PROCESSING OF TENSE MORPHOLOGY AND FILLER-GAP DEPENDENCIES BY CHINESE SECOND LANGUAGE SPEAKERS OF ENGLISH

by

Zhiyin Renee Dong

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Linguistics and Cognitive Science

Winter, 2014

UMI Number: 3617879

UMI®
Dissertation Publishing

UMI 3617879

ProQuest®

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

PROCESSING OF TENSE MORPHOLOGY AND FILLER-GAP DEPENDENCIES BY CHINESE SECOND LANGUAGE SPEAKERS OF ENGLISH

by

Zhiyin Renee Dong

Approved: _____
Benjamin Bruening, Ph.D.
Chair of the Department of Linguistics and Cognitive Science

Approved: _____
George H. Watson, Ph.D.
Dean of the College of Arts and Sciences

Approved: _____
James G. Richards, Ph.D.
Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Gabriella Hermon, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Arild Hestvik, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Peter Cole, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Roberta M. Golinkoff, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

Undertaking this Ph.D. has been the most challenging and fulfilling life experience I have ever had. It is a journey that I could not have completed without the help and support of many wonderful people in my life. First and foremost, I would like to thank my advisors, Dr. Arild Hestvik and Dr. Gabriella Hermon for their guidance, support, and friendship. Dr. Hestvik introduced me to the wonderful world of neuro-psycholinguistics. I am forever indebted to him for training me to become an experimental psycholinguist and for his contributions of time and ideas that made my thesis project possible. I also wish to express my heartfelt gratitude to Dr. Hermon for her mentorship; her invaluable advice on the theoretical side of my work, and for helping me to complete my thesis and pushing it to the next level. I also thank Dr. Peter Cole and Dr. Roberta Golinkoff, for reviewing my thesis and providing many helpful comments.

I gratefully acknowledge the two funding sources for my thesis project: the University of Delaware College of Art and Sciences Faculty Research Grant and the University of Delaware Alumni China Award. My deepest appreciation also goes out to the professors and fellow students in the linguistics department, without whom I could not have survived the Ph.D. program. In particular, I would like to thank Rebecca Jia and Jane Chandlee, who not only gave me insightful advice regarding my work, but also provided the much-needed emotional support.

I will forever be thankful to my dearest friends Dorothy and Stuart Sharkey, who love and take care of me and my family as their own. I am also indebted to my

<p style="text-align: center;">**TABLE OF CONTENTS**</p>

Appendix

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

There is an ongoing debate in the field of Second Language Acquisition (SLA) concerning whether there is a fundamental difference between the native language (L1) and adult second language (L2) online processing of complex syntax and morpho-syntax. While some scholars maintain that L1 and L2 processing share basically the same system (e.g., Schwartz & Sprouse, 1996; Sabourin & Stowe, 2004, 2008; Bates & MacWhinney, 1987; MacWhinney, 2008), others claim that L2 processing is qualitatively different (e.g., Weber-fox & Neville, 1996; Pakulak & Neville, 2011; Hawkins & Chan, 1997; Clahsen & Felser, 2006a, b). Among the accounts holding the second view, the Shallow Structure Hypothesis (SSH) (Clahsen and Felser, 2006a, b) argues that L2 speakers build a "shallower" syntactic representation with less hierarchy and fewer details than L1 speakers, and therefore L2 speakers cannot make effective use of purely structural principles in online processing. Instead, L2 speakers rely mainly on semantic and lexical information to process sentences. The primary goal of this thesis is to evaluate such claims by examining how L2 speakers process complex filler-gap dependencies and inflectional past tense morphology in real time. Specifically, I test the following two predictions of the SSH: (a) L2 learners cannot posit abstract syntactic traces in filler-gap (relative clause) constructions (Experiment I), and (b) L2 learners cannot use morphological decomposition rules when processing tense morphology (Experiment II). The method

of Event Related Potentials (ERP), suitable in this context because its indexes reveal the nature of the underlying processing mechanism (i.e., syntactic vs. semantic), is used to compare the brain responses of L2 learners to those of native speakers. In addition, I examine how L2 proficiency, L1 interference, and working memory (WM) capacity interact with L2 processing patterns. The results indicate that the learners are able to produce native-like brain responses for tense morphology processing. However, their ERP indicators differ drastically from the native controls in resolving filler-gap dependencies in relative clauses, a construction with highly abstract elements. As the SSH predicts, the L2 speakers resort to semantic and lexical information without building syntactic traces. Furthermore, this parsing pattern is not affected by proficiency and working memory differences. While the empirical evidence is largely in line with the SSH, a different theoretical conclusion than that of the SSH was reached. Assuming the view that the parser and the grammar are integrated into one system, I conclude that L2 shallow parsing is limited to constructions with highly abstract elements with no overt reflex and persists only when this mechanism does not interfere with successful meaning computation.

Chapter 1

INTRODUCTION

1.1 Overview

An important research topic in Second Language Acquisition (SLA) concerns whether adult (or late[1]) second language learners adopt a fundamentally different parsing mechanism from that of native speakers to process complex syntax and morpho-syntax in real time. While one group of researchers maintain that first language (L1) and second language (L2) processing largely share the same system (e.g., Schwartz & Sprouse, 1996; Sabourin & Stowe, 2004, 2008; Bates & MacWhinney, 1987; MacWhinney, 2008), others argue that L1 and adult L2 processing differ qualitatively (e.g., Weber-fox & Neville, 1996; Pakulak & Neville, 2011; Hawkins & Chan, 1997; Clahsen & Felser, 2006a,b). The Shallow Structure Hypothesis (SSH, e.g., Clahsen & Felser, 2006a, b), as one of the accounts that hold the latter view, states that adult L2 learners are unable to compute deep syntactic representations with full details during online comprehension; their processing is largely guided by semantic, pragmatic, and contextual information instead of

---

[1] By "late" I mean adults who acquire the target language after childhood and in a non-immersion environment. In this thesis the terms "late second language speakers" and "L2 speakers" are used inter-changeably.

structural principles. In addition, the over-reliance on non-structural information is a basic property of the L2 parser and is therefore not influenced by L2 factors such as proficiency, working memory capacity, and native language influence. The purpose of this thesis is to evaluate such claims and to further explore the nature of L2 parsing by reporting two Event Related Potentials (ERP) experiments. The method of ERP is advantageous in this context, because its indexes clearly reveal how different information sources (syntactic or semantic) are being used in processing. Furthermore, this thesis examines how proficiency, working memory, and L1 interference interact with L2 processing patterns.

1.2 Processing as the Center of (Second) Language Acquisition

Early research on SLA largely focused on the *properties* of the L2 grammar and whether it differs from that of the L1. At the center of the debate was the state of L2 knowledge and how and to what extent Universal Grammar (Chomsky, 1965, 1981) is involved in L2 grammar construction. However, more recent SLA theories, especially those from a cognitive approach, stress that it is not sufficient to just explain the nature of the L2 grammar (or the inter-language at various stages of acquisition) in a static state. As Gregg (1996, 2003) argued, a *transition* component explaining how the L2 grammar evolves must be included in any successful SLA model. Furthermore, it is also necessary to understand the underlying mechanisms which drive such transitions and how these mechanisms differ from those of the L1. It has been widely established that this transitional aspect of language acquisition relies crucially on the

2

language input, in both the L1 and L2 context (Van Pattern, 1996, 2006; Carroll, 2001). However, how the input triggered the acquisition was not clearly defined. Fodor (1995), Fodor & Inoue (1998) noted that processing plays a central role in developing the grammar and that the input we refer to is in fact *the processed input*. There will be no language, and hence no language acquisition, if the processing of the unanalyzed segments of sound or visual stimuli did not take place. To be more specific, acquisition *is* the process of parsing, in which representations of the target grammar are constructed and then adjusted, as more input is parsed. In addition, while the parser plays this critical role in building the grammar, it is at the same time constrained by the existing grammar. The L2 parser, for example, is constrained by possibly multiple components, including the L1 grammar, at least at the beginning. Such a parser-grammar interaction should explain the transition component of SLA (i.e., how the L2 grammar evolves). Understanding L2 parsing, therefore, directly bears on the research of L2 acquisition, and whether there is a fundamental difference between L1 and L2 parsing also addresses the question of the L1-L2 acquisition difference in general. It is against this backdrop that the above-mentioned debate began and the issues that motivated this thesis are explored.

1.3 The Shallow Structure Hypothesis

According to the researchers who maintain that L1 and L2 online processing of syntax share the same system (i.e., one-system accounts), native-like processing patterns are attainable by L2 learners. Although at times L2 parsing may appear to be

less automatic and in-depth than that of L1, such deficiencies can be explained by various factors other than the age of acquisition. These factors include the level of proficiency (e.g., Larson-Hall, 2006; Ojima, Nakata, & Kakigi, 2005), L1 influence (e.g., Weber & Cutler, 2004; Jeong et al., 2007; Sabourin, Stowe, & de Haan, 2006; Tokowicz & MacWhinney, 2005;), and a higher demand for cognitive resources such as working memory and attention in L2 processing (e.g., Bialystok & Hakuta, 1999; McDonald, 2006). In contrast to these "one-system accounts", the Shallow Structure Hypothesis (SSH) (Clahsen & Felser, 2006a, b) maintains that L2 processing operates differently from that of L1 even at the most advanced stage of acquisition, regardless of the L1 background. Specifically, the SSH holds that in real time parsing L2 speakers cannot construct detailed and complete syntactic representations as the native speakers do. In addition, based on findings of the two studies conducted by Felser, Roberts, Gross & Marinis (2003) and Papadopoulou & Felser (2003), the SSH claims that L2 speakers primarily resort to semantic, lexical information and pragmatics in real-time comprehension and under-use syntactic rules (Clahsen & Felser, 2006a, b; Clahsen & Neubauer, 2010; Neubauer & Clahsen, 2009; Silva & Clahsen, 2008).

Both of the studies examined L2 speakers' preferences for relative clause attachment. Consider the following sentence:

(1)   Someone shot the servant of the actress who was standing on the balcony.

The relative clause (RC) *who was on the balcony* can plausibly modify the first NP (*the servant*) or the second NP (*the actress*). It has been shown that native speakers of English prefer to attach the RC to the second NP based on phrase structure rules (e.g., Carreiras & Clifton, 1999). The L2 speakers tested in Felser et al. (2003) and Papadopoulou & Felser (2003), however, indicated no attachment preferences; that is, they attached the RC to either *the servant* or *the actress* at the rate of chance. These L2 speakers, all at an advanced level of proficiency, were from typologically different language backgrounds. In their native languages they had different relative clause attachment preferences (either NP1 or NP2). It was only when they processed the second language did they consistently show no attachment preferences. Clahsen and Felser (2006 a, b) argued that this pattern was due to the "semantics and pragmatics-first" strategy of the L2 speakers and concluded that L1 influence had no significant effect on the L2 parsing strategies. In addition, the SSH states that while speakers at a higher proficiency level deliver better L2 performances, this doesn't change the basic properties of the L2 processing mechanism, especially at the sentence level.

Since the original formulation of the SSH in 2006, there have been a number of studies testing its predictions (e.g., Rodriguez, 2008; Omaki & Schulz, 2011), but the results are far from conclusive. Whether the L2 learners use a separate parsing system remains an issue under active investigation in the developing L2 processing field. Given the importance of this topic, it is clear that much work is needed, particularly with electrophysiological and neuroimaging techniques. What has also become increasingly evident is that the difference between L1 and L2 processing might

interact with the type of syntactic construction under investigation. While the L2

studies on relative clause attachment preference tend to generate non-native-like

patterns as predicted by the SSH (e.g., Jegerski, 2010; Rah & Adone, 2010), other

studies on constructions like gender and number agreement have produced more

consistent, native-like results (e.g., Ojima et al, 2005; Downes, T. Guo, J. Guo, Barber

& Carreiras 2011; Downes, Vergara, Barber & Carreiras, 2010). The current study

examines inflectional morphology and filler-gap dependencies, which are particularly

important and suitable for experiments for a number of reasons. First, the properties of

these structures make them ideal test cases for the nature of L2 processing. FG

dependencies are complex hierarchical structures involving the use of highly abstract

syntactic rules such as movement and trace posting, hence they directly test L2

speakers' ability to use abstract structural information to parse online. Similarly, in the

area of morpho-syntax, inflectional morphology processing also demands the effective

use of abstract morpho-syntactic rules (i.e., grammatical agreement, decomposition for

morphologically complex words). Second, both structures have been found to be

difficult to process by L2 learners (e.g., Juff, 2005, 2006; Jiang, 2004, 2007). Lastly,

the existing L2 processing data on these two structures are both limited and inclusive

(see section 2.3.1 and 4.4, for a detailed review). This dissertation thus focuses on the

online parsing of these two constructions as representative of syntax processing.

To evaluate the claims of the SSH, this dissertation (1) examines the L2 brain

responses to complex filler-gap dependency structures in an Event Related Potential

(ERP) experiment. ERP is a particularly suitable test for the SSH, because its

6

components reveal the nature of the underlying mechanisms of parsing. For example, the ELAN component reflects highly automatic, first-pass syntactic processing, and the N400 component indicates semantic anomalies (see 3.2.1 for details). In the first experiment, sentences like *The zebra that the hippo kissed the money ran far away*, as opposed to its grammatical counterpart *The zebra that the hippo kissed on the nose ran far away,* were presented to the subjects in auditory form. Assuming a trace is posited immediately after the verb *kissed*, the extra NP *the monkey* can't fit into the direct object position because it is taken by the trace. Consequently, the (EL)AN, an ERP component indicating phrase structure violations, should be obtained. This was confirmed in (Hestvik, E. Bradley & C. Bradley, 2012) in which native speakers processed such violations, suggesting that they indeed posit an abstract structural trace. The L2 learners, being unable to build such detailed and abstract syntactic structures according to the SSH, are expected to identify the violation as semantic in nature and generate the N400, or else fail to produce any syntactic related component. With the same method, this thesis also studies whether the L2 speakers can effectively apply morpho-syntactic decomposition rules for processing English past tense. An ERP experiment modeled after the study by Hestvik, Maxfield, Sshwartz & Shafer (submitted) will be reported to show how advanced Chinese speakers of English react to tense violations such as *Yesterday I walk to school* and *Yesterday I eat a banana*. In addition, as previously mentioned, a few factors such as proficiency, L1 language background, and working memory capacity have been proposed to interact with L2 acquisition to various degrees (e.g., Ojima et al. 2005; McDonald et al., 2006;

Tokowicz & MacWhinney, 2005). It is thus critical to inspect their roles in L2 parsing as well. Both experiments take L2 proficiency into consideration by correlating carefully measured proficiency scores with the L2 brain responses. Lastly, since working memory has been found to affect the processing of long-distance dependencies in an L1 (e.g., McDonald, 2006; Hestvik, et al. 2012), Experiment I on filler-gap dependencies also explores the effect of WM on L2 parsing. The results obtained will not only directly test the SSH and address the one- vs. two-system controversy, they will also help to build a more accurate L2 processing model and contribute to our understanding of SLA in general.

1.4   The Organization of the Dissertation

This thesis is organized as follows. Chapter 2 provides the necessary background information and addresses a few important issues in L2 sentence processing, especially that of long-distance filler-gap dependencies (FG). It first reviews several L1 processing models related to the SSH and then discusses the key findings about L1 filler-gap dependency processing. Issues in L2 sentence processing are then presented, followed by a detailed discussion of the SSH. Section 3 of Chapter 2 reviews a few key variables known to affect second language processing, with a special focus on the role of proficiency and working memory (WM) capacity. Chapter 3 reports an experiment on the processing of long-distance filler gap dependencies. This chapter starts off with an introduction to the ERP methods, its correlates in filler-gap dependency processing, and a list of predications of the SSH for FG processing in

8

ERP terms. The next section reports on (1) the L2 working memory test, (2) the paper-and-pencil acceptability test designed to measure the subjects' off-line grammaticality knowledge, and (3) the Versant English test for the proficiency measurement. The last part of Chapter 3 includes the details of the ERP component of the experiment and a discussion of the results. Chapter 4 examines the topic of how L2 learners process morpho-syntax in real time. A literature review on the issue of decomposition and the existing L2 morphology processing findings are presented first. Following that, the predictions of the SSH in the context of Experiment II are laid out. The rest of Chapter 4 reviews and discusses the design, implementation, and results of Experiment II. Finally, Chapter 5 summarizes the results of both experiments and addresses the differences in their findings. In particular, the implications for the SSH, L2 processing, SLA models in general, and language instruction pedagogies are discussed. The thesis concludes by identifying a few limitations of the current study and potential future research directions.

Chapter 2

L1 AND L2 PROCESSING OF FILLER-GAP DEPENDENCY

2.1   Introduction

This chapter provides some necessary background information about L1 and

L2 sentence processing. It reviews the Shallow Structure Hypothesis (SSH) as a

theoretical construct and presents the issues that motivate Experiment I. The first part

of this chapter discusses a central issue in L1 parsing about how different information

sources (i.e., syntax, lexical semantics, and pragmatics) are utilized in sentence

processing. This issue has direct implications for L2 processing research and the SSH.

I then review research on the processing of filler-gap dependencies, which are of

particular importance here as one of the ideal testing grounds for sentence parsing.

Much of the supporting evidence for the claims of the SSH has been generated in

studies investigating this construction. Next, the role of syntactic traces is examined

from the processing perspective to lay the foundation for the more in-depth discussion

of the SSH's claims in section two. I then discuss Good Enough Representations or

GERs (Ferreira, Ferraro & Bailey, 2002; Ferreira & Patson, 2007), a form of shallow

processing in L1, and how they relate to the SSH. The second part of this chapter

focuses on L2 sentence processing. It begins with a literature review on how L2

parsing has been found to be similar and/or different from that of L1. Next, the SSH as

a L2 parsing model is examined in detail, followed by a summary of the issues and previously untested SSH predictions that will be addressed by this thesis. In part three I review a few factors that are known to potentially affect L2 acquisition and processing, in particular, L2 proficiency and working memory capacity. I discuss how these two factors are important in evaluating L2 models of processing and why their accurate measurement and interpretation are needed when exploring their roles in L2 parsing.

2.2   L1 Sentence Processing

2.2.1   The nature of L1 sentence processing

Human online processing involves the real-time integration of different types of information (e.g., structural, lexical, and pragmatic). The issue of focus in L1 sentence processing is when and how these different sources of information come into play. This question is equally critical to L2 processing research and the SSH in particular, as it bears directly on the claims of the SSH that L2 learners under-use syntactic cues but over-use semantics and pragmatics to process, as mentioned in the introduction. What are considered "syntactic" as opposed to other non-structural information sources here are rules and principles such as phrase structure rules (e.g., S-> NP VP), structural traces, and verb argument structural requirements (e.g., whether the verb takes a direct or indirect object, whether it requires a prepositional phrase, etc.). Semantic information, however, refers to the meaning and the lexical

properties of words. In the case of verbs, this includes the thematic assignment

requirements such as whether the verb takes an animate or inanimate argument, or

how a given verb is biased to take certain arguments.

With regard to how the different sources of information affect online parsing,

two opposing views have been formulated in the L1 processing literature, mostly

through examination of structures with temporary ambiguity. One view assumes a

modular perspective, in which syntax is considered as its own module and is used first

for parsing without any other information (e.g., Frazier, 1987). This is also called the

serial model or the two-stage model, because two consecutive phases are assumed:

initially, structural cues dominate the first-pass building of the representation, while

other information only comes into play in a later phase. One representative proposal in

this category is the well-known garden-path model (Frazier, 1983, 1987; Frazier &

Rayner, 1982), through which some influential syntactic processing principles were

put forth. Based on the widely accepted assumption that there are limited resources

available to the human parser (e.g., Frazier, 1999; Mitchell, 1994), Frazier (1987)

proposed that the parser builds the simplest structure possible with the fewest

branching nodes, using only structural information (Minimal Attachment). Thus, as

the parser processes *The horse raced past the barn fell* (Bever, 1970), *raced* is

typically treated initially as a main verb in the matrix clause instead of as a past

participle, because the main-verb analysis entails a much simpler syntactic

representation. Similarly, the strategy of Late Closure was formulated such that any

incoming word is immediately attached to the last item processed in an attempt to

build a structure as soon as possible. For example, in a sentence like *When Anna dressed the baby played in the crib*, when the parser encounters *the baby*, it immediately attaches it to the matrix verb *dressed*. Only after it processes *played* does it revise the structure. The application of such a strategy reflects another established characteristic about online processing, that is, it is highly incremental.

Contrary to the modular or "syntax-first" perspective, the lexical or interactive approach posits that non-structural information is in full play from the very beginning of the parsing process (e.g., MacDonald, 1999; Trueswell, Tannenhaus & Garnesy, 1994). Specifically, this perspective assumes that lexical, pragmatic, and discourse information, as well as other non-syntactic factors such as frequency, are all processed in parallel and immediately influence the building of a structure (Boland 1997, MacDoanld, Just & Carpenter, 1992; Trueswell et al; 1994, Trueswell 1996). Of particular interest here is the parallel processing account put forth by Townsend & Bever (2001). In this model, a "quick and dirty" representation is computed first based on lexical information, segmentation of phrases, and possible argument relations between the heads of these phrases. Later, a full and complete structure is computed to check this initial "shallow representation" and to make necessary revisions or reanalysis. This model highlights the important role of non-structural information in initial parsing and resonates with other accounts of shallow processing for both L1 and L2. For L1 speakers, the most well known shallow parsing account in relation to parallel processing is the Good Enough Representations (GERs) proposed by Ferreira

et al. (2002) and Ferreira & Paston (2007), which will be discussed in detail in section 2.2.3.

Although the issue of what and how information is used in the initial stage of parsing is not fully resolved (e.g., Boland 1993; Clifton et al., 2003; Garnsey, Pearlmutter, Myers & Lotocky, 1997), sentence processing models, regardless of whether the serial/modular view or the lexical/interactive view is assumed, converge on the point that a full and detailed syntactic representation is built during online processing. This is important when considering the shallow processing accounts proposed for both L1 and L2 speakers. In the next section, I will turn to another structure that has generated much interest in both the L1 and L2 processing research: filler-gap dependencies.

2.2.2   L1 Filler-Gap processing and the status of abstract traces

Filler-Gap dependencies (FG) refer to the relation between a sentence constituent that appears in a non-canonical position (mostly as a result of movement) and its originating position, typically where a verb assigns this item its thematic role (Phillips & Wagers, 2007). FG dependencies exist in structures such as relative clauses, topicalization, and *wh*-interrogatives. The following example illustrates a FG dependency in English:

(2)   The Lady that the doctor treated__ yesterday for a minor cut was from

        England.

14

In (2), it is assumed that the NP *the Lady* was extracted from its original position after the verb *treat*, which assigns it the patient role in the underlying structure. *The Lady* therefore is the dislocated item or the *filler* that needs to be restored to the position following *treat,* or the *gap,* for successful meaning computation. A FG dependency can be long-distance, and there might be multiple potential gap sites. Consider (3)

(3)   Who did you invite__to visit the campus with__?

The parser might be tempted to fill the gap as soon as it encounters the first verb *invite*. This is confirmed by a parsing mechanism called Active Filler Strategy (AFS) (Clifton & Frazier, 1989), which has been identified in many studies on FG dependencies. With this strategy, the parser restores the filler at the nearest potential gap position, regardless of its semantic suitability. Consider the following sentence (4):

(4)   Which magazine did the old lady say__ that she read _with great pleasure?

By the Active Filler Strategy, the parser would attempt to fill the dislocated item *magazine* at the first potential gap position after the verb *say*, even though it is implausible to "say a magazine". AFS has been well tested in experimental settings

and has received strong empirical support in the research on FG dependencies (e.g.,

Williams, Möbius & Kim 2001; Williams, 2006).

Another much-explored issue in FG dependencies critical to this thesis

concerns the nature of the relation between the filler and the gap, specifically, whether

an abstract trace is involved in this relation. There are two contending models in this

regard: the direct association hypothesis (DAH) maintains that the FG relation is

established by directly associating the dislocated item with its verb subcategorizer

without positing an abstract trace (Pickering & Barry, 1991; Sag & Fodor 1994). That

is, verb argument structure, thematic role assignment, and other semantic information

are used to resolve the dependency. According to this account, any verb after the filler

in the sentence, if semantically appropriate, is linked by the parser to the filler. In

other words, the filler remains activated until the point at which the subcategorizing

verb is found. On the other hand, the trace reactivation hypothesis (TRH) claims that

an abstract trace must be posited at the gap position (Nicol & Swinney 1989; Swinney,

Ford & Bresnan, 1989). Under this view, the parser registers the filler in working

memory upon first recognition. It then proceeds until it processes the trace, which

triggers the reactivation of the filler for resolution of the dependency. Thus, only verbs

with potential trace positions will be considered for the integration of the filler.

Various studies with different methodologies have attempted to differentiate

the above two views. Love (2007), Nicol (1993), and Roberts, Marinis, Felser &

Clahsen (2007) all used cross-modal priming to examine how adults and children

process long distance FG dependencies. Their findings show a priming effect at the

16

gap site in contrast to the non-gap site and were interpreted as evidence for trace

activation. However, it was argued that these findings could also be compatible with

the direct association view (Nicol, 1993; Sag & Fodor; 1994; Miller, 2011), due to

methodological issues that will be discussed in detail in section 2.3.2. A Self-Paced-

Reading (SPR) study conducted by Gibson and Warren (2004) examined sentences as

in (5):

(5)   a. The manager who the consultant claimed *<who>#1* that the new proposal
        had pleased <who>#2 will hire five workers tomorrow.

        b. The manager who the consultant's claim about the new proposal had
        pleased <who> #1will hire five workers tomorrow.

By the trace-reactivation account, an intermediate trace is posited at *<who>*#1

in (5a). A delay in reading time at this position is predicted due to the additional

integration cost in comparison to (5b), which has no intermediate trace. The results of

Gibson and Warren (2004) confirmed this prediction and lent support to the trace-

driven FG dependency accounts.

Furthermore, several FG processing experiments examined verb-final

languages (Aoshima, Phillips & Weinberg, 2004; Fieback, Schlesewsky & Friederici,

2002) and demonstrated that filler integration began before the verb, a point at which

the argument structure information couldn't have been available because the parser

has not processed the verb yet. Those findings are in line with the trace reactivation

account. Finally, in an Event Related Potential (ERP) study, Hestvik et al. (2007,

17

2012) had native speakers of English listen to the following ungrammatical stimuli

(6a) as opposed to their grammatical counterparts (6b):

(6)   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.

In (6a), the parser is hypothesized to attempt to fill *the zebra* at the potential

gap site right after the verb *kissed*, only to find that the gap position has already been

filled. This kind of "filled gap" effect generated an ELAN, an ERP component

typically found for syntactic violations that suggests sensitivity to structural violations,

as opposed to the N400 component indicative of semantic violations (details on these

ERP components will be reviewed in 3.2). Such a finding is compatible with the trace-

reactivation account, for if the FG association were based on the verb's arguments and

semantics, a semantic N400 would have been obtained. To summarize, the above

reviewed evidence from L1 FG processing studies seems to support that an abstract

trace is posited and used to resolve FG dependencies by native speakers.

2.2.3   Shallow processing in L1: the Good Enough Representations (GERs)

As mentioned in Clahsen & Felser (2006b), the SSH is not a processing

mechanism unique to L2 learners. Several similar accounts of "shallow processing"

(e.g., Fodor, 1995; Ferreira, et al., 2002, Ferreira & Paston, 2007; Sanford & Sturt,

2002) have been proposed for native speakers as well. Among these accounts the

"Good Enough Representations" account (GERs) (Ferreira et al., 2002, Ferreira & Patson, 2007) can be related to the SSH. According to Ferreira and colleagues, the human parser does not always construct accurate, complete, and detailed representations for the language input. Instead, the comprehenders merely create a "good enough" understanding due to time, capacity, or input ambiguity constraints. Specifically, for example, some simple heuristics rather than compositional algorithms are sometimes used for sentence construction. Consider a sentence such as *A fox shot a poacher*. It was found in an Event Related Potential experiment (van Herten, Kolk & Chwilla, 2005) that subjects who heard such a sentence didn't generate the N400 component, which again is typically obtained for semantic anomaly (Kutas & Hilliard, 1980a, b, c), possibly because they were heavily biased by their world knowledge and computed the semantically more plausible interpretation *A poacher shot a fox*, thus ignoring the straightforward structure and clear word order. Such use of plausibility heuristics was taken as evidence that the processing of the structure was clearly "shallow". Similarly, it was observed that for garden path sentences such as *When the gardener bathes his poodle joins him*, the listeners often mistook *the poodle* as the recipient of the verb *bath* (Ferreira & Henderson, 1991), not just temporarily, but even after the sentence was finished. When presented with follow-up comprehension questions, although the participants correctly acknowledged that the poodle bathed with the gardener, they also erroneously agreed that the gardener bathed the poodle as well. In this case, it was the over-use of structural heuristics such as the Minimal

19

Attachment principle (Fodor & Inoue, 1998) that led the participants to derive a semantically shallow/erroneous interpretation.

However, it needs to be clarified that the scholars who proposed the L1 shallow processing account also suggested that whether such an algorithm was used depended on the demands of the comprehension task (Sanford & Strut, 2002) . In two studies carried out by Swets, Desmet, Hambrik & Ferreira (2007) and Swets, Desmet, Clifton & Ferreira (2008), the participants were asked to read sentences like those in (7):

(7)   a. The maid of the princess who scratched herself in public was terribly humiliated.

        b. The son of the princess who scratched himself in public was terribly humiliated.

        c. The son of the princess who scratched herself in public was terribly humiliated.

In (7b) and (7c), the ambiguity in (7a) due to the relative clause attachment position is avoided by the gender information encoded in the reflexive pronouns. The participants were divided into two groups: one group answered *RC detail* questions such as *did that maid/son/princess scratch in public?* And the other answered *superficial* questions like *was anyone humiliated?*. It was found that those who answered *RC detail* questions spent more time reading overall, particularly at the reflexives when compared to the group who answered only *superficial* questions. These findings suggest that although the parser opts for shallow processing at times

20

for the sake of saving memory costs, it can engage in full analyses when the situation calls for them.

Additional empirical evidence from ERP experiments revealed that seemingly shallow processing is only temporarily shallow. For example, van Heurten, Chwilla & Kolk (2006) conducted an experiment in which the participants read sentences such as (8) and (9):

(8)   a. The painter who climbed the ladder suddenly fell.

        b. The ladder that climbed the painter suddenly fell.

(9)   a. The squirrel that climbed the tree looked cute.

        b. The apple that climbed the tree looked juicy.

It was predicted that if the parser prioritizes plausibility heuristics, then no N400 indicative of semantic anomaly would be detected for (8b) compared to (8a). That is, the worldly knowledge of the participant simply takes over and the structural information is ignored when interpreting the sentence. However, in the case of (9), the plausibility heuristics are not expected to interfere, because neither *the apple* nor *the tree* can climb. Thus, a N400 is expected for (9b) but not (9a). Van Heurten et al. found no N400 for (8b) as expected. However, a P600 indicative of syntactic reanalysis was detected later, suggesting that syntactic information was incorporated after all and conflicted with the semantically built representation. Additionally, a N400 of a small amplitude was found for (9b), followed by a greater P600 than that

observed for (8b). The P600 component thus clearly suggests that structural

calculation was conducted, albeit later. Similar evidence was produced in another ERP

study (Nieuwland & Van Berkum, 2005), in which no N400 but a P600-like effect was

found for sentences like *The woman told the suitcase that she thought he looked

really trendy* with a context of a man and a woman talking at luggage check-in. In

short, the evidence reviewed above suggests that what seems to be shallow processing

in L1 might occur only in the initial stage of parsing, to be followed up by more

complete syntactic analysis later.

Although Good Enough Representations (GERs) and their supporting evidence

contradict the syntax-first accounts, heuristic-based algorithms were not proposed to

always replace structural analysis for L1 speakers. Rather, in GERs semantics and

other non-syntactic information is used interactively with the structural information,

although it is not entirely clear how the syntax- and semantics-guided mechanisms

work together. While the SSH and GERs converge in various aspects, and the shallow

processing suggested by both are very likely based on the same mechanism, L1

shallow processing seems to differ crucially from the SSH in that (1) L1 processing is

only temporarily shallow, while in the case of L2 parsing the structural details are

never built, and (2) native speakers only shallow process when necessary. Although

the exact triggering circumstances have not yet been clearly defined by the GERs

proposal, native speakers can prioritize either semantics or syntax for shallow parsing

depending on the task, as shown in previous examples. In contrast, L2 learners,

according to the SSH, are largely restricted to shallow parsing and must rely

exclusively on semantics and pragmatics, regardless of their individual differences.

Thus, there are substantial differences between the two types of "shallow processing"

proposed by the GERs proposal and the SSH, and these differences must be taken into

consideration when comparing L1 and L2 processing patterns, in order to evaluate

whether L1 and L2 processing mechanisms differ qualitatively.

2.3   Issues in L2 Sentence Processing

2.3.1   L2 processing and Filler-Gap processing

Although L2 online sentence processing is still a relatively new research area,

it is growing fast and the results generated so far have provided us with some

understanding of the subject. Along the line proposed by the one-system proponents,

who believe that L1 and L2 parsing share largely the same system, it has been found

that L2 learners have access to complex structural representations (Juff, 2005, 2006)

and can make use of many parsing routines available to L1 speakers in similar ways.

For example, Williams et al. (2001), Williams (2006) tested whether advanced L2

speakers of English could use the Active Filler Strategy and structural and non-

structural (plausibility) information together in a Stop-Making-Sense (SMS) task. In

SMS experiments, the subjects read sentences one word (or a segment of a sentence)

at a time by pushing a button, much like in the Self-Paced Reading (SPR) task. When

they encounter a word that makes the sentence unacceptable, they push a special

"rejecting" button. The response time is recorded at the "rejecting" point, and helps

23

identify the location of incongruity or processing difficulties. The following stimuli sentences were presented to the participants in Williams (2006)'s experiment:

(10)   a. Which girl did the man push the very noisy bike into late last night?

       b. Which river did the man push the very noisy bike into late last night?

The rationale is that the dislocated item *which girl/river* prompts the parser to look for a gap to fill in. If the Active Filler Strategy (AFS) is online, then the parser should posit a gap at the earliest position available, namely, after *push*. However, in (10b) it is not plausible to *push a river*, which might lead the reader to consider the sentence as not sensible at that point. In contrast, since it is plausible to *push the girl* in (10a), the reader should read on to find the true gap position. These expectations were borne out in both Williams et al. (2001) and Williams (2006), in that both the native speakers and the L2 participants[2] made more "reject" decisions right before the *very noisy bike* in (10b) only, suggesting that the AFS is applied in L2 online processing, and that plausibility information was actively incorporated into structural decisions[3] just like the native speakers.

---

[2] These subjects were intermediate high to advanced English speakers from various L1 backgrounds (Korean, German, Chinese).

[3] Plausibility information was found to be effectively used by the L2 speakers in Williams (2006) only.

Related to how verb subcategorizaton information affects L2 online processing, Frenck-Mestre & Pynte (1997) examined L2 speakers' eye movements while reading sentences such as *They accused the ambassador of espionage/Indonesia but nothing came out of it*, in which the PP *of espionage/Indonesia* can be attached to either the verb *accused* or the NP *the ambassador*. It was found that among the native speakers the attachment preference was influenced by the subcategorization requirement of the verb, such that if the verb is di-transitive, then the VP attachment is preferred, and if the verb is mono-transitive, then NP attachment is preferred. This preference was reflected in a shorter total first-pass reading time in the target region. The eye movement patterns (mostly first-pass reading duration) of the advanced L2 English speakers of French showed that their attachment decisions were influenced by verb subcategorization information the same as the native speakers, although the learners had some difficulty attaching the PP to the VP, as reflected by their significantly greater regression times after VP attachments. Frenck-Mestre & Pynte (1997) conclude that their data don't support the claim that the L2 parsing strategy is qualitatively different from that of L1.

Additionally, learners were found to process structures with temporary ambiguities such as reduced relative clauses and garden path sentences in a similar way as native speakers (Juffs 1998; 2004). Juffs (1998) investigated the processing of sentences like *The bad boys criticized almost every day were playing in the park* by L2 speakers of English with diverse L1 backgrounds. Given that reduced relative clauses are resolved by quickly integrating various verb information, including

25

subcategroization requirements, three different kinds of verbs were tested: transitive

verbs like *criticize*, optional transitive verbs like *watch*, and verbs with irregular past

participle forms such as *see* (seen). The last type suggests the presence of the reduced

RC more that the first two. In addition, two types of time adverb phrases were

positioned after the verb, ones like *almost every day*, which tend to suggest the

preceding verb is a main verb, and others like *during the morning*, which rule out a

direct object and suggest the preceding verb is not a main verb. Thus the adverb

phrase helps with the reduced RC interpretation. Juffs (1998, 2004) found that the

advanced L2 speakers could incorporate various cues (subcategorization and non-

structural) and produced a reading time pattern similar to that of the native speakers

despite their L1 background, indicating that L1 and L2 processing are not qualitatively

different.

However, it has been argued that late learners are qualitatively different from

native speakers in real-time parsing (e.g., Weber-Fox & Neville, 1996; Pakulak &

Neville, 2011; Hawkins & Chan, 1997). For example, in an ERP study testing phrase

structure violations, Pakulak & Neville (2011) found that native speakers produced a

LAN followed by a P600 component. The LAN is usually found 300-500ms after the

onset of the offending item, reflecting earlier syntactic analyses such as morpho-

syntactic processing (e.g., Friederici, 2002), and sometimes phrase structure building

(Hestvik et al., 2007, 2012). The P600 occurs later, in the 600-800ms time window

after the violation, and is typically associated with second-pass, controlled syntactic

analyses such as those triggered by Garden Path sentences (Hagoort et al., 1993;

Friederici, 2002; Kann, Harris, Gibson & Holcomb, 2000). The advanced L2 English speakers, however, only generated the P600. The lack of a LAN component is usually interpreted as reduced sensitivity to structural violations. Pakulak and Neville therefore concluded that their results, which were further supported by a previous fMRI study using the same paradigm (Pakulak & Neville, 2011), showed a different neural organization for L1 and L2 syntactic processing. It has also been observed that L2 learners tend to rely more on semantics and plausibility information to process (Williams et al., 2001; Felser et al., 2012) than native speakers. These experimental results gave rise to the two-system accounts of L2 processing such as the SSH. In the next section, I will discuss this hypothesis in detail and review a few important studies that constituted evidence for the SSH's claims.

2.3.2 The SSH and sentence processing

The SSH states that late L2 learners use a fundamentally different online parsing mechanism. As Clashen and Felser (2006 a, b) noted, while L1 speakers build a full, hierarchical syntactic representation from both the top down and bottom up, L2 learners can only manage a shallow representation and resort to thematic relations and simple heuristics to derive meaning. Specifically, for complex structures such as long distance FG dependencies, C & F claim that L2 learners try to semantically integrate the filler directly with the verb subcategorizer, instead of positing abstract traces as native speakers do. These claims were formulated partially based on Townsend & Bever (2001)'s lexical-interactive proposal on sentence parsing. Under this view, there

27

exists two parsing routes for every parser: a syntax-dominant parsing, which builds a full structure with all the abstract details, and a semantics-driven shallow parsing, which relies on lexical, semantic, and pragmatic information to achieve a still adequate understanding. These two parsing routes are both readily available to native speakers, and at times they attempt shallow parsing initially without building the full structure (Ferreira et al., 2002). The SSH, however, states that L2 parsing is limited to only the shallow processing route during online parsing. Furthermore, such a limitation will not change with an improvement in proficiency.

