不知道|哪个讲座|学校|在|新学期|开始|时|要举办|活动|去|宣传|。|	学校要举办活动去宣传一个讲座吗？	是
9	long	implausible	学生|不知道|哪位教授|学校|在|新学期|开始|时|要举办|活动|去|宣传|。|	学校要举办活动去宣传一个讲座吗？	否
10	short	plausible	在|假期|结束|后，|经理|在询问|哪位员工|公司|指派了|小李|去|调查|。|	小李是要去调查一项支出吗？	否
10	short	implausible	在|假期|结束|后，|经理|在询问|哪项支出|公司|指派了|小李|去|调查|。|	小李是要去调查一项支出吗？	是
10	long	plausible	经理|在询问|哪位员工|公司|在|假期|结束|后|指派了|小李|去|调查|。|	小李是要去调查一项支出吗？	否
10	long	implausible	经理|在询问|哪项支出|公司|在|假期|结束|后|指派了|小李|去|调查|。|	小李是要去调查一项支出吗？	是
11	short	plausible	在|观影|结束|后，|观众|在猜测|哪位演员|组委会|要邀请|大家|来|评判|。|	大家被邀请去评判一位演员吗？	是
11	short	implausible	在|观影|结束|后，|观众|在猜测|哪部电影|组委会|要邀请|大家|来|评判|。|	大家被邀请去评判一位演员吗？	否
11	long	plausible	观众|在猜测|哪位演员|组委会|在|观影|结束|后|要邀请|大家|来|评判|。|	大家被邀请去评判一位演员吗？	是
11	long	implausible	观众|在猜测|哪部电影|组委会|在|观影|结束|后|要邀请|大家|来|评判|。|	大家被邀请去评判一位演员吗？	否
12	short	plausible	在|发布会|开始|前，|记者|不了解|哪个新产品|厂商|要制造|舆论|去|宣传|。|	厂商要去宣传一个新服务吗？	否
12	short	implausible	在|发布会|开始|前，|记者|不了解|哪个新服务|厂商|要制造|舆论|去|宣传|。|	厂商要去宣传一个新服务吗？	是
12	long	plausible	记者|不了解|哪个新产品|厂商|在|发布会|开始|前|要制造|舆论|去|宣传|。|	厂商要去宣传一个新服务吗？	否
12	long	implausible	记者|不了解|哪个新服务|厂商|在|发布会|开始|前|要制造|舆论|去|宣传|。|	厂商要去宣传一个新服务吗？	是
13	short	plausible	在|专访|节目|前，|观众|不确定|哪篇文章|作者|要发表|评论|去|解析|。|	作者要去解析一篇文章吗？	是