The SSH also appeals to Ullman (2001a; 2001b; 2005)'s Declarative/procedural (DP) model for theoretical support from the neuropsychological perspective. Ullman explains that two mental components are involved in language use: the lexicon, which records all the words and everything specific/idiosyncratic about the language, and the grammar, which handles the "regular" part of the language via rule making and application (e.g., basic phrase structure rules). These components are distinctly associated with two long-term memory systems: the declarative memory system and the procedural memory system. The declarative memory provides critical support to the lexicon and is largely available to conscious control. It handles the learning, representation, and use of knowledge about facts and events, and it is especially important for language functions such as word learning. The procedural memory, on the other hand, is an implicit memory system that supports certain parts of the grammar that are rule-based, combinational, and pertain to structure building. These two systems interact and can

work cooperatively or competitively (e.g., the use of one suppresses the functionality of the other) in both language learning and other mental domains. Crucially, both systems are sensitive to certain pharmacological and endocrine changes (i.e., estrogen has been found to stimulate the declarative memory system), and those changes are more significant at certain ages in the human life span (i.e., puberty and menopause). Thus, the functionalities of these two systems are age modulated. For instance, while the procedural memory system is fully matured at an early age (before puberty), the declarative memory system continues to develop into early adulthood. Late L2 language learners therefore must rely exclusively on the declarative memory system, at least in the beginning, to develop their L2 system. Furthermore, as one system's functionalities increase, the other tends to be depressed. This account corresponds nicely with the fact that L2 learners rarely have difficulties with lexical acquisition but struggle more with certain abstract, rule-based grammar functionalities. Ullman's DP model thus provides a suitable neuropsychological foundation for SLA theories that posit a fundamental difference between L1 and L2 acquisition, including the SSH. However, the DP model also maintains that although the procedural memory is not as applicable to L2 learners as the declarative system, it does not mean that there is no room for enhancement, and native-like attainment of the grammar functionality is possible with sufficient quality exposure to the L2.

Turning now to the empirical evidence cited by the SSH camp, I first discuss the case of intermediate traces in the L2 processing of FG dependencies. Clahsen and Felser (2006b) claim that L2 speakers are incapable of building hierarchical abstract

syntactic elements online by referring to the results obtained in Marinis et al. (2005),

which concluded that L2 speakers fail to build an intermediate trace in establishing

long-distance FG dependencies. Marinis et al. (2005)'s Self-Paced Reading

experiment is a L2 replication of the Gibson & Warren (2004) study reviewed above

in section 2.2.2. Subjects were intermediate-high to advanced learners of English from

various L1 backgrounds such as German, Greek, Chinese, and Japanese. The latter

two are *wh*-in-situ languages, so possible native language interference was expected to

occur. The following are sample test sentences presented in the experiment:

(11)   a. The manager who the consultant claimed *<who>#1* that the new proposal
        had pleased *<who>*#2 will hire five workers tomorrow.

        b. The manager who the consultant's claim about the new proposal had
        pleased *<who>*#1will hire five workers tomorrow.

        c. The manager thought the consultant claimed that the new proposal had
        pleased the boss in the meeting.

        d. The manager thought the consultant's claim about the new proposal had
        pleased the boss in the meeting.

Assuming the movement account of generative grammar theory (e.g.,

Chomsky, 1986), the <who>#1 trace in (11a) originated from the <who>#2 trace,

which is in the base position. Because a one-step movement from *<who>*#2 is

prohibited by principles such as Subjacency (Chomsky, 1986; Huang, 1982), which

stipulates that movement cannot cross a clausal boundary, a middle step is needed,

hence the intermediate trace (<who>#1). In contrast, there is no need for the

intermediate trace in (11b), as there is no clausal boundary in between the trace and the filler *the manager*.

Marinis et al. (2005) reported that both native speakers and L2 learners showed an increased reading time at the ultimate gap sites (after *had pleased*) for (11a) and (11b), in comparison to the controls (11c) and (11d), indicating that both groups integrated the filler at its canonical position. However, only the native speakers had an increase RT at the proposed intermediate site, that is, the *<who>*#1 after *claimed* in (11a) in comparison with (11c). Furthermore, the RT at the gap position in the condition with an intermediate trace, (11a), was shorter than that in the condition without an intermediate trace, (11b), presumably because the activation in the middle alleviated the integration cost at the final gap site. It was also reported that the L2 learners were highly accurate with the off-line comprehension test, suggesting that their off-line performance was comparable to that of the native speakers. The difference between the L1 and L2 online data was taken by Marinis et al. and Clashen & Felser (2006b) as evidence that the intermediate trace was not built in the L2 subjects' syntactic representations during online computation, because the L2 speakers are incapable of constructing such an in-depth structure.

Several issues were raised by other researchers in response to Marinis et al. (2005)'s study (Rodriguez, 2008; Dekydtspotter, Schwartz & Sprouse, 2006). Rodriguez (2008) redesigned the Marinis et al. (2005) study with a few

methodological changes[4] and tested advanced Chinese and Spanish speakers of English. A different picture emerged from his findings: the L2 learners generated native-like RT patterns showing a delay at both the intermediate trace site and the gap site. However, only the native speakers showed a reduced RT at the ultimate gap site, reflecting integration "alleviation" brought on by the additional filler activation at the intermediate trace position. Rodriguez explained that it was actually due to a spill-over effect caused by the difficult genitive construction in (11b), which negatively affected the L2 learners more than the L1 learners and increased the RT at the gap site. Such an explanation is certainly plausible; however, it cannot be verified without another replication experiment with easier stimuli. Moreover, it is difficult to interpret RT differences in the given paradigm in relation to the underlining *nature* of the mechanism of the FG dependency formation, because RT delays can result from both direct association and trace reactivation. To better answer the question of whether L2 speakers rely on non-structural information, it is necessary to examine the issue with a different paradigm.

Additional evidence that L2 speakers under-use syntactic information in online processing comes from Felser and Roberts (2007), who tested adult advanced Greek learners of English. This study replicated Roberts et al. (2007)'s study on English

---

[4] For example, the response data reported in Rodriguez (2008) was residual time instead of raw time, which was used in the original Marinis, et al. (2005) study. According to Rodriguez (2008), residual data is adjusted for lexical differences among stimuli, so it is appropriate for the Marinis et al. (2005) materials.

native speaker adults and children (age 5-7). Using the method of Cross Modal

Priming, Roberts et al. (2007) presented stimuli with an indirect object relative clause,

as in (12)

(12)   Fred chased the squirrel to which the nice monkey explained the game's [#1]
         difficult rules [#2] in class last Wednesday.

In cross-modal priming experiments for FG dependencies, participants listen to

stimuli sentences and make some kind of judgment about visual probes (e.g., live or

not alive) shown at certain critical position (e.g., the hypothesized gap position such as

[#2] in (12)) by pushing a button. The target visual probe is usually related to the filler

semantically and is supposed to elicit a faster reaction time (i.e., priming effect)

compared to the control probes that are unrelated to the filler, if a trace is posited and

being reactivated. For control, there are also probes at non-gap positions such as [#1]

in (12), and these probes don't show a response time difference between the related-

to-filler ones and the non-related ones, because at a non-gap position no re-activation

occurs due to the absence of the structural trace. Roberts et al. (2007) found the

predicted pattern for the children and adults with high working memory capacity, but

not for the low working memory subjects. In contrast, Felser & Roberts (2007)'s L2

learners didn't differentiate the gap (#2) and non-gap (#1) locations: they took less

time to process the related (identical in this case) probe picture than the unrelated

picture in *both positions*, regardless of their working memory capacities. This pattern,

according to Felser and Roberts, suggests that the filler was kept active throughout

instead of being re-activated at the trace location. Note that these L2 learners are advanced in terms of their proficiency level and that Greek and English share an almost identical structure for indirect object relative clauses like (12), so the L1 and L2 difference observed couldn't have been due to low proficiency nor to an interference from the learners' native languages. Felser and Roberts thus concluded that the learners followed a different parsing routine, that is, using direct association based on verb argument requirements instead of posting a trace in the gap position. Primed by the animals they had most recently seen in the sentence, they picked the same animal quicker.

How L2 learners comprehend complex syntactic structure in real time can also be examined by studying whether they observe so-called island constraints. Island constraints refer to the syntactic phenomenon in which a FG dependency formation is blocked in certain structures such as complex NPs (relative clauses). Consider the following sentences in (13):

(13)   a. The book that the author wrote_? regularly and with great dedication
           about__was named after an explorer.

        b. The book that the author $_{RC}$ [who wrote__? regularly and with great
           dedication] saw_ was named after an explorer.

In (13b), there is a relative clause island boundary between the filler *the book* and the gap site after *saw*. Assuming the Active Filler Strategy (Frazier & Clifton, 1989), the parser will attempt to restore a dislocated item to the nearest gap site. In

(13a), for example, the parser will attempt to place the filler *the book* after the verb

wrote, which is the closest gap location, but will find out later at the true gap position

about that the initial posited gap is wrong. However, such initial gap positing is

blocked for the native speakers (Traxler and Pickering, 1996) in (13b), because of the

island constraint. The native parser becomes aware of the island constraint as it

encounters the RC marker *who*. These kinds of sentences provide an ideal testing

ground for the claims of the SSH, because islands usually appear in long and complex

structures and their computation requires detailed and hierarchical syntactic

representations.

Very few studies on L2 island sensitivity have been carried out; three of them

(Omaki & Schulz, 2011; Cunnings, Batterham, Felser & Clahsen, 2009; Felser et al.,

2012) will be reviewed here. All three studies showed native-like island sensitivity by

the L2 learners. However, the authors interpreted the results differently. Omaki and

Schulz (2011) conducted a reading time experiment with advanced Spanish speakers

of English by adopting the same paradigm as Traxler and Pickering (1996)[5], in which

the learners read the following sample sentences (14):

(14)   a. The book that the author wrote_ unceasingly about _ was named after an
         explorer. (non-island condition; plausible to "write a book")

         b. The city that the author wrote_ unceasingly about _ was named after an
         explorer. (non-island condition; implausible to "write a city")

---

[5] Omaki and Schulz (2011) used slightly modified stimuli from those of Traxler & Pickering (1996).

c. The book that the author _RC_ [who wrote unceasingly] saw_ was named after an explorer. (island condition; plausible to "write a book")

d. The city that the author _RC_ [who wrote unceasingly] saw_ was named after an explorer. (island condition; implausible to "write a city")

They found that both L1 and L2 speakers showed a plausibility effect (delayed RT at the adverb) for the non-island stimuli like (14a) and (14b), such that the implausible condition (14b) was read more slowly than (14a) at the verb *wrote* (the first potential gap position) and at the spill-over adverb *unceasingly* after the verb. Such an effect was not obtained for the island stimuli like (14c) and (14d), indicating that L2 speakers were sensitive to the RC island and avoided positing gaps inside of it. Omaki and Schulz (2011) concluded that their findings were not compatible with the claims of the SSH, for if the L2 learners were aware of island constraints in moment-by-moment processing, their syntactic representations must be hierarchical and have sufficient detail.

Cunnings et al. (2009) tested intermediate-high to advanced-high German and Chinese speakers of English in the same Traxler and Pickering (1996) paradigm but used eye-tracking instead of the self-paced reading method of Omaki & Schulz (2011). Although the L2 group avoided positing gaps in the RC islands, they differed from native speakers in that they showed no reliable evidence for reanalysis indicative of filler integration at the ultimate gap position. In addition, the Chinese speakers took significantly longer in the island condition in general. These findings were taken to support the SSH in that the L2 speakers had difficulty establishing FG dependencies

36

because of the intervening island, hence they are deficient in building complex structures in real-time processing. To explain this L2 sensitivity to RC islands, Cunnings et al. adopted the processing account of island constraints (Kluender, 2004), which claims that island phenomena are not syntactic, but can be explained by the increased referential processing and memory load at island boundaries. They further argued that if *wh*-islands were indeed a grammatical constraint, then they would have found a difference between the German and the Chinese groups. This is because Chinese is an in-situ language with no *wh*-movement (hence no island constraints), and this L1 property should have affected the processing of the target language. The fact that there was no indication of interference from the L1 grammar seemed to be in line with the notion that island effects are due to processing overload and are not grammatical in nature.

Felser et al. (2012) replicated and extended the Cunnings et al. study by adding a second experiment, in which the implausible gaps in the stimuli like (14b) and (14d) above were replaced by an ungrammatical "filled gap". Below is a sample set of their stimuli:

(15)   a. No island, Gap

        Everyone liked the magazine that the hairdresser read _quickly and yet
        extremely thoroughly before going to the beauty salon.

        b. No island, Filled gap

        Everyone liked the magazine that the hairdresser read <u>articles</u> with such strong
        conclusions before going to the beauty salon.

c. Island, Gap

Everyone liked the magazine that the hairdresser $_{RC}$ [who read quickly and yet extremely thoroughly] bought__ before going to the beauty salon.

d. Island, Filled gap

Everyone liked the magazine that the hairdresser $_{RC}$ [who read articles with such strong conclusions] bought__ before going to the beauty salon.

The SSH assumes that two processes are involved in gap filling. One concerns the "semantic goodness-of-fit", following Townsend & Bever (2001), as tested in Traxler and Pickering (1996)'s plausibility paradigm (Experiment I). The other process is about the structural requirements of the gap, meaning whether there is an abstract trace being posited (Experiment II), as tested with the stimuli set in (15). The rationale is that a "filled gap" is a structural violation since there is a conflict between the extra noun phrase and the already posited trace, and this should cause a delay in processing. Felser et al. intended to examine the nature of the dependency formation (semantic or syntactic) by comparing and contrasting the results from both experiments. They found the following eye movement patterns for the L2 speakers. First, they avoided positing a gap in the RC island in both experiments. Again, the processing account was assumed to explain this sensitivity. Second, the L2 speakers showed a semantic plausibility effect, reflected in their first-pass reading indices, *earlier* than the native speakers, whose plausibility effect was only evident in their rereading times in Experiment I. In contrast, in Experiment II the learners had a *delayed* "filled gap" effect, indicated by longer fixation times in the spill-over region

only, whereas the native speakers showed the effect earlier in the offending noun phrase region. Felser et al. conclude that the results of the two experiments in combination suggested that L2 learners are quick and accurate when using plausibility but are not as automatic in their responses to structural mistakes. Thus, their *wh*-dependency formation is based on semantic feature matching rather than structure as in native comprehension.

Felser et al. (2012)'s and Cunnings et al. (2009)'s conclusions directly contradict the interpretation of Omaki & Schultz, who argued that regardless of the nature of the island phenomena, the parser still builds a complex syntactic representation for the island *before* realizing it is either too costly to process (by the processing account), or is simply a structure that doesn't allow gaps (by the grammar account). That L2 speakers can build such a representation argues against the SSH. In fact, if the SSH was correct, the L2 speakers would have used the semantic "goodness-of-fit" of the sentence as their first guide and would have attempted to fill the gap inside of the island when it is semantically possible to do so. Furthermore, Cunnings et al. argued against the grammar account of island constraints based on the fact that the German and Chinese groups were both sensitive to the RC islands and no L1 inference effect was found. The underlying assumption is that Chinese speakers, coming from a *wh*-in-situ language background, shouldn't be sensitive to islands due to L1 grammar transfer. However, such an assumption is questionable. Although Chinese lacks overt movement, there is strong evidence suggesting island constraints are active in the

Chinese grammar (Huang, 1982; Huang, Li & Li, 2009). It is therefore not sound to deny the grammar account based on the absence of a L1 transfer effect.

Nevertheless, the lack of evidence for reanalysis by the L2 learners at the "real" gap position, as found by Cunnings et al., seems to question L2 speakers' ability to process complex filler gap dependencies (those with intervening islands) at a native level. In addition, the different timing of the island effects as shown in the Felser et al. study suggests that the nature of the dependency formation might be different for L1 and L2 groups. Unfortunately, neither fixation times from eye-tracking nor response times from self-paced reading can make clear the syntax vs. semantics distinction. Thus further exploration is called for with different paradigms and methods, such as ERPs.

In summary, the findings to date regarding the SSH are mixed. In addition, in its current formulation the SSH as a L2 processing model is still very vague. It is evident that extensive empirical research with different experimental methodologies is needed for a full evaluation. In particular, it is critical to adopt an experimental paradigm and methodology that can clearly differentiate the use of different information sources. Moreover, the SSH states that shallow parsing is an inherent property of L2 parsing and will not change as a function of other individual factors such as working memory, L1 background, and proficiency level. Yet in all studies cited by the SSH camp, only one addressed the issue of working memory, and none included multiple proficiency levels. Consequently, it is important that future research includes these factors in the evaluation of this proposal. In the studies reported by this

thesis, two individual difference factors, proficiency and working memory, are carefully measured and interpreted with the L2 processing data. In addition, the factor of L1 transfer is also addressed. I will now turn to these factors and examine how they have been found to interact with L2 processing.

2.4   Proficiency in L2 Processing and Acquisition

It has been widely acknowledged that many factors could potentially affect L2 acquisition and processing. Among these, Age of Acquisition (AOA), L1 interference, proficiency level, and working memory capacity have all received considerable attention. AOA, for example, was considered the single most important factor in L2 acquisition and processing in the early days of SLA research. It also led to the development of the Critical Period Hypothesis (CPH), which stipulates a biologically determined time in a human's life (i.e., the onset of puberty) beyond which language acquisition will be increasingly difficult (e.g., Lenneberg, 1967; Penfields & Roberts, 1959). Similarly, a vast amount of L2 research has been devoted to studying the influence of the learner's first language (Gass & Selinker, 2001), because the L1 is an important source of knowledge for both the L2 grammar and processing strategies, at least in the initial stage of acquisition. These studies look for the linguistic differences/similarities between the L1 and L2 to explain the variation in L2 outcomes such as order of acquisition and ultimate attainment state. SLA models such as the Competition Model (MacWhinney, 1989; 2006) have emerged to account for the L1 transfer effect (see 4.2.1 for details). This thesis will also address how L1 influence is

41

implicated in L2 processing when interpreting the experimental results, but the main focus is on the other two well-known SLA variables, proficiency level and individual working memory. I will start with the role of overall proficiency in L2 processing.

2.4.1 Proficiency level in L2 processing

Although almost every L2 experimental study reports on the learners' proficiency levels, and many of them specifically examined its role in L2 processing (see van Hell & Tokowicz (2010) for a review), the effects of proficiency in L2 processing are surprisingly unclear. While some evidence suggests that learners at a high level of proficiency can attain native-like processing patterns, others indicate that L2-specific mechanisms persist even at the most advanced stage of acquisition (see the studies cited by the SSH supporters reviewed above in section 2.3.2). For the L2 processing of filler-gap dependencies, Hopp (2006) examined L2 learners of German (L1 Dutch and English) with subject/object ambiguity structures. It was revealed that only the near-native L2 speakers showed native-like sensitivities; the advanced learners did not. Frenck-Mestre (1997, 2002) investigated prepositional phrase and relative clause attachment preferences among intermediate and advanced L2 English speakers of French and observed that the learners at a high level of proficiency can switch to the target language preference in spite of the L1 transfer effect. The same pattern was replicated in Dekydtspotter, Donaldson, Edmonds, Fultz & Petrush (2008), who tested the same structure with classroom-taught learners who were only moderately proficient. These findings directly contradict the results obtained by Felser

(2003) and Papadopoulou & Felser (2003) reviewed above in 2.3.2. For phrase structure violations, the two studies conducted by Hahne (2001) and Rossi, et al. (2006) again produced different results. While Hahne (2001) found different patterns for native speakers and advanced Russian learners of German, Rossi et al. (2006) obtained a native-like processing profile for intermediate to advanced late learners (L1 English, L2 Italian and French). Similarly, no compelling evidence has been found in the area of morpho-syntactic processing. Ojima et al. (2005) tested Japanese learners of English at high and low proficiency levels and found that highly skilled learners' ERP responses to agreement violations, but not those of low proficiency learners, were qualitatively similar to those of the native speakers. In a longitudinal study, Osterhout, McLaughlin, Pitkänen, Frenck-Mestre & Molinaro (2006) observed a non-native to native-like shift in low to intermediate English learners of French in their responses to agreement violations. On the other hand, Hahne, Mueller & Clahsen (2006) failed to obtain native-like ERP patterns among very proficient Russian learners of German in a study testing participle inflection, which led them to the conclusion that L2 processing must be fundamentally different in morpho-syntax processing as well.

These contradicting results can be contributed to, at least partially, the fact that very few L2 studies directly compare multiple levels of proficiency, and there is not enough empirical data to draw any decisive conclusion. In addition, proficiency could interact with many other factors, such as L1 interference and the structures being investigated, and in turn such factors could further interact with each other. The role of proficiency thus must also be examined in the context of these interactions.

43

Furthermore, the measurement of proficiency has been inconsistent across studies. In many studies, self-evaluation is the only measure (e.g., Jackson & van Hell, 2011; Hahne & Friederici, 2001). Other studies used scores from standardized tests such as the TOFEL and TOEIC, which were taken at the beginning of college/graduate school possibly a few years before the experiment (e.g., Chen, Shu, Liu, zhao & Li, 2007; Ojima et al., 2005). Additionally, as these tests are well known and study resources are abundant, students might have practiced the test too much before taking it. As a result, the scores don't always reflect true proficiency, especially at the advanced or near-native level. Secondly, proficiency measures differ in terms of the specific skills being measured. For example, an "advanced" level designation based on the results of a C-test [6] with no listening or oral component could be different from the "advanced" level determined by the TOFEL, which tests all four skills. More importantly, the test results were not interpreted with a common, established set of foreign language proficiency guidelines such as the American Council of Teaching Foreign Languages (ACTFL) or the Common European Framework of Reference for Languages (CEFR). Thus the results from these studies cannot be interpreted with respect to proficiency. In sum, there is little doubt that proficiency level should be taken into account by any sufficient SLA model or well-designed experimental study. However, the real issue with proficiency is that in order to adequately account for its role, multiple levels need

[6] The C-test is a reading test commonly used for second language learners of English. The test is in the cloze format with no word banks.

to be compared directly in the same paradigm, and its interactions with other SLA factors must also be considered. Furthermore, proficiency needs to be carefully measured with reliable tests and the outcomes should be interpreted with well-established proficiency guidelines.

2.5   Working Memory and Sentence Processing

2.5.1   Working memory and language

Working memory has been a research focus in a wide variety of fields, including cognitive psychology and psycholinguistics. In particular, it has been studied extensively in the context of various language aspects such as performance, processing, and acquisition and has been identified as a major factor affecting these areas (e.g., Miyake & Shah, 1999; Oberauer, Süß, Wilhelm & Wittman, 2003). It has also been related to (L2) language proficiency and aptitude (VanDen Noort, Bosch & Hugdahl, 2006; Robinson 2005) and has been found to correlate highly with measures of general intelligence (IQ) (Colom, Shih, Flores-Mendoza & Quiroga, 2006b). In this section I will first briefly review the conceptualization of working memory and its current models. I then discuss how working memory is implicated in L1 word learning and sentence processing. Lastly, I discuss the relevance of WM for L2 processing and address the pros and cons of a widely used working memory test, the Reading Span Test (RST) by Daneman & Carpenter (1980), as it is the foundation of the WM test used in this thesis.

45

Working memory as a theoretical construct was first put forth by Baddeley and Hitch (1974). In this model the working memory is a site where both storage and processing take place simultaneously. There are four components in Baddeley and Hitch's influential model. The first three are the Phonological Loop, the Visual-spatial Sketchpad, and the Central Executive. The fourth component, the Episodic Buffer, was added later (Baddeley, 2000). Figure 2.1 below illustrates the structure of the model.
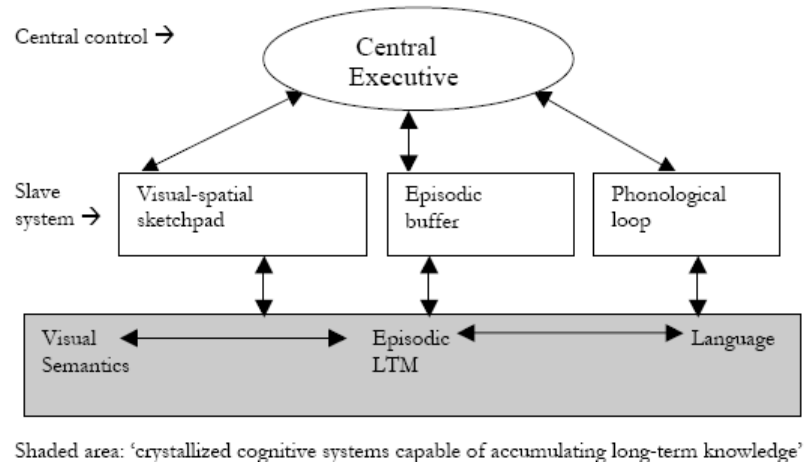
Figure 2.1:   Baddeley's WM model (2000) from Juffs & Harrington (2011)

The Phonological Loop holds and handles audio input, while the Visual-spatial Sketchpad codes visual and spatial information. The Episodic Buffer functions to integrate and store various types of information and coordinates with the long-term memory. The Central Executive is responsible for controlling the above domains and allocating attention to them. Although the current working memory models differ in

the extent to which WM interacts with the long-term memory (e.g., Öztekin &

McElree, 2007; Szmalec, Verbruggen, Vandierendonck & Kemps, 2011), they

converge on the point that the Phonological Memory (PM), sometimes known as the

verbal working memory (VWM), as opposed to the Visual-spatial Memory, is of

central importance to language performance and acquisition.

Baddeley (1998) were among the first to propose that the verbal working

memory *is* a language-learning device. Such a conclusion was supported by a large

amount of evidence demonstrating a strong correlation between VWM and word

learning (e.g., Baddeley, 1998). For example, Bowey (2001) measured the verbal

working memory capacity of children of various ages by using the method of Non-

Word Repetition (NWR), in which the subject is asked to repeat multi-syllable non-

sense words upon reading or hearing them. They found that children with a greater

verbal working memory capacity developed a larger vocabulary. Furthermore, verbal

working memory has been identified as a predictor for language aptitude in general

(Sawyer & Ranta 2001; Robinson 2005).

2.5.2 WM and L1 processing

As discussed before, the overall processing of sentences involves rapid

integration of multiple processes such as individual word recognition, the building of

syntactic structure, and the incorporation of discourse information. It is reasonable to

assume that during this complex task the verbal working memory is needed to store

the lexical items heard/read while computing the syntactic representation (Swets et al.,

2007; Vallar & Baddeley, 1984). This hypothesis is indeed supported by findings

from studies in the past twenty years, which have mostly involved structures that are

complex or need reanalysis (King & Just, 1991; Just & Carpenter, 1992; Just et al.

1996, Hestvik et al., 2012). In addition, it has been observed that the parsing strategy

changes as a function of working memory capacity (McDonalds et al., 1992; Just &

Carpenter, 1992), and working memory interacts with the type of structure being

processed (Just, Carpenter, Keller, Eddy & Thulborn,1996; Miyake et al., 2000).

However, when the target structure is not as complex, the WMC doesn't seem to

affect sentence processing, at least not in the way it affects word learning. Waters &

Caplan (1996 a, b, c) observed that certain individuals with brain damage who have

close to zero verbal working memory were still able to process syntax. They also

failed to replicate a few studies (e.g., McDonald et al. 1992; King & Just 1991) that

claimed to find a correlation between WMC and reading time when more thorough

statistical analyses were conducted. Based on these results, Caplan and Waters (1999)

proposed a dedicated resource account for WM in which it has two sub-components.

One handles meaning extraction through word recognition, structure building,

thematic role assignment, and semantic/prosodic representation building. The other is

dedicated to post-interpretation tasks such as reasoning and connection to long-term

memory. The first component, which is where sentence parsing happens, cannot be

measured by the traditional methods (including the reading span task developed by

Daneman & Carpenter, 1980, which will be discussed in detail later) used to measure

verbal working memory. Water and Caplan's claims triggered an ongoing debate in

the field of sentence parsing and working memory research (see Lauro, Reis, Cohen, Cecchetto & Papagno, 2010, for a recent installment), and no clear conclusion has been reached yet. Nevertheless, it is agreed upon that WM still plays a significant role in accounting for individual differences in parsing sentences with complex syntax, such as those involving long-distance dependencies.

To accurately measure working memory capacity, it is important that the appropriate WM test is chosen. In the rest of this section I will review a widely used working memory test for sentence processing, the Reading Span Test (RST) developed by Daneman & Carpenter (1980). A L2 version of this test (Harrington, 1992) is used in Experiment I. To justify this choice, the advantages and limitations of the original Daneman & Carpenter RST need to be discussed carefully. Following the WM model proposed by Baddeley et al. (2000) above, the RST measures *both* working memory's storage and processing abilities simultaneously by asking the respondent to both read sentences out loud and remember the last word of the sentence. Typically, the respondent reads the sentences in increasingly larger sets, e.g., the beginning level contains sets of two sentences, and the last level contains six-sentence sets. In each level there are five sets. At the end of each set, the respondent is asked to recall the last words. A modified version of the original RST also asks the respondent to answer true/false questions based on the content of the sentences to ensure that they are processing the sentences while storing the last words. Later, a listening version of the RST Test (Desmette, Hupet, Schelstraete & Van der Linden, 1995) and a few other variants of the original RST were also developed (e.g., Kondo

49

& Osaka, 2004; LaPointe & Engle, 1990). It was discovered that the RST correlates with other established reading performance measures such as the reading part of the Scholastic Aptitude Test (SAT) much better than traditional WM tests like digit span and word span tests (e.g., Non-Word Repetition Span Test) (Daneman & Merikle, 1996). Additional studies that adopted or reviewed the RST indicated its high correlation with various assessments of language abilities in addition to reading, such as pronoun interpretation and verbal fluency (e.g., Carpenter & Just, 1989; Daneman, 1991; Just & Carpenter, 1992). The RST was found to be a predictor for fluid intelligence as well (e.g., Ackerman, Beier & Boyle, 2005; Engle, Tuholski, Laughlin & Conway, 1999). However, the RST and its variants are not without limitations, as its administration and scoring are sometimes problematic (e.g., Waters & Caplan 1996b). For instance, absolute span scoring was used in the earlier days of the RST and was found to be not differentiating enough in the following years. As mentioned before, the RST divides the stimuli sentences into 5 levels, each with 5 sets of sentences. The number of sentences in the sets at each level increases, e.g., level 2 (there is no level 1) has two sentences in a set, level 3 has three sentences in a set, and so on. An absolute span score assigns the respondent 1point when he or she recalls correctly all sentence-final words in a set, for at least 2 out of the 5 sets in the level. However, the test stopped if the respondent failed on 3 out of the 5 sets for any given level. This scoring method produces a very narrow range of score values, usually from

50

2-6, and is often not sensitive enough to yield significant results in correlation analyses.[7] In addition, the test can terminate too early for a respondent due to non-WMC reasons. For example, a subject can be stopped at level 3 when he or she fails on 2 sets, but maybe the reason for those failures was that s/he was temporarily distracted by one set of stimuli. Such an individual could very well pass level 4. In addition to the scoring issue, the administering procedure can be inconsistent across studies. In the early days of the RST, the participants read sentences written on index cards that were given to them manually by the experimenter. The recall response time or whether the reader was rehearsing before recall was not always monitored. It is therefore necessary that the RST task is given on a computer, supervised by an experienced experimenter in a maximally controlled lab environment.

Despite its limitations, the Daneman and Carpenter RST (or its listening version) remains the most appropriate measure of working memory capacity for research in sentence processing (Conway et al., 2005; Juffs & Harrington, 2011). As mentioned above, the working memory test adopted by Experiment I is a L2 version of the Daneman and Carpenter test (Harrington & Sawyer, 1992). The scoring and delivery of the test have been improved to address the limitations of the original RST test as described above. The details of the Harrington & Sawyer (1992) working memory test will be discussed in the following section.

---

[7] An alternative way of scoring the RST was developed which calculates the total number of words recalled correctly.

2.5.3   WM and L2 processing

The past two decades have produced a growing body of work investigating the role of WM in accounting for the outcomes of L2 studies. WM effects were tested for various aspects of L2 acquisition, including word and grammar learning, sentence processing, and general L2 aptitude (for a detailed review, see Juffs & Harrington, 2011). Some of these studies have shown, for example, that verbal working memory is critical (e.g., Baddeley,1998; Papagno Valentine & Baddeley, 1991) in L2 vocabulary learning. WM also predicts more successful grammar learning (e.g., Service 1992; Ellis & Sinclair, 1996) and has been found to directly correlate with L2 reading comprehension (e.g., Harrington & Sawyer, 1992; Leeser, 2007), speaking ability (FortKamp, 1999), and writing skills (Kormos & Sáfár, 2008). Furthermore, a number of experiments concluded that a larger WMC promotes a higher overall L2 proficiency (e.g., French & O'Brien, 2008; O'Brien, Segalowitz & Freed 2006; O'Brein, Chiaravalloti, Goverover & DeLuca, 2008). Notably, these results converge in that L2 processing is more resource demanding. In particular, the low-level lexical processes are much more costly in L2 parsing and can take limited resources away from sentence-level processing (Fender, 2001).

In the area of L2 processing, a number of studies have also demonstrated that WMC may influence morpho-syntatic processing such as subject-verb agreement (e.g., Reichle, Tremblay & Coughlin, 2013; Dowens et al. 2010; 2011). However, relatively few studies have been devoted to WM and the L2 processing of complex syntactic structures with highly abstract elements, and the limited results generated are not

without contradictions. In three self-paced reading (SPR) studies conducted by Juffs (2004, 2005, 2006), Chinese, Japanese, and Spanish L2 learners of English read sentences that are either structurally complex (long-distance *wh*-movement) or temporarily ambiguous due to reduced relative clauses or a Garden Path effect. The participants all took the Daneman and Carpenter (1980) test for working memory capacity. While WM was reliably found to affect the processing of these sentence types of L1 speakers, it didn't correlate with L2 reading times at critical regions in all three of Juffs' studies. In the Felser and Roberts (2007) study reviewed above on how Greek L2 speakers of English process complex relative clauses, no interaction between the subjects' reading span scores and their response times in different experimental conditions was obtained. Again with the method of SPR, Rodriguez (2008) tested the processing of various complex structures. He administered (to Spanish subjects) a Spanish reading span task based on D&C (1980). [8] However, no WM effect was found for the reading times obtained at critical regions of three target structures.

On the other hand, some evidence has been found that links WMC to the processing of certain resource-taxing structures. A Self-Paced-Reading study conducted by Dussias & Piñar (2010), in which stimuli with subject/object *wh*-movement asymmetries were presented to Chinese L2 speakers of English. Plausibility

---

[8] Due to the lack of a Chinese L1 WM test, Rodriguez (2008) didn't test the Chinese speakers of English in the study for WMC.

was also manipulated. They found that only the high span learners, as identified by the Waters & Caplan (1996b) reading span test, were able to show reading time subject/object and plausibility differences. Sagarra & Herschensohn (2010) studied how English L2 speakers of Spanish process agreement errors in a SPR experiment. It was found that WMC, measured by the Waters & Caplan (1996b) test, affected only the off-line accuracy scores among high-intermediate learners but not the lower proficiency groups. There was no relation between WMC and the online measures across proficiency groups.

The absence of compelling evidence is partially due to the different WMC measures adopted in the above studies. Because of the scoring and the procedure problems discussed above in section 3.2.3, it is sometimes difficult to ensure accuracy in measurement and to interpret the collected WMC data. For instance, the above reviewed Juffs (2004, 2005, 2006) studies used the RST (1980), but the test was delivered manually and the scores were calculated with the absolute span method. This might have led to their failure to relate WMC to reading time, since the WMC scores obtained were too narrow in range. Other methods such as the Waters & Caplan (1996b) test seem to be more successful in establishing a relation between WMC and reading times, as shown in Dussias & Piñar (2010) above. However, Sagarra & Herschensohn (2010) adopted the same WM measurement but didn't find the anticipated WM effect on online processing.

It is important to stress that the primary concern when measuring L2 WM is the issue of language confound, especially when the learners are not advanced

speakers of the target language. Given that reliable WM measures are often not available in the subjects' first languages, a sensible option is to adopt a simplified version of a reliable L1 WM test. Very few such tests are available, but one was developed by Harrington and Sawyer (1992). This test uses fewer, shorter, and less complex sentences as well as a simpler vocabulary than the Daneman and Carpenter tests. Moreover, the secondary task in the Daneman & Carpenter variants is to ask the respondent questions about the content of the presented sentence, which is about random facts in any subject from European cuisine to Biochemistry. Many sentences are hard to judge for those L2 learners who are not familiar with the "common knowledge" of western culture, hence these questions are too distracting or discouraging for them to perform well. In contrast, in Harrington & Sawyer (1992)'s test, half of the sentences presented are ungrammatical due to a reversed word order between the middle and end part of the sentence. The subjects are asked to judge whether the sentence is acceptable or not (Turner & Engle, 1989). Such a secondary task incorporates the processing component by focusing on syntax rather than world knowledge, making it thus more suitable to L2 learners. The Harrington & Sawyer (1992) test was found to correlate significantly with the reading and grammar sections of the TOFEL test. A listening version of it was used in other studies (e.g., Martin & Ellis, 2012) in which correlations between WMC and vocabulary and grammar learning were found.

In conclusion, the research on how WM affects L2 online sentence comprehension is still in its beginning stage. Given its theoretical importance,

however, it is necessary to expand the currently limited empirical work, especially with methodologies other than Self-Paced-Reading, such as eye-tracking ERP and fMRI. In addition, reliable WM tests that are suitable for L2 learners must be administered carefully in order to ensure maximum measurement accuracy.

2.6   Summary

In this chapter I presented several key processing models and the related experimental findings for the L1 and L2 online processing of complex syntax, focusing on filler-gap dependencies. For L1 processing, I first reviewed the debate over how syntactic and other non-structural information is utilized in online processing. Although the issue is not completely resolved, a few accounts have emerged as the theoretical basis for the SSH. Next, the empirical findings regarding whether a trace is posited in filler-gap resolution were discussed. These findings largely suggest that L1 speakers do posit such an abstract element in the syntactic representations they construct online. In addition, a proposal suggesting shallow syntactic processing among native speakers, the Good Enough Representations (Ferreira et al., 2002; Ferreira & Patson, 2007), was presented and discussed in relation to the SSH. Against this background, I reviewed a few studies on L2 processing and examined the claims of the SSH in detail, along with the evidence for and against these claims. In the third part of this chapter, I turned to a few factors that could potentially affect L2 processing and discussed two of these, proficiency level and working memory capacity, in detail. In particular, I addressed the need for them to

be accurately measured and interpreted in the L2 context. I conclude that the L2

processing findings to date are far from conclusive regarding whether there is a

principled difference between L1 and L2 parsing and that much research is still

needed. Given that the research questions at hand concern the nature of the underlying

mechanism in L2 parsing and whether it is guided mostly by non-structural

information, it is critical to adopt different methodologies that can clearly distinguish

between the use of syntactic vs lexical/semantic information. The Event Related

Potential methodology meets this requirement. In chapter 3 I will report on an ERP

experiment on how Chinese learners of English process FG dependencies, specifically,

whether they posit a trace in the online syntactic representation.

Chapter 3

L2 FILLER-GAP PROCESSING

3.1   Introduction

In the previous chapter I considered a few issues of the L2 online processing of complex sentences and discussed the Shallow Structure Hypothesis (Clahsen & Felser, 2006a,b), particularly, its claim that late L2 learners process filler-gap dependencies with a fundamentally different mechanism from the L1 processor in real time comprehension. In this chapter, I report on an Event-Related Potential (ERP) study examining how Chinese L2 speakers of English react to "Filled Gap" violations such as *The camel that the zebra kissed <u>the giraffe</u> (*extra noun phrase*) ran far away*. The L2 brain responses are compared to those of L1 speakers from Hestvik et al. (2012). The current experiment sought to test the SSH's claims and to contribute to the understanding of L2 real-time sentence parsing. Specifically, the current study attempts to answer the following research questions:

(1)   Is there a fundamental difference between L1 and L2 parsing algorithms for the real-time processing of complex FG dependencies?

(2)   How does the L2 online sentence parsing utilize various information sources? Is the L2 parser capable of building sophisticated syntactic representations with sufficient details as the native speakers do?

58

(3)   How do individual proficiency level and working memory capacity

affect the L2 online parsing of complex FG dependencies?

As mentioned in previous chapters, a more in-depth understanding of the

nature of L2 online syntactic processing necessitates the use of a variety of

methodologies, specifically, those that clearly indicate the use of distinct information

sources. I start by explaining how the methodology of ERP and its relevant

components can serve that purpose and is hence an ideal method for the research

questions at hand. Next, I describe the design of the current experiment and lay out the

predictions of the SSH in terms of ERP indexes. Before moving on to the details of the

experiment, I address the issue of a potential L1 transfer effect. In what follows I

describe the findings of the study, which showed a non-native-like brain response

pattern for the L2 participants. I argue in the discussion section that the data obtained

is best explained by the absence of syntactic traces in the L2 syntactic representation

that is built online, and that the L2 parsing strategy is mainly semantic and meaning-

based. I thus conclude that the results from Experiment I lend support to the claims of

the SSH.

3.2   ERP and Language Comprehension

The Event-Related Potential (ERP) method measures online brain responses to

perceptual and cognitive processes in the form of electroencephalogram (EEG)

signals, recorded at certain time intervals by electrodes attached to the scalp. These

EEG signals represent voltage changes in electrical brain activity that are time-locked

to the target event. In the context of language comprehension, these events can be the presentation of a word or a syntactic violation in a sentence. ERP components, which are a series of positive and negative signals, are derived via signal filtering and averaging of the multiple EEG segments elicited by the events (van Hell & Tokowicz, 2010, Schremm, 2012). The timing, scalp location, and polarity (positive or negative in terms of voltage shift) of the ERP components are used to index functionally different neurocognitive processes (Frisch, Hahne & Friederici, 2004). Compared to behavioral measures such as Self-Paced-Reading, ERP is more sensitive to highly automatic and sometimes subconscious processes due to its excellent temporal resolution (Luck, 2005).

Because ERP components are usually elicited by illicit language forms relative to the grammatical forms, most ERP studies on language processing adopt a violation paradigm, in which ERPs are time-locked to language stimuli with target syntactic, morpho-syntactic, or semantic violations. The ERP method is excellent for psycholinguistic studies, because in addition to its high timing sensitivity, it is particularly helpful when we need to know if a process is syntactic or semantic in nature.[9] ERP is therefore highly suitable as a methodology for the studies in this thesis, as the above mentioned SSH claim draws a crucial lexical/semantics and structural/syntax distinction regarding the nature of L2 processing.

---

[9] Given conventional assumptions about the nature of the various ERPs and what they index.

Of particular interest here are four ERP components, the first of which is the Early Left Anterior Negativity (ELAN). Often found in the frontal region of the scalp and more pronounced on the left hemisphere, the ELAN is a negative-going potential that occurs in the 150-250 millisecond (ms) range after the target language stimulus. It is related to phrase structure violations as in *The scientist criticized Max's proof of the theorem* (Neville et al., 1991, p.156) and has been obtained for multiple languages (e.g., Friederici, Pfeifer & Hahne, 1993; Isel et al., 2007; Lau, Stroud, Plesch & Phillips, 2006; Neville, Nicol, Barss, Forster & Garrett, 1991). The ELAN is assumed to index structure building that occurs extremely early (first-pass processing), quickly, and highly automatically, and is hence most difficult to observe in L2 settings (van Hell & Tokowicz, 2010; Kotz, 2009). In addition, the ELAN is known to be affected by predictions concerning the word category of an upcoming item (Lau et al., 2006).

Related to the ELAN is the Left Anterior Negativity (LAN), which is the same as the ELAN in terms of polarity but occurs later, between 300-500 ms after the onset of a target violation. It is also obtained in the anterior position, commonly on the left side, but sometimes bi-laterally or even on the right side more than on the left. In addition, the LAN distribution is wider than that of the ELAN, as the latter is more focused on the lower, left frontal region of the scalp (Friederici, Hahne & Mecklinger, 1996; Kluender & Kutas, 1993). The LAN effect is related to morpho-syntactic rule violations and has been found to indicate problems with morphological processing (Friederici, 2002; Gross, Say, Kleinger, Clahen & Münte,1998; Penke et al., 1997; Rodriguez-Fornells, Clahsen, Lieo, Zaake & Münte, 2001; Weyerts, Penke, Dohrn,

Clahsen & Munte, 1997). A detailed review of these violations will be presented in Chapter 4. In addition, previous studies have found that the LAN is evoked when there is an increased working memory load in sentences with long-distance dependencies (details will follow in section 3.3).

Another syntax-related component is the P600, a positive-going voltage wave obtained between 500-600 ms and 800-900 ms post-onset of the stimulus in the parietal region of the scalp. The P600 is often observed for various syntactic anomalies including phrase structure violations and morpho-syntactic violations (e.g., Hagoort et al., 1993). Complex syntactic structures such as filler-gap dependencies and "reanalysis" triggered by Garden Path sentences can also elicit a P600 (Hagoort et al., 1993; Friederici, 2002; Kaan, Harris, Gibson & Holcomb, 2000; Osterhout & Holcomb, 1992, 1993). It is also detected with subcategorizing verbs in well-formed, long-distance *wh*-movement (Kaan et al.,2000; Phillips, Kazanina & Ababa, 2005), reflecting syntactic integration difficulties (Kaan et al., 2000; Fiebach et al., 2002; Felser et al., 2003; Ueno & Kluender, 2003; Phillips et al., 2005).

Lastly, the fourth ERP component relevant here is the N400. It is a central-parietal negative-going voltage shift typically associated with semantic and pragmatic processing, which usually starts at 250-500 ms and peaks at 400 ms after the offending word (Kutas & Federmeier, 2000; Van Petten, Coulson, Rubin, Plante & Parks, 1999). The N400 indexes the integration of meaning and world knowledge (e.g., Kutas & Hillyard, 1980a,b,c; Hagoort, Harld, Bastiaansen & Petersson, 2004) and has been obtained with various manipulations such as word frequency, vocabulary class, and

62

predictions based on context (e.g., Schremm, 2012). The N400 component is typically observed with semantic incongruities and violations associated with verb argumentation structures (e.g., Kutas & Hillyard, 1980 a, b, c; Frisch et al., 2004)

3.2.1 ERP and L2 sentence processing

ERP experiments in the field of L2 syntactic and morpho-syntactic processing fall largely into two categories. A sizeable number of ERP studies focus on agreement and occasionally inflectional morphology (see Chapter 4.2 for details). Another cluster of ERP studies examine how L2 learners process phrase structure violations (see Van Hell & Tokowicz, 2010 for a review). For native speakers, phrase structure violations typically elicit the ELAN component, sometimes followed by the P600 (e.g., Friederici et al., 1993; Isel et al., 2007; Lau et al., 2006; Neville et al., 1991). However, the L2 findings are much less consistent. Isel (2007) tested proficient German speakers of French and Rossi et al. (2006) tested proficient German learners of Italian and Italian learners of German. Both studies obtained the ELAN and the latter the P600 as well. In contrast, Hahne (2001) ran proficient Russian learners of German but found only the P600. Hahn and Friederici (2001) tested moderately proficient Japanese speakers of German but found hardly any component for syntactic violations. Additionally, Weber-Fox and Neville (1996) tested adult Chinese learners of English whose ages of acquisition ranged from 1 year old to older than 16 and observed a P600 or a delayed P600 for subjects whose AOA was before 13. However, they found no ELAN across all groups. At first glance, L1 interference and proficiency

might appear to be potential factors in accounting for these inconsistent results, because in both cases where the ELAN was completely absent the learners had different L1 backgrounds (Chinese and Japanese) from the target language (English and German, respectively), and they were only moderately proficient. However, L1 interference and proficiency fail to explain why the very proficient Russian speakers of German couldn't generate the ELAN in Hahne (2001), yet even the moderately proficient speakers coming from different L1 backgrounds in Rossi et al. (2006) produced an ELAN + P600. In short, existing ERP data are unequivocal in their findings, especially for syntactic indexes such as ELAN/LAN.

In comparison to the syntactic violations, L2 speakers seem to generate the N400 in response to semantic-related anomalies in a more comparable way to native speakers (e.g., Friederici, 2011; Ojima et al., 2005), although the timing and amplitude of the N400 might show some discrepancies between the language groups. For example, Hahne (2001) observed that the L2 N400 had a smaller amplitude and a longer latency.

In regard to the SSH, it seems that some of the brain response patterns reviewed above appear to support the claims of the SSH, while others do not. More explicitly, the N400 indicative of sensitivity to semantic violations has been found among L2 learners, similar to native speakers, while the ELAN, which indexes highly automatic and "first-pass" structure building, is missing in the majority of ERP L2 studies. In the middle is the P600 component. It is more reliably observed among L2 learners than the ELAN for phrase structure violations. Such a discrepancy between

the P600 and ELAN is consistent with the SSH, as the P600 is usually related to more controlled, "second-pass" syntactic processing, and thus it is more achievable by L2 learners than the highly automatic structure building indexed by the ELAN. On the other hand, the results reviewed so far clearly suggest that AOA, proficiency level, and L1 background seem to affect the L2 processing outcomes to various extents. This pattern contradicts the view of the SSH, which stipulates minimal roles for these variables in L2 parsing. The findings to date are thus not conclusive with regard to the proposals of the SSH. In the next section, I will focus on the current ERP research on filler gap dependencies, a structure of question in terms of the SSH. I then will review two studies carried out by Hestvik et al. (2007, 2012) and describe how they provide an excellent paradigm for further testing the claims of the SSH.

3.2.2 ERP correlates in FG dependencies

A number of ERP studies have investigated FG dependencies in various structures such as *wh*-questions (e.g., Fiebach et al., 2002; Phillips et al. 2005), relative clauses (e.g., King & Kutas, 1995; Ueno & Garnsey, 2008), and scrambling (Ueno & Kulender, 2003). These studies mostly focused on the processing costs of FG dependencies, and they have identified the following two ERP correlates: an anterior negativity potential (L)AN and/or a Sustained Anterior Negativity (SAN) at the filler and the P600 component at the Gap site.

The negative-going potential has been observed at the filler or immediately after it (within 200-400 milliseconds) in various studies (King & Kutas, 1995; Phillips,

2005). Sometimes it is found to continue up to the point where the gap is (e.g.,

Fiebach et al., 2002; Phillips et al., 2005), hence the name Sustained Anterior

Negativity (SAN). Such a component is more commonly obtained in long-distance FG

dependencies than in the shorter ones, and as Fiebach et al. (2002) found, the

amplitude of the component increased as the distance between the filler and the gap

lengthened. It was assumed that this component indicated the additional costs of

keeping the filler in working memory (e.g., King & Kutas, 1995), although it is not

always clear what specific features of the filler are being maintained (e.g., Fiebach et

al., 2002; Wagers & Phillips, 2009). A few studies also examined how the (S)AN

interacted with subjects' working memory capacities, and the results showed that WM

capacities affected the distribution and latency of the (S)AN (Fiebach et al. 2002; Vos,

Gunter, Kolk, Mulder, 2001), suggesting that the (S)AN indeed indexes WM cost in

FG dependency resolution.

The P600 component was elicited at the gap site in several studies

investigating grammatical long-distance dependencies in comparison to the condition

with no gaps. (e.g., Gouvea, Phillips, Kazanina & Poeppel, 2010; Fiebach et al., 2002;

Phillips et al., 2005). It was proposed that this P600 indicated an integration cost

instead of the reanalysis cost commonly found for Garden path sentences (Kaan et al.,

2000). Such a P600 is not modulated by WM capacities (e.g., Fiebach et al., 2000)

Very few ERP studies directly test the trace-reactivation vs. direct association

accounts of FG dependencies mentioned above regarding the nature of the relation

between the filler and its subcategorizing verb. Hestvik et al. (2007) investigated this

issue by presenting the following auditory test sentences:

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

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

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

Inserting an extra noun after the subcategorizing verb *kissed* fills the assumed

gap site. When the parser attempts to fill the gap with the NP *the zebra*,

ungrammaticality is recognized and the corresponding ERP component should result.

If the direct association hypothesis is correct, and the relation between the verb and the

filler is formed based on argument structure, thematic role assignment or other

semantic information, then the violation should generate a N400 at the filled gap site

in comparison to the control condition. However, no N400 was found by Hestvik et al.

(2007). Instead, an ELAN was found at the offending extra NP. Based on the

neurophysiological time course model for syntactic processing proposed by Friederici

and her colleagues (Feriederici 1995; Freiderici et al., 1996), Hestvik et al. (2007)

argued that multiple ERP components could result for any grammatical violation, but

the timing in which they are detected is crucial for the interpretation. Recall from

previous discussion that the ELAN is typically observed only 100-200 ms after the

offending item, for problems in first-pass, highly automatic structure building such as

phrase structure violations. At this early stage, the parser only attends to the minimal

structural information such as word category specifications. Evaluation of the verb

argument structure and semantic fit would take place in the second stage, in which the

related violations would elicit a LAN and/or a N400. In the last stage, the parser

reanalyzes (if necessary) and conducts the final consolidation of information

(integration), so the P600 usually appears in this stage. The building of a trace should

occur in the earliest stage as part of phrase structure building, as it is an identical copy

of the moved element that is not phonetically realized (Chomsky, 1986, 2002). In the

set-up of Hestvik et al. (2007), the parser posits a trace as soon as the verb *kissed* is

processed, filling in the direct object position projected. It then expects the next item

to be of a different word category than NP. However, as it moves on to the extra NP

the camel, a violation of expectation for word category occurs and the ELAN results.

If the violation occurs in the later stage, when the verb's arguments and thematic role

assignment are evaluated, then the N400 should appear. In addition, studies

investigating ERP correlates with combined syntactic/semantic violations showed that

problems in the early stage such as word category violations often block further

syntactic processing. Consequently, in cases with both phrase structure violations and

semantic anomalies, only the ELAN will be obtained, and not the N400 (Friederici et

al., 1996). In sum, Hestvik et al. (2007)'s results are consistent with the trace-

reactivation account but not the direct association model of FG dependency

processing.

Based on Hestvik et al. (2007), Hestvik et al. (2012) explored the relation between gap-filling and working memory capacity. In addition, a change was made to the experimental materials. In the earlier study the ungrammatical sentence was compared to two control conditions: the first is the grammatical object condition as shown in (17b), repeated from (16b), and the second is the grammatical sentence with a trace, as illustrated below in (17c), repeated from (16c):

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

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

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

There is a potential complication in the ERP components when comparing (17a) and (17b), as Hestvik et al. (2007) pointed out. Based on the previous findings reviewed in section 3.1, a SAN might result as an indication of the additional costs of filler-keeping in WM in a FG dependency resolution when compared to a non-gap filling construction like (17b). If a LAN or AN is obtained at the critical verb, it is unclear whether the component was elicited by the ungrammaticality or by the increased WM costs. Therefore, a second control (17c) (trace condition) was added to address that issue. In addition, Hestvik et al. (2007) found negativity when the ungrammatical condition (17a) was compared to both control conditions (17b) and (17c). However, there was an issue with the trace condition (17c) as well. Since the

69

items after the critical verb *kissed* in (17c) are not the same (one is a NP and the other a PP) in these two sentences, the effect (ELAN) could have been due to the word category difference between these two phrases. In Hestvik et al. (2012), the ungrammatical sentence was compared to an ADJUNCT condition, in which the item after the critical verb is identical to the one in the ungrammatical condition, and the words leading up to the verb are maximally matched as well. Consider the following examples in (18a) and (18b):

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

b. The day that the hippo kissed * the camel on the nose it was humid.

Hestvik et al. (2012) found an early bilateral negativity (functionally interpreted as equivalent to early left anterior negativity or the ELAN) for the high working memory group at 100-200 ms after the onset of the offending NP, which continued as an AN (bilateral version of the LAN) starting at 200 ms and peaking at 500 ms. For the low WM span group, no early AN was obtained, but an AN starting at 400 ms was detected. In addition, both the high and the low WM group generated a P600 component indicative of integration difficulties, although the P600 for the high WM group started 200 ms earlier than the low WM group. In sum, Hestvik et al.

(2012) argued that the low WM group had a delayed gap filling, suggesting that their

parsing is less automatic than the high WM group.[10]

3.3   Experiment I

3.3.1   Rational and predications

Hestvik et al. (2007, 2012) thus provide an ideal paradigm for testing the

claims of the SSH. Recall that the SSH maintains that L2 learners under-use syntactic

cues and rely exclusively on semantics and pragmatics to build a shallower structure in

the online processing of syntax. More specifically for the processing of FG

dependencies, the learners can only resort to verb argument, thematic role assignment,

and other semantic information to form the dependency, without positing details like

an abstract trace. If this line of thinking were followed, then in the Hestvik et al.

paradigm the L2 speakers should produce a N400, indicating a failed attempt to

integrate an extra argument, instead of the early or delayed anterior negativity that

would indicates a phrase structure and word category conflict as a result of the abstract

gap positing. Alternatively, if an anterior negativity or comparable onset and

distribution were obtained, then it would suggest that the L2 processer is capable of

---

[10] The AN for both the high and the low WM group could indicate the working memory cost as mentioned. However, given that the P600 of the low WM group was also later at around 200 ms, it is conceivable to posit that the AN obtained for the low WM group was not just an indication of additional WM cost, but of a delayed phrase structure violation.

building syntactic representations with structural details like a gap in a similar way as the native speakers.

Assuming this rationale, Schremm (2012) carried out an ERP study based on Hestvik et al. (2007) to assess the SSH by examining how advanced Swedish speakers of English react to the "filled gap" violation in online comprehension. In addition to the 3 conditions used in Hestvik et al. (2007), three more conditions with semantic anomalies were included to see if the participants generated a N400 as expected. The complete set of the experimental sentences are listed below:

(19)   a. Filled Gap
The receptionist that the painter scared the reporter by accident answered the phone.

b. Grammatical gap
The receptionist that the painter scared by accident answered the phone

c. Grammatical object
The receptionist said that the painter scared the reporter by accident and   then answered the phone.

d. Semantic anomaly
The receptionist said that the painter scared the document by accident and then answered the phone.

e. Correct subject-relative
The receptionist that scared the painter by accident answered the phone.

f. Semantically anomalous subject-relative
The receptionist that scared the freezer by accident answered the phone.

Schremm didn't find any ELAN, LAN, nor N400 at the extra NP location for the ungrammatical condition in relation to the two control conditions (19b) and (19c).

72

Only a delayed P600 was obtained, and it was taken as evidence that the L2 learners initially attempted to analyze the extra NP as the direct object of the verb, but then later revised this analysis. Schremm argued that this finding goes in part against the SSH in that the L2 learners engaged in some syntactic analysis, albeit at a later stage, as evidenced by the late P600.

However, a few concerns can be raised about her study. First, only 14 subjects were tested, a number that might suggest inadequate statistical power in the data analysis. It was also reported that among these subjects, 3 of them produced the ELAN, 4 subjects produced the N400, and the rest generated the P600. Such high variability among subjects questions the reliability of the data interpretation. The second issue is that of the material. Schremm's study was a replication of Hesvik et al. (2007), in which the control conditions were seen as problematic (see section 3.2.2 above) and were replaced in a better version of the experiment which produced similar results in the more recent Hestvik et al. (2012). Thirdly, the P600 obtained and illustrated in the paper was an average of a few isolated electrodes, instead of a group of continuous electrodes as is typical for P600 sites. Finally, while a N400 was obtained for the semantically anomalous subject relative in comparison to the correct subject relative condition, Schremm failed to produce the desired N400 for the semantic anomaly condition (12d) as compared to the grammatical object condition (19c). Note that the two conditions are identical, except that the NP follows the verb in the latter.

The present study, a L2 replication of Hestvik et al. (2012), corrected these problems with Schremm (2012). First, the control condition is the adjunct condition following Hestvik et al. (2012). A total of 53 subjects were run in order to ensure the results could be interpreted with sufficient statistical power for the between-subject variables of proficiency and working memory capacity. In addition, since it is important that the factors known to affect L2 acquisition and processing, such as proficiency and working memory capacity, be taken into account, the present study also tests the subjects' WM listening span and includes subjects at three levels of proficiency. Critically, I examined the ERP elicited at the extra NP *the camel* in sentences such as *The zebra that the hippo kissed the camel on the nose* in relation to its control *The weekend the hippo kissed *the camel on the nose.* The precise predictions of the SSH in terms of ERP indexes are summarized in Table 3.1:

Table 3.1: Possible outcomes for the L2 learners in experiment I and the SSH predictions

ERP at *the camel*	Component I	Component II	Implication
Outcome 1	ELAN	With or without P600	Contradicts the SSH
Outcome 2	LAN	With or without P600	Contradicts the SSH
Outcome 3		P600 only	Partially contradicts the SSH
Outcome 4	N400	With or without P600	Supports the SSH

If outcome 1 is obtained, then the L2 learners are essentially the same as the native speakers, even if the P600 is absent. This is because in Hestvik et al. (2007), the P600 was not found for the native speakers either. Outcome 2 for the L2 learners would indicate that they may be similar to those native speakers whose working memory capacity is lower, as found by Hestvik et al. (2012). Given that L2 processing is memory taxing to start with, such a result is quite expected for the L2 speakers. These two outcomes would argue against the claims of the SSH. If only a P600 is obtained, then the SSH is partially correct in that there is a significant difference between L1 and L2 parsing, because at least one component is completely missing. However, while such a result indicates that the trace is absent, it doesn't necessary suggest that semantics are used exclusively and that L2 learners can't make effective use of some "second-pass" syntactic rules. The last outcome (N400 with or without P600) would provide direct evidence in support of the SSH in that the violation is recognized as semantic in nature, suggesting the absence of a trace. A syntactic trace would have caused a structural conflict indexed by a (E)LAN type of component, which would have blocked the subsequent N400. A P600 following the N400 means that some reanalysis occurred after the initial semantics-guided parsing, but it cannot be used to argue that the L2 parser is similar in a qualitative way to the native parser.

In this study, the ERP outcomes are evaluated in correlation with the L2 speakers' individual characteristics such as proficiency and WM capacity. According to the SSH, L2 processing algorithms are distinct from those of L1 in that variables such as proficiency, L1 interference, and working memory don't change the semantic-

dominant properties of L2 parsing. If the SSH were correct, then neither individual variable (proficiency or working memory) should show a significant correlation with the ERP measures. Contrarily, if L1 and L2 processing only differ quantitatively, then we should expect to see these variables interact with the potential outcomes described above. Specifically, we should see more native-like ERPs as proficiency level and WM capacity increase.

3.3.2   Participants

A total of 56 subjects (37 female and 19 male) participated in Experiment I. One subject was excluded from all data analyses because of data collection errors. Two more subjects were excluded from the ERP data analyses for high bad trial percentages (see section 3.6). The majority of the participants were students[11] recruited from the University of Delaware. The average age is 24.4 years old (SD= 2.53, Range=18-30). They are all late Chinese learners of English who studied English mostly in classroom settings with traditional teaching methods. The average age of the first exposure to English (English class) is 10 (SD=3.73, Range 3-15). None of them had lived in an all-English environment prior to age 14. Prior to the experiment, they had lived in English-speaking countries for an average of 35.9 months (SD=19.5,

---

[11]   Three participants are from outside of the University of Delaware: one is a college student from the University of Maryland, one is a high school senior (20 years old) from the St. Andrews school in Middletown, Delaware, and the last one is a junior teacher at St. Andrews.

Range=7-96). The average length of formal English instruction is 10 years and 2 months. The background questionnaire including their experiences with languages can be found in Appendix D. None of the Chinese subjects have any neurological impairment, and all except one are right-handed. They were paid $20-$40 for their participation[12] and gave informed consent before the experiment.

3.3.3   FG dependencies in Chinese and L1 influence

This section addresses the issue of L1 (Chinese) interference. As discussed in section 2.4, L1 interference is a potential factor affecting second language acquisition and processing. A L2 parser, at least in the initial stages of acquisition, must rely on certain L1 grammatical rules and parsing routines to anticipate and construct L2 syntactic structures in online processing (Fodor, 1995; Lau et al., 2006). If it so happens that similarities exist between the native language and the target language, it is expected that L2 sentence comprehension will be easier because of the "facilitating" influence of the L1 (positive transfer, see Gass & Selinker, 1992). The opposite happens when grammatical differences between the L1 and L2 hinder L2 comprehension (negative transfer). For instance, it has been suggested that word order differences between the L1 and L2 (e.g., SVO vs. SOV) affect online *wh*-question and relative clause processing (Juff, 2004, 2005; Kanno et al., 2007). More specifically, English is a head-initial language in which verb information is relied on heavily in

[12] Depending on their proficiency level.

online parsing. Japanese or Turkish learners, whose native languages are head-final

and therefore verb information does not become available to the parser until later,

might find English sentences difficult to comprehend. While it has been proposed that

L1 background doesn't change the fundamental nature of L2 sentence parsing (e.g.,

Marinis et al., 2005), there is empirical evidence for the modulating effects of L1

interference (e.g., Hashimoto, 2009). It is therefore necessary to examine the potential

L1 interference effects. Of particular relevance here for Experiment I is the

grammatical nature of Chinese filler-gap dependencies and the related parsing

strategies. The next two sections discuss how these constructions are similar and

different from those of English.

Although Chinese is considered typologically different from English in that it

is a *wh*-in-situ language (e.g., Li and Thompson, 1981), whether movement is

involved in the derivation of Chinese *wh*-questions is still under active investigation

(Huang, Li, & Li, 2009). Proponents of the movement account argue that movement

takes place in Chinese *wh*-questions just like in English, albeit at spell-out (e.g., LF,

Huang, 1982). More importantly for Experiment I, Chinese has overt filler-gap

dependency structures with a dislocated item. The most common are the topic

structure and the relative clause (RC) constructions. Consider the following sentences

in (20) and (21):

(20). Shu, Wo yijin   mai __ le
        Book I   already buy ___Asp[13]
        I have bought (the) book(s).'

(21). a. Subject-gapped RC

        [__ xihuan xiaogou de] nuhai
            like   dog   DE girl
        'the girl that likes the dog'

        b. Object-gapped RC

        [nuhai xihuan __ de] xiaohou
         girl   like   DE   dog
         'the dog that the girl likes'

The first sentence is a topical structure with a fronted item, i.e., *shu* (the book), which is almost identical to English topicalization sentences such as *That pizza, I would never want to eat.* There are also overtly filled gap positions in both subject and object relative clauses as shown in (21). The Chinese relative clauses slightly differ from English in terms of branching directionality, such that the Chinese RC is left-branching (i.e., head-final). The parser usually recognizes the RC construction upon processing the (mostly) obligatory RC marker DE (see Hsu, 2006 for a detailed discussion).

---

[13] Chinese aspect marker for completed actions. Other abbreviations adopted in this thesis for Chinese examples: CL: counter, DE: relative clause marker

A movement analysis has been put forward for both Chinese topic structures and relative clauses (e.g., Shyu, 1996; Hsu, 2006, 2008), mainly based on the fact that island constraints, an important diagnosis for movement-based syntactic phenomena, are active in both constructions. The following examples in (22) and (23) show that topicalization and relativization out of a complex NP are both not permitted:

(22)   Chinese Relative Clause Island Effect

a. Wo   xihuan [[na   ge   ren chuan t$_j$ de]] yifu$_j$.
   I   like   that CL person wear DE cloth.
   I like the cloth the person wears.

b. * [[Wo xihuan [[e$_i$ chuan t$_j$]de] yifu$_j$]de   na   ge ren$_i$.
       I   like   wear   DE cloth DE that CL person
     The person who I like the cloth he/she wears.

(23)   Chinese Topicalization Island Effect

* Wo de qian, [[qiang ___de xiaotou] pao le].
  My   money, rob   DE thief ran ASP
  My money, the thief that took ran away.

This island effect, however, appears to have some exceptions. For example, subject-modifying RCs occasionally violate island constraints. This led to another account (e.g., Xu and Langendoen, 1985) arguing that *pro* is base-generated in the gap position, and the filler-gap relation is formed based on thematic assignment and semantic matching instead of being derived by movement. To account for the inconsistency with island phenomena, the pro-movement accounts such as Hsu (2006, 2008) and Shyu (1996) note that when an island constraint is ignored, the matrix

predicate is always [+stative], based on the fact that Chinese has two semantically-distinct predicates, one [+stative] and the other [-stative]. They argue that [+stative] predicates license a dislocated item as a "major subject" (Shyu, 1996), which is base-generated outside of the island so that its movement does not violate any island constraints. Such an account also receives empirical support from topicalization in other languages such as Japanese and Korean. Thus, these movement analyses provide an empirically-supported unified account. In comparison, the base-generated accounts, though adequate in explaining why sometimes island constraints are violated, fall short in accounting for the fact that most of the time the island constraints *are* obeyed in these structures.

There has been limited research done on the online processing of FG dependencies in Chinese. Huang and Kaiser (2008) conducted a SPR study on Chinese topic structures with parasitic gaps. They aimed to investigate whether the Chinese parser actively searches for a gap after encountering a filler, and if such a filler-gap dependency is formed under syntactic constraints (i.e., strong island and weak island constraints licensed by parasitic gaps; for details, see Phillips et al., 2005). Their findings suggest that (1) the Active Filler Stagey is fully in play in Chinese topic structures, and (2) that FG dependency formation is sensitive to island constraints. They conclude that their results are compatible with the movement analysis for Chinese topic structures. A slightly larger number of studies have been conducted on the processing of Chinese relative clauses, but with a focus on the subject vs. object processing asymmetry (i.e., whether a subject RC or object RC is more costly to

process). Nevertheless, it has been observed that gap searching in Chinese RCs is similar to that in other languages including English (Lin & Garnsey, 2011). In an ERP study by Packard, Ye and Zhou (2011), a larger P600 component was found at the integration sites (either the gap or the RC marker DE next to the purported gap) of subject RCs compared to object RCs, suggesting that a higher integration cost was involved in the former construction. Packard et al.'s findings, although they provide no direct evidence that a trace is posited at the gap site (see 2.3), are consistent with the results obtained by filler-gap dependency studies in English (e.g., Kaan et al., 2000; Phillips et al., 2005), suggesting that relative clause processing is based on similar mechanisms in Chinese.

In summary, there are overt filler-gap dependency constructions in Chinese just as in English. Evidence has been reported on both the theoretical and empirical fronts demonstrating that these structures are derived from syntactic movement and therefore involve a trace posited in the gap position. The existing online processing data, although limited, also suggests that similar parsing strategies are applied to Chinese FG dependencies as in other languages. While the different branching direction of Chinese RCs might cause a slight interference effect (Hashimoto, 2009), FG dependency processing in Chinese is generally similar to that of English, and no significant negative transfer effect is expected.

3.3.4   The English Proficiency Test

The English proficiency level of the participants was determined by the results

of the Versant English Test (Pearson Plc) (for a review, see Downey, Farhady,

Present-Thomas, Suzuki, & Van Moere, 2008; Chun, 2008). Versant English is a fully

automated spoken English test delivered over the phone or computer. The test taker is

asked to read the given materials, listen to English speech about various topics, and

give linguistically, socially, and pragmatically appropriate responses at a native pace.

Versant English has been widely used around the world for admission and job

placement purposes. The validity of Versant English is high (e.g., Bernstein & Cheng,

2007). It aligns with The Common European Framework of Reference (CEFR) and its

U.S. counterpart the American Council on the Teaching of Foreign Languages

(Bernstein & De Jong, 2001; Tannebaum & Caroline, 2005) [14] and is well correlated

with established proficiency tests such as the TOEFL iBT Speaking Test.

The subjects scored an average of 59.2 out of 80 points on this test (SD= 8.5,

Range=47-80), indicating that on average they are advanced-low speakers of English.

The ACTFL Proficiency Guidelines define advanced-low speakers as "are able to

handle a variety of communicative tasks. They are able to participate in most informal

and some formal conversations on topics related to school, home, and leisure activities.

---

[14] Versant English does not directly align with the ACTFL proficiency guidelines,
which are more transparent than the CEFR standards in determining the actual
proficiency level of L2 speakers. To translate the Versant English score into the
ACTFL proficiency scale, this thesis adopts the alignments summarized by Baztán
(2008).

They can also speak about some topics related to employment, current events, and matters of public and community interest" (ACTFL, 2012). Three proficiency groups were constructed based on their Versant scores: the high proficiency group, whose members have reached at least the advanced-mid proficiency level by the CFER/ACTFL standards, consists of 16 subjects with scores equal to or above 68. Two subjects scored 80 and were classified as at near-native proficiency. The middle proficiency group included 24 subjects at the advanced-low level of proficiency, with their Versant scores falling between 58-67. The low-proficiency group included 16 subjects whose Versant scores are equal to or lower than 57, and they are considered as having intermediate-high proficiency, though two subjects were at the intermediate-mid level (with scores equal to or below 48 points). Table 3.2 below shows the alignment between the proficiency guidelines of the CEFR and ACTFL:

Table 3.2:   Versant English scores and L2 proficiency levels by the ACTFL and CEFR standards

Versant English Score	CEFR	ACTFL	Proficiency Index in this study	Number of L2 participants in Exp. I at this level
78-80	C2	Advanced-high/Superior	HIGH	2
69-78	C1	Advanced-mid	HIGH	14
58-68	B2	Advanced-low	MID	24
48-57	B1	Intermediate-high	LOW	14
38-47	A2	Intermediate-mid	LOW	2

3.3.5 The working memory test

The working memory test used in current study is an audio version of the Harrington and Sawyer (1992) reading span test. As explained in Chapter 2, the Harrington and Sawyer (1992) test is a reliable L2 version of the most widely used WM test, the Self Paced Reading Test (RST) by Daneman and Carpenter (1980). In a pilot study, most of the participants complained that the Danemen and Carpenter listening span task was too difficult. However, to the best of my knowledge, there is no reliable working memory test available in Chinese. In addition, Hestvik et al. (2012) used Daneman and Carpenter (1980) for working memory measurement. To facilitate the comparisons between the L1 and L2 ERP and behavior results of that study and the current one, it is best that the working memory measures are as close as possible.

A total of 42 sentences, divided into 4 levels with three sets of sentences at each level, were used as stimuli. There are thus 4x3=12 sets of sentences. The number of sentences contained in each set increases over the levels, such that the first level has two sentences per set, the second level has three, and the last level has five in a set. These sentences are simple active sentences that are 10-13 words long. Out of the 42 sentences, half of them have a disrupted word order at the middle or end part of the sentence. An example is *Favorite foods in are melon the summer and sweet corn*. The subjects are supposed to process the sentences to identify the ungrammatical ones while also remembering the sentence-final words, so the WM with regard to both

processing and storage is evaluated. The grammatical and ungrammatical sentences were randomly distributed across levels. The complete list of sentences is in Appendix C. All stimuli were read by a female native speaker of English at a normal pace and were digitally recorded using 16 bit resolution and a 22,050 kHz sampling rate.

The WM test was set up in E-prime (Schneider, Eschman, & Zuccolotto, 2002) to collect data for response time, accuracy of acceptability judgment, and the number of words correctly recalled. After the subjects signed the consent form and completed a background questionnaire, they were directed to take the WM test. Instructions were given carefully by the experimenter, who advised the subjects to not use any strategies to memorize the sentence-final words. A practice section was then given, and questions and concerns were addressed before the test began.

. The test proceeded as follows. The subject heard a sentence and was then prompted to press the corresponding button for a judgment of acceptable or unacceptable. The subject then pressed the space bar to proceed to the next sentence; this repeated until all the sentences in a set were delivered. The experimenter monitored the subject closely to make sure they didn't wait too long to press the space bar to move on to the next sentence.[15] The subject was then prompted to recall the final word of each sentence in the set. The experimenter recorded the subject's answers on both paper and computer by keying in the number of correctly recalled

---

[15] If the subject was waiting too long (more than 5 seconds), the experimenter pressed the space bar to move on to the next item.

words. When the recall portion was finished, the experimenter pressed the space bar to move on to the next level.

At the end of the test, the total number of final words correctly recalled was computed for each subject. As mentioned above in section 2.6, this is the most reliable way to probe an L2 learner's storage and processing capacity. On average, the participants recalled 31.6 words out of the total 42 words (SD=0.6, Range=19-41). The median listening span test score was 33. The participants were assigned to the low WM group if their score was lower than 33 and to the high WM group if their score was equal to or higher than 33. This procedure resulted in 27 low WM subjects and 29 high WM subjects.

3.3.6 Experiment I overall procedure

The Filled Gap experiment has 4 components, administered to the subjects in the following order: (1) the Versant English proficiency test, if the subject had not already completed it by the EEG session appointment, (2) the working memory test, (3) the ERP task vocabulary drill and practice run, followed by electrode net application and the EEG recording session, and (4) the Paper-and-Pencil Acceptability Judgment Test (see 3.3.6). The Acceptability Judgment Test was given after the ERP session to prevent the subject from being aware of the ungrammatical sentences on the test while completing the ERP tasks. The entire experimental session with all four tasks lasted for approximately two and a half hours. Occasionally, the subject took the Versant proficiency test last, due to lab scheduling issues.

3.3.7 The Acceptability Judgment Test

As discussed above, in order to better interpret the online ERP measures, it is important to examine the L2 participants' off-line grammatical knowledge of long-distance filler-gap dependencies. In particular, it is necessary to (1) confirm that they are sensitive to phrase structure violations in relative clauses in a no-pressure, no-time-constraint environment, and (2) to test how individual factors (WM, proficiency) affect these off-line behavioral results. A Paper-and-Pencil acceptability task in the format of a questionnaire was therefore administered to the L2 participants after the ERP session, in which they rated the sentences' acceptability on a 7-point scale (1 being completely not acceptable and 7 being perfectly acceptable). A group of native English speakers also filled out the questionnaire for control purposes.

3.3.7.1 The Acceptability Judgment Task: native English speaker participants

In addition to the 57 L2 speakers of English, 37 native English speakers (22 female, 15 male) were recruited to be the Acceptability Judgment Task controls. All of them are monolingual undergraduate students from the University of Delaware. Their average age is 19.8 (SD=2.2, Range=18-29). All of them gave informed consent and completed the background questionnaire. For their participation, they were awarded a small amount of extra credit in their first-year Chinese foreign language course.

3.3.7.2 Materials and data analysis

The Acceptability Judgment Task consists of a total of 30 sentences, in which 12 of them are the target items. These sentences are structurally identical to the stimuli used in the following four conditions of the ERP tasks, but with different vocabulary:

Table 3.3: Paper-and-Pencil Task sample sentences by experiment condition

	Condition	Sample sentence
A.	Ungrammatical	The customer that the waitress greeted *the gentleman* only comes on Thursdays evenings.
B.	Adjunct	The night that the policeman caught the thief, it was extremely cold.
C.	Object	My sister thought that I would be home for the rest of the day so she left without her keys.
D.	Trace	The test Jane took last week had more essay questions than she expected.