13	short	implausible	在|专访|节目|前，|观众|不确定|哪个人物|作者|要发表|评论|去|解析|。|	作者要去解析一篇文章吗？	否
13	long	plausible	观众|不确定|哪篇文章|作者|在|专访|节目|中|要发表|评论|去|解析|。|	作者要去解析一篇文章吗？	是
13	long	implausible	观众|不确定|哪个人物|作者|在|专访|节目|中|要发表|评论|去|解析|。|	作者要去解析一篇文章吗？	否
14	short	plausible	在|案件|发生|后，|侦探|在推断|哪个嫌疑人|警方|要扣留|目击者|去|指认|。|	目击者会去指认一个现场吗？	否
14	short	implausible	在|案件|发生|后，|侦探|在推断|哪个现场|警方|要扣留|目击者|去|指认|。|	目击者会去指认一个现场吗？	是
14	long	plausible	侦探|在推断|哪个嫌疑人|警方|在|案件|发生|后|要扣留|目击者|去|指认|。|	目击者会去指认一个现场吗？	否
14	long	implausible	侦探|在推断|哪个现场|警方|在|案件|发生|后|要扣留|目击者|去|指认|。|	目击者会去指认一个现场吗？	是
15	short	plausible	在|期末|大会|上，|同学们|很好奇|哪个老师|校长|要安排|助理|去|介绍|。|	助理会去介绍一个老师吗？	是
15	short	implausible	在|期末|大会|上，|同学们|很好奇|哪个观点|校长|要安排|助理|去|介绍|。|	助理会去介绍一个老师吗？	否
15	long	plausible	同学们|很好奇|哪个老师|校长|在|期末|大会|上|要安排|助理|去|介绍|。|	助理会去介绍一个老师吗？	是
15	long	implausible	同学们|很好奇|哪个观点|校长|在|期末|大会|上|要安排|助理|去|介绍|。|	助理会去介绍一个老师吗？	否
16	short	plausible	在|全国| 各个|地区，|大家|在关注|哪个政策|政府|实施了|新计划|去|宣传|。|	政府会去宣传一个劳模吗？	否
16	short	implausible	在|全国| 各个|地区，|大家|在关注|哪个劳模|政府|实施了|新计划|去|宣传|。|	政府会去宣传一个劳模吗？	是
16	long	plausible	大家|在关注|哪个政策|政府|在|全国| 各个|地区|实施了|新计划|去|宣传|。|	政府会去宣传一个劳模吗？	否
16	long	implausible	大家|在关注|哪个劳模|政府|在|全国| 各个|地区|实施了|新计划|去|宣传|。|	政府会去宣传一个劳模吗？	是
17	short	plausible	在|网上|课程|平台，|小赵|在学习|哪个方案|教授|设计了|新范式|去|研究|。|	教授会去研究一个方案吗？	是
17	short	implausible	在|网上|课程|平台，|小赵|在学习|哪个群体|教授|设计了|新范式|去|研究|。|	教授会去研究一个方案吗？	否
17	long	plausible	小赵|在学习|哪个方案|教授|在|网上|课程|平台|设计了|新范式|去|研究|。|	教授会去研究一个方案吗？	是
17	long	implausible	小赵|在学习|哪个群体|教授|在|网上|课程|平台|设计了|新范式|去|研究|。|	教授会去研究一个方案吗？	否
18	short	plausible	在|董事|会议|之后，|小周|在猜测|哪个产品|公司|要研发|新程序|去|支持|。|	公司会去支持一个新团队吗？	否
18	short	implausible	在|董事|会议|之后，|小周|在猜测|哪个团队|公司|要研发|新程序|去|支持|。|	公司会去支持一个新团队吗？	是

18	long	plausible	小周|在猜测|哪个产品|公司|在|董事|会议|之后|要研发|新程序|去|支持|。|	公司会去支持一个新团队吗？	否
18	long	implausible	小周|在猜测|哪个团队|公司|在|董事|会议|之后|要研发|新程序|去|支持|。|	公司会去支持一个新团队吗？	是
19	short	plausible	在|课程|开始|前，|同学们|很好奇|哪本书|助教|要读完|论文|去|买|。|	助教读完论文会去买一本书吗？	是
19	short	implausible	在|课程|开始|前，|同学们|很好奇|哪个设备|助教|要读完|论文|去|买|。|	助教读完论文会去买一本书吗？	否
19	long	plausible	同学们|很好奇|哪本书|助教|在|课程|开始|前|要读完|论文|去|买|。|	助教读完论文会去买一本书吗？	是
19	long	implausible	同学们|很好奇|哪个设备|助教|在|课程|开始|前|要读完|论文|去|买|。|	助教读完论文会去买一本书吗？	否
20	short	plausible	在|学术|会议|前，|教授|在考虑|哪篇论文|研究生|要写|报告|去|修改|。|	研究生要去修改一个实验吗？	否
20	short	implausible	在|学术|会议|前，|教授|在考虑|哪个实验|研究生|要写|报告|去|修改|。|	研究生要去修改一个实验吗？	是
20	long	plausible	教授|在考虑|哪篇论文|研究生|在|学术|会议|前|要写|报告|去|修改|。|	研究生要去修改一个实验吗？	否
20	long	implausible	教授|在考虑|哪个实验|研究生|在|学术|会议|前|要写|报告|去|修改|。|	研究生要去修改一个实验吗？	是
21	short	plausible	在|施工|开始|前，|工程师|在评估|哪种材料|技术员|要采购|新设备|去|测试|。|	技术员要去测试一个新材料吗？	是
21	short	implausible	在|施工|开始|前，|工程师|在评估|哪个场地|技术员|要采购|新设备|去|测试|。|	技术员要去测试一个新材料吗？	否
21	long	plausible	工程师|在评估|哪种材料|技术员|在|施工|开始|前|要采购|新设备|去|测试|。|	技术员要去测试一个新材料吗？	是
21	long	implausible	工程师|在评估|哪个场地|技术员|在|施工|开始|前|要采购|新设备|去|测试|。|	技术员要去测试一个新材料吗？	否
22	short	plausible	在|运动会|到来|前，|教练|很关心|哪个器械|运动员|要购买|新课程|去|练习|。|	运动员要去练习一个动作吗？	否
22	short	implausible	在|运动会|到来|前，|教练|很关心|哪个动作|运动员|要购买|新课程|去|练习|。|	运动员要去练习一个动作吗？	是