Out of the twelve target sentences, six of them are in the ungrammatical condition (A) and resemble the UNGRAMM condition sentences in Hestvik et al. (2012). These sentences should be rated as low acceptability if the rater is sensitive to the "filled gap" violation. The other six items are grammatical ones identical to the ADJUNCT, TRACE, and OBJECT conditions (see 3.6 for details) in the ERP task. In addition, 18 filler sentences of various acceptability (Sprouse and Almeida, 2012) were incorporated for two purposes: (1) to distract the subjects' attention from

the target sentences, and (2) to calibrate the participants' use of the rating scale, as the varying acceptability of these sentences is based on a large number of native speakers' ratings (Sprouse & Almeida, 2012). The target sentences were counter-balanced and randomly distributed among the filler sentences. The Acceptability Judgment Test can be found in Appendix C.

After the data were collected for both the L1 and L2 groups, repeated measures ANOVAs (Grammaticality x Proficiency/WM Groups) were conducted for the Chinese group first to exam (1) whether there is a significant difference between the ratings of the grammatical and ungrammatical sentences, and (2) how proficiency and WM group affect these ratings. After that, the Chinese ratings were compared to the English ratings by adding an additional between-group factor, Group (Chinese vs. English).

3.3.7.3   Results

Figure 3.1 below shows that the L2 group rated the grammatical condition sentences much higher than the ungrammatical ones:
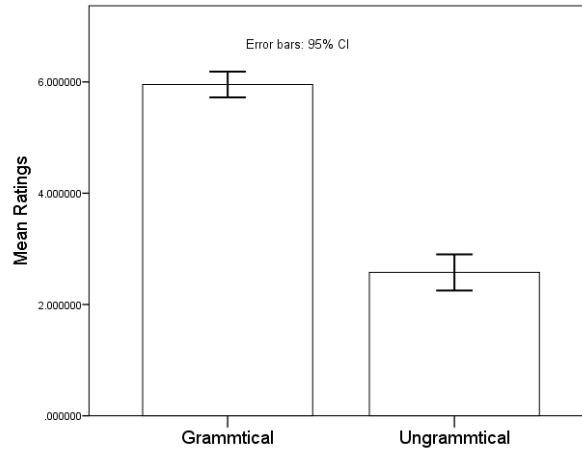
Figure 3.1:   Acceptability judgment data for the L2 group

The average rating for the grammatical sentences is 5.95 (SD=0.87), and the average for the ungrammatical ones is 2.59 (SD=1.24). The ANOVA revealed a main effect of grammaticality: $F(1,56)= 296.7$, $P<0.001$, confirming that the learners are sensitive the violations caused by a filled gap. Further examination of the proficiency group variable showed a significant interaction between proficiency group and grammaticality, $F(2,54)=3.775$, $p<0.05$, such that the higher the proficiency, the bigger the difference in ratings between grammatical and ungrammatical sentences. This indicates that proficiency is a function of sensitivity to grammaticality, as illustrated in Figure 3.2:
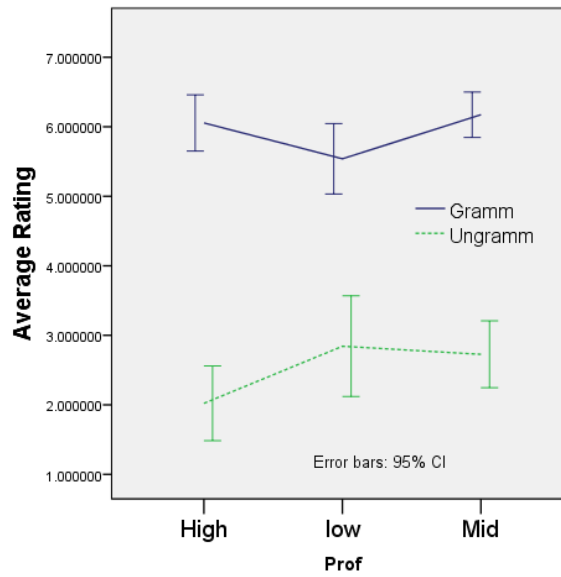
Figure 3.2:   L2 Group acceptability judgment ratings by proficiency group

The L2 ratings were marginally affected by WM capacity, as the ANOVA revealed a marginally significant interaction between grammaticality and WM group: $F (1,55)=3.65$, $p=0.061$.

When compared to the native speakers, the L2 group rated the sentences similarly, as seen in Figure 3.3 below. For the L1 group, the mean acceptability rating for the grammatical sentences is 6.09 (SD=0.48), and for the ungrammatical sentences the mean rating is 1.95 (SD=0.08).
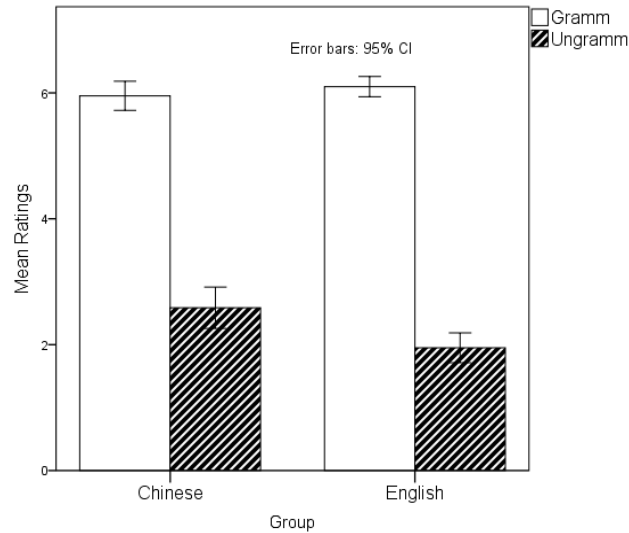
Figure 3.3:   Acceptability judgment data for the native-speaker (L1) group and the L2 group

The native English-speaking group rated the ungrammatical sentences slightly lower than the L2 learners. A between-group repeated measures ANOVA compared the two groups and revealed a significant interaction between group (English vs. Chinese) and grammaticality: $F(1,92)=8.5$, $p=0.004$. Such an interaction, together with the absolute mean ratings listed above, suggests that the English native speakers are more sensitive to the violations than the Chinese group. However, no significant between-group effect was found by the ANOVA: $F(1,92)=3.09$, $p=0.083$. To further examine the difference between the L1 and L2 ratings, the L2 subjects with low proficiency were removed. The remaining 40 L2 subjects and the English L1 subjects were compared again, and the ANOVA reported a non-significant interaction between groups and grammaticality: $F(1,75)=3.57$, $p=0.07$. When only the high proficiency L2

learners were compared to the native speakers, it was revealed that the ratings of the
two groups are identical, as the grammaticality x group interaction is highly
insignificant: $F_{(1,50)} = 0.15$, $p=0.7$. To summarize, these results confirm that the L2
group has the necessary grammatical knowledge to process the complex filler-gap
dependency structures. Their acceptability ratings are not identical to the native
speakers when they are treated as a homogeneous group. However, once proficiency is
controlled, the L1 and L2 judgments of grammaticality are highly comparable.

3.3.8   The ERP Experiment

3.3.8.1   Materials

The stimuli and the design of the ERP experiment are based on those of
Hestvik et al. (2012), with a slight modification to the comprehension question
manipulation. A complete list of the stimuli sentences can be found in Appendix B.
Four experimental conditions were included, as summarized in Table 3.4 below:

Table 3.4:   Example sentence from each experimental condition

	Condition	Sample sentence
A.	Ungrammatical	The zebra that the hippo kissed *the camel on the nose ran far away.
B.	Adjunct	The weekend that the hippo kissed the camel on the nose, it was humid.
C.	Object	The zebra said that the hippo kissed the camel on the hose and then ran far away.
D.	Trace	The zebra that the hippo kissed on the nose ran far away.

The critical comparison is between the ungrammatical condition (hereafter UNGRAMM) and the adjunct (ADJUNCT) condition. The sentences in both of these conditions contain relative clauses, and the parser is supposed to look for a gap as soon as the first noun phrase is processed. The two conditions are identical before and in the critical region (the verb *kissed* and the noun immediately after it), except that in the UNGRAMM condition the extra noun *the camel* causes a "filled-gap" ungrammaticality. The object (OBJECT) and trace (TRACE) conditions are included as fillers and distractors so that not all of the stimuli contain relative clauses (OBJECT) and not all of the object relative clauses are ungrammatical (TRACE). Thirty-two sentences were constructed in each condition for a total of 128 stimuli, which constituted script A. Then, for each set of 4 sentences in script A, a corresponding set for script B was made by switching the agent (*the hippo* in Table 3.4) and the patient (*the zebra*) in the script A sentence with those of another sentence for counter-balancing purposes. In addition, the part after the critical region was changed slightly to prevent the subject from detecting the pattern. Thus, script B was constructed from a different set of 128 sentences. This results in 8 lists (4 in script A and 4 in script B) of 32 sentences each, for a total of 256 experimental sentences. The order of the sentences was counter-balanced across subjects: half of the subjects listened to script A in the first half of the experiment and script B in the second half, and the other half of the subjects followed the opposite order. Lastly, the order of

95

presentation of the sentences within each list and the selection of sentences from each condition across lists were both randomized.

Each sentence presentation was followed by a comprehension question. There were four kinds of comprehension questions: an "easy" Yes-No question such as "Did you hear the word humid?", a "content" Yes-No question such as "Did the hippo kiss the zebra?", an object *wh*-question such as "Who did the zebra kiss?", and a subject *wh*-question such as "Who chased the squirrel?". For the Yes-No questions, the choices "YES" and "NO" were displayed on the screen. For the *wh*-questions, two pictures were displayed on the screen: one picture of an animal, and another of a cartoon person with a question mark over their head. Two sample pictures are shown elow in Figure 3.4, adopted from Hestvik et al. (2012):

ALTERNATIVE 1   ALTERNATIVE 2
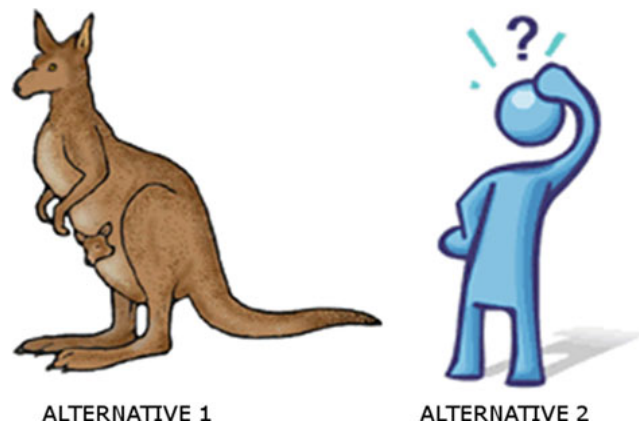
Figure 3.4: Example of picture choices for comprehension questions

The subject was instructed to select either "YES" or "NO" for a Yes-No question, or the correct animal for a *wh*-question. If the correct animal was not

96

pictured, then the subject was to select the "blue man" picture. For example, the subject may have heard the sentence *The winter that the ostrich raced the giraffe down the road, it was humid*, followed by the comprehension question *Who did the ostrich race*?. Then the subject would see the above two picture choices. Given that the ostrich raced a giraffe, which is not one of the picture choices, the correct answer is alternative 2, the "blue man". In Hestvik et al. (2012), the correct answer for the UNGRAMM condition was always the "blue man", as the condition is ungrammatical and technically no answer should be considered "correct". However, this set-up renders the UNGRAMM condition different from the other conditions regarding the instructions given to the subject. In other words, if the subject followed the instructions consistently, he or she would be okay in the other conditions but would miss the correct answer in the UNGRAMM condition every so often. This was not an issue for the native speakers, as confirmed by their behavioral data. However, in the pilot study the L2 subjects reported that the instructions were too confusing to follow, and they were distracted from the task. Since the resource strain on the L2 learners is already heavy due to the type of sentences used in the experiment, the additional confusion over the instructions, though it might be insignificant in the case of the native speakers, might disrupt the online processing of the L2 learners. For this reason, both picture choices were coded as correct for the UNGRAMM condition, and thus the L2 subjects always answered the comprehension question correctly for stimuli in this condition.

Lastly, the auditory stimuli (which were from Hestvik et al. 2012) were recorded by a professional linguist in a lab setting, with the intonation carefully controlled to avoid the prosodic cues for FG sentences found in everyday speech. Additionally, the speech rate for both the stimuli and the comprehension questions (which were recorded by a different female speaker) were matched to a typical conversational pace.

3.3.8.2   ERP EEG procedure

The experimenter first reviewed the ERP task instructions with the subject and had them practice on a PC without the ERP net. They were then drilled on some of the vocabulary that was considered difficult by participants in the pilot study. After they were considered proficient with the task procedure and the vocabulary, the electrode net was applied and the subject was seated in a chair with a table top in a sound-attenuating booth. A PST serial response box was placed on the tabletop, with its buttons clearly labeled. There were also a computer screen and two speakers placed on a desk approximately three feet away, facing the subjects. The participants were instructed to listen to each sentence and the comprehension question following it. The two pictures described above then showed up at the left and right lower corners of the computer screen. The subjects were asked to indicate their picture selection by pressing the buttons in the corresponding locations on the response box (e.g., the left button for the picture choice on the left, or the right button for the choice on the right). The experiment was programed using the E-Prime software (Schneider et al., 2002)

and was divided into four blocks of 64 sentences randomly presented to the subjects. The subject was offered a break between blocks. The entire EEG recording session took about one hour and fifteen minutes.

3.3.8.3   Data collection and EEG recording

The EEG was recorded with a 128-channel EGI 300 system (Hydrocel HCGSN 100 v.1.0, Geodesics, U.S.A), with a sampling rate of 250 Hz. Eye movements and blinks were monitored with electrodes placed under each eye. Cz online was used as a reference, and the electrode impedances were kept below 50kΩ. The continuous EEG was divided into epochs of 1400 ms for each trial by time locking to the onset of the critical noun phrase (at the beginning of the article *the*). Baseline correction was performed using a 200 ms baseline period (before the onset of the noun phrase) as a reference signal value. For artifact correction, the bad channels were replaced, eye blinks were subtracted, and then the eye movements were corrected using ICA Dien (2010). This order of procedures could potentially cause the eye blink channels (some with substantial voltage fluctuation) to be registered as bad channels and replaced instead of corrected. The data was therefore re-examined by conducting the artifact detection in the reverse order. The grand average patterns resulting from these two orders were determined by visual inspection to be identical. Thus, the voltage data derived via the initial data processing procedure was kept.

The artifact correction and bad channel replacement resulted in the removal of an average of 22.5% of the trials per subject (SD=0.034). Subjects with more than

30% bad trials were not included in the ERP analysis, leaving 54 subjects for the ERP

data analyses. Table 3.5 shows how the remaining 54 subjects were distributed among

the proficiency groups and WM groups.

Table 3.5:   WM and proficiency groups of subjects used in ERP analyses.

Group	Level	Number of participants
Proficiency group	High	16
	Mid	21
	Low	16
Working Memory group	High	28
	Low	25

The data for all subjects were included in the behavioral data analyses, and

trials that were not responded to correctly were also included in order to prevent

excess loss of trials and power. The ERP average was computed for each condition,

each subject, and each relevant electrode site and referenced to the average voltage of

all electrodes.

3.3.8.4   Data analyses

For the behavioral data, the accuracy of the responses was recorded via E-

Prime (Schneider et al., 2002) and submitted to a 2x2 mixed factorial repeated

measures ANOVA, with condition (3)[16] as a within-subject measure. When the Chinese behavioral data were interpreted in comparison to the English results reported in Hestvik et al. (2012), the Mann-Whitney U test, the non-parametric version of the regular independent t test, was used because of a normality issue and the significantly different sample sizes across studies.

For the voltage data, the entire set of electrodes was divided into regions based on initial visual inspection, such that an average over those regions could be computed and used as one of the dependent measures in the ANOVA. The following three regions were identified based on the initial visual inspection of the grand average line plot: one at the frontal-central region of the scalp, which shows a positivity (Frontal-Central), another showing a widely distributed negativity in the mostly central region with some extension to the posterior area (Central), and a third in the posterior-inferior region (Posterior-Inferior). Figure 3.5a shows the arrangement of electrodes on the electrode net used:

---

[16] The UNGRAMM condition was excluded because in the case of the L2 experiment, the answer to the comprehension question following those stimuli was considered correct regardless of which picture the subject chose.

Frontal-Central Positivity	Central Negativity (N400)	Posterior Positivity
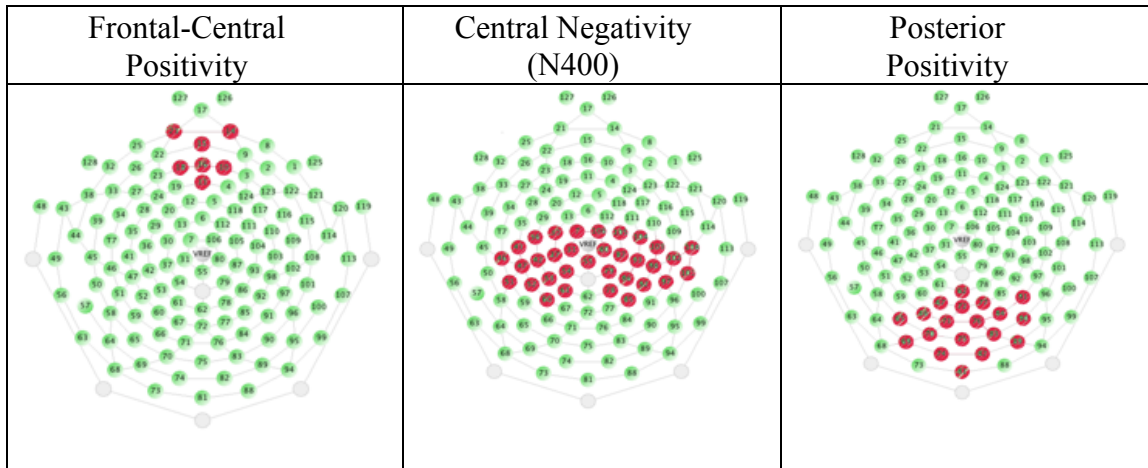
Figure 3.5: The electrode distribution in three critical regions (highlighted) of the
128- Channel HCGSN v.1.0 net

The Frontal-Central region contains the following electrodes: 10, 11, 14, 15,

16, 18, and 21. The Central region contains 33 electrodes: 7, 30, 31, 36, 37, 41, 42, 46,

47, 51, 52, 53, 54, 55, 60, 61, 78, 79, 80, 85, 86, 87, 92, 93, 97, 98, 101, 102, 103, 104,

105, 106, 108, and Cz. Lastly, the Posterior-Inferior region includes 19 electrodes: 62,

65, 66, 67, 69, 70, 71, 72, 74, 75, 76, 77, 81, 82, 83, 84, 89, 90, and 91. These

electrodes are highlighted in the regions' respective graphs in Figure 3.5a.

Another dependent measure for voltage data analyses is time window. A total

of seven 200 ms time bins were constructed for the 1400 ms epoch starting from the

onset of the critical noun phrase. The mean amplitude over these time bins was

computed for each electrode region for each subject and condition. The average

amplitude obtained was submitted to a mixed-factorial repeated measures ANOVA

with Time x Condition (2) (and laterality (2) when necessary) as within-subject

factors. To measure the effect of working memory capacity and proficiency level, WM group (2, high and low) and proficiency level (3, high, mid, and low) were included as between-group factors. When appropriate, *p*-values were adjusted by using Greenhouse-Geisser (1959) correction for violation of the assumption of sphericity. In addition, significant interactions between condition, time, and electrode region were followed up by planned orthogonal contrast analyses. Lastly, regression analyses were conducted between voltage data and Versant score and working memory score.

3.3.8.5 Behavioral results

As explained above, comprehension accuracy was computed based on the three conditions other than the UNGRAMM condition, for which the Chinese subjects obtain 100% accuracy because both picture options were coded as correct. The overall mean accuracy (without UNGRAMM) for the Chinese group is 79% (SD=7.7%). There is a significant difference among conditions, as the ANOVA yielded a main effect of condition: $F_{(1,55)}=87.34$, $p<0.00$. A post-hoc Scheffe test showed that the accuracy rate in the OBJECT condition is significantly lower than the other two conditions, which do not differ significantly from each other.

The Chinese subjects were further divided into two working memory groups based on their WM recall score (see section 3.3.6). The mixed-factorial repeated measures ANOVA 2 (WM group) x 3 (condition) showed that there was a main effect of WM group ($F_{(1, 55)}=11.002$, $p=0.002$) and condition ($F_{(2, 110)}=42.49$, $p<0.001$), as well as a significant interaction between WM group and condition ($F_{(2, 110)}$

103

=5.25, p=0.007). The following figure shows how WM group interacts with accuracy rate for each condition:
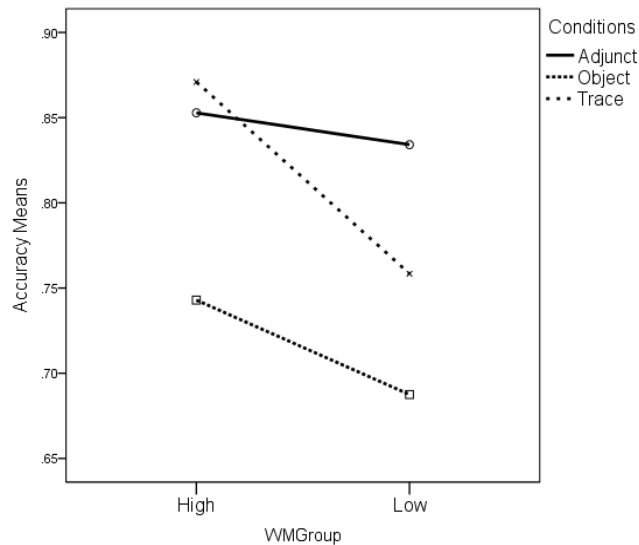
Figure 3.6:   WM group interaction with accuracy rate by condition

In order to examine how proficiency interacts with the accuracy scores, three proficiency groups were formed as explained above: High (16), Middle (23), and Low (16). A main effect was found for proficiency group (F (2, 54) =7.603, P=0.001), as well as for condition (F (2, 108)=39.26, P<0.001). Further regression analyses confirmed that there was a weak but significant correlation between WM score and accuracy rate (R^2=0.25, F (1, 55)=18.3, P=0.001). There was also a significant correlation between Versant score (proficiency) and accuracy rate (R2=0.26, F (1,55)=18.87, P=0.000).

The overall accuracy rate of the Chinese subjects (79%, SD= 7.7%) is slightly lower than that of the native English speakers, who showed 86% (SD=5%) accuracy in the three conditions combined.[17] The direct comparison between the two groups is illustrated in Figure 3.7:
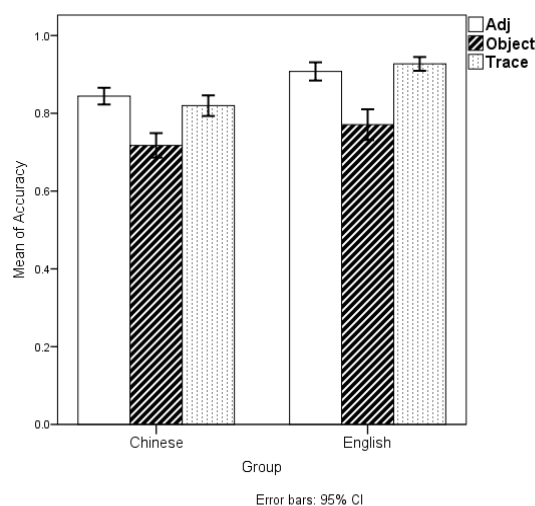
Figure 3.7:   Comprehension question accuracy rate by condition for Chinese and English groups

A non-parametric version of the independent samples t test, the Mann-Whitney U test, was used for the following reasons. First, the sample sizes are very different for the native speakers (30) and the learners (56), and the ratio exceeds the 1.5 that is recommended for regular t tests. Secondly, the distribution of the accuracy scores

---

[17] The behavioral data for 30 native English speakers were obtained from Hestvik et al. (2012) and were compared directly with the Chinese speakers.

among the native speakers group is not normal (a Shapiro-Wilk test yielded a *p*-value of 0.02). The Mann-Whitney U test rendered a Z value of -1, p=0.3, indicating that the difference between the native speaker and the learner groups is not significant. Another similarity shared by the native speaker and L2 groups is that both groups were significantly less accurate in the OBJECT condition than in the other two conditions, which are not very different from each other.

3.3.8.6 ERP voltage results

Visual inspection for a grammaticality effect revealed very different ERPs from the L2 subjects than those from the native speakers. As described in section 3.3.3.8, a widespread N400-like negativity is evident in the central posterior region. The ungrammaticality also elicited positivity in the front-central region of the scalp, slightly spreading into the anterior regions of both hemispheres. A very weak, late positivity trend after 800 ms in the posterior-inferior region was also observed. There was no anterior negativity discovered in either hemisphere. The following Figure 3.8a illustrates the above-described patterns, with the red regions representing positive potentials and the blue regions representing the negative potentials. To show the contrast between native and L2 brain responses, I include the topo plots from Hestvik et al. (2012) showing the brain responses generated by the native speakers (Figure 3.8b):
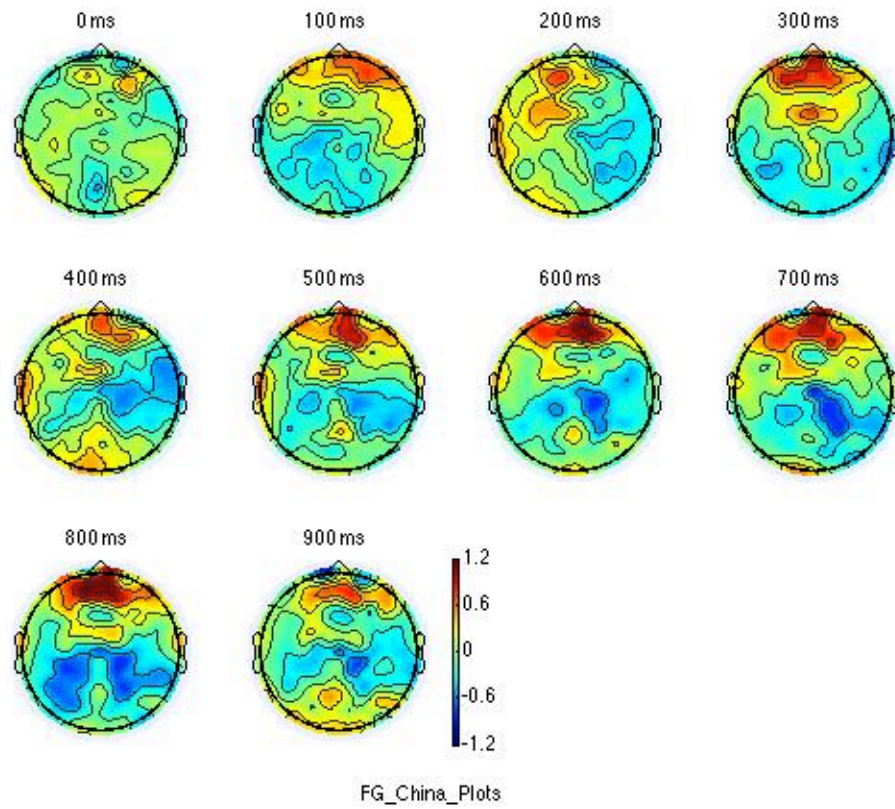
Figure 3.8a: Grand average topo plot for the grammaticality effect: Chinese subjects
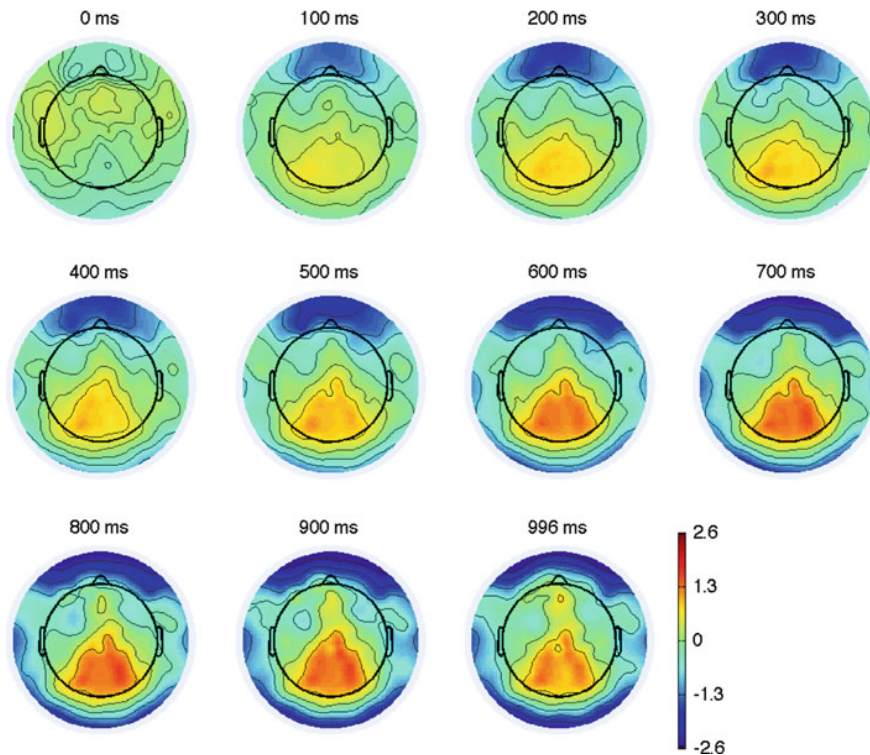
Figure 3.8b: Grand average topo plots for the grammaticality effect: native speakers from Hestvik et al. (2012)

The above two topo plots show the Chinese subjects' brain response patterns are significantly different from those of the native speakers, I will present the analyses for the Chinese subjects separately in this section. I begin by examining the left anterior region.

Left anterior negativity

As we can see from the above grand topo plot (Figure 3.8a), no anterior negativity was observed in either hemisphere. In fact, the ungrammatical condition waveform is more positive in comparison to the grammatical control. To rule out that potential ELAN/LAN effects were masked by the variability in proficiency level and

working memory capacity, additional graphs by proficiency and WM group were generated, based on the mean voltage of the electrodes in the left anterior region (see section 3.3.3.8) computed for each 100 ms time window starting from 300 ms to 1000 ms. As Figures 3.9a and 3.9b below illustrate, there was no component resembling the ELAN or LAN even when proficiency level and WM capacity were separated. Instead, we see that the ungrammatical condition elicited positivity in all three proficiency groups in largely the same pattern over the time span, with the low proficiency group a bit slower in showing the effect. Additionally, the low WM group demonstrated more positivity than the high WM
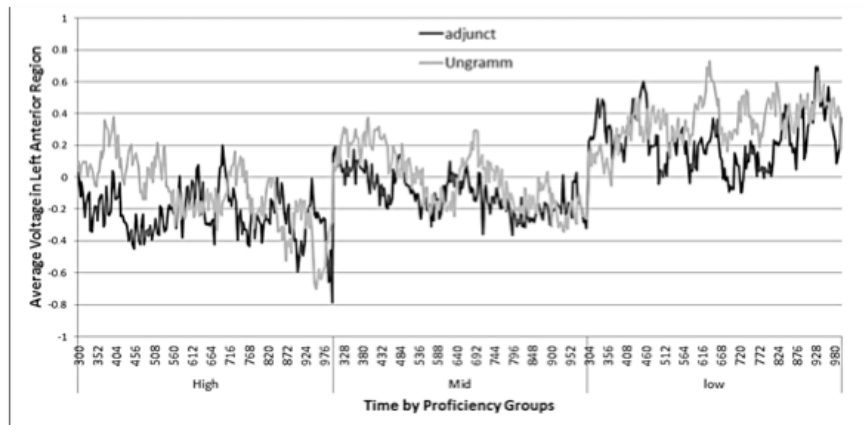
Figure 3.9a: Left anterior region by proficiency group
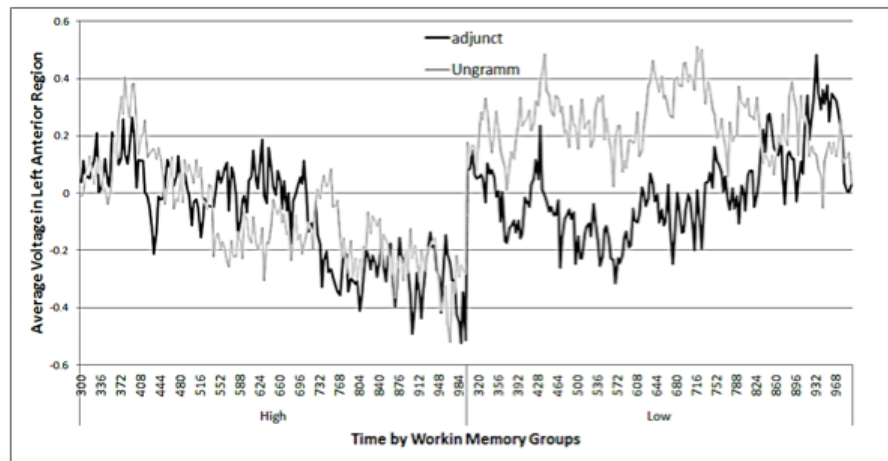
Figure 3.9b: Left anterior region by WM group

group. However, a repeated measures ANOVA with Time (7, 200 ms-900 ms) x

Grammaticality (2) as within-subject factors and working memory group as a

between-subject factor yielded no significant group effect (F (1,52)=1.47, p=0.231),

no interaction between WM group and grammaticality (F (1, 52)=0.683, p=0.412), and

no three-way interaction among WM group, grammaticality, and time (F (6,

312)=2.044, p=0.06), suggesting that the difference in WM capacity doesn't change

the positivity obtained in the LAN area. Therefore, it is confirmed that the L2 learners

didn't produce anterior negativity regardless of their proficiency level or working

memory capacity.

The Central Negativity (N400)

As Figure 3.10 below shows, the negativity in the Central region starts at

around 250-300 ms after the onset of the offending extra noun phrase, peaks at around

500 ms, and is sustained into the later time windows.

110
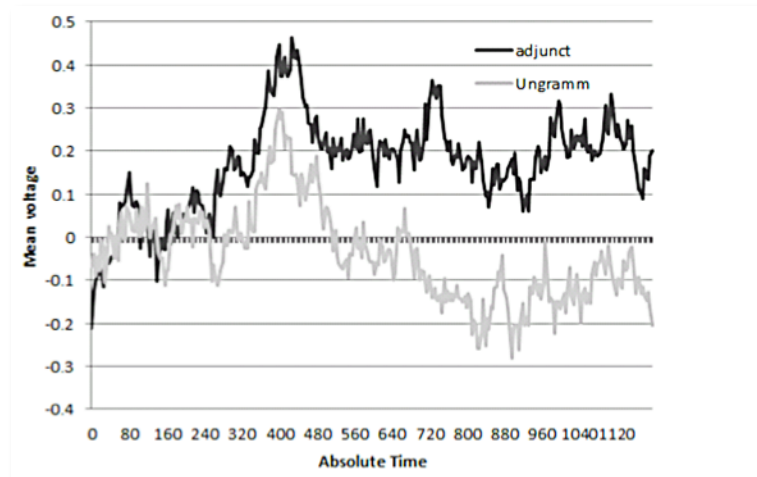
Figure 3.10: The N400 waveforms for grammatical (adjunct) and ungrammatical
(Ungramm) conditions

Although the observed negative component seems to maximize slightly later
and lasted longer than the typical N400 found in native speakers (Kutas & Fedemeier,
2011), it has been reported that L2 learners and bilinguals tend to have a delayed N400
(e.g., Aldal, Danald, Meuter Muldrew & Luce (1990)). Its more central as opposed to
central-posterior distribution is expected for auditory stimuli and tasks with picture
identification (Kutas & Fedemeier, 2011). Therefore, this ERP can be functionally
considered to be the N400. Again, the mean voltage of the electrodes in the Central
region (see section 3.3.3.8) was first computed for each 100 ms time window starting
from 300 ms to 1000 ms for each subject and each condition. After that, a repeated
measures mixed-factorial ANOVA was run on the mean voltage data, with Condition
(2) x Time (7) as within-subject factors, which revealed a main effect of condition (F
$(1, 53)=11.11$, $p<0.001$), suggesting that there is a significant difference between the

111

grammatical and ungrammatical conditions from 300 ms on. To inspect whether this

negativity is significant in each 100 ms time window, planned orthogonal tests were

conducted, which reported significance in every time window starting at 300 ms

(t=3.02, p<0.001 for the 300-400 ms window). The mean of the difference voltage

(voltage for the UNGRAMM condition minus the voltage for the ADJUNCT

condition) was computed and submitted to a pairwise t test, which generated the

Figure 3.11. As we can see, the negativity continues to be significant in every

subsequent time window.
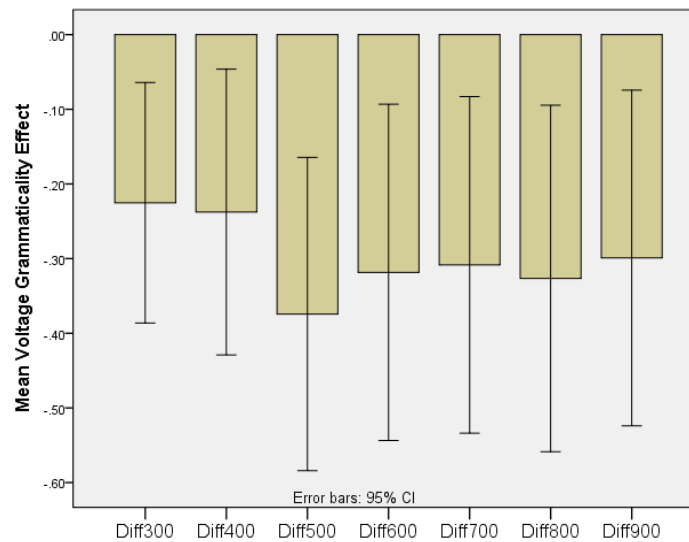
Figure 3.11: The N400 in difference waveform by 100 ms time windows

To inspect how proficiency level affected the N400, the between-subject factor

Proficiency (3) was added to the above ANOVA, but no significant interaction was

112

found between the N400 and proficiency group (F (2, 51)=0.78, p=0.4). Figure 3.12

illustrates the difference waveform for each time window by proficiency group.
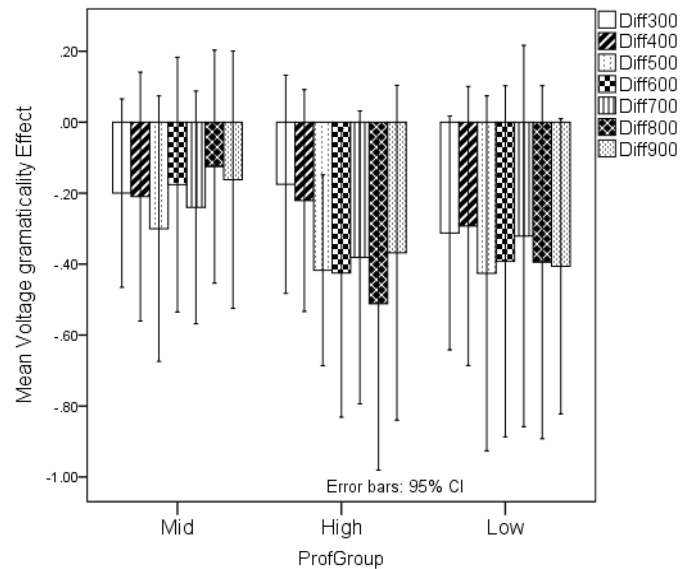
Figure 3.12: The N400 difference waveform by proficiency group

By visual inspection, the N400 seems to be the most prominent for the High

proficiency group, especially in the 500-700 ms time window, and least visible for the

Mid proficiency group participants. Next, proficiency (Versant) scores were run with

the voltage average and no significant correlation was found (r=0.021, F (2, 53)=0.02,

p=0.88), as Figure 3.13 shows:
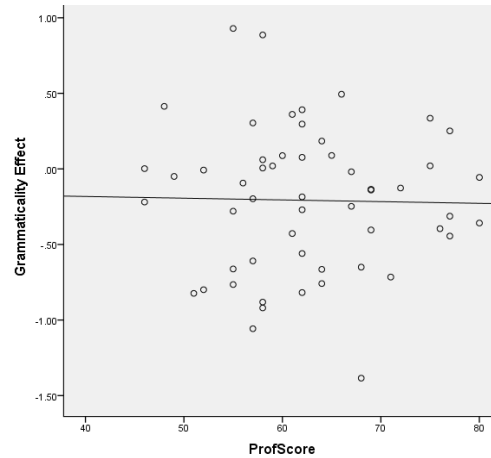
Figure 3.13: Proficiency scores regressed on voltage changes for the N400

The relation between working memory capacity and the N400 amplitude was examined in the same way. An ANOVA with WM as a between-subjects factor revealed no interaction between working memory group and the negativity (F (1, 52) =1.38 p=0.25), as shown in Figure 3.14:
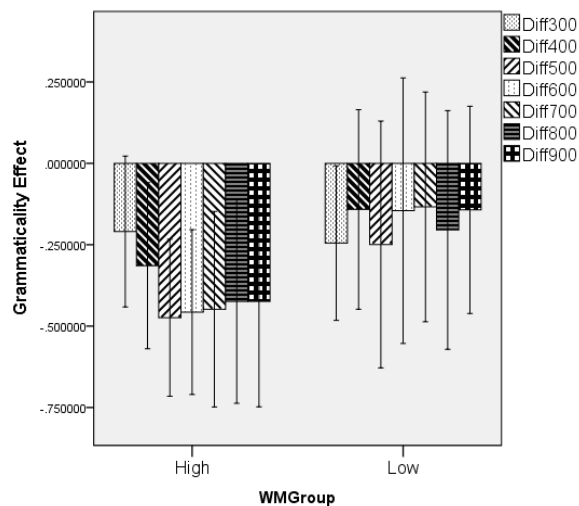
Figure 3.14: The N400 by working memory group

Note again that as Figure 3.14 shows, the N400 seems to be mostly carried by the group with a high working memory capacity, in spite of the fact that it didn't reach statistical significance. Regressing the working memory scores on the amplitude of the negativity revealed a R^2=0.024, (F $(1, 53)$=1.367, p=0.248), confirming that WM scores don't predict the voltage changes. Figure 3.15 shows the correlation between WM scores and the grammaticality effect:
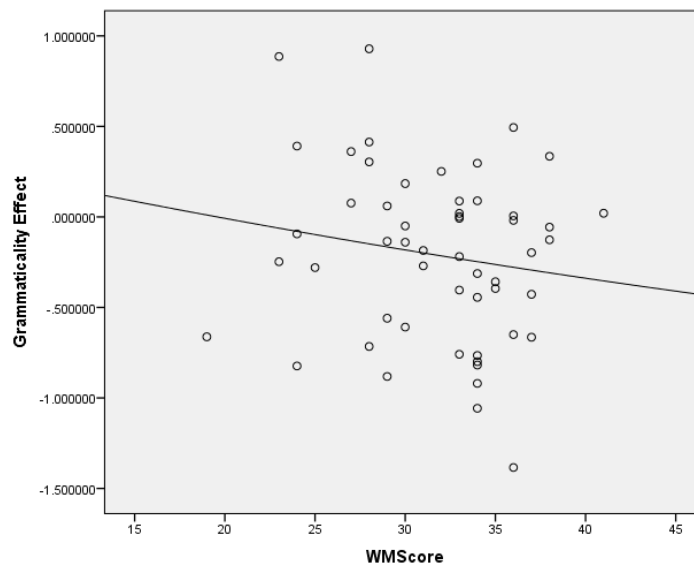
Figure 3.15   WM Scores regressed on voltage changes for the N400

The Frontal Central Positivity (FCP)

A clear positivity in the UNGRAMM condition was observed in the frontal-central region between 300 ms to 1,000 ms, as shown in Figure 3.16 below. A mixed-factorial repeated measures ANOVA yielded significant results for the grammaticality

115

factor (F (1, 53)=4,677, p=0.035), suggesting that there is a significant difference between the grammatical and ungrammatical conditions from 300 ms on. In addition, the interaction between time and grammaticality is also significant (F (6, 52)=2.308, P=0.034), demonstrating that the ungrammaticality effect changes over time. Planned orthogonal contrast tests revealed that the positivity became significant during the 500-600 ms time window (t=-2.286, p=0.006), approached significance during the 600-700 ms time window (t=-1.94, p=0.057), became significant again during the 700-800 ms time window (t=-2.126, p=0.03), and then fell slightly under significance during the 800-900 ms time window (t=-1.96, p=0.055).
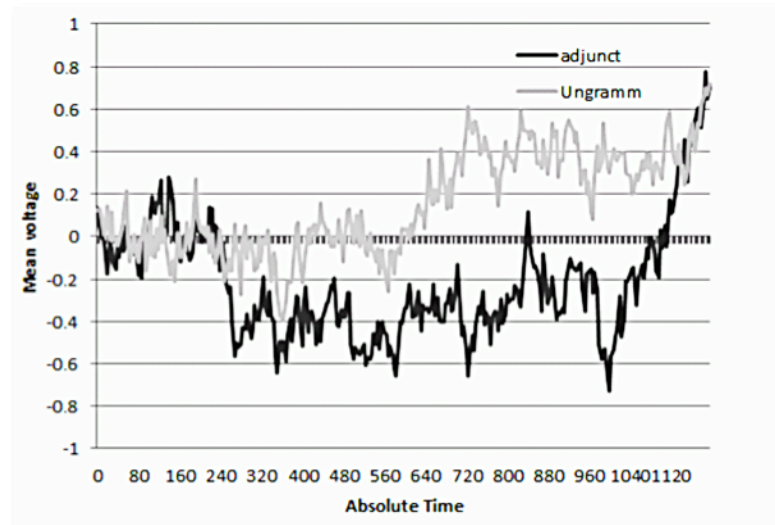
Figure 3.16: Frontal central positivity (FCP) waveforms

The positivity could be summarized as largely significant from 500-900 ms after the onset of the extra noun phrase. Figure 3.17 shows pairwise t tests (two-tailed) for voltage difference scores for each time window.
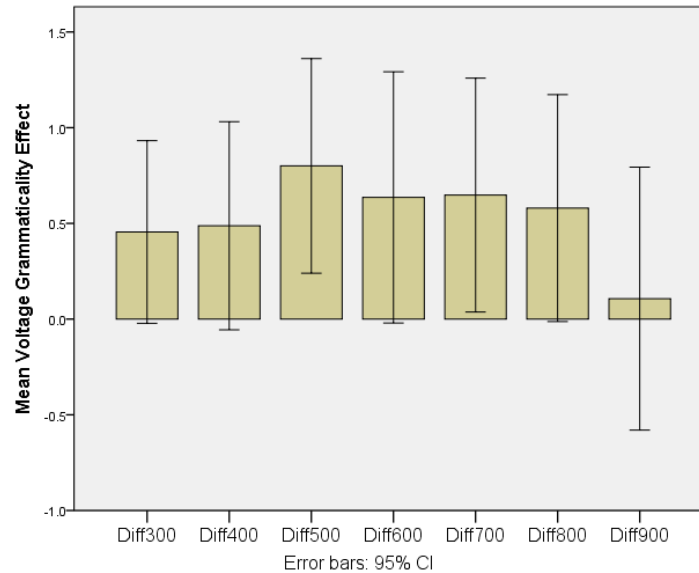
Figure 3.17: Frontal central positivity by time window

Adding WM as a between-subjects factor, repeated measures ANOVA reported no significant interaction between WM and the grammaticality effect (F (1, 52)=1.416, p=0.24), although Figure 3.18 shows that the high working memory group displayed more positivity than the low working memory group.

117
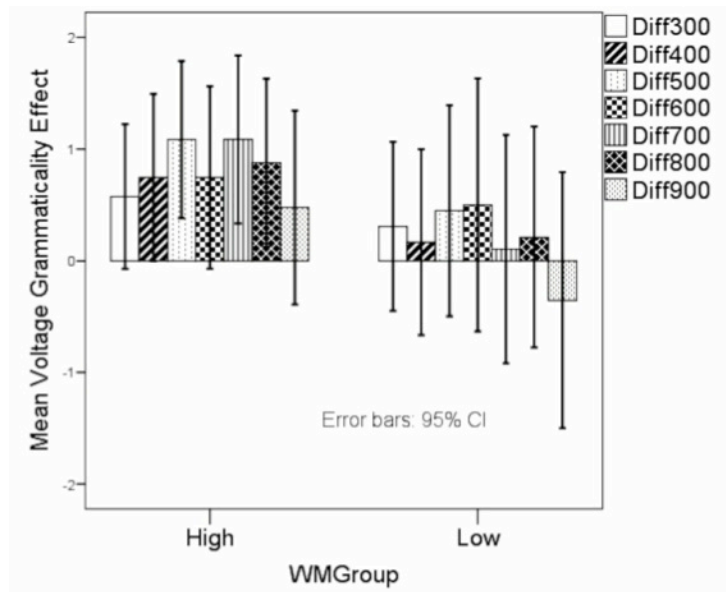
Figure 3.18: Frontal central positivity by WM group.

Similarly for the proficiency effect, although visual inspection of Figure 3.19 below seems to indicate that the positivity is mostly carried by the High and Low proficiency groups, and not so much by the Mid proficiency group, no significant interaction between proficiency group and grammaticality was found ($F_{(2, 51)}$=0.879, p=0.21).
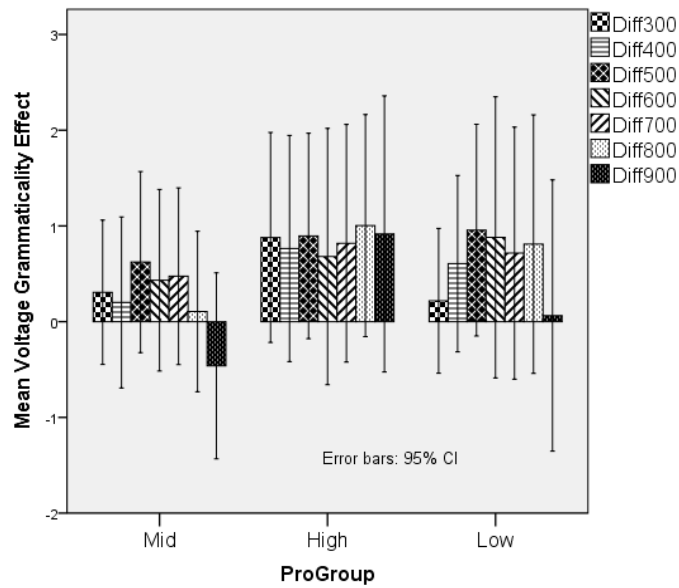
Figure 3.19: Frontal positivity by proficiency group

Furthermore, no significant associations were found between either WM score or proficiency score (Versant score) and the ERP measures, as the results of regression analyses show ($R^2=0.007$, $p>0.05$; $R^2=0.02$, $p>0.05$, respectively).

The Posterior Positivity

Lastly, I discuss the weak positivity discovered in the lower posterior region. As Figure 3.20 illustrates, the average waveforms for the grammatical and ungrammatical conditions overlap until around 900 ms after the violation.
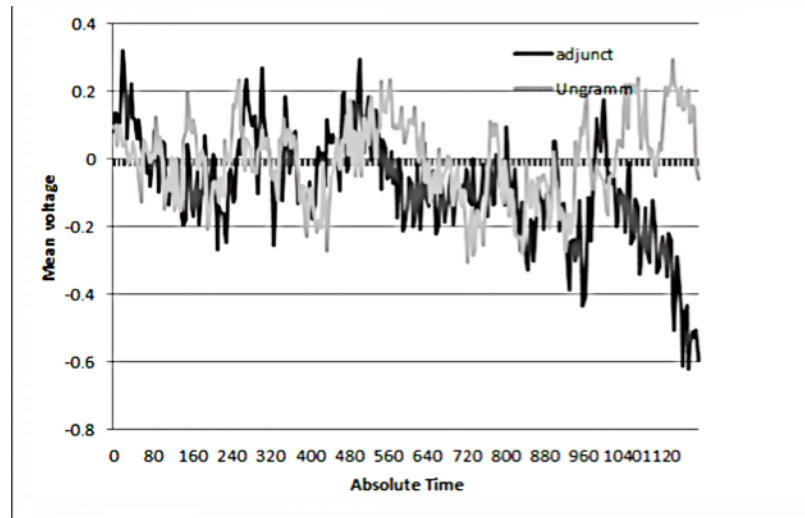
Figure 3.20: Posterior positivity waveforms

A very late positivity started to emerge between 950 ms and 1000 ms. This pattern is clearly different from that of a typical P600, which starts around 400-500 ms after the onset of the violation and peaks around 600 ms. Voltage data in the typical P600 time window from 600 ms on were submitted to a repeated measures ANOVA, which reported no significant time effect ($F_{(3, 53)}=0.862$, $p=0.462$) and no grammaticality effect ($F_{(1, 53)}=1.284$, $p=0.262$). A planned orthogonal contrast analysis showed that the positivity is significant only in the last time window of 900-1000 ms ($t=-2.238$ $p=0.029$), as illustrated in the difference voltage bar chart in Figure 3.21. Note that it is difficult to interpret such a late effect in relation to the P600.
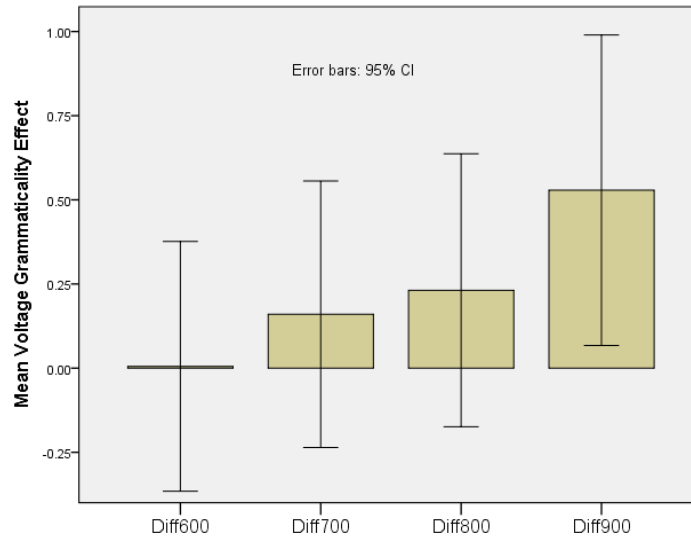
Figure 3.21: Posterior positivity difference in 600-1000 ms time window

When working memory is factored into the picture, a significant between-groups effect was found by an ANOVA (F (1, 52) =4.63, p=0.036), indicating that the two groups generated different brain responses regarding this late positivity. This is for four time windows combined (600-1000 ms). As reflected in Figure 3.22, the two groups differ in that there was hardly any effect for the high working memory group for the earlier time windows, while the low working memory group showed a sizable positivity in the window of 700-800 ms. However, in the later time windows, from 900 ms on, the amplitude of the positivity of the high working memory group increased sharply. Yet, no interaction between WM and proficiency and grammaticality was found (F (1, 52)= 0.181, p= 0.672).
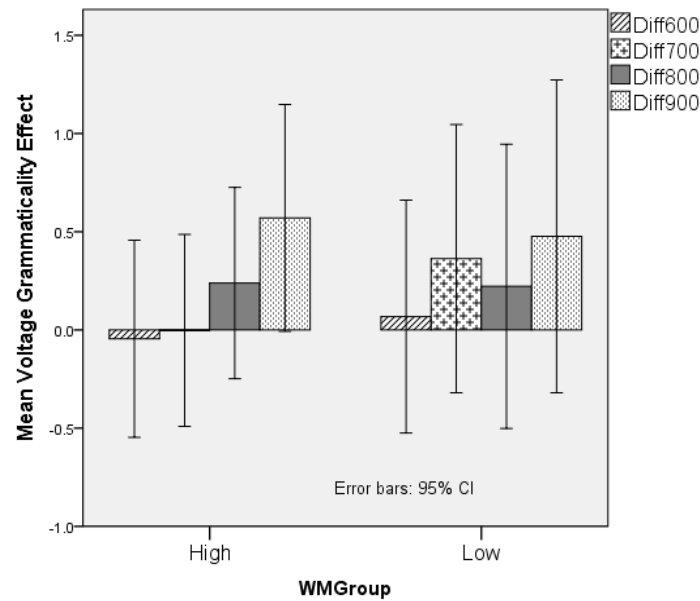
Figure 3.22: Late posterior positivity by working memory group

Turning now to how proficiency affects the late posterior positivity, an ANOVA didn't find any significant interaction between proficiency group and the grammaticality effect ($F_{(2, 51)}=0.19$ $P=0.828$). There was also no between-group effect ($F_{(2, 51)}=0.21$, $p=0.811$). Figure 3.23 below illustrates the posterior positivity for the three proficiency groups. We see that the high proficiency group has the largest positivity throughout the time windows. The positivity of the low proficiency group, on the other hand, is delayed, visible only in the last time window (900-1000 ms). Regression analyses indicated no significant correlation between either Versant score ($R^2=0.015$, $F_{(2, 53)}=0.012$, $p=0.9$) or working memory score ($R^2=0.039$, $F_{(1, 53)}=0.083$, $P=0.73$) and voltage changes.
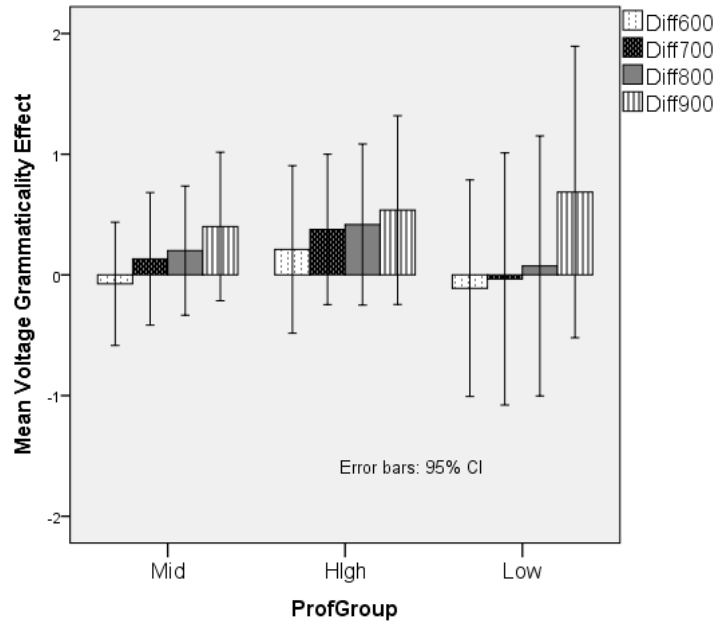
Figure 3.23: Posterior positivity by proficiency group

3.4   Discussion

Before the above results of the learners can be discussed and compared to those of the native speakers, it is necessary to revisit the goals, rationale, and predictions of the current study. This experiment set out to explore how late L2 learners process filler-gap dependency structures and to test the claims of the SSH, which maintains that the L2 learners adopt fundamentally different strategies from native speakers during online parsing by relying exclusively on semantics and other non-structural cues. Specifically, the SSH stipulates that the learners are incapable of building syntactic structures in real time with details such as the abstract trace posited for filler-gap dependencies. The current experiment replicates Hestvik et al. (2007,

2012), in which such a trace was found to be posited by native speakers at the gap

position in the following stimuli sentences:

(24). *The zebra that the hippo kissed (purported gap position) _the camel_ on the nose
ran far away.

If the L2 learners are similar to the native speakers and posit a trace after the

verb _kissed_ in (24), the following extra NP _the camel_ would conflict with the trace

already occupying the object position and cause a phrase structure violation, typically

indexed by the ERP components ELAN/LAN. A P600 suggesting reanalysis or

integration difficulty for the same violation is expected to follow as well, as was

produced by the native speakers. If, however, the SSH is correct about L2 learners'

inability to posit syntactic traces online, neither the ELAN nor the LAN should be

generated. In addition, if L2 learners indeed use semantics/lexical information to guide

their online parsing, as the SSH stipulates, a N400 should be obtained, suggesting the

filler-gap dependency is formed with meaning-based cues such as verb

argumentation/thematic role information. Furthermore, the SSH states that although

L2 performance varies by factors such as proficiency, L1 transfer, and the learners'

working memory capacity, the underlying L2 parsing strategies don't change due to

these factors. The current study also evaluates this particular claim by examining how

proficiency level and working memory capacity interact with learners' brain

responses. If no significant correlation is obtained between these factors and the

learner's brain responses, then the SSH's position would be supported. Otherwise, it

would suggest that L2 parsing is modulated by individual factors and could potentially become native-like.

3.4.1 Grammaticality Judgment Task and behavioral results

The current study included an off-line grammaticality judgment task to test the subjects' grammatical knowledge of filler-gap dependencies as in relative clauses. The participants were asked to rate the sentences such as the illicit (24) and its various grammatical counterparts. Both the native speakers and the L2 learners were able to distinguish between the well-formed and ill-formed FG dependency structures and judged the ungrammatical items as highly unacceptable. Thus, the L2 learners showed that they have the appropriate grammatical knowledge to process long-distance FG dependencies in an off-line context. Although the L2 group as a whole is not as sensitive to the violations as are the native speakers, it was found that as their proficiency level increased, the grammaticality ratings became more native-like. Statistical analysis showed that when the low proficiency L2 speakers were excluded from the data analyses, there was no difference between the L1 and L2 ratings.

In addition, the accuracy rate of the comprehension questions in the ERP experiment was also analyzed. A strong positive correlation between proficiency level and accuracy rate was found, and a relatively weaker correlation between working memory capacity and accuracy rate was also found. Both findings were expected and confirm the validity of the WM/proficiency tests. The overall accuracy of the L2 learners is relatively low (79%), due to the difficult and resource-demanding materials

and comprehension questions used in the ERP task. Note that even the native speakers were only 86% accurate. Crucially, statistical analysis reported no significant difference between the native speakers and the L2 learners. It is also observed that the L1 and L2 participants produced similar accuracy patterns in regard to the conditions; namely, both found the OBJECT condition questions to be the hardest. To summarize, proficiency-matched L2 participants were found to be no different from the native speakers in both their off-line and online behavioral responses.

3.4.2 Voltage results

The ERP patterns of the L2 participants, however, are dramatically different from those of the native speakers. First, there was no ELAN nor LAN found for the L2 speakers in the anterior region of both hemispheres, regardless of their proficiency level and working memory capacity. In addition, no P600 was obtained for the L2 speakers either. Although there was a very late positivity obtained in the posterior region, it didn't become statistically significant until around 900 ms, hence it was too late to be considered a P600. Therefore, both components found among the native speakers indexing the phrase structure violations and reanalysis are missing in the L2 learners' brain responses.

More importantly, a negative-going potential at the central-posterior region was detected starting at 250-300 ms after the offending extra noun phrase and peaked during the 500-600 ms window. As explained earlier, this potential fits into the profile of the N400 generated among L2 learners/bilinguals for tasks involving picture

126

judgments and auditory stimuli. Based on the off-line and behavioral data, it is very unlikely that the L2 participants didn't recognize the violations in the UNGRAMM condition. The absence of the ELAN/P600 and the N400 in combination thus seems to suggest that the L2 learners simply didn't treat the violation as syntactic in nature. Instead, what possibly happened was the L2 learners formed the filler-gap dependency based on semantic-driven heuristics, specifically, verb argumentation requirement, without positing the abstract trace, which is precisely what the SSH predicts.

Furthermore, when these components were analyzed together with proficiency level and WM capacity, no reliable interactions nor correlations were found, suggesting that these two factors have no affect on L2 learners' ERP measures. In fact, visual inspection (pre-statistical analyses) revealed that the N400 was more prominent among those who are at a high proficiency level and have a greater working memory capacity. If it were true that a high proficiency level and large working memory promote more native-like processing patterns, then the opposite pattern would have been obtained. This pattern contrasts sharply with that of the behavioral results reviewed above, which indicated that the L2 performance is indeed affected by individual differences. Taken altogether, the combined brain response patterns and the lack of a modulation effect from WM capacity/proficiency render direct support for the claims of the SSH.

What is relatively novel about the ERP results obtained in Experiment I is the frontal central positivity. As described above in section 3.3, this component was found largely from 400-800 ms, approaching significance in the 400-500 ms and 600-700 ms

time windows, and was significant at the 500-600 ms (peak) and 700-800 ms time windows. It emerged after the onset of the N400, and its amplitude appears to correspond with that of the N400 as well. A similar frontal positivity component has been reported in the literature in association with the N400. It was named the frontal Post-N400 Positivity (PNP) (See Van Petten & Luka, 2012 for a detailed review), and two situations appear to give rise to such a potential. In one context the frontal PNP is elicited by a highly unlikely *final* word (high vs. low cloze probability) (e.g., Fedemeier & Kutas, 2005). The other context is when lexical predictions are not confirmed (Delong, Urbach, Groppe & Kutas, 2011; Fedemeier, 2007). For instance, a frontal PNP was obtained when a highly expected noun appeared with a gender marking incongruent with that of its preceding article in Spanish, a morphologically rich language (Delong, Urbach & Kutas, 2005). It was also observed when an unexpected lexical item was encountered (Thornhill & Van Petten, 2012), as in sentences such as *On his vacation he got some much needed rest/sun* (from Thornhill & Van Petten, 2012). In this case, the NP *sun* elicited a frontal PNP following a N400 when compared to *rest,* because it was much less expected, although still congruent in that position. I argue that the frontal positivity obtained in the current experiment is best explained by the "unexpected lexical item" account, which is also in line with the above conclusion that the learners used semantic/lexical information to resolve the filler-gap dependencies.

It might be argued that the categorically different L2 brain responses could be explained as an L1 inference effect. However, as discussed in the previous section,

although Chinese *wh*-questions are in-situ, Chinese has an abundance of filler-gap dependency structures such as RCs and topic structures with overt, dislocated items. To posit that the categorically different L2 brain response is exclusively due to L1 transfer is to assume all Chinese FG dependencies are resolved non-structurally (i.e., the filler and gap are base-generated and linked to each other by semantic fit and verb argument specifications). However, such an assumption goes against both theoretical analyses of Chinese FG dependency structures (i.e., they are movement-derived and contain an abstract trace) and the empirical processing data reviewed above. Since the experimental stimuli don't involve *wh*-questions, the only cross-linguistic difference relevant in this experimental design is the directionality of the RC, which has been proven to have only a minor interference effect in a cross-linguistic study of RC processing (Hashimoto, 2009). If L1 transfer was indeed involved, it couldn't have caused such a dramatic effect as to change the fundamental nature of the processing mechanism. Lastly, if L1 transfer is assumed, then its interaction with other factors such as proficiency would be highly expected, as it has been found in the literature that L1 influence attenuates as proficiency increases (e.g., Hashimoto, 2009)). Such an interaction was not observed in the results of the current study. Thus, though there might exist some transfer effect, which could only be confirmed by testing another carefully matched group of L2 learners from an English-like language background in the same design, it could not by itself have been responsible for the categorical differences found between the L1 and L2 parsing profiles.

3.4.3 Conclusion

In trying to answer the research questions laid out at the beginning of the chapter in relation to the claims of the SSH, this chapter reported an ERP study testing whether L2 learners can process FG dependency structures with a phrase structure violation in the same way as native speakers do. The findings reviewed above suggest a qualitative difference between the L1 and L2 brain response patterns. Specifically, the two components indexing syntactic violations found for the native speakers were completely missing in the responses generated by the L2 learners, suggesting that the latter failed to posit abstract traces in their syntactic representations during online parsing of the target filler-gap dependency structures. Additionally, a reliable N400, typically found for semantic anomalies, was observed instead, indicating that the L2 learners employed a semantic/pragmatic-driven strategy to resolve the filler-gap dependencies. Lastly, I showed that while individual factors such as working memory and proficiency level positively correlate with L2 performance, neither of these two factors was found to affect the L2 parsing characteristics indexed by ERPs. The lack of correlations between these individual difference factors and the L2 parsing profile provides further proof that the L2 parsing mechanism is distinct from that of L1 and hence does not change with memory limitations and proficiency. Thus, the empirical findings of Experiment I fully support the claims of the SSH regarding complex long-distance dependency structures. In the next chapter, the focus will shift to another aspect of the grammar: morpho-syntactic processing at the sentence

130

Chapter 4

L2 MORPHO-SYNTACTIC PROCESSING

4.1 Introduction

The Shallow Structure Hypothesis (SSH) (Clahsen & Felser, 2006a, b)

extended the proposal of L2 shallow processing to the area of morpho-syntax as well[18]

(e.g., Neubauer & Clahsen, 2009; Silva & Clahsen, 2008; Clahsen & Neubauer, 2010).

In regard to grammatical agreement dependencies, the SSH attributes the well-known

difficulties of L2 speakers (e.g., Montrul, 2004; Montrul, Foote & Perpiñán, 2008;

White, 2003) to their inability to effectively use the relatively abstract morpho-

syntactic rules. Of particular interest here is their claim about verb inflectional

morphology, namely, that while native speakers decompose regular forms such as

walked into stem + -*ed* affix by rule application, the L2 speakers rely exclusively on

semantics-based whole-form memorization (i.e., they memorize *walked* as one piece),

---

[18] In the earlier version of the SSH, it was proposed that morpho-syntactic processing
among L2 speakers may follow a native-like pattern, given that a high-enough
proficiency level is obtained, or that the first language of the learner shares a lot of
similarities with the target language (Clahsen & Felser, 2006b). However, such a view
was changed in the more recent publications by SSH proponents (Silva & Clahsen,
2008; Neubauer & Clahsen, 2009; Clahsen & Neubauer, 2010) such that much like
long-distance dependency sentence processing, L2 morpho-syntactic processing is
independent of the influences of individual factors like L1 transfer and proficiency,
and hence is fundamentally different from that of native speakers as well.

due to their incomplete access to morphological structure and their inability to decompose complex forms. To assess the validity of such a claim, this chapter presents another ERP study aiming to examine how advanced Chinese L2 speakers of English process past tense violations. The specific research questions being asked are:

(1)   Is there a fundamental difference between the L1 and L2 online parsing of morpho-syntax in context (i.e., at the sentence level)?

(2)   Can L2 learners decompose morphologically complex words during real-time comprehension?

(3)   How do proficiency level and L1 transfer affect late learners' online parsing of morpho-syntax?

This chapter is organized as follows. Section 4.2 provides some background information on the L1 processing of inflectional morphology and morpho-syntax, focusing on the issue of morphological decomposition, which is of great importance to the SSH claim under investigation. Section 4.3 begins with an overview of L2 morpho-syntactic processing. Since the overwhelming majority of L2 morpho-syntax studies are about grammatical agreement processing, the findings to date are first summarized and discussed in relation to the SSH. Next, the SSH claim about L2 tense morphology processing is explained in more detail, accompanied by a discussion of the related experimental findings. The third part of this section spells out the predictions of the SSH in terms of the ERP indexes. Section 4.4 reports the design, implementation, and results of Experiment II, in which the Chinese L2 speakers of English showed native-like processing patterns suggesting the online use of

132

morphological decomposition. Lastly, a summary section discusses the implications of these findings.

4.2   Morphological Decomposition in L1 Morpho-syntactic Processing

One of the important issues in the field of morphology and morpho-syntactic processing concerns how morphologically complex words are stored and processed. In particular, how verb type (regular vs. irregular) interacts with the processing of tense morphology (especially for English) has been under intense investigation for the past few decades (see Gor, 2010 for a review). One group of scholars adopts a dual-route approach and argues that the regular forms are formed by rule, namely, attaching the *–ed* suffix to the stem (e.g., Penke et al, 1997; Pinker & Prince, 1994; Pinker & Ullman, 2002). Under this view, the stems of the regular forms are stored separately from the tense suffix. In contrast, the irregular forms are stored whole in memory, and the online processing of these forms does not involve any decomposition. The other view, the single-route view, maintains that all inflected forms are processed by some form of pattern association (i.e., sing-sang, ring-rang, spring-sprang all follow the same phonological pattern) (e.g., McClelland & Patterson, 2002; Bybee, 1995). For the past few decades, these two opposing views have often been assessed via the test of a frequency effect, which is indicated by a reduced response time or higher accuracy rate for more frequent words/stems in, for example, a lexical decision task. If the dual-route model were correct, then a whole-form frequency effect should be observed for the irregular forms only, but not for the regular forms, except for the very

133

few which are considered so highly frequent that it has been proposed that they too are memorized whole (Gor, 2010). However, although frequency effects have been widely observed for irregular forms, the findings for regular forms are inconclusive. This is partially due to an inherent methodological problem of frequency effect testing, namely, the frequency of the stem (lemma) can't be perfectly matched with the frequency of the whole form (stem+suffix) (e.g., Baayen, Dijkstra & Schreuder,1997). In addition, it has been argued that English inflectional morphology is quite different from many other languages with rich morphology in that the English regular forms are the norm and are more frequent than the irregular forms. It is thus difficult to judge whether the observed difference between the verb types is truly due to the verb types themselves, and not to other reasons.

For languages with rich morphology, the distinction between regular and irregular verbs is less clearly drawn, and there are often multiple classes of verbs with different inflectional patterns. Research based on verb regularity thus seems implausible. Due to the methodological issue with frequency tests and cross-language differences, the research focus in recent years has shifted to how decomposition vs. full-form access interacts with various factors such as verb class, frequency, and the degree of stem allomorphy (Gor, 2010). With decomposition taking center stage, it was proposed that all inflected forms are decomposed (e.g., Taft & Forster, 1975; Marslen-Wilson & Tyler, 2007). For English, this means that the irregular forms are decomposed as well (e.g., Halle & Marantz, 1994; Embrick & Halle, 2005), contrary to the dual-route view under which only the regular forms are decomposed. For

example, the Distributed Morphology account (DM, Halle & Marantz, 1994) states

that inflected forms are formed by combing the stem (or Root, see Halle & Marantz,

1994) with an abstract past tense morpheme. Thus, *ate* consists of √ eat + [past]. At

spell-out, morpho-phonological rules (partially idiosyncratic) are applied to change the

stem and realize the suffix as zero for the irregular verbs. The regular verbs, however,

will surface as the Root + *ed*. Experimental evidence in support of this full

decomposition view was found by a wide range of studies such as Kielar, Joanisse and

Hare (2008), and Pastizzo and Feldman (2002), in which a priming effect between

irregular forms and their base forms was observed, suggesting decomposition occurs

for these forms just like the regular forms. Additional neurophysiological findings

(e.g., Kielar & Joanisse, 2010; Stockall & Marantz, 2006; Solomyak & Marantz,

2010) also support the hypothesis that decomposition is involved in both forms, at

least during the initial stages of lexical access (Amenta & Crepaldi, 2012). Crucial to

this thesis is the ERP study conducted by Hestvik et al. (submitted), which the current

study replicates for L2. In this study, subjects listened to sentences such as *Yesterday

I walk to school*, or **Yesterday I eat a banana*. The rationale is that as the parser

processes the time adverb *Yesterday*, the tense feature is anticipated and the correct

tense marking is actively looked for by the parser. The violation is thus recognized

when the parser encounters the illicit bare stem form. If the irregular form is indeed

stored whole, then the violation should be recognized as a lexical error (i.e., an

incorrect word), resulting in an N400 indicative of a lexical retrieval error (Ullman,

2001 a, b). Contrary to this prediction, Hestvik et al. (submitted) found a LAN

component, typically observed for morpho-syntactic violations (e.g., Friederici, 2002),

for both regular and irregular forms, suggesting that both forms were decomposed

(details of this study will be discussed in 4.3). Although no consensus has been

reached regarding the issue of verb type and decomposition[19], in this thesis the full

decomposition account is assumed to facilitate the comparison between L1 and L2

processing patterns.

4.3   L2 Morpho-syntactic Processing, the Use of Morphological Decomposition, and the SSH

4.3.1   Overview of L2 morpho-syntax processing

The L2 acquisition of morpho-syntax has always received attention because

one of the most observed deficiencies in L2 performance pertains to the use of

inflectional morphology, in particular, grammatical agreement (Montrul, 2004;

Montrul, et al., 2008; White, 2003). On the processing front, a body of L2 research has

also indicated that L2 speakers are less sensitive to morphological cues in online

comprehension and often leave out the markings in production (Jiang, 2007; Dewaele

& Veronique, 2001; Franceschina, 2005). Before addressing the issue of the L2

processing of tense morphology and use of decomposition, an overview of L2

morpho-syntax processing will be given to establish a general understanding of the

---

[19] There is also experimental evidence in support of the Dual-Route view (i.e., Stanner, Neiser, Hernon & Hall, 1979; Weyerts et al., 1996; Rodriguez-Fornell et al., 2002).

field. Although many of the findings reviewed deal with grammatical agreement, they are very relevant because they also directly address the research question of the fundamental L1-L2 difference and thus have important implications for the evaluation of the SSH. As in the previous chapter, the previous ERP studies are given special attention.

In general, the L2 processing research on morpho-syntax, especially gender and number agreement, appears to show that L2 speakers are more successful in attaining native-like patterns in this area than in other aspects of grammar (i.e., FG dependencies), at least L2 speakers with higher proficiency (see van Patten & Tokowicz, 2010; Tolentino & Tokowicz, 2011 for a review). In addition, the individual factors such as proficiency and L1 interference seem to play a more significant role in modulating L2 processing behavior (Kotz, Holcomb & Osterhout, 2008; van Patten & Tokowicz, 2010; Tolentino & Tokowicz, 2011). To begin the review of some representative studies, it is helpful to know that a few ERP signatures have been reliably elicited during morpho-syntax processing. For instance, a P600 is expected at the offending item for morpho-syntactic violations (e.g., Barber & Carrieiras, 2005; Hagoort, 2003), and sometimes it is preceded by a LAN component (Gunter, Friederici & Schriefers, 1997; Molinaro, Vespignani & Job, 2008). Such a bi-phasic (LAN+P600) pattern was obtained with native speakers of Spanish in an ERP study carried out by Dowens et al. (2010, 2011), in which number and gender agreement violations were presented visually in two positions: within and across phrases. The within-phrase agreement stimuli, such as *La suelo está plano y bien

137

acabado (The$_{\text{fem-sing}}$ floor$_{\text{masc.-sing}}$ is flat and well finished) are less resource

demanding as the agreeing features are linearly close to each other, while across-

phrase agreement might require more working memory, as in * *El suelo está plana y*

bien aca bado (The$_{\text{masc.-sing}}$ floor$_{\text{masc.-sing}}$ is flat$_{\text{fem.-sing}}$ and well finished). The second

language subjects in this experiment were highly proficient native English learners of

Spanish. They generated a similar LAN+P600 ERP component for the within-phrase

agreement violations, but only a P600 for the across-phrase agreement violations,

suggesting that L2 working memory demands may constrain parsing ability. There

were also differences between the L1 and L2 in terms of ERP effect size and onset

timing, such that the L2 speakers were slower in responding to the agreement

violations and showed a smaller effect size. In a follow-up study identical in design

but with proficient Chinese learners of Spanish, Dowens et al. (2011) observed a

slightly different pattern: only the P600 was obtained with the Chinese subjects, and

there was no difference between the within- and across-phrase conditions. Note that

both Chinese and English are typologically different from Spanish in terms of

agreement features, but Chinese is much more so since the language is

morphologically deprived and has no overt morphological agreement of any kind.