22	long	plausible	教练|很关心|哪个器械|运动员|在|运动会|到来|前|要购买|新课程|去|练习|。|	运动员要去练习一个动作吗？	否
22	long	implausible	教练|很关心|哪个动作|运动员|在|运动会|到来|前|要购买|新课程|去|练习|。|	运动员要去练习一个动作吗？	是
23	short	plausible	在|媒体|见面会|前，|大家|很好奇|哪个演员|经纪公司|雇了|新团队|去|宣传|。|	经纪公司会去宣传一个演员吗？	是
23	short	implausible	在|媒体|见面会|前，|大家|很好奇|哪首歌曲|经纪公司|雇了|新团队|去|宣传|。|	经纪公司会去宣传一个演员吗？	否
23	long	plausible	大家|很好奇|哪个演员|经纪公司|在|媒体|见面会|前|雇了|新团队|去|宣传|。|	经纪公司会去宣传一个演员吗？	是
23	long	implausible	大家|很好奇|哪首歌曲|经纪公司|在|媒体|见面会|前|雇了|新团队|去|宣传|。|	经纪公司会去宣传一个演员吗？	否
24	short	plausible	在|分享会|结束|后，|妈妈|想知道|哪首歌|孩子|听了|建议|去|下载|。|	孩子会去下载一个程序吗？	否
24	short	implausible	在|分享会|结束|后，|妈妈|想知道|哪个程序|孩子|听了|建议|去|下载|。|	孩子会去下载一个程序吗？	是
24	long	plausible	妈妈|想知道|哪首歌|孩子|在|分享会|结束|后|听了|建议|去|下载|。|	孩子会去下载一个程序吗？	否
24	long	implausible	妈妈|想知道|哪个程序|孩子|在|分享会|结束|后|听了|建议|去|下载|。|	孩子会去下载一个程序吗？	是

Appendix F
Experiment 3 ERP Experimental Stimuli

set	item	filler plausibility	gap licensor	sentence	question	answer
1	a	plausible	vt	Jane wondered which customer the manager **informed** the intern to find ___ after the meeting.	Should the intern find a customer?	Y
1	b	implausible	vt	Jane wondered which document the manager **informed** the intern to find ___ after the meeting.	Should the intern find a document?	Y
1	c	plausible	prep	Jane wondered which customer the manager **informed** the intern about ___ after the meeting.	Did the manger inform the intern?	Y
1	d	implausible	prep	Jane wondered which document the manager **informed** the intern about ___ after the meeting.	Did the manger inform the intern?	Y
2	a	plausible	vt	Amy wondered which patient the doctor **warned** the nurse to check ___ before the surgery.	Did the doctor warn the patient?	N
2	b	implausible	vt	Amy wondered which medicine the doctor **warned** the nurse to check ___ before the surgery.	Did the doctor warn the patient?	N
2	c	plausible	prep	Amy wondered which patient the doctor **warned** the nurse about ___ before the surgery.	Did the doctor warn the patient?	N
2	d	implausible	prep	Amy wondered which medicine the doctor **warned** the nurse about ___ before the surgery.	Did the doctor warn the patient?	N
3	a	plausible	vt	Bill asked which assistant the doctor **instructed** the nurse to follow ___ at the hospital.	Were they at the hospital?	Y
3	b	implausible	vt	Bill asked which protocol the doctor **instructed** the nurse to follow ___ at the hospital.	Were they at the hospital?	Y
3	c	plausible	prep	Bill asked which assistant the doctor **instructed** the nurse about ___ at the hospital.	Were they at the hospital?	Y
3	d	implausible	prep	Bill asked which protocol the doctor **instructed** the nurse on ___ at the hospital.	Were they at the hospital?	Y
4	a	plausible	vt	Carly guessed which artist the gallery **invited** the committee to evaluate ___ after the show.	Did the committee evaluate a singer?	N
4	b	implausible	vt	Carly guessed which painting the gallery **invited** the committee to evaluate ___ after the show.	Did the committee evaluate a singer?	N
4	c	plausible	prep	Carly guessed which artist the gallery **invited** the committee for___ after the show.	Did the committee invite a singer?	N

4	d	implausible	prep	Carly guessed which painting the gallery **invited** the committee for ___ after the show.	Did the committee invite a singer?	N
5	a	plausible	vt	David confirmed which teacher the principal **asked** the candidate to meet ___ in the office.	Should they meet in a café?	N
5	b	implausible	vt	David confirmed which criteria the principal **asked** the candidate to meet ___ after the interview.	Did David confirm criteria beforehand?	N
5	c	plausible	prep	David confirmed which teacher the principal **asked** the candidate about ___ in the office.	Did David confirm it in a café?	N
5	d	implausible	prep	David confirmed which criteria the principal **asked** the candidate about ___ in the office.	Did David confirm it in a café?	N
6	a	plausible	vt	Emma knew which detective the police chief **dispatched** the officers to follow ___ after the accident.	Were the officers dispatched?	Y
6	b	implausible	vt	Emma knew which case the police chief **dispatched** the officers to follow ___ after the accident.	Were the officers dispatched?	Y
6	c	plausible	prep	Emma knew which detective the police chief **dispatched** the officers for ___ after the accident.	Were the officers dispatched?	Y
6	d	implausible	prep	Emma knew which case the police chief **dispatched** the officers for ___ after the accident.	Were the officers dispatched?	Y
7	a	plausible	vt	Frank revealed which actress the director **advised** the reporter to describe ___ during the celebration.	Was the celebration about a song?	N
7	b	implausible	vt	Frank revealed which movie the director **advised** the reporter to describe ___ during the celebration.	Was the celebration about a song?	N
7	c	plausible	prep	Frank revealed which actress the director **advised** the reporter about ___ during the celebration.	Was the reporter advised about a song?	N
7	d	implausible	prep	Frank revealed which movie the director advised the reporter about___ during the celebration.	Was the reporter advised about a song?	N
8	a	plausible	vt	George knew which student the dean **called** the professor to find___ after the exam.	Did the dean call the professor?	Y
8	b	implausible	vt	George knew which file the dean **called** the professor to find ___ after the exam.	Did the dean call the professor?	Y
8	c	plausible	prep	George knew which student the dean **called** the professor about ___ after the exam.	Did the dean call the professor?	Y