Dowens et al. (2010) and Dowens et al. (2011) conclude that the L1-L2 processing

differences are not qualitative in nature, but are affected by a number of factors,

mostly L1 inference, working memory constraints, and proficiency.

     The significance of proficiency in L2 online processing has been highlighted in

L2 morpho-syntax processing, as shown in the findings of Mclaughlin et al. (2010), in

which a few longitudinal studies on various aspects of L2 grammatization were reviewed. Most of the findings revealed a clear shift from non-native-like processing patterns to native-like ones as the participants progressed in proficiency. In particular, Mclaughlin et al. (2010) reported that while some of their low-proficiency English-speaking learners of German produced an N400 or an N400 followed by a P600 in response to agreement violations, the high proficiency group showed a P600 just as the native-speaker controls did. A similar transition was observed, although slower, in another number and gender agreement study with English-speaking learners of French, where a more significant L1 interference was expected.

Morgan-Short, Sanz, Steinhauer and Ullman (2010) investigated online L2 morpho-syntax processing by using an artificial language and considering the type of input given in L2 acquisition (i.e., input by explicit instruction in a traditional classroom setting vs. input from immersion instruction) and revealed some interesting L2 processing patterns. To use an artificial language as the L2 in an experimental setting has a few advantages. For example, the factors of proficiency level and L1 interference can be better controlled than when the L2 is a natural language. The quality of the second language exposure and the type of instruction (implicit vs. explicit) has also been extensively studied in the field of SLA and has been suggested to affect ultimate attainment in both performance and processing (see section 5.2 for details). Morgan-Short et al. found that for the low proficiency group, gender agreement violations yielded the N400 for the explicitly trained group but not the implicitly trained group. However, among the high proficiency group, noun-adjective

139

agreement violations elicited an N400 for both the implicitly taught and explicitly taught groups, while the noun-article agreement violations elicited a P600 for both groups. The results from Morgan-Short et al. (2010) thus painted a more complex picture in which linguistic characteristics of the target structures, exposure type, and proficiency are all in play to modulate the state of L2 processing.

The data reviewed so far contradict the SSH but are largely in accordance with the "one-system" accounts, which explain L2 processing with L2-specific factors such as proficiency, L1 interference, resource and memory constraints, and the interactions among these factors. For example, according to McDonald (2006) and Bialystock and Hakuta (1999), cognitive resources such as attention and working memory are not as readily available to L2 learners, especially those whose proficiency level is low. Learners are thus not as sensitive to morpho-syntactic cues, because they have a weaker decoding ability, slower processing speed, and smaller working memory reserve when they are working in the target language. As we see in the results of Dowens et al. (2010) and Dowens et al. (2011), the distance between the agreeing elements in the dependency, which indexes the working memory demands, impacts L2 speakers' processing more than that of native speakers. Furthermore, the cognitive constraints account predicts that as the L2 speakers gain proficiency, more online cognitive resources will be freed to processing and the online parsing pattern should become more and more native-like. Such predictions are also born out in the McLaughlin et al. (2010) study reviewed above.

In addition, the combined results from both Dowens et al. (2010) and Dowens et al. (2011) highlight the role of L1 interference, which has increasingly been seen as relevant to L2 morpho-syntactic processing (e.g., Frenck-Mestre, Foucart, Carrasco-Ortiz & Herschensohn, 2009; Sabourin & Stowe, 2008; Tokowicz &MacWhinney, 2005). It is also considered in the current experiment, as the target language's (English) tense morphological features are absent in the native language (Chinese). The L1 influence in L2 acquisition has been formally addressed by many researchers (e.g., Selinker, 1969, 1972; Gass & Selinker, 1992). One of the most prominent processing models that addresses L1 influence is the Competition Model (CM) (MacWhinney & Bates, 1989), and its latest installment the Unified Competition Model (UCM) (MacWhinney, 2008). According to the CM/UCM, sentence parsing is guided by various linguistic cues or features such as word order, morphological markings, and semantic features (i.e., animacy), which differ in weight across languages. For example, gender agreement is a prominent feature in Spanish that would weigh significantly more as a linguistic cue than in English. Thus, while native speakers of a morphologically rich language such as Spanish tend to rely on agreement markings to parse, speakers of other languages such as English and Chinese don't tend to rely on this information in online comprehension. When adult learners acquire a second language, they are heavily influenced by their L1 parsing strategies and rely on them to process the target language, at least in the beginning. L1 negative transfer or inhabiting effects obtain when (1) both the L1 and L2 share certain grammatical features but use them differently (i.e., SVO vs. SOV word order), and (2) when the L2

cue or grammatical feature is absent in the native language. In the first case

competition between different uses of a cue is likely to result and cause more

difficulty in processing. In addition, the transfer effect is also modulated by cue

weight. A prominent, consistent cue in the target language (i.e., gender agreement in

Spanish) is predicted to be easier to acquire than those cues that are either less

prominent or inconsistent. The more native-like processing pattern generated by the

native-English L2 speakers of Spanish in comparison to that of the Chinese L2

speakers in Dowens et al. (2010, 2011) thus can be explained by the UCM, as Chinese

has very little surface morphological markings (Li & Thompson, 1981).

　　　　While the L2 speakers appear to be more successful in attaining a native-like

profile for grammatical processing, whether fundamentally different parsing routines

are adopted by L1 and L2 speakers in tense morphology processing is much less clear.

In the next section, I discuss L2 tense processing and the issue of whether the

decomposition of morphologically complex words is actively used in L2 online

comprehension.

4.3.2　The SSH and L2 tense morphology processing

　　　　Relatively less research has been carried out examining the processing of tense

morphology in the late L2 setting. While there is little doubt that the learners can

achieve near-native performance with inflectional morphology, they lack in speed and

automaticity when it comes to morphological and morpho-syntactic rule application

(Gor, 2010). The SSH maintains that the L2 parser is significantly less sensitive to

morphological structure and under-uses morphological rules (Neubauer & Clahsen, 2009). Specifically, the L2 parser resorts to full-form lexical memorization rather than morphological decomposition in processing inflectional morphology. In the case of the English past tense, for instance, the SSH states that while native speakers decompose regularly inflected forms into the stem plus –*ed* suffix, L2 speakers tend to memorize these forms as single units, regardless of their proficiency level and L1 background. In this way the SSH follows the dual-route model (Pinker & Prince, 1994: Ullman, 2001a, 2001b, 2004, 2006), which is one of the two contending theories in the field of L1 morphological processing reviewed above. Since the dual-route model stipulates that the regular forms are generated by rule and the irregular forms are directly accessed as full lexical items, the SSH asserts that both L1 and L2 speakers process inflected irregular forms via full-form memorization.

To illustrate the precise claims of the SSH, we will review two experiments cited by Clahsen & Neubauer (2010) to support their claim that L2 speakers use storage instead of decomposition to process inflectional morphology. Neubauer and Clahsen (2009) reported findings of a masked priming experiment, in which L1 (native German speakers) and L2 (Polish L2 learners of German) groups produced the same priming pattern for irregulars. The results indicated that both the L1 and L2 speakers used full-form memorization. However, for regular participles with the suffix –*t*, only the L1 group showed a reliable full-stem priming effect, which is indicative of decomposition. The L2 speakers were therefore claimed to have utilized strategies other than decomposition, such as full-form memorization, for the regular forms.

143

Similar results were obtained in a masked priming study done by Silva and Clahsen (2008) with English. However, it has been argued that the above patterns could be interpreted differently (Gor, 2010). In a masked priming study, the prime item is first displayed very shortly (60 ms), followed by the target item. The subjects are asked to carry out some judgment tasks (i.e., a lexical decision task), and the response time is recorded. The priming effect is reflected in a shorter response time. Gor (2010) suggests that it is possible that the stimuli presentation time of 60 ms was too short for the L2 learners (but not for the L1 speakers) to fully decompose the inflected form. To address this concern, Clahsen, Schutter & Cunnings (2013) conducted yet another masked priming study based on Silva and Clahsen (2008) with a group of advanced Arabic learners of English, and an additional 200 ms was included between the prime item and the target item.[20] The same non-native pattern was obtained with the L2 learners and was taken as evidence that the learners are not able to use decomposition rules effectively with regular forms.

Another study reported by Neubauer and Clahsen (2009) also involved a lexical decision task. In lexical decision studies, subjects are instructed to determine whether or not a presented item is a word. Shorter production latencies or lexical decision times for high-frequency forms are generally interpreted as effects of memory storage. It was discovered that the L2 learners exhibited a frequency effect (i.e., shorter response time for highly frequent words) for both regulars and irregulars, while

---

[20] The subjects saw a black screen between the prime item and its target for 200 ms.

144

the L1 group did so for irregulars only, suggesting that while L1 processing relies on memory storage for the irregular forms only, L2 processing relies on it for both irregular and regular forms. However, Neubauer and Clahsen (2009)'s finding is contradicted by results from two studies using the same paradigm (Portin, et al., 2007; Portin, Lehtonen, & Laine, 2007), in which bilingual Swedish-Finnish speakers decomposed inflected Finnish words across the entire frequency range, indicating a L2 sensitivity to morphological structure. These two studies also found that the Chinese speakers of Swedish, on the other hand, used whole-word representations for both high and low frequency words, suggesting that L1 background is a more critical factor in determining L2 processing strategies. Additionally, Gor and Cook (2010) conducted a lexical decision priming task using auditory stimuli on how highly proficient English-speaking learners of Russian process inflected infinitives. They found that the L2 learners fully decomposed both regular and irregular forms.

In addition to such mixed results, a common problem among all of these studies is that the inflected forms were presented in isolation, which is not the natural way in which they are processed (Paradise, 2004). This issue in part motivated another study on the L2 processing of regular and irregular verb morphology conducted by Pliatsikas & Marinis (2011). In this self-paced-reading study, highly proficient Greek speakers of English read sentences containing *morphological violations* (i.e., an irregular stem received a regular marking and a regular stem received an irregular marking, see section 4.5 for details). It was found that the L2 learners could decompose the regular forms like the native speakers. Their findings contradict the

SSH and suggest that stimuli presentation in this context may affect the experimental results. Thus, the current study adopts an experimental paradigm in which the inflected forms are presented in naturally occurring contexts rather than as "mixed paradigm violations". In addition, the method of ERP will be used due to the various advantages explained in the chapter on Experiment I.

4.3.3 The Event Related Potential (ERP) method and tense morphology processing

Before moving on to the details of Experiment II, let us examine some important correlations between ERP components and the various processes of online morphological parsing. As mentioned above, the LAN effect is obtained when morpho-syntactic rules are violated and has been found to indicate problems with morphological processing (e.g., Friederici, 2002, Kluender & Kutas, 1993, Rösler, Putz, Friederici & Hahne, 1993). In regard to tense morphology, LAN-type effects were found in the *morphological rule violation* paradigm, in which incorrect affixes are attached to the stems, when the *regular* affixes are incorrectly applied to the irregular stems (e.g., *bringed) (Weyerts et al., 1997; Penke et al., 1997; Morris & Holcomb, 2005). It was also discovered that such LAN effects vary in scalp location, such that sometimes it is found on both the left and the right side of the brain (e.g., Gross et al., 1998; Rodriguez-Fornells et al., 2001). This LAN (or AN) pattern is taken to indicate that the regular forms are decomposed, and the affix attachment is rule-based. In another paradigm, the *contextual violations paradigm*, subjects are presented

146

with stimuli containing correct tensed verb forms, but in an inappropriate context, as in *Yesterday I *walk/walked to school*. The LAN effects were also obtained in this experiment for the regular forms (Newman, Ullman, Pancheva, Waligura & Neville, 2007, Stenhauer & Ullman, 2002a). The N400 component is typically associated with semantic anomalies, and usually peaks at 400 ms after the offending word (Kutas & Federmeier, 2000, Van Petten et al., 1999). In addition, it has been found to reflect declarative memory processes such as lexical access (Ullman, 2001b). For instance, Weyerts et al. (1997) conducted a study on German noun inflections and found the N400 when irregular suffixes were applied to regular stems, although sometimes such a component was absent in other similar experiments with different languages (Penke et al., 1997; Gross et al., 1998). Lastly, the P600 was also reported in relation to morpho-syntactic violations (Dowens et al., 2010; Carreiras, Pattamadilok, Meseguer, Barber & Devlin, 2012). In particular, it was observed in the contextual violation paradigm (e.g., Newman et al., 2007; Stenhauer & Ullman, 2002a; Allen, Badecker, &Osterhout, 2003). However, like the N400, it is not always found in the morphological rule violation studies (Gross et al., 1998; Penke et al., 1997;Weyerts et al.).

Very few ERP studies have been conducted on the L2 processing of tense inflectional morphology (for a review, see van Hell & Tokowicz, 2010). Among these studies, Hahne et al. (2006) was cited to support the SSH. They examined the ERP effects generated by both native and late L2 (first language Russian) speakers of German when regular German participles were applied incorrectly to irregular stems

147

(Hahne, Mueller & Clahsen, 2006). While the L1 speakers generated an earlier and focal LAN, the L2 speakers produced a bilateral anterior negativity followed by a P600 effect. Hahne et al. also tested German noun plural forms in the same paradigm. For the corresponding incorrect noun plural forms, the L1 group produced a focal LAN followed by a P600, while the L2 speakers had only a reduced P600 effect. It was concluded that the L2 speakers are less sensitive to morphological structure and under-use decomposition for regularly inflected forms. However, objections can be raised to such an interpretation. First, the precise nature of anterior negativity (whether it is bilaterally distributed, for example) is not well known; sometimes it is considered qualitatively the same as the left focused LAN (Friederici, 2002). Furthermore, the bilateral/less focused LAN has been observed in ERP studies using the same paradigm with native speakers in various languages (Gross et al., 1998; Rodriguez-Fornells et al., 2001). It is therefore not convincing to claim that the L1 and L2 processing of morphology are qualitatively different in terms of the use of decomposition based solely on the results of Hahne et al.'s study. Secondly, it has been argued that the morphological violation paradigm itself is problematic (Ullman, 2001a) in that it creates unnatural processing conditions with mixed effects from both verb classes. Lastly, Hahne et al.'s study didn't accurately measure the proficiency level of the L2 speakers. All the L2 speakers were claimed to be highly proficient, but the only qualitative measure reported was a self-evaluation. Given that proficiency level may interact significantly with the use of native-like processing strategies, it is critical to

148

measure subjects' proficiency level in a more objective and consistent way and to analyze the results along with the other online measures.

4.4   Experiment II

To test the SSH's claim that L2 speakers rely mostly on whole-word memorization instead of decomposition when processing tense morphology, I replicated the ERP experiment carried out by Hestvik et al. (submitted) with a group of advanced Chinese speakers of English. The subjects listened to sentences such as *Yesterday I *walk /walked to school* and *Yesterday I *eat/ate a banana.*[21] The contextual violation paradigm provides a more natural processing condition and bypasses the potential problems discussed above in the single word/morphological violation paradigm. For sentences like **Yesterday I walk to school,* the listener expects a past tense marker by the time the verb is processed. If decomposition occurs in regular inflected verbs, the missing *–ed* would violate the affixation rule and a LAN-type effect would result. That is precisely what Hestvik et al. observed for the native speakers. Furthermore, Hestvik et al. found the same LAN effect for the irregular verbs, suggesting that decomposition occurs for both verb types. Their results support Single Route Models such as Distributed Morphology (Halle & Maranz, 1994)

---

[21] The testing material in Hestvik et al. is similar to that used by Newman et al. (2007), but was presented in auditory form.

and are not compatible with the predictions made by the SSH for native speakers, namely, a LAN effect for regular verbs and an N400 for irregular verbs.

Given this paradigm, let us now reconsider the SSH's claims: if L2 learners decompose much less or not at all and rely mostly on whole-word memorization, they should NOT produce any LAN effects for regular verbs. Instead, the SSH predicts that L2 speakers might produce an N400, indicative of violations of lexical access, if the regular verb past tense forms were indeed memorized. For irregular verbs as in *Yesterday I *eat/ate a banana*, it is expected by the SSH that the L2 speakers will generate an N400, since storage is the primary processing mechanism for both regular and irregular forms. The following table illustrates the possible outcomes of the current study and their implications for the predictions of the SSH.

Table 4.1: Possible outcomes for the L2 learners in Experiment II and the SSH predictions

	Regular form	Irregular form	Implication
Outcome 1	N400	N400	Supports the SSH
Outcome 2	LAN	N400	Partially contradicts the SSH
Outcome 3	LAN	LAN	Contradicts the SSH

As previously discussed, morpho-syntactic processing has been found to be affected by language similarities and differences (see Tolentino & Tokowicz, 2011 for

a review). Thus I have chosen L2 speakers of English whose native language is Mandarin Chinese, because Chinese is known to have very limited overt morphology (Jiang & Zhou, 2009; Ye, et al., 2006). Different from most of the Indo-European languages, the tense information in Mandarin Chinese is usually marked by aspect markers, adverbs, and noun phrases encoded with temporal information. Mandarin Chinese therefore is an ideal test ground to study how L1 knowledge and processing patterns affect those of a L2. Specifically, since the tense feature in English is represented in a radically different way in Chinese, the native Chinese speakers must overcome this linguistic difference or "negative" L1 interference to process English. However, if we obtain similar brain response patterns from these Chinese subjects, we will have very strong evidence that native-like processing patterns can develop in spite of L1 interference. Lastly, to assess the role of proficiency, the Chinese subjects' English skills were measured carefully via multiple measures. In addition to a language background questionnaire, all L2 subjects took the same Versant English Proficiency Test (Pearson, Plc), used in Experiment I, and the scores obtained were correlated with the behavioral and brain voltage data.

4.4.1 Participants

Thirty-two second language speakers of English (9 male and 20 female) from the University of Delaware participated in this experiment. One subject decided to not continue before the ERP recording started. The data of two other participants were excluded because of experiment errors. The average age of the remaining participants is 23 years (SD=2.56, Range=20-30). Fourteen of these subjects also participated in Experiment I. All of the subjects are native speakers of Mandarin Chinese and

151

acquired English mostly via classroom teaching. The average age of first exposure to English is 10.5 years (SD= 2.7, Range=6-15). None of them lived in an English-speaking country prior to the age of 16. By the time of the experiment, they had lived full-time in the U.S and/or other English-speaking countries for an average of 34 months (SD=16, Range=8-79 months). All subjects reported to have normal hearing and normal to corrected vision. None of them have any neurological impairment, and all except one are right- handed. They were paid $20 for their participation and gave informed consent before the experiment. Each L2 speaker's proficiency level was examined by administering the Versant English Proficiency Test (Versant English) reviewed above in section 3.3.4. All scored above 49 out of 80, with an average score of 62.6 (SD=7.4. Range=47-77). By the proficiency guidelines of the CEFR and its U.S. counterpart the American Council on the Teaching of Foreign Languages (ACTFL), all L2 participants are proficient speakers of English. Among them 21 have achieved an advanced level of proficiency by the guidelines established by the ACTFL, and 8 are considered intermediate-high to advanced-low speakers of English. Table 4.2 shows the alignment:

Table 4.2:   Versant English scores and L2 proficiency levels by the ACTFL and
              CEFR standards

Versant English Score	CEFR	ACTFL	Number of L2 participants in this experiment who reached this level
69-78	C1	Advanced-mid	6
58-68	B2	Advanced-low	15
47-57	B1	Intermediate-high	8

Additionally, two proficiency groups (High and Low) were created based on the Versant scores. The median score 62 was taken as the cut-off score, resulting in a High proficiency group with 16 subjects and a Low group with 13 subjects.

4.4.2 Materials

The stimuli for this experiment consisted of 320 simple declarative English sentences, in which 56 regular verbs and 56 irregular verbs were used. These are the same stimuli used in Hestvik et al. (submitted). All the verbs are monosyllabic in both their stem and past tense forms and are matched in a one-to-one manner for both stem and surface frequency. In addition, the regular and irregular verbs were chosen to be as close as possible in phonological complexity (see Newman et al. 2007 for details). A typical sentence starts with the time adverb *Yesterday*, followed by the pronoun *I*, then the verb and the rest of the sentence, e.g. *Yesterday I walked to school.* The verb

type (irregular vs. regular) and tense (past vs. present) creates a 2 x 2 design, illustrate in Table 4.3:

Table 4.3:   Tense x VerbType design with Yesterday

	Verb type	
Tense	Irregular	Regular
Past (grammatical)	Yesterday, I ate a banana.	Yesterday, I walked to school.
Present (ungrammatical)	*Yesterday, I eat a banana. (* Indicates ungrammaticality)	*Yesterday, I walk to school

Crucially, this study sets out to examine the ERP responses to the ungrammatical sentences, i.e., how the brain reacts to *Yesterday I ate/walked to school*, compared to **Yesterday I eat/walk to school* at and after the critical verb *eat/walk*. If the SSH is correct that L2 speakers store inflected verbs as single units, then we should expect that in the case of **Yesterday, I walk to school* the native speakers will generate a LAN after the verb *walk*, but the second language speakers will generate an N400. For irregular verbs as in **Yesterday, I eat a banana,* the SSH predicts that the L2 speakers will generate an N400 as well.

To control for the possibility that the difference in ERP signatures is due to the difference in tense (present versus past), instead of ungrammaticality, a third factor, Context (Null Context versus Yesterday) was added. The null context stimuli are all

154

grammatical, i.e., *I walk/walked to school* and *I eat/ate a banana*. I will compare the

ERP signatures to *I walk to school* to those of *I walked to school*, for example, to see

how the brain reacts to the tense effect alone. This allows us to confirm that the ERP

in the ungrammatical cases is truly elicited by the tense expectation violations and not

by the difference between tenses. Table 4.4 provides a null context sample set.

Table 4.4:   Null context sample stimulus set; * indicates ungrammaticality

Verb type		
Tense	Irregular	Regular
Past	I ate a banana.	I walked to school.
Present	*I eat a banana.	I walk to school.

4.4.3   Procedures

After signing the consent form, the subjects first took the 15-minute Versant

English test over the phone (for the details of the Versant English Test, see 3.3.4).

Then the ERP net was applied and the subjects were seated in a chair with a table top

in a sound-attenuating booth. On the table top there was a PST serial Response Box.

There were also a computer screen and two speakers placed on a desk three feet away,

facing the subjects. The participants were instructed to listen to each sentence and

determine whether the sentence described an event in the past, in the present, or didn't

make sense at all. They were asked to press button 1 (left-most) on the response box for past actions or events, button 2 (middle) for present actions or events, and button 3 (right-most) for the ungrammatical sentences. The numbers are clearly labeled on the response box. On the computer screen in front of the subjects, an image of the response box with helpful visual hints was displayed during the experiment.

The 320 sentences were distributed over 4 lists, and a given critical verb appeared only once in each list. In any given list, the verb appeared in one of the four possible combinations of tense and context. The stimuli were pseudo-randomized so that there were no more than 2 consecutive ungrammatical sentences in any given list. Verb type and grammaticality were counter-balanced across lists, and all subjects heard the sentences in the same order. The participants first practiced on a set of six trial sentences. They had to reach 75% accuracy with their judgments before they were allowed to start the experiment session. The 320 experimental sentences were presented successively, in 4 blocks of 20 trials. Between the blocks there was a brief pause. The subjects were offered a break after 2 blocks. The entire experiment took around 40-45 minutes to complete.

The stimulus sentences were delivered through the two speakers positioned on the desk in front of the subject. Every trial started with the sound of a bell, followed by a 300 ms pause, then the sentence itself. There was a 1000 ms pause after the critical verb. The subjects were given 2000 ms to respond. All subjects used their right hand to press the buttons. There was a 1500 ms pause between trials.

4.4.4 Data collection and analyses

For the behavioral data, the accuracy and reaction time (RT) of the behavioral responses were collected via E-Prime (Schneider et al., 2002), which was also used to present the stimuli. Both the accuracy and RT data were submitted to a mixed-factorial repeated measures ANOVA, with verb type (2), tense (2), and context (2) as within-subject measures. When the Chinese data were interpreted in comparison with the English results reported in Hestvik et al. (submitted), an additional measure group (2; Chinese vs. English) was used as the between-subjects measure.

For the voltage data, EEGs were recorded with 128 channels mounted on a net (Hydrocel HCGSN 100 v.1.0, Geodesics, U.S.A), the same system used in Experiment I. Eye movements and blinks were monitored with electrodes placed under each eye. The EEG was sampled continuously at a rate of 250Hz. Again, Cz online was used as a reference, and the electrode impedances were kept below 50kΩ. The continuous EEG was divided into epochs of 1400 ms for each trial by time-locking to the onset of the critical verb. Baseline correction was performed by using a 200 ms baseline period (before the onset of the verb) as a reference signal value. Following the baseline correction, epochs with artifacts were rejected, which resulted in the average removal of 21.3% of the trials per subject (SD=19.7%, Range=0.6%-45.9%). Subjects with more than 50% bad trials were not included in the ERP analysis. However, trials that were not responded to correctly were kept, in order to prevent excess loss of trials and power. The ERP average was computed for each condition, each subject, and each electrode site.

The analysis of the voltage data were conducted similar to Experiment I: the entire set of electrodes was divided into regions, such that an average over those regions could be computed and used as one of the dependent measures in the ANOVA. In the current study, a total of eight electrode sites were used based on the factors of ANTERIORITY (anterior vs. posterior electrodes), LATERALITY (left vs. right hemisphere, excluding the midline electrodes), and DORSALITY (inferior vs. superior electrodes). As Figure 4.1 below of the electrode distribution illustrates, the left anterior inferior region contains electrodes 18, 21, 22, 23, 25, 26, 27, 32, 33, 34, 38, 39, 40, 43, 44, 45, and 48 (red-colored), electrodes 127 and 128 monitored eye activity, and the left anterior superior region contains 7, 12, 13, 19, 20, 24, 28, 29, 30, 35, and 36. (grey-colored)
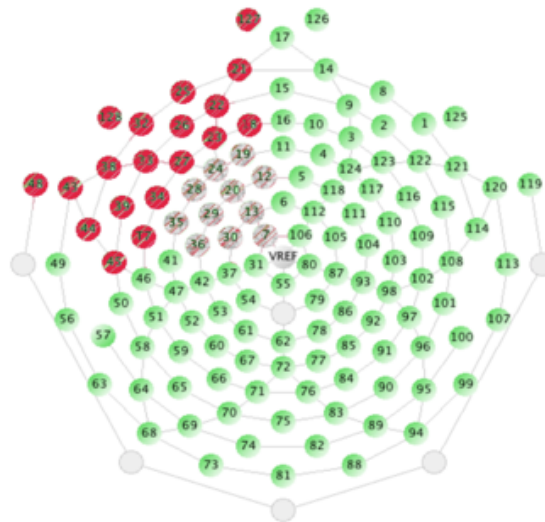
Figure 4.1:   The LAN region as shown with the highlighted electrodes

The 8 electrode sites were collapsed into 4 major regions of left/right hemisphere x anterior/posterior. A total of seven 200 ms time bins were constructed for the 1400 ms epoch starting from the onset of the critical verb. The mean amplitude over these time bins was computed for each electrode region for each subject and condition. The average amplitude obtained was submitted to a mixed- factorial repeated measures ANOVA with Time x Region x Condition (Tense, Verb type) as within-subject factors. To examine the difference between the L1 and L2 speakers, the data set of the native English speakers from the experiment conducted by Hestvik et al. was included, and another measure, Group, was created and used in the above ANOVA as a between-subjects factor. In the Left-Anterior region only, an additional factor DORSALITY (inferior vs. superior electrodes) was also included in the ANOVA for better resolution of the locus of the LAN effect. When appropriate, p-values were adjusted by using Greenhouse-Geisser (1959) correction for violation of the assumption of sphericity. In addition, significant interactions between conditions, time, and electrode region were followed up by either pairwise t tests or planned orthogonal contrast analysis. Lastly, two separate sets of analyses were carried out for the ERP data. The conditions with context (i.e., with the adverb *yesterday*) were analyzed separately from the conditions with a null context (i.e., without *yesterday*).

4.4.5 Results

4.4.5.1 Behavioral results

The Chinese group was highly accurate with their responses to the comprehension questions and yielded an average of 87.9% correct (SD=0.04, Range=0.18). A repeated measures ANOVA with Tense (2) x Context (2) x VerbType (2) showed that the accuracy rate for the Chinese group was affected by VerbType and Tense, but not by the presence or absence of context, as Table 4.5 illustrates.

Table 4.5:   ANOVA results for Chinese Group's accuracy rate

	F	p
VerbType	78.45	**
Tense	15.79	**
Context	--	--
VerbType * Tense	73.55	**
Tense * Context	26.32	**

Only significant F and p values are reported; * =p<0.05; ** =p<0.01

In addition, the Chinese speakers were most accurate with regular verbs in the past tense (96%) and least accurate with past tense irregular verbs (82%).

For the response time (RT), the Chinese group took an average of 722.63 ms (SD=213 ms, Range=862) to give their judgments. To understand the factors affecting RT, an ANOVA with Tense, VerbType, and Context as within-subject measures was run and yielded a significant main effect of Context ($F_{(1, 28)}$=11.73; $p<0.01$) such that it took longer to judge the null context than the adverb context stimuli. A

160

significant main effect of Tense (F (1, 28)=38.27; p<0.01) was also observed, such that the ungrammatical present tense took longer than the grammatical past tense stimuli, and also for Verb Type (F (1, 28)=7.84; p<0.01) such that irregular verbs took longer than the regular verbs. VerbType interacts with both Context and Tense (VerbType x Context: F (1,28)=8.36; VerbType x Tense: F (1,28)=56.32, both p values <0.01), such that it took longer for the Chinese participants to respond to the regular present stimuli (806 ms) than to the irregular past stimuli (583 ms).

A further investigation into how the grammaticality of the test sentences affected the behavioral data yielded interesting results. The performance of the Chinese participants was negatively affected by ungrammaticality, as expected, but especially so with the regular verbs. In the case of accuracy, the Chinese group made significantly more mistakes in the ungrammatical condition (84% vs. 90%; F (1, 28)=38.95, p<0.01), and this difference is more pronounced for the regular verbs, as indicated by an interaction effect (Grammaticality x VerbType; F (1, 28)=19.59, p<0.01), such that the responses to regular verbs in the ungrammatical condition were only 86% accurate, compared to 94% accuracy in the grammatical condition. Accuracy for irregular verbs, however, was almost the same regardless of grammaticality (83% correct for the ungrammatical condition and 84% accurate for the grammatical condition). In the case of the response times, the ANOVA showed that the Chinese participants were reliably slower only for the regular verbs in the ungrammatical condition, but not in the grammatical condition (761 ms vs. 684 ms; F (1, 28)=15.33, p<0.01).

In sum, the Chinese participants were highly accurate with their judgments. The performance of the Chinese group was affected by the Tense and VerbType factors, and the interaction between them and with Context. More importantly, the Chinese group was negatively affected by the ungrammatical condition in both RT and accuracy, but mostly in the case of regular verbs.

4.4.5.2 ERP results

The ERP results will be presented in the following way: I first address the conditions with context (the adverb *yesterday*) by comparing the waveforms from the grammatical condition in which the tense feature of the verbs matched the context, as in *Yesterday I ate a banana*, to those from the ungrammatical condition, as in *Yesterday I eat a banana*. Given the *a priori* hypothesis, we expect LAN, N400, and P600 effects for the grammaticality violations. Following that, the ERP effect locus on the scalp, its onset, and its amplitude were analyzed. I then examine the null context condition, which was used to control for the past vs. present tense effects, to make sure any effects obtained in the context condition are due to grammaticality and not confounded by the simple tense difference.

Before the analyses are reviewed, it is important for critical time windows to be established. Recall that the entire length of the EEG recording is 1400 ms, divided into seven 200 ms time bins. The ERP is time-locked to the onset of the verbs, and the average verb duration is 548 ms (SD=89 ms) for the regular verbs, and 513 ms (SD=95 ms) for the irregular verbs. This means that any LAN effect, if found, would

not show up until 813 ms (513 ms+300 ms) for the irregular verbs, and until 848 ms (548 ms+30 ms) for the regular verbs. The relevant time windows are therefore 800-1000 ms, 1000-1200 ms, and 1200-1400 ms. In addition, it is expected that the LAN effect for irregular verbs should be seen earlier than that for the regular verbs, since the tense markers for irregular verbs are in the middle of the string rather than at the end of the verb.

For the stimuli with the adverb *yesterday*, visual inspection reveals a clear negativity in the brain responses to the tense violation with respect to the grammatical controls in the left anterior region, as shown in the upper left panel of Figure 4.2:

Left anterior region
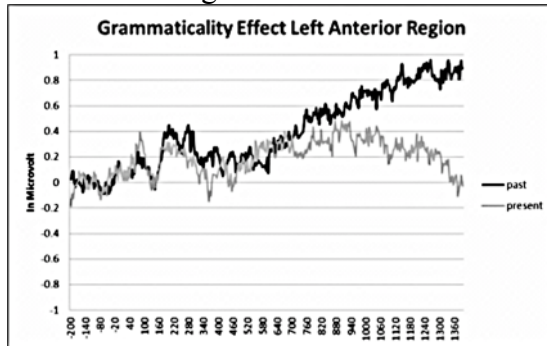
Right anterior region
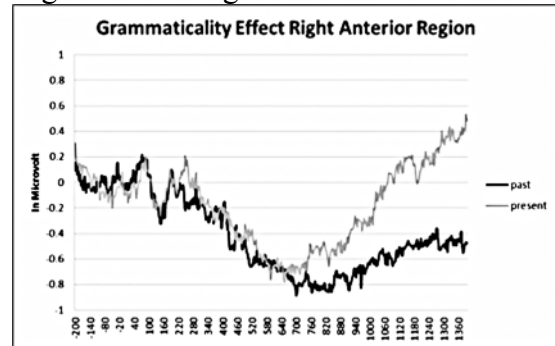
Left posterior region
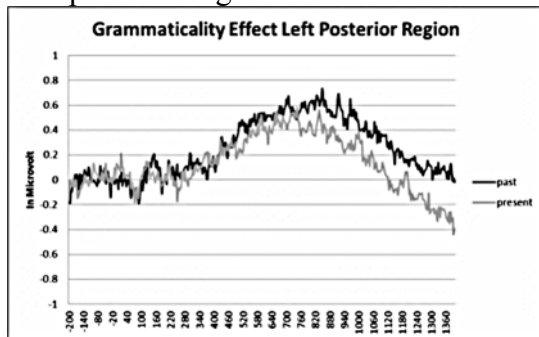
Right posterior region
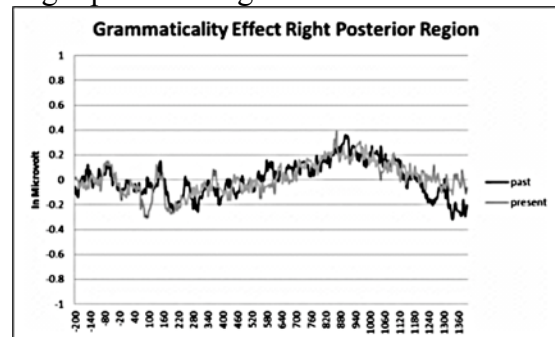
The X-axis is the time course

163

Figure 4.2:   Average brainwaves across four scalp regions for the context condition
              (with *Yesterday*), both verb types combined

This negativity starts at around the 850-1000 ms time window, which is about 300-350 ms after the violation, and is consistent with a typical LAN effect. In the right anterior region, positivity for the present tense (ungrammatical condition) over the past tense (grammatical condition) is observed. Such a reversed pattern to the brainwavess on the left anterior side is commonly found with the LAN and was obtained for the native English group as well in Hestvik et al.'s study. The brain responses of the Chinese speakers didn't show any pattern similar to an N400 or P600.