8	d	implausible	prep	George knew which file the dean **called** the professor about ___ after the exam.	Did the dean call the professor?	Y
9	a	plausible	vt	Helen asked which seminar the faculty **held** an event to introduce ___ during the summer.	Was the event in the winter?	N
9	b	implausible	vt	Helen asked which researcher the faculty **held** an event to introduce ___ during the summer.	Was the event in the winter?	N
9	c	plausible	prep	Helen asked which seminar the faculty **held** an event about___ during the summer.	Was the event in the winter?	N
9	d	implausible	prep	Helen asked which researcher the faculty **held** an event for ___ during the summer.	Was the event in the winter?	N
10	a	plausible	vt	Ivan guessed which employee the manager **told** the secretary to investigate ___ after the meeting.	Did Ivan tell the secretary?	N
10	b	implausible	vt	Ivan guessed which payment the manager **told** the secretary to investigate ___ after the meeting.	Did Ivan tell the secretary?	N
10	c	plausible	prep	Ivan guessed which employee the manager **told** the secretary about ___ after the meeting.	Did Ivan tell the secretary?	N
10	d	implausible	prep	Ivan guessed which payment the manager **told** the secretary about ___ after the meeting.	Did Ivan tell the secretary?	N
11	a	plausible	vt	Jay revealed which scientist the professor **persuaded** the students to investigate ___ before the class.	Did Jay reveal something?	Y
11	b	implausible	vt	Jay revealed which theory the professor **persuaded** the students to investigate ___ before the class.	Did Jay reveal something?	Y
11	c	plausible	prep	Jay revealed which candidate the professor **persuaded** the students on ___ before the class.	Was it Jay who revealed something?	Y
11	d	implausible	prep	Jay revealed which theory the professor **persuaded** the students about ___ before the class.	Did Jay reveal something?	Y
12	a	plausible	vt	Kristin revealed which product the company **made** an advertisement to promote ___ in the summer.	Did the ad for a product?	Y
12	b	implausible	vt	Kristin revealed which actor the company **made** an advertisement to promote ___ in the summer.	Was the ad for an actor?	Y
12	c	plausible	prep	Kristin revealed which product the company **made** an	Was the ad for a product?	Y

				advertisement for ___ in the summer.		
12	d	implausible	prep	Kristin revealed which actor the company **made** an advertisement for ___ in the summer.	Was the ad for an actor?	Y
13	a	plausible	vt	Lucy identifies which article the author **wrote** a column to describe ___ in the newspaper.	Did the author write a column?	Y
13	b	implausible	vt	Lucy identifies which woman the author **wrote** a column to describe ___ in the newspaper.	Did the author write a column?	Y
13	c	plausible	prep	Lucy identifies which article the author **wrote** a column about ___ in the newspaper.	Did the author write a column?	Y
13	d	implausible	prep	Lucy identifies which woman the author **wrote** a column about ___ in the newspaper.	Did the author write a column?	Y
14	a	plausible	vt	Mike questioned which suspect the police **detained** the informant to identify ___ during the interrogation.	Was the informant released?	N
14	b	implausible	vt	Mike questioned which evidence the police **detained** the informant to identify ___ during the interrogation.	Was the informant released?	N
14	c	plausible	prep	Mike questioned which suspect the police **detained** the informant with ___ during the interrogation.	Was the informant released?	N
14	d	implausible	prep	Mike questioned which evidence the police **detained** the informant for ___ during the interrogation.	Was the informant released?	N
15	a	plausible	vt	Nancy knew which professor the organizer **asked** the expert to introduce ___ before the conference.	Should the expert introduce a professor?	Y
15	b	implausible	vt	Nancy knew which paper the organizer **asked** the expert to introduce ___ before the conference.	Should the expert introduce a paper?	Y
15	c	plausible	prep	Nancy knew which professor the organizer **asked** the expert about ___ before the conference.	Did the organizer ask something?	Y
15	d	implausible	prep	Nancy knew which paper the organizer **asked** the expert about ___ before the conference.	Was the expert asked about a paper?	Y
16	a	plausible	vt	Owen confirmed which athlete the coach **reminded** the runners to imitate ___ during the training.	Did Owen remind the runners?	N
16	b	implausible	vt	Owen confirmed which action the coach **reminded** the runners to learn ___ during the training.	Did Owen remind them?	N

16	c	plausible	prep	Owen confirmed which athlete the coach **reminded** the runners of ___ during the training.	Did Owen remind them?	N
16	d	implausible	prep	Owen confirmed which action the coach **reminded** the runners of ___ during the training.	Did Owen remind the runners?	N
17	a	plausible	vt	Paul wondered which doctor the woman **met** the neighbor to visit ___ in the afternoon.	Did the woman meet the neighbor?	Y
17	b	implausible	vt	Paul wondered which store the woman **met** the neighbor to visit ___ in the afternoon.	Did the woman meet the neighbor?	Y
17	c	plausible	prep	Paul wondered which doctor the woman **met** the neighbor with ___ in the afternoon.	Did the woman meet the neighbor?	Y
17	d	implausible	prep	Paul wondered which store the woman **met** the neighbor at ___ in the afternoon.	Did the woman meet the neighbor?	Y
18	a	plausible	vt	Ryne wondered which student the teacher **texted** the message to find ___ after the exam.	Did the teacher make a call?	N
18	b	implausible	vt	Ryne wondered which book the teacher **texted** the message to find ___ after the exam.	Did the teacher make a call?	N
18	c	plausible	prep	Ryne wondered which student the teacher **texted** the message about ___ after the exam.	Did the teacher make a call?	N
18	d	implausible	prep	Ryne wondered which book the teacher **texted** the message about ___ after the exam.	Did the teacher make a call?	N
19	a	plausible	vt	Sarah asked which assistant the girl **consulted** the librarian to find ___ in the library.	Did the girl consult the librarian?	Y
19	b	implausible	vt	Sarah asked which bookshelf the girl **consulted** the librarian to find ___ in the library.	Did the girl consult the librarian?	Y
19	c	plausible	prep	Sarah asked which assistant the girl **consulted** the librarian with ___ in the library.	Did the girl consult the librarian?	Y
19	d	implausible	prep	Sarah asked which bookshelf the girl **consulted** the librarian about ___ in the library.	Did the girl consult the librarian?	Y
20	a	plausible	vt	Dr. Sabourin confirmed which paper the student **wrote** the report to improve ___ for the degree.	Was the report for a funding?	N
20	b	implausible	vt	Dr. Sabourin confirmed which experiment the student **wrote** the report to improve ___ for the degree.	Was the report for a funding?	N
20	c	plausible	prep	Dr. Sabourin confirmed which paper the student **wrote** the report for ___ in the lab.	Was the report for a funding?	N

20	d	implausible	prep	Dr. Sabourin confirmed which experiment the student **wrote** the report for ___ in the lab.	Was the report for a funding?	N
21	a	plausible	vt	Tony asked which material the technician **purchased** new equipment to test ___ before the construction.	Was the new equipment purchased?	Y
21	b	implausible	vt	Tony asked which location the technician **purchased** new equipment to test ___ before the construction.	Was the new equipment purchased?	Y
21	c	plausible	prep	Tony asked which material the technician **purchased** new equipment for ___ before the construction.	Was the new equipment purchased?	Y
21	d	implausible	prep	Tony asked which location the technician **purchased** new equipment for___ before the construction.	Was the new equipment purchased?	Y
22	a	plausible	vt	William questions which neighbor the woman **bothered** the mayor to visit ___ during the event.	Did the woman bother the president?	N
22	b	implausible	vt	William questions which church the woman **bothered** the mayor to visit ___ during the event.	Did the woman bother the president?	N
22	c	plausible	prep	William questions which neighbor the woman **bothered** the mayor about ___ during the event.	Did the woman bother the president?	N
22	d	implausible	prep	William questions which church the woman **bothered** the mayor about ___ during the event.	Did the woman bother the president?	N
23	a	plausible	vt	Suzy wondered which actor the company **hired** a team to promote ___ in the summer.	Did the company hire a team?	Y
23	b	implausible	vt	Suzy wondered which song the company **hired** a team to promote ___ in the summer.	Did the company hire a team?	Y
23	c	plausible	prep	Suzy wondered which actor the company **hired** a team for ___ in the summer.	Did the company hire a team?	Y
23	d	implausible	prep	Suzy wondered which song the company **hired** a team for ___ in the summer.	Did the company hire a team?	Y
24	a	plausible	vt	Anna wondered which museum the girl **visited** the city to explore ___ during the vacation.	Did the girl visit the city on workdays?	N
24	b	implausible	vt	Anna wondered which idea the girl **visited** the city to explore ___ during the vacation.	Did the girl visit the city on workdays?	N
24	c	plausible	prep	Anna wondered which museum the girl **visited** the city for ___ during the vacation.	Did the girl visit the city on workdays?	N

24	d	implausible	prep	Anna wondered which idea the girl **visited** the city for ___ during the vacation.	Did the girl visit the city on workdays?	N