The mean voltage was computed by region, time, and tense (grammaticality) and then was submitted to a mixed-factorial ANOVA with the following factors as within-subject measures: Time (3) x Region (4) x Grammaticality (2). The Tense (Grammaticality) x Region interaction turns out to be significant ($F_{(3,78)}=4.89$, $p<0.01$), indicating that the tense effect differs by region. Inspection of Figure 4.2 above suggests that this interaction is driven by the negative, present tense waveform in the left-anterior region. The three-way interaction Time x Region x Grammaticality also turns out to be significant ($F_{(6,156)}=3.38$, $P<0.01$), reflecting that the above effect also changes as a function of time, i.e., it becomes larger over the course of the 800 ms to 1200 ms window, as indicated by visual inspection. Given the *a priori* hypothesis, orthogonal contrast analysis revealed that the LAN effect was statistically significant in the 1000-1200 ms time window (400 ms after the violation) ($t=6.09$,

p=0,02) and the 1200-1400 ms time window (t=9.16, p=0.005), but not in the initial

800-1000 ms window, as shown in Figure 4.3. Since the negativity in the 800-1000 ms

time window is not statistically significant for the Chinese speakers, the onset of the

Chinese LAN is considered relatively late (400-500 ms after the detection of the

violation).
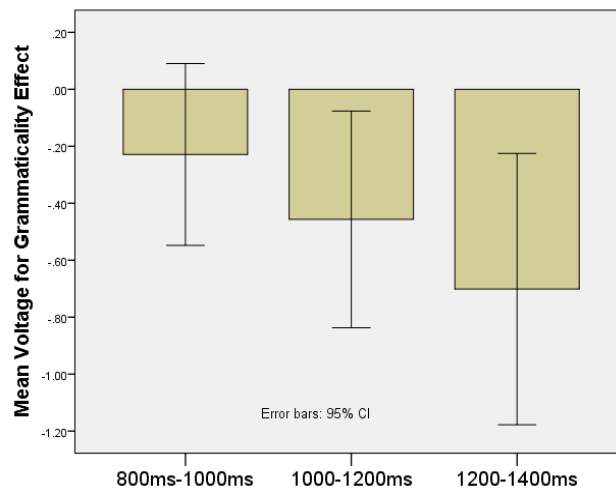
Figure 4.3:   The LAN effect in difference wave form (verb types combined) in three
              critical time windows for the Chinese group

When verb type is taken into consideration, we see that the LAN effect, which

is the difference between the positive-going wave (grammatical, past tense condition)

and the negative-going wave (violation, present tense condition), is bigger in terms of

voltage difference for the irregular verbs than the regular verbs. Figure 4.4 show the

grammaticality effect in voltage difference (ungrammatical condition minus

grammatical condition) by verb type:
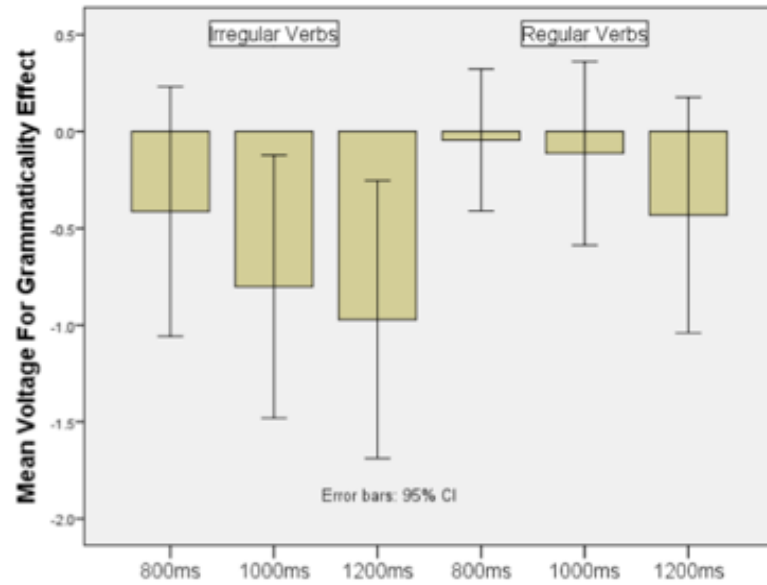
Figure 4.4:   The LAN effect in voltage difference by verb type for the Chinese group

Topoplots in Figure 4.5 (regular verbs) and Figure 4.6 (Irregular verbs) confirm this pattern and indicate that the onset of the LAN for the regular verbs is later (in the 1200 ms window) than for the irregular verbs (800 ms time window):
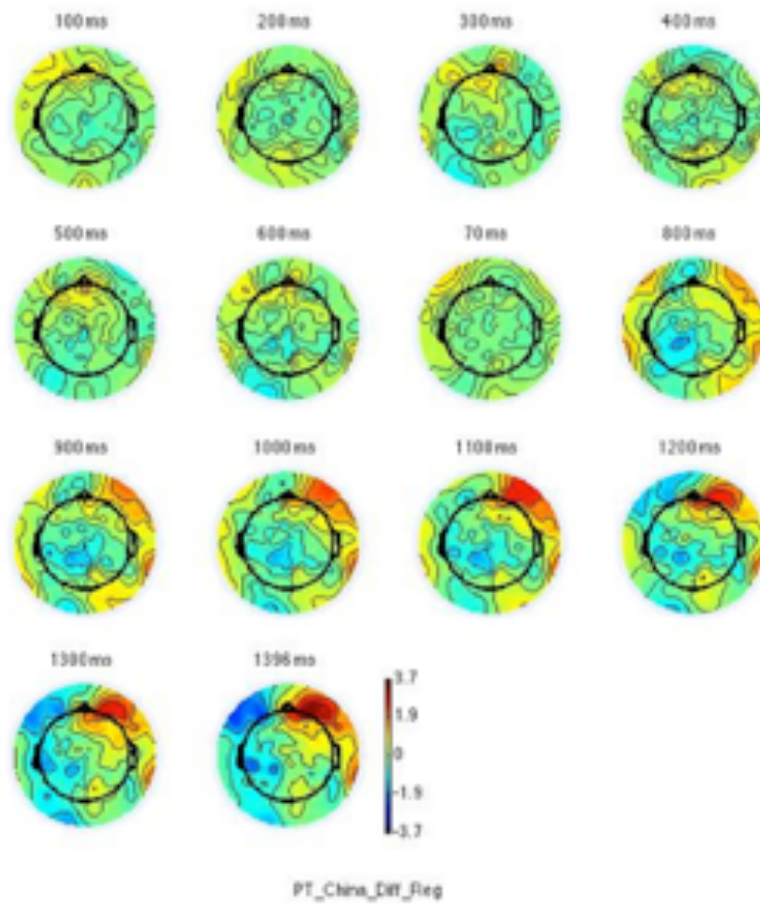
Figure 4.5:   The LAN effect for regular verbs: ungrammatical minus grammatical
condition voltage difference waveform topo plot, past (with yesterday)
context
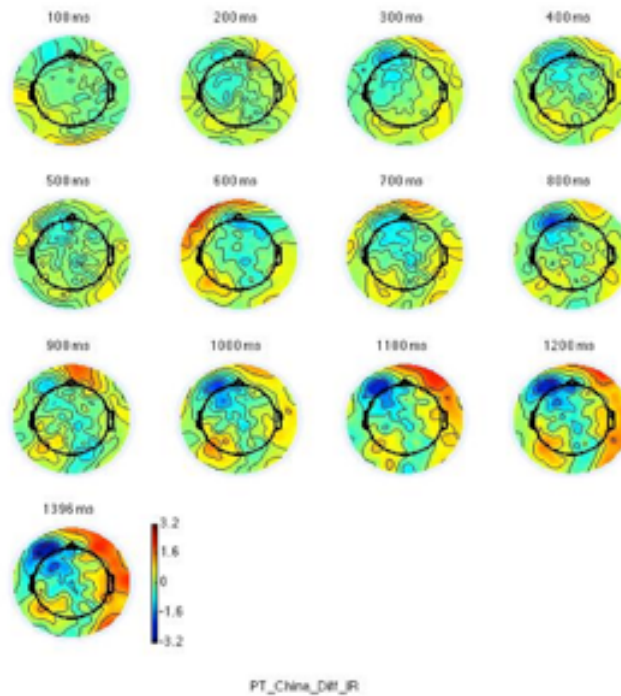
Figure 4.6:   The LAN effect for irregular verbs: ungrammatical minus grammatical condition voltage difference waveform topo plot, past (with *Yesterday*) context

The LAN effect in the 800 ms time window for the irregular verbs was not significant, as a t test yielded a t-value of 1.319, and a *p*-value of 0.09. A significant LAN effect was obtained in the last two time windows starting from 1000 ms (1000 ms: t=-2.43, p=0.02; 1200 ms: t= -2.78, p=0.01). For the regular verbs, however, we see that the shape and time course of the negativity waves for regular verbs look exactly like a typical LAN, though it didn't reach statistical significance for all time windows according to pairwise t tests. In the last time window, the pairwise t test

shows near-significant values: t=1.46, p=0.07. Interestingly, an ANOVA didn't yield a statistically significant interaction between verb type and tense (F (1,26)=3.38, p=0.07) for the Chinese group.

It is also of interest to examine the precise locus of the negativity induced by ungrammaticality inside the left anterior region, since the LAN indicative of morpho-syntactic violations is typically found in the inferior (lower, closer to left eye) region of the left-frontal hemisphere. This is accomplished by comparing the brain responses obtained in the inferior region to those in the superior (higher, closer to the crown) region of the left hemisphere. Figure 4.7 illustrates the comparison of these difference waveforms.
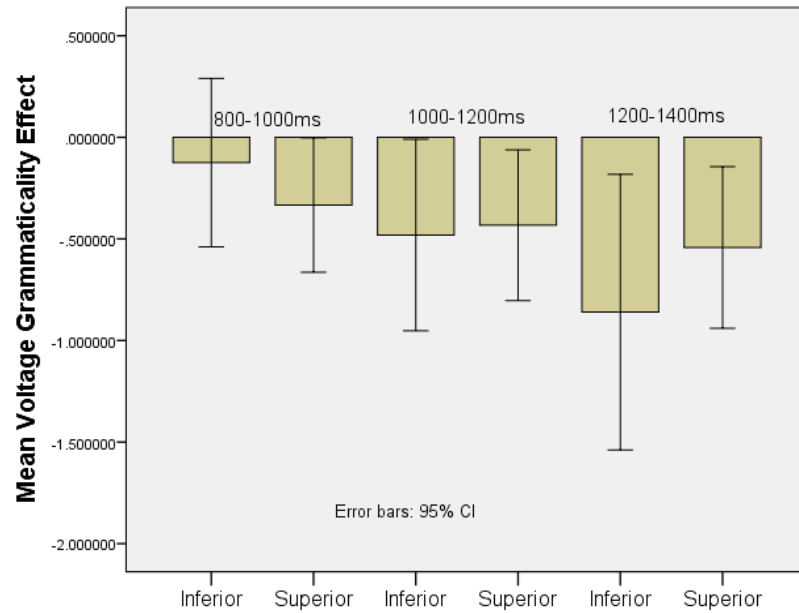
Figure 4.7: The LAN effect in voltage difference in the inferior and superior sub-regions of the Left Anterior region, both verb types combined, Chinese group

By visual inspection, we see that in the 800-1000 ms window the Chinese group has a slightly bigger effect in the superior sub-region. In the 1000-1200 ms and the 1200-1400 ms time windows, the pattern is reversed as the effect becomes bigger in the inferior sub-region. However, an ANOVA (Sub-region x Time x Tense) yielded no significant interaction between Sub-regions and Tense ($F_{(1,26)}=0.066$, $p=0.8$). This is probably because the LAN effect sizes in the two sub-regions are too small in the first two time windows to show any differences between them.

The null-context condition

The null-context condition was inspected separately to ensure the negativity patterns obtained for the context condition are due to grammaticality alone. We found that the Chinese participants showed a different brain response pattern between past and present verbs for the null context sentences. Figure 4.8 below illustrates the continuous EEG waves for the Context vs. Null Context conditions in the left anterior region. When there is no adverb *yesterday*, both stimulus sentences are grammatical and the corresponding brainwaves are largely the same with no detectable trends, as shown in the right panel. Orthogonal contrast analyses yielded no statistical differences between these two waveforms in any of the time windows starting from 600 ms. This is different from the context condition with *yesterday*, in which the violation generates a clear and statistically significant negativity at 1000 ms and after, as mentioned above.
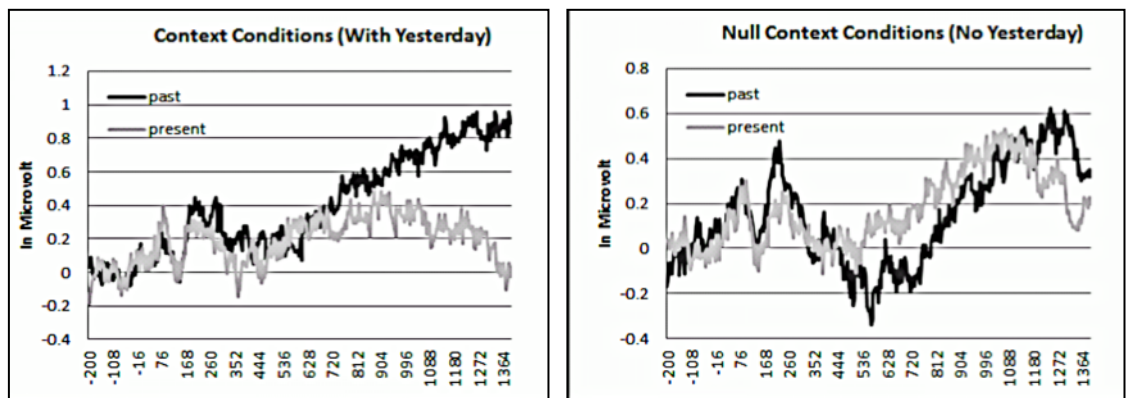
Figure 4.8: Brainwaves for context (*Yesterday*) vs. no context (no *Yesterday*) conditions in Left Anterior region

Examining the brainwaves separately for the irregular and regular verbs confirms that the context/no context conditions are different for both verb types, as shown in Figure 4.9.
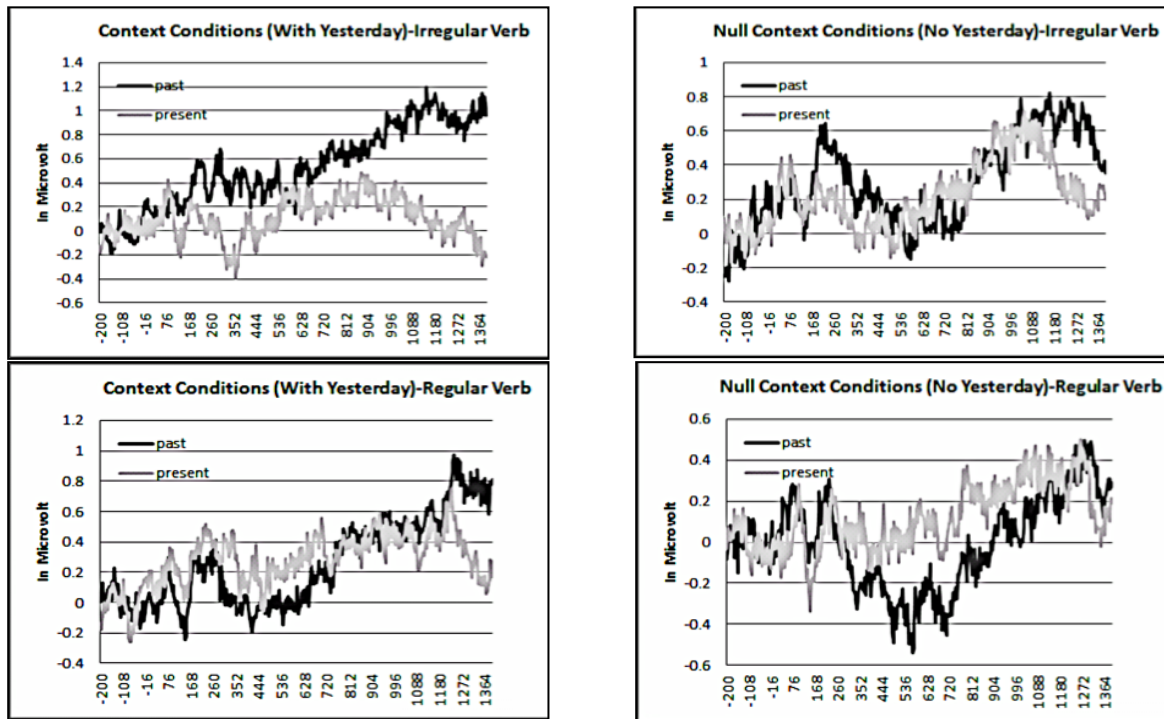
Figure 4.9: Brainwaves for context (*Yesterday*) vs. no context (No *Yesterday*) conditions in Left Anterior region by verb type

Proficiency and ERP Results

To understand the role of the English proficiency of the Chinese speakers of English, we examined how proficiency group relates to the LAN effect size (by voltage difference score). As Figure 4.10a shows, the High proficiency group shows a

172

different pattern than the Low proficiency group in that the LAN (grammaticality

effect) starts about 400 ms later for the Low proficiency group.
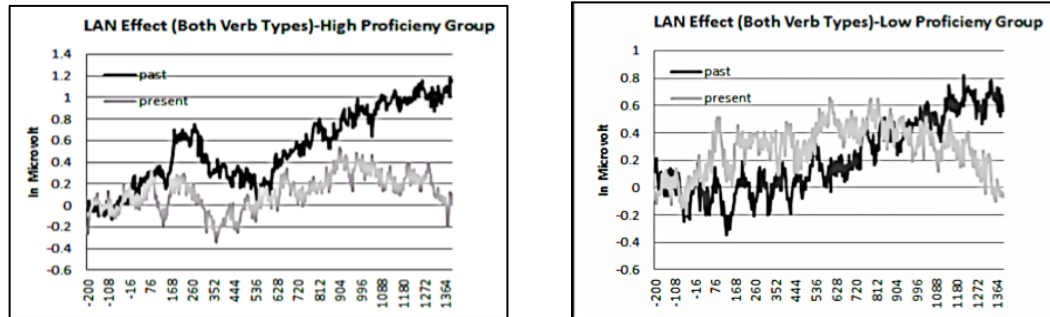
Figure 4.10a: The LAN effect by proficiency group (both verb types)

Separate graphs for regular and irregular verbs, as in Figure 4.10b, revealed

that the High-Low proficiency group difference lies primarily in the processing of

regular verbs.
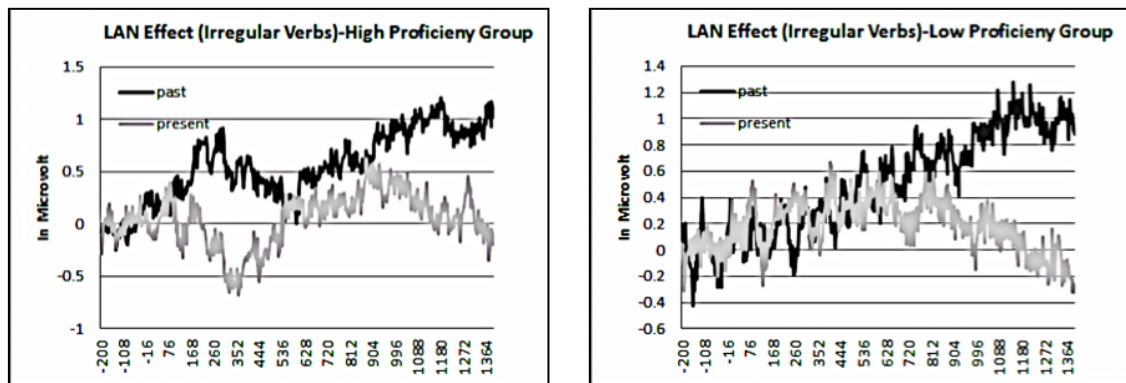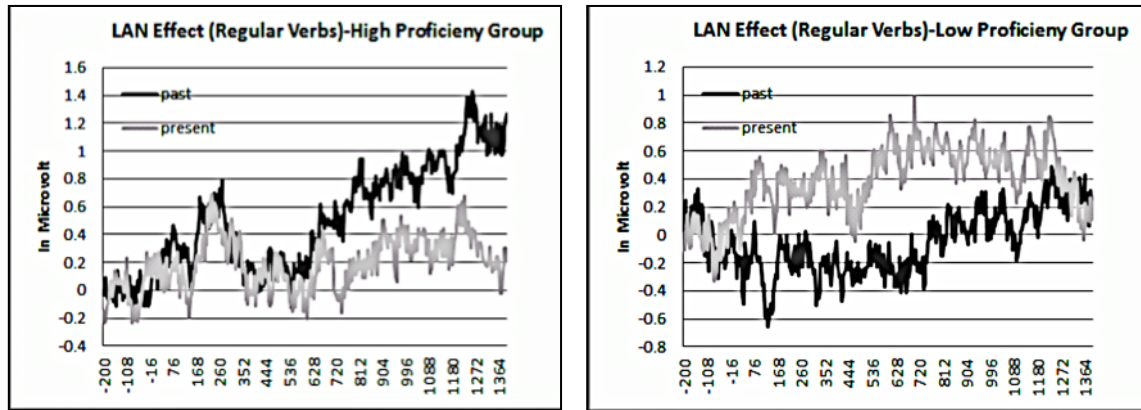
Figure 4.10b: The LAN effect by proficiency group for regular and irregular verb

As indicated by the lower right panel, for the Low proficiency group, the waveform for the ungrammatical condition is positive and the waveform for the grammatical condition is negative at the beginning, which has obviously shrunk the size of the LAN when both groups are combined. This pattern is responsible for the statistically insignificant LAN in the 800-1000 ms time window as described above. Taken together, it appears that the Low proficiency group is responsible for the relatively small size of the LAN for regular verbs.

However, when a repeated measures ANOVA with Tense and Time as within-subject factors and Proficiency Group as the between-subject factor was run on the voltage difference data from 800-1400 ms, no significant between-group effect was found (F (1, 25)=0.07, p<0.05), and no Proficiency Group vs. Grammaticality interaction was obtained (F (1, 26)=1.14, p<0.05). A simple regression analysis of

Versant score on the LAN size for all time windows combined also yielded no significant correlation between the two factors (R^2=0.132, F=2.122, p>0.1), both when verb types are combined and separated, as shown in Figure 4.11:
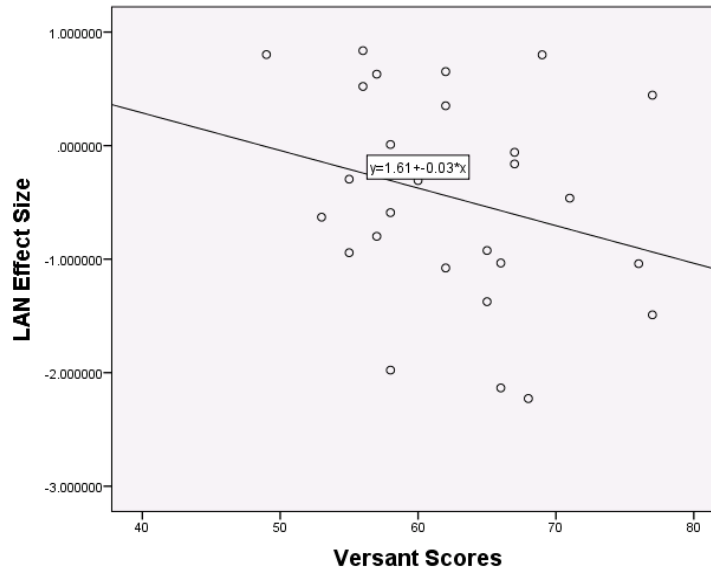
Figure 4.11: Regressing versant scores on the LAN effect size (verb types combined)

To summarize the results, the Chinese group generated significantly negative brain responses typically reported as a LAN to the past tense violations, and the LAN was bigger for the irregular verbs than for the regular verbs. However, the negativity generated due to the violations with regular verbs didn't reach statistical significance. Visual inspection of the LAN elicited by the High and Low proficiency groups revealed that this pattern was caused by the Low proficiency group (Figure 4.10), suggesting a modulating effect of proficiency level, although this was not statistically

confirmed. Within the typical LAN region, the Chinese group showed a concentration of effects in the inferior sub-region descriptively, but the difference didn't reach statistical significance. The Chinese participants also didn't generate any effect similar to the N400 or P600 (Figures 4.5 and 4.6).

4.5   Discussion

Experiment II used EEG recording to examine the online processing of tense violations in English sentences by 29 proficient late Chinese learners of English, to test the claim of the SSH that L2 speakers rely exclusively on full-form memorization rather than decomposition for tense morphology processing. In other words, they adopt a parsing strategy that is categorically different from that of native speakers. Before I discuss the Chinese results by comparing them to those of the native English speakers reported by Hestvik et al., it is necessary to review the predictions the SSH makes for this study. First, the SSH predicts that L2 speakers will generate the ERP N400 component for the irregular verb violations. This is because the irregular verbs are claimed to be memorized as single lexical items. When an incorrect irregular form in context is encountered by the parser, i.e., when *eat* appears instead of *ate*, an ERP effect similar to that of retrieving an incorrect lexical item in Declarative Memory should result, hence the N400. For the regular verbs, the SSH states that the native speakers process regular past tense by applying a rule of decomposition (affix removal). The LAN effect, which is widely reported for violations of morpho-syntactic rules, should result if the regular verb is missing the affix expected due to the

176

context. In contrast to the native speakers, the SSH predicts that L2 learners will generate an N400 for the regular verbs, assuming they under-use syntactic rules and rely on storage for both regular and irregular forms. An alternative, weaker version of the SSH predicts that L2 learners will generate an N400 for the irregular forms, but no or a reduced LAN for the regular forms. In the following sections, the behavioral data and the ERP signatures of the two groups will be compared to evaluate the above predictions of the SSH.

4.5.1 Behavioral data comparison

The two groups performed virtually the same in terms of the overall accuracy of their responses, with the Chinese group (average 87.9% accuracy) performing slightly worse than the English group (average 89.7% accuracy). An omnibus ANOVA (Tense x Context x VerbType x Group) yielded no significant three-way interaction among the first three factors, and no four-way interaction when Group was included, confirming that there is no statistically significant difference between the groups. However, interactions between the Context and VerbType main effects and Group (Context x Group $F (1, 50)=10.86$, $p<0.01$; VerbType x Group $F (1, 50)=22.1$, $p<0.01$) suggest that Context and VerbType affect accuracy differently between the groups. When the groups are examined individually with separate ANOVAs, we see that the accuracy rate for the Chinese group varies more across factors, especially by VerbType and Tense, but not as much by Context. This is shown in Table 4.6.

Table 4.6:   ANOVA results for English and Chinese groups' accuracy rate

	Chinese		English	
	F	p	F	p
VerbType	78.45	**	6.661	*
Tense	15.79	**		
Context			9.98	**
VerbType * Tense	73.55	**		
Tense * Context	26.32	**		

Only significant F and *p*-values are reported; * =p<0.05; ** =p<0.01

As expected, the Chinese group took longer (average 722.63 ms) than the
English group (average 566.4 ms) to give judgments. Pairwise t tests confirmed that
the difference across groups is significant (t=12.462, p<0.01). ANOVAs yielded a
significant main effect of Context for both groups (Chinese: F (1, 28)=11.73; English:
F (1,22)=29.27), such that the null context condition took longer than the context
condition. A significant main effect of Tense (Chinese: F (1, 28) =38.27; English: F
(1,22)=22.98) was also observed, such that the present tense took longer than the past
tense condition, and a significant main effect of Verb Type (Chinese: F (1,28) =7.84;
English: F (1,22) =11.26) was also found, such that the irregular verbs took longer
than the regular verbs (all *p*-values <0.01). There is no interaction between or among
these three factors.

As reviewed in the results section above, the Chinese group performed much
worse on the ungrammatical sentences for both RT and accuracy, mostly with the
regular verbs. In contrast, the English group was much less affected by

grammaticality. In the case of accuracy, a between-group ANOVA with Group (2) x Grammaticality (2) confirmed that the groups are affected by grammaticality differently, as there was a significant interaction between grammaticality and group (F $(1, 50)=7.05$ $p<0.05$). For response time, grammaticality again impacts the groups differently, although an ANOVA only yielded a marginally significant interaction between Group and Grammaticality (F $(1,50)=4$; $p=0.051$). However, when VerbType is considered, we found a very strong interaction effect between Grammaticality and VerbType (F $(1, 50)=15.22$, $p<0.01$), which further interacts with Group (F $(1,50)$ $=6.68$, $p<0.05$). Separate ANOVAs confirmed that the Chinese group was reliably slower *only for the regular verbs* in the ungrammatical condition than in the grammatical condition (761 ms vs. 684 ms; F $(1, 29)=15.33$, $p<0.01$), whereas the English group RT didn't change.

In short, the Chinese participants were highly accurate with their judgments, as was the English group, though they were significantly slower in response time, which is normal for L2 learners. The accuracy of the Chinese group was more affected by the Tense and VerbType factors, and by the interaction between these factors and Context. In addition, the Chinese group was negatively affected by the ungrammatical condition in both RT and accuracy when compared to the English group, but mostly in the case of regular verbs.

4.5.2 Chinese and English ERP compared[22]

As Hestvik et al. (submitted) reported, the English native speakers had clear

LAN effects for both regular and irregular verbs in the ungrammatical condition. The

present tense ungrammatical sentences such as *Yesterday I eat a banana* elicited a

negative brainwave compared to the grammatical control *Yesterday I ate a banana*

after the critical verb *eat*. Such a LAN effect focused around electrode AF7 (EGI 14)[23]

and started to develop 900-1000 ms past the verb onset for the regular verbs (300-400

ms past verb offset) and 800 ms past the verb onset (200 ms past verb offset) for the

irregular verbs. Both onsets and durations are consistent with a classic LAN. The

effect sizes are smaller for the regular than for the irregular verbs, but no statistically

significant interaction between grammaticality and verb type was found. In addition,

when the null context condition was examined, Hestvik et al. found that the present

tense grammatical sentences such as *I eat a banana* generated a *positive-going*

brainwave when compared to past tense sentences such as *I ate a banana*. It was

concluded that the negative-going brainwaves observed in the with-

context/ungrammatical condition arose from the ungrammaticality alone, and not from

the difference between the past and present tense. This contrast converges with the

Chinese results, as the Chinese participants also generated different ERP patterns for

---

[22] The direct comparison is between the Chinese group and the data of 28 English-speaking subjects from Hestvik et al. (submitted).

[23] In Hestvik et al.'s study, an electrical Geodesics 200 system with a 65 channel Geodesic Sensor Net was used.

the Context vs. Null Context conditions. Furthermore, neither an N400 nor a P600 was

detected in the English speakers' brain responses for all conditions, which is also

replicated in the present study for the Chinese group.

As presented in the results section above, the Chinese participants generated a

significant LAN in response to the violation condition (e.g., *Yesterday I eat a banana

and *Yesterday I walk to school*) with a slightly delayed onset. When the Chinese ERP

signatures are compared to those of the English group, the two groups demonstrate a

very similar LAN pattern in terms of the time course and the direction of the

corresponding waveforms, as shown in Figure 4.12 below. In the following graphs, the

downward-going columns represent the voltage difference (i.e., the ungrammatical

condition voltage is subtracted from that of the grammatical condition).
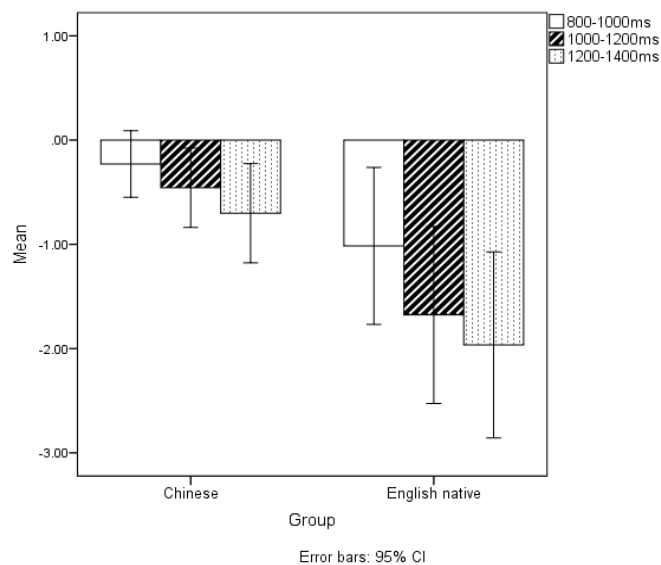
Figure 4.12: LAN effect in voltage difference for both the English and Chinese
groups, verb types combined

An ANOVA showed a statistically significant interaction between Group and Grammaticality (F (1, 53)=6.062, p<0.01), suggesting that the English group produced a stronger LAN effect than the Chinese group, probably due to a greater sensitivity to the violation.

In terms of how the different verb types affected the LAN across groups, the following topoplots (Figure 4.13-Figure 4.16) compare the two groups by verb type and show the brainwave changes over the entire recording time in all scalp regions. We see that there is no N400 or P600 for either group for either verb type.
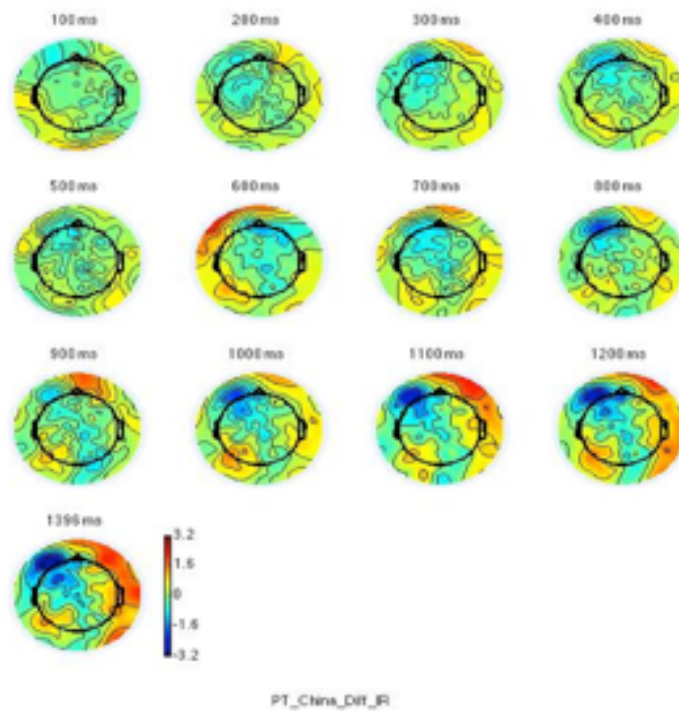
Figure 4.13: Chinese group voltage difference waveform (ungrammatical-grammatical) for irregular verbs
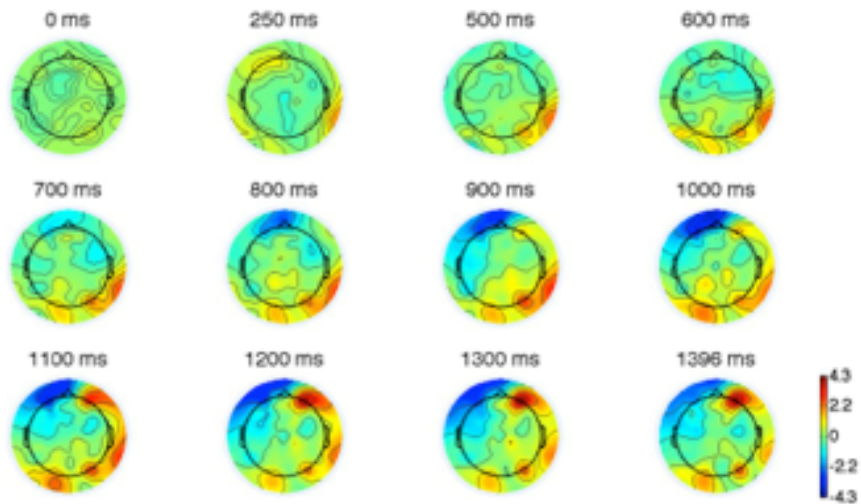
Figure 4.14: English group voltage difference waveform (ungrammatical-grammatical) for irregular verbs (from Hestvik et al. (submitted))

The above two topoplots for the irregular verbs show that the left anterior negativity indicated by the blue region started at around 800 ms past the violation and developed over time in an almost identical way for both groups. However, the following topoplots for the regular verbs show that the negativity starts earlier and is more focal for the English group than the Chinese group:
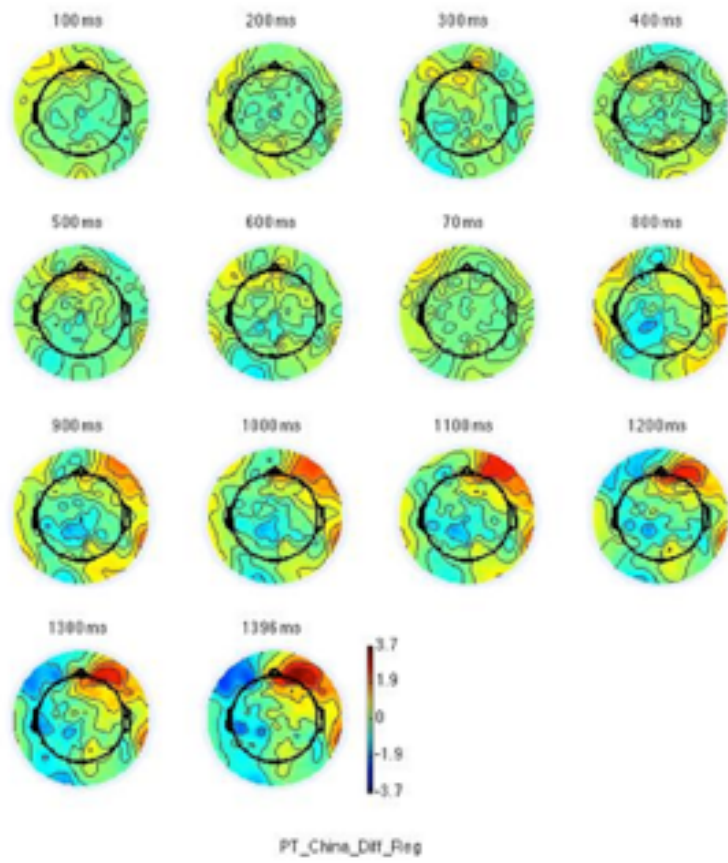
Figure 4.15: Chinese group voltage difference waveform (ungrammatical-
grammatical) for regular verbs
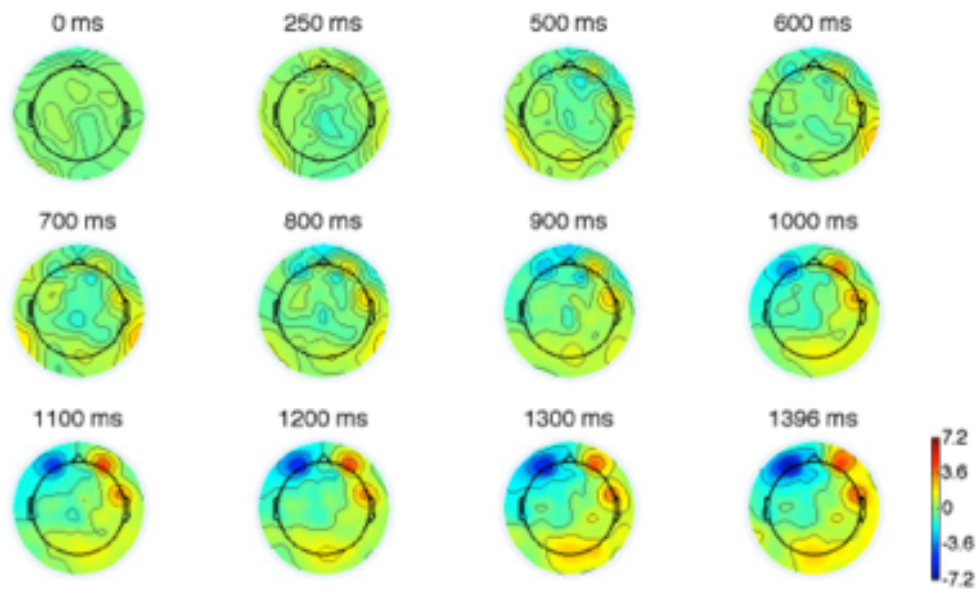
Figure 4.16: English group voltage difference waveform (ungrammatical-grammatical) for regular verbs (from Hestvik et al. (submitted))

A similar pattern was observed for both groups such that the LAN effect is bigger for the irregular verbs than the regular verbs, as illustrated below in Figure 4.17. In the case of the Chinese group, the LAN effect didn't reach statistical significance for regular verbs in all three windows, but was significant for the irregular verbs in the last two time windows. The English group had a reliable LAN effect for all three time windows for both verb types.
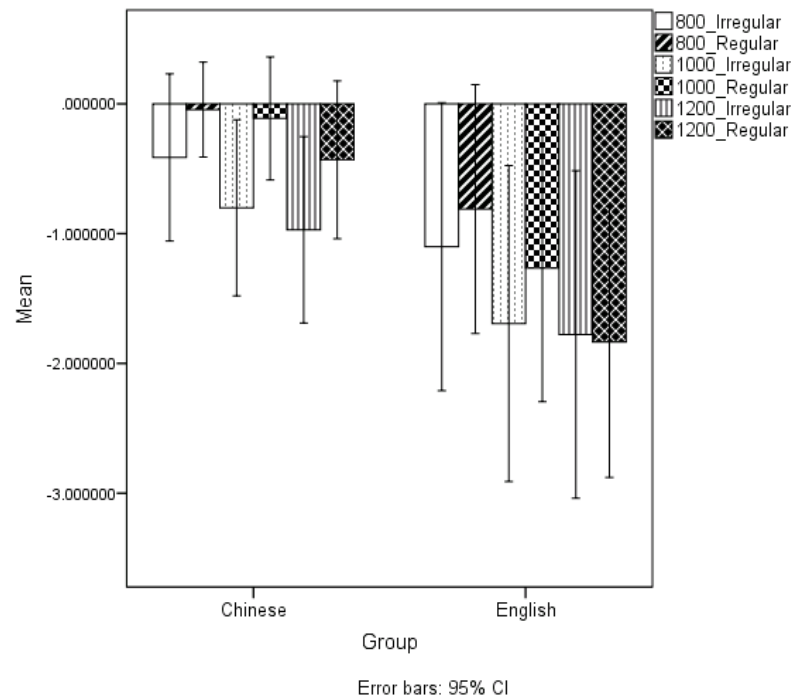
Figure 4.17: The LAN effect in voltage difference score by verb type for the Chinese and English groups

Interestingly, however, the ANOVA didn't yield a statistically significant interaction between VerbType and Tense (F (1, 26) =3.38, p=0.07) for the Chinese group, and the same was reported for the English group. The ANOVA across groups produced no statistically significant interaction between VerbType and Tense, or among VerbType x Tense x Group, suggesting that verb type didn't affect the violation responses. Thus there is no reliable difference between the groups in that regard.

Lastly, I examined whether the two groups are comparable in terms of the LAN distribution inside the left anterior region. The purpose is to see if the Chinese

186

LAN has a different locus than that of the English group. Figure 4.18 illustrates the comparison of the voltage difference waveforms.
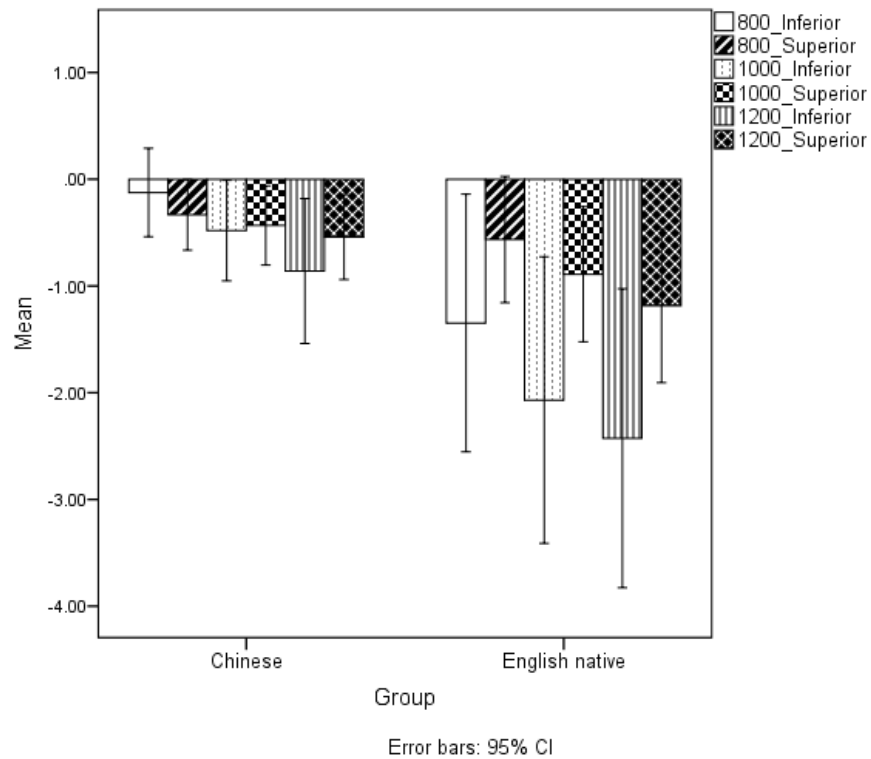
Figure 4.18: The LAN effect in the inferior and superior sub-regions, both verb types combined, English group vs. chinese group

By visual inspection, the Chinese group shows a reversed pattern from that of the English group in the 800-1000 ms window. However, in the 1000-1200 ms window and the 1200-1400 ms window, both the Chinese group and the English group have bigger effects in the inferior sub-region than in the superior sub-region. Pairwise t tests revealed that in the case of the English group, the average voltage difference

187

between the inferior and superior sub-regions is statistically significant in all three time windows (800 ms time window t=-2.1, p<0.05; 1000 and 1200 ms time windows: t=-2.51, p<0.01; t=-2.6, p<0.01). For the Chinese group, an ANOVA (Sub-region x Time x Tense) yielded no significant interaction between Sub-region and Tense (F (1,26) =0.066, p=0.8). As mentioned in the results section, this pattern is due to the small LAN effect sizes in the two sub-regions. When the ANOVA was run across groups with Time (3) x Grammaticality (2) x Sub-region (2) as within-subject factors, there was a significant interaction between Sub-region and Grammaticality (F (1, 53)=6.84, p=0.012) and Sub-region x Grammaticality x Group (F (1,53)=5.83, p=0.019), suggesting that the locus of the effect is different between groups such that it is more concentrated in the inferior sub-region than in the superior sub-region for the English group.

To summarize, we found that the ERP patterns of the late L2 learners are in general similar to those of the native English speakers, but not entirely identical because there is a main effect of group on the LAN amplitude and onset. Most importantly, there is evidence showing that the L2 speakers indeed take advantage of decomposition just like the native speakers. More specifically, a reliable negativity was found for the Chinese speakers for tense violations with irregular verbs starting at the 1000 ms window (450 ms past the violation) in the left anterior region.

Although smaller in amplitude, and with a later onset, this effect fits the LAN profile obtained for morpho-syntactic violations in previous ERP studies and is comparable to that of the native speakers. There was also a statistically significant

LAN found for both verb types combined, again with a slightly later onset and smaller amplitude. Our findings therefore contradict the SSH's prediction, in both its strong and weaker forms, that L2 speakers don't take advantage of decomposition. The present study tested inflectional morphology in a context in which the tense features need to be checked and matched over a long distance between the adverb and the verb. Such a structure resembles the gender and number agreements argued by the SSH to be harder to process than inflected forms in isolation. The fact that we nevertheless obtained LAN effects proves that L2 morphological processing is fully capable of effective syntactic rule application. Furthermore, we found no N400 effect to support the SSH's claim that L2 speakers use full-form memorization to process both verb types.

In addition, the LAN effect found for the Chinese speakers is located in the classic left anterior region. This finding is different from that of Hahne et al. (2006). They found a bilateral LAN for the non-native speakers in a morphological violation paradigm, which was taken as evidence of the ineffective use of morpho-syntactic rules (Clashen and Neubaur, 2010). It was observed that the average voltage amplitude was bigger in the inferior sub-region for the English speakers, reflecting a frontal (lower) locus. Such a pattern was observed for the Chinese speakers descriptively, but it didn't reach statistical significance, possibly due to the smaller amplitude of the LAN in both sub-regions. Given that the more frontal area in the left anterior region is often associated with linguistic processes that are highly automatic (Friederici, 2002), it is conceivable that the L2 speakers are not as fast and automatic as the native

189

speakers, which is also reflected in their slower reaction times in the behavioral data. While such a pattern is unique to the L2 learners, a slightly more diffuse LAN distribution can only show quantitative (not qualitative) differences between the two systems.

Lastly, the LAN effect obtained is modulated by verb type in a similar way for both groups, such that the irregular verbs produced a stronger LAN than the regular verbs, although no statistically significant difference was reached for either group in the ANOVAs. Additionally, no reliable difference was found between groups either. We didn't obtain a statistically significant LAN for the regular verbs alone for the Chinese group. However, the location, time course, and direction of the Chinese regular verb waveforms match those of a typical LAN trend. It is only the amplitude that was too small to reach statistical significance. There are a number of possible explanations for the small amplitude obtained with the regular forms. First, the regular forms are less "learnt" than the irregular forms. Note that the amplitude difference between the verb types was also obtained for the native speakers. The irregular verbs, though proven to be fully decomposed by the results discussed here, still have a stronger memory component than the regular verbs in that the brain has learned to match the different affixes to the correct stems in a more individual fashion (Halle & Maranz, 1994). Therefore, irregular verbs are inherently more reinforced than regular verbs, for both native and non-native speakers alike. It then follows that the violations can be expected to generate a larger effect. Secondly, the regular verb tense marker (and hence the violation of it) is harder to detect acoustically. Given the increased

cognitive load involved in processing an L2 (McDonald, 2006), it is very likely that

the lack of acoustic saliency and reinforcement make the processing of regular verb

violations extra challenging for the late learners, which resulted in a smaller violation

effect. Lastly, the detection of violations is slower for the regular verbs because the

affix doesn't appear until the end of the stem; therefore, the L2 learners might need

more time to react to the violation. The behavioral data goes along with the above

analysis in that the L2 speakers performed significantly worse in the violation

condition in both accuracy and response time, but more so for the regular verbs than

the irregular verbs, suggesting that processing ungrammatical sentences in the target

language takes a significant toll on the L2 system, especially with the regular verbs.

As far as proficiency level is concerned, the L2 learners range from

intermediate-high to advanced-mid by established foreign language assessment

standards. The results obtained clearly indicate that native-like processing is

nevertheless possible at just the intermediate-high level. Although not confirmed by

statistical analyses, it was also observed by visual inspection that the High proficiency

group generated a LAN of bigger size and earlier latency than the Low proficiency

group, suggesting that as proficiency improves L2 learners become more native-like.

This again favors the one-system accounts and contradicts the SSH, which argues that

native-like processing is not attainable even at the end stage of L2 acquisition and is

not affected by factors such as proficiency.

Another significant finding of this present study is that negative L1 transfer

doesn't seem to inhibit the attainment of native-like processing patterns for the L2

learners at a high proficiency level. The late learners of English tested in this study come from a language background in which the target structure doesn't exist, but they were capable of generating brain responses quite similar to the native speakers in real time. This in part supports L1 transfer accounts such as the CM/UCM (MacWhinney, 2008), as they claim that when the target grammatical feature is absent in the native language, the L2 learners may be relatively successful in processing it due to minimum competition between the L1 and L2. However, to understand the precise role of L1 interference, we need to test another group of late L2 learners whose native language inflectional morphology bears some resemblance to English. For example, advanced Spanish speakers of English should be tested and compared to both the native English group and the Chinese group.

4.6 Conclusion

Returning to the research questions proposed in the introduction of this chapter, it is clear that, first, the late L2 learners tested in Experiment II share the same system with the native speakers for processing tense morphology in context. This conclusion follows from the fact that the same ERP component indicative of syntactic rule application, the LAN, was obtained for both groups with a very similar time frame and distribution. No other ERP components were detected for both groups. The only major difference between the groups is that of the amplitude of the LAN effect, especially for the regular verbs, suggesting that the difference between L1 and L2 processing is quantitative in nature. In addition, the L2 speakers in our study didn't

show signs of over-reliance on storage, as indicated by the lack of an N400

component. Secondly, the LAN obtained with the L2 learners argues for their ability

to decompose in the online parsing of morphology. As for the roles of L1 transfer and

proficiency, our results demonstrate that (a) a native-like pattern of morpho-syntax

parsing is attainable among even intermediate-high learners , and (b) while the L2

cognitive manipulation of features absent in their first language may not be as

automatic as that of native speakers, the L2 learners nevertheless are capable of

processing their second language in a native-like way. Altogether, the results from

Experiment II go against the SSH's claims regarding L2 online morphological

processing.

Chapter 5

SUMMARY AND CONCLUSION

5.1 Overview of the Study and Summary of Results

The nature of L2 online parsing and whether it is qualitatively different from parsing in L1 has been an important issue of debate in the field of SLA. Motivated by empirical L2 processing data that appeared to indicate non-native properties (e.g., Felser & Roberts, 2007; Felser et al., 2003; Marinis et al., 2005), the Shallow Structure Hypothesis (SSH, Clahsen & Felser 2006a, b) argued that L2 online processing doesn't compute full syntactic representations, under-uses syntactic rules, and over-relies on non-structural cues. Hence it is fundamentally different from the processing of an L1. Since the initial proposal of the SSH, a body of research has attempted to test this model, but no conclusive evidence has been produced (see 2.3.2 and 4.4). In trying to answer the question of whether there exists a fundamental L1-L2 processing difference and to test the claims of the SSH, this thesis set out to (1) assess late L2 learners' ability to use morpho-syntactic rules and to compute filler-gap dependencies during online parsing, and (2) determine what sources of information (syntactic vs. non-syntactic) are recruited in the real-time processing of L2 speakers in comparison to native speakers. In addition, L2 parameters proposed to affect L2 processing, such as proficiency, working memory capacity, and L1 interference, were

194

also carefully measured and examined. The goals of this work were achieved by taking advantage of the methodology of ERP, which reveals the nature of the processing mechanism, and by adopting paradigms in which the test materials target the use of morpho-syntactic rules. In the specific case of the L2 processing of filler-gap dependencies, the research question is whether abstract syntactic traces are built online. This question was addressed with the results of Experiment I, in which 56 intermediate to advanced Chinese learners of English listened to ungrammatical sentences like, *The zebra that the hippo kissed the camel on the nose ran far away*. For native speakers, a syntactic trace is posited in the direct object position typically filled by a NP as soon as the verb *kissed* is processed. The parser then expects a different word category than a NP. When it instead encounters the extra noun phrase *the camel*, this expectation is violated and an ELAN component indicating an unexpected word category results (Hestvik et al., 2012). According to the time course model of syntactic parsing proposed by Friederici et al. (1996) and Friederici (2002), the initial mapping of the syntactic structure takes place within 100-200 ms after auditory recognition. Violations that occur in this time region, such as to word category expectations and phrase structure rules, are typically associated with an ELAN, as evident in Hestvik et al. (2007, 2012). The next time window of 300-500 ms is when argument structure satisfaction, semantic role assignment, and morpho-syntactic feature matching take place. Violations involving these processes generate a N400 (argument structure and semantic roles) and a LAN (morpho-syntax processing). Note, however, that no additional N400 component is generated by native speakers in

195

the filled gap experiments. This is predicted in the paradigm of Hestvik et al. (2007, 2012), because the initial structural mapping and trace building saturate the verb argument requirements and stop the parser from further analysis. As a result no N400 is predicted to occur (and indeed none was found in Hestvik et al. (2007, 2012). A second ERP component in a later time window (500-800 ms) that is observed with native speakers is a P600, typically found for structural integration anomalies and difficulties (Fridederici et al, 1996; Friderici,1995; Hestvik et al., 2007; 2012).

Contrary to the native speakers, the L2 learners in this study showed completely different brain responses, in spite of having native-like proficiency in both off-line grammatical judgment and online behavioral tasks. Neither ELAN/LAN nor a P600 was found, suggesting that the L2 participants detected no structural violation. Instead, a N400 component, indicative of semantic anomalies and verb argument structure violations (Kutas & Hillyard, 1980a, b, c) was obtained at the offending extra noun phrase. This N400 indicates that the L2 speakers used verb argument structure information to establish the filler-gap dependencies. To be exact, no trace was posited during the early syntactic structure building when the parser processed the verb. Therefore, when the extra noun phrase was encountered, no word category violations were detected (and hence no ELAN was generated). But in the 300-500 ms time region when the verb's argumentation structures were evaluated, the L2 parser detected the conflict between the filler and the extra noun phrase competing for the same argument position, and the N400 was generated. Additionally, the N400 was accompanied by a frontal positivity referred to as the Post-N400 Positivity (PNP) (e.g.,

Van Petten & Luka, 2012), which is sometimes observed with an N400 component due to unexpected lexical items or semantic interpretations. The PNP further corroborates that the dependency formation is meaning- instead of structure-based. Lastly, there was no reliable correlation between the brain responses and the two individual factors of proficiency and working memory capacity. A numerical trend (though not statistically significant) was found such that the higher the proficiency level and the greater the working memory capacity, the larger the N400 component. Taken together, Experiment I provides evidence that L2 learners, regardless of their proficiency level and working memory capacity, resort to verb argument relations or other semantic-based mechanisms to resolve complex FG dependencies online, without constructing abstract syntactic elements such as traces.

In contrast, the L2 learners exhibited native-like processing patterns when processing tense inflectional morphology, as shown by the results of Experiment II, a L2 replication of Hestvik et al. (submitted). The focus of the second study was to test whether the L2 learners can effectively use a decomposition rule to process English past tense forms. The SSH predicts that while native speakers process regular inflected forms such as *walked* by decomposing them into the stem (e.g., *walk*) and the past tense marker *–ed*, the learners memorize these forms as one piece, due to their insensitivity to morphological structure and their deficiency on using morpho-syntactic rules. In Experiment II, 29 intermediate to advanced Chinese learners of English listened to sentences like **Yesterday I walk to school/eat a banana* as well as their grammatical controls. In response to the morpho-syntactic violation induced by

197

the mismatch between the adverb *yesterday* and the lack of past tense marking on the verb, the native speakers generated LANs for both the regular and irregular forms, suggesting full decomposition regardless of verb type. Similarly, a LAN was found for the L2 learners with a slight onset delay and a more diffused distribution, which is expected due to the learners' slower and less automatic processing. The amplitude of the LAN-like component for the regular forms was too small to statistically qualify as a typical LAN, possibly due to the later violation detection, weaker acoustic cues, and reduced saliency of the regular forms in comparison to the irregular forms. Crucially, no N400 component indicative of over-reliance on full-form memorization was found. Thus, the learners didn't memorize the regular forms as predicted by the SSH. In addition, descriptive statistics revealed that proficiency modulated the amplitude of the LAN such that the more proficient the learner, the larger the LAN amplitude, although this pattern did not generalize to the population. Lastly, it is important to stress that there exists a significant L1-L2 difference, as Chinese lacks morphological markings. The fact that the Chinese learners could replicate the native processing pattern regardless of this language difference highlights the ability of L2 learners to internalize novel, non-L1-like morpho-syntactic rules and use them online. In general, the findings of Experiment II contradict the claims of the SSH regarding the L2 processing of morpho-syntax and are more in line with the one-system accounts and Ullman's L2 processing model following his Declarative/Procedural view (2001a; 2001b, 2005).

5.2 General Discussion and Implications for Research

5.2.1 Individual differences and L1-L2 processing

Although the Chinese subjects produced native-like processing patterns for tense morphology, the findings of the two experiments in combination are problematic for one-system accounts, which attribute the L1-L2 difference to L2-specific factors other than age of acquisition, such as proficiency, L1 transfer, and resource constraints. We saw that proficiency cannot explain the discrepancies between the current experiments as the subjects in these two studies were matched in proficiency level. And although working memory was not specifically measured for Experiment II, the chances of the participants in Experiment II having a greater WM capacity than those in Experiment I are very small, given the large sample sizes of both studies. In addition, working memory capacity as a potential modulating factor for processing is most relevant in regard to the distance between the target sentence elements. For the two experiments presented here, the distance between the filler and the gap in Experiment I is only slightly longer (three lexical items apart vs. one lexical item apart) than the distance between the time adverb and the verb in Experiment II. It is unlikely that the extra length in Experiment I overwhelmed the learners' memory capacity to the point of forcing an altogether different parsing algorithm. Crucially, in the FG dependency experiment, we saw that the qualitatively different L2 brain responses were not affected by proficiency and working memory capacity.

199

As for L1 interference, recall from chapter 3 that it could also be argued to cause the non-native brain responses of the learners (see 3.3.3). However, while it is true that Chinese and English relative clauses are not entirely the same, the L1-L2 difference regarding morphological features as tested in Experiment II is far more drastic, yet the Chinese subjects managed to overcome the potential negative native language influence to achieve native-like processing. L1 transfer thus cannot be the decisive factor in determining the nature of L2 processing. Note that although the current findings suggest fundamental L1-L2 processing differences in some cases, they are not necessarily in line with all the "two-system" accounts. For example, Hawkins & Chan (1997) proposed the Failed Feature Hypothesis (FFH), which assumes that the initial state of L2 acquisition is the L1 grammar. According to the FFH, any target language properties not instantiated in the L1 will not be acquired. Such a notion is not supported by this work, as we see that in Experiment II the Chinese speakers of English are quite successful in processing tense morphology, a feature not available in their native language. The qualitative difference between L1 and L2 processing thus cannot be explained by L1 interference alone.

Another potential factor is the quality of L2 lexical access, which has been proposed to be less automatic/complete than that of native speakers. It was found that L2 learners tend to have difficulty processing at the sentence level due to the unfamiliar vocabulary (e.g., Dekydtspotter et al., 2006; Koda, 2005). The vocabulary used in Experiment I included only high frequency words, and the same materials have been used to test children (Epstein & Hestvik, to appear) who produced the

bilateral AN. While it is nevertheless possible that some animal names might be unfamiliar to the adult L2 learners, the vocabulary used in Experiment II is not any less difficult/infrequent, especially the past participle forms (e.g., spin, spun). Therefore, under-routinized lexical access cannot be the reason for the qualitatively different ERP patterns between the two experiments.

Lastly, it is necessary to consider the effect of limited L2 processing speed and resources. The results of Experiment II indicate that the L2 processing differs from that of L1 in a quantitative way, specifically, that the L2 ERP components have a later onset and smaller amplitude when compared with the L1 ERPs. One might wonder if the L2 parser is just "slower" and builds the complex structures and syntactic details in a *later* stage for constructions like FG dependencies. The L2 ERP patterns obtained in Experiment I, however, are not consistent with such a proposal. Assuming the neurophysiological time course model of syntactic parsing proposed by Friederici et al. (1996) and Friederici (2002), three stages that crucially differ in timing are involved in sentence processing. While it is possible that the L2 learners in the FG experiment miss the "first-pass" syntactic analysis due to a slower processing speed, and hence fail to produce an ELAN, it is impossible for them to generate the N400 in the 300-500 ms window but not produce the P600 in later time windows, if the L2 computation is not completed due to time constraints. Note that it has been found that the P600 can coexist with the N400 for combinatory syntactic/sematic anomalies and that its indices (onset, amplitude, location) are not affected by those of the N400 (Hagoort, 2003). Furthermore, the L2 patterns obtained here contrast with the indices

observed during L1 shallow processing (i.e., Good Enough Representations or GERs) as proposed by Ferreira et al. (2002) Ferreira & Paston (2007). Recall the native speakers sometimes calculate less grammatically accurate representations for efficiency or due to structural ambiguity. However, native speakers are not restricted to such a parsing strategy and are able to engage in full 'deep' processing and build complete structures in other contexts (see Section 2.2.3 for details). Crucially, several ERP studies (e.g., Ven Henten et al., 2005,) presented stimuli sentences like *A fox shot a poacher* to native speakers and found no N400, due to an over-reliance on the pragmatics and world knowledge. The P600 components were detected instead, suggesting that the structures were first misinterpreted to make them plausible and then later reanalyzed. The point is that native speakers, unlike L2 speakers, are still accurate in their second-pass syntactic analysis even after semantics-first shallow processing is conducted to achieve economy with online resources.

Lastly, consider the lack of the ELAN. If it is caused by a slower speed and memory constraints, we should expect a later version of it (anterior negativity in the time window of 300 ms and on), as was produced by the native speakers with smaller WM capacity (Hestvik et al., 2012). However, no such anterior negativity was found, and the L2 brain wave patterns in the anterior region do not correlate with the working memory scores.

To conclude, the L2 ERP evidence in the FG Experiment cannot be explained by online processing speed and resources limitations, or by proficiency and L1

interference. Instead, they clearly suggest a non-native processing mechanism guided primarily by non-structural information.

5.2.2 Toward a more complete L2 processing model

5.2.2.1 The parser-grammar distinction

As discussed in the previous section, the findings of this thesis are not compatible with the "one-system" accounts. In particular, the results from Experiment I highlight the L1-L2 fundamental differences and lend strong support to the SSH claims that L2 speakers (1) cannot compute detailed syntactic structures (i.e., they lack traces in FG dependency formation) and (2) exhibit an over-reliance on semantic and pragmatic information to process complex syntax. However, the evidence from both experiments painted a much more complex picture of L2 processing than that depicted by the SSH. In particular, we see that L2 shallow processing does not apply to all constructions. Moreover, the discrepancies between the native-like L2 behavioral performance (at near-native proficiency), which is often taken to indicate implicit grammatical knowledge, and the non-native-like brain responses raise an issue of important theoretical consequence. According to Clahsen and Felser (2006b), this performance/processing brain response contrast shows that L2 learners have a complete L2 grammar but cannot use it during online computation. Such a claim is evident in the following quotes from Clahsen and Felser (2006b):

> …The L2 learners performed at native-speaker levels in the judgment
> task and also achieved high proficiency scores. The differences

between native-speakers in their on-line task therefore cannot be
attributed to incomplete acquisition of the Greek grammar… (p. 44)

What is assumed here is that grammar and the parser can be fully

disassociated, a view that also shared by many other psycholinguists in the field of

SLA research (see White, 2012 for a review). In recent years, a line of SLA research

has studied the persistent difficulties experienced by L2 learners at the "Interfaces"

where either core grammar components (i.e., syntax/semantics), or these components

and grammar-external components (i.e., syntax/discourse) interact. For example, it

was observed that while near-native L2 speakers function like native speakers with

respect to the syntactic properties of pro-drop at the sentence level in Italian, they

couldn't master the discourse requirements imposed on the use of null pronouns (e.g.,

Sorace, 2004, 2009). This proposal has recently been extended to include the interface

between grammar and processing, thus viewing processing as another "external

domain" with which the grammar also interfaces (White, 2011). The L2 learners' non-

native behavior and processing patterns can therefore be attributed to problems with

integrating the grammar; the latter presumably can be fully native-like (White, 2011,

2012). In this scheme of things, the SSH maintains the claim that shallow parsing is an

inherent property of L2 processing, or in other words, L2 processing is "always and

forever" shallow, while a native-like grammar could somehow be developed.

However, the view that grammar and parsing can be disassociated becomes

problematic once we carefully consider how the grammar and the parser are

developed. Given the fact that the parser *always* mediates the grammar (i.e., all input used to develop the grammar is filtered through the parser) and the grammar in return informs the parser (e.g., Gregg, 2003), it is unclear how a globally shallow processor can be used to build the target grammar with all its richness. Equally unexplainable is why and how the complete grammar fails persistently to develop a fully functional parser. Thus the notion endorsed by the proponents of the SSH that a complete grammar could be acquired with a (always) shallow parser is simply not tenable.

In the L1 context, where the issue of the parser-grammar distinction was first discussed, convincing arguments and strong evidence were presented in support of the integrated parser-grammar view. For instance, as Lewis and Phillips (2013) point out, although in the L1 context the parser occasionally produces representations that are not licensed by the grammar (e.g., Garden Paths, Good Enough Representations), which could be taken to suggest that a discontinuity exists between the grammar and the parser, the separate grammar-parser account offers little explanation as to how (1) all the richness and idiosyncrasies of the grammar are fully instantiated in online computation in a highly incremental fashion, and (2) the outputs of the parser *are* identical to the grammar the majority of the time (see Lewis and Phillips 2013 for a detailed review of these issues). In contrast, if the grammar and the parser are one system, then there is no need to account for the close alignment between these two components. Lewis and Phillips (2013) also explain that the difference in online and offline data (for adult native speakers) simply represents different stages and aspects of a given process, as captured by different experimental methods. The misalignments

between the parser and the grammar (for example the simple representations or so called "Good Enough Representations") are due to time and resource limitations, specific properties of cognitive control, and temporary ambiguities caused by the linear nature of online processing. To illustrate the latter point, they systematically analyzed all the parser-grammar misalignments and account for them without involving a separate grammar and parser, thus strengthening their point that the integrated grammar-parser view is much more compelling than the alternative from both a theoretical and an empirical perspective.

Returning to the current research questions for L2 processing, we must also assume that the L2 parser and grammar share one system. Due to the heavy resource demands of L2 processing, there are more misalignments between the parser and the grammar in the L2 context, which results in non-native processing patterns and performances. However, L1 and L2 processing do not differ qualitatively here. These L2 deficiencies should correlate with indices of resource and cognitive control capacity and disappear as proficiency improves. We have seen that this is the case with L2 morpho-syntactic processing. The L1-L2 fundamental difference proposed by this work, however, concerns those non-native L2 processing patterns that are not indicative of parser-grammar misalignments, but rather reflect their perfect synchronization. That is, what the online computation lacks (i.e., traces) is simply not present in the grammar either. Such a case is well illustrated by the FG experiment reported here, where we see no correlation between proficiency/WM and the L2 ERP indices. In sum, while the SSH is correct that L2 learners process differently, the

206

theoretical conclusion about L2 parsing is questionable. Let us now turn to the second important issue raised by the results of this thesis, that is, when does L2 shallow processing occur, and why?

5.2.2.2   L2 shallow processing: scope and motivation

Having explained that the parser and the grammar must be intimately connected and that the grammar is an abstract representation produced by the parser, it is not feasible to claim that the L2 parser shallow-processes globally and that the semantics-first mechanism is a permanent property of the L2 parser. Instead, it would be reasonable to propose that L2 shallow processing, at least at the end stage of acquisition, is limited to only a few situations, perhaps where inadequate structure building does not compromise meaning computation.

It is still premature to draw any conclusions based on the existing ERP studies regarding what the L2 learners can and cannot do for various syntactic structures, mostly because only a limited number of structures have been examined and only a small portion of these studies carefully controlled for factors like proficiency. However, the literature to date does show that L2 learners with high proficiency can achieve native-like patterns at least for some structures. As discussed in chapter 4, a large portion of L2 ERP sentence processing studies deal with morpho-syntactic processing, specifically with grammatical agreement. The high proficiency L2 learners with L1s that are similar to the target language are fairly successful in producing at least partially native-like patterns. That is, they generate a P600 a majority of the time

and sometimes even a LAN (see Tolentino & Tokowicz, 2011 for a review). Outside

of morpho-syntax, a relatively restricted range of syntactic constructions (phrase

structure and verb subcategorization) has been tested with ERP in the L2 context.

Table 5.1 gives a summary of the L2 processing patterns for these structures. Only the

studies that tested fully proficient learners are included.

Table 5.1:   Summary of ERP processing patterns by proficient L2 learners

Syntactic Structure	Studies	Native pattern	L2 pattern
Phrase structure violation	Hahn (2001) L1 Russian, L2 German	ELAN+P600	P600 with delayed peak onset
	Isel (2007) L1 German, L2 French	ELAN+LAN +P600	ELAN+LAN No P600
	Rossi et al. (2006) L1/L2 German-Italian	ELAN+P600	ELAN +P600
	Pakulak & Neville (2011)	ELAN + P600	P600
Syntactic ambiguity (induced by phrase structure rules and verb subcategorization information	Kotz et al., (2009) L1 Spanish, L2 English	P600	P600

These syntactic studies converge with the morpho-syntactic studies in that they show that it is possible for proficient L2 learners to conduct complex syntactic computation, at least in the more controlled, "second-pass" stage reflected by the P600. In some cases, the learners even appeared to conduct the highly automatic, "first-pass" analyses by producing the ELAN with phrase structure violations. These findings and the different results from the two current studies help to illuminate the question of *what* is processed shallowly by L2 speakers. In addition, now that we know that only certain constructions are shallow processed, the reasoning provided by the SSH regarding *why* L2 speakers shallow process is no longer compatible with our general understanding of how processing takes place in L2. Recall that the SSH attributes L2 non-native processing patterns to the acquisition age constraint. Specifically, it appeals to the Declarative/Procedural model and its L2 extension proposed by Ullman (2004, 2005) for theoretical support. The D/P model states that adult L2 learners over-rely on semantics because the memory component (Procedural Memory) that handles linguistic structure building and rule application attenuates after a certain age, but the Declarative system that underlies semantics and the lexicon is still accessible. While this view indeed accounts for L2 reliance on meaning-based parsing and predicts that rules, even those as universal as trace building, are not available to L2 learners, it fails to provide a reason for why *certain* syntactic structures trigger qualitatively different parsing strategies in L2 speakers while others do not.

I propose that 1) the distinct linguistic features of the structures and 2) whether shallow processing has any consequence for the accuracy of the semantic computation

determine the L2 parsing strategy. As for why sometimes the L2 learners are capable

of native-like processing, recall that although the Procedural memory declines with

age in terms of its ability to induce structure, hence leading to shallower processing in

the case of adult L2 speakers, it does not stop functioning entirely in adulthood under

the D/P model. It was argued that with sufficient exposure, structural rules could be

acquired (e.g., Ullman, 2005). Now consider the cognitive processing of tense

morphology: it requires a relatively transparent feature matching operation similar to

the Agree operation used for grammatical agreement. To be more precise, assuming

that decomposition always takes place, this operation is triggered by some surface

cues (e.g., morphological marking on the subject in subject-verb agreement, a time

adverb for tense information as in Experiment II) that syntactically link the different

sentence constituents (Molinaro et al. 2013; 2011). Since those cues are overt and

meaning-bearing (i.e., -*ed* can be related to events/actions that happened in the past),

the learners could "notice" these cues or, in the case of communication break-down,

the absence of them. Such a mechanism is one of the critical components of

acquisition (e.g., Gregg, 2003; Bley-Vroman, 1989). As the learners relate these cues

to the corresponding morpho-syntactic rule (such as memorizing a list of lexical

variations attached to certain structures via the Declarative memory), they are able to

compensate for the declined functioning of the Procedural system by more extensively

recruiting the use of the Declarative system. In contrast, structures like filler-gap

dependencies require the use of an abstract trace for the complete computation of the

syntactic representation. These operations and elements are (1) completely devoid of

semantic meaning, and (2) opaque at the surface level (i.e., they lack a phonological reflex). Crucially, the meaning computation is still accurate without the trace and the complex structure building[24], and the learners cannot "notice" the use of the rule and relate it to any sort of meaning-form mismatch. In fact, the learners might be motivated by efficiency to use a meaning-based routine. Thus, the processing of such constructions could remain permanently shallow. Under this view, it is explainable why proficiency sometimes predicts a more native-like parsing profile (e.g., for morpho-syntax processing, as shown in McLaughlin et al. (2010), Steinhauer, White & White (2009)) but at other times (e.g., for FG dependencies as in Experiment I) it does not affect L2 processing.

Following this line of thinking, some predictions of L2 processing for various structures could be made. The general guideline for such predictions is that structural details tend to be ignored by L2 speakers in real time processing when their omission cannot be detected in surface forms and no meaning miscomputation results. It is thus

---

[24] For example, consider the sentence *[The purse [$_{cp}$ Op$_i$ that [IP Allison misplaced t$_i$]]] is very expensive.* There is a person whose name is Allison. She misplaced a purse and the purse is expensive. While the native speakers build a hierarchical structure with a null operator and the trace in the gap position, the L2 speakers, with a declined ability to induce structure due to the age constraint, compute the meaning by segmenting the incoming information into chunks via thematic role assignment and by associating modifiers to semantically appropriate head phrases. These chunks are integrated into the existing semantic representation in an incremental fashion (Clahsen & Felser, 2006). Therefore, as soon as *misplaced* is processed, the Agent (*Allison*) and the Theme (*the purse*) are identified. The L2 speakers then work out that Allison misplaced a purse, and later that the purse is expensive (not *Allison* because that would not make sense). Thus, the meaning of the sentence is successfully computed, albeit via a non-native mechanism.

predicted that, similar to FG dependencies, constructions involving VP ellipsis could trigger shallow processing among L2 speakers as well. This is because the elided VP is syntactically represented throughout the stages of the derivation, but has no phonological value (e.g., Sag, 1976). The L2 speakers, not capable of building full abstract syntactic structures, would process the covert elements via semantic reconstruction (such as computing a pro form) instead of building structure inside the elided VP. Nevertheless, the L2 speakers still effectively compute the meaning of the sentence. For instance, in the case of sentences like *Mary likes apples and Betty does [VP] too*, L2 speakers could get the interpretation that both Mary and Betty like apples without building internal structure for the elided VP. To be more specific about how to test such predictions about the L2 parsing of VP ellipsis, consider the following two sentences in (25):

(25) a. John defended himself and Bill did too.

b. John defended himself better than Bill did.

For native speakers, two readings are possible for both sentences. The first is the *sloppy reading* (e.g., Carnie, 2012), in which the elided VP contains an element co-indexed with the local antecedent, such that *John* defended *John* and *Bill* defended *Bill*. The second reading, the *strict reading,* allows the elided elements to co-index with an antecedent beyond their clause boundary, so that both *John* and *Bill* defended *John*. However, the strict reading is much harder to obtain in (25a) than in (25b)

(Hestvik, 1995). Hestvik (1995) explains the contrast as follows. Given that the elided VP has internal syntactic structure and the elements in this structure are subject to Binding Principles (Binding Principle A in this case), the strict reading is easier to obtain in (25b) because only in this case does the NP *John* c-command the elided VP and hence is eligible to be an antecedent. Crucially, the reading contrast between (25a) and (25b) is not expected for L2 speakers, as they do not build internal structure for the elided VP, but rather parse via semantic reconstruction. Thus, the L2 speakers should allow both the strict and sloppy readings for both sentences, without being affected by the structural difference between them. Contrary to the trace positing in FG dependencies and VP ellipsis, verb subcategory information (like agreement and tense decomposition) could be processed native-like by L2 learners, due to the overt cues evident in the surface form (e.g., the presence of prepositions, for instance). To conclude, L2 learners do not always shallow process. Whether they resort to a qualitatively different parsing routine depends on two conditions: (1) the extent to which abstract elements are involved in the target structure, and (2) whether shallow processing results in inaccurate meaning computation or is evident in the surface form.

5.3   Limitations and Future Research Directions

A limitation of this thesis is that only two structures, though representative, were tested. To validate the point that the L2 parsing strategy is construction-specific, more structures need to be investigated in future studies. For example, it would be helpful to see how L2 learners process other types of filler-gap dependencies such as

213

topicalization and wh-constructions to see whether they too fail to elicit a structure-induced ERP response. In addition, as mentioned above, how verb subcategory information (e.g., transitivity) affects L2 online parsing could also be informative (Rodríguez, 2008). Referential dependencies are another promising test ground, their computation by native speakers is based on structural principles like c-command. If indeed non-native speakers can only engage in shallow processing, we would predict that c-command violations will not necessarily generate the LAN plus P600 components and will instead only result in a semantic violation (an N400 response).

In the current thesis only auditory stimuli were used. It would be beneficial to try to replicate the current results with other testing modalities. Previous research has suggested that the testing modality could affect L2 processing (e.g., Miller, 2011). More specifically, auditory stimuli may be harder for some late L2 subjects to process, as many of them are trained in a traditional style that emphasizes reading and writing. Given the relatively long and complex sentences used in Experiment I, it is possible that the L2 learners' parsers were overwhelmed and gave up on the structural analysis. It is therefore desirable to test subjects with reading materials (with ERP) as well.

Another promising research avenue is to examine the effect of L2 instruction type, a topic that has been explored extensively in the field of second language acquisition and teaching (e.g., Krashen, 1982; Norris & Ortega, 2001), but is rarely considered in neurolinguistic and psycholinguistic studies (Batterink & Neville, 2013). With regard to exposure type, it has been proposed that naturalistic exposure, defined as an unlimited target language environment with limited formal instruction, is more

effective in achieving high language competence than classroom exposure, in which the learning of the foreign language takes place via formal instruction only (see Flege, 2009 for a review). A few online L2 processing studies have explored this issue, and there is evidence suggesting that naturalistic exposure affects L2 parsing strategies (Dussias, 2003; Frenck-Mestre, 2002, Pilatsikas & Marinis, 2011). A further distinction can be made in the type of instruction learners received, namely, whether they received explicit, memorization-based training or implicit, immersion-style instruction. In the former environment, the foreign language is taught in the native language, and the instruction relies mostly on grammar rule/vocabulary memorization with a strong focus on reading and writing. In comparison, implicit instruction, especially with immersion, is delivered 100% in the contextualized target language. Syntactic rules are supposed to be derived by the learners via the processing of naturalistic input that is carefully chosen to focus on the target structure, and lexical items are taught mostly via visual cues and context instead of translation into the native language. Recently, variation in instructional style as a L2 parameter has started to attract attention from researchers in experimental psycholinguistics. Morgan-Short, Steinhauer, Sanz & Ullman (2012) trained subjects to learn an artificial language (Brocanto2), using both an implicit and an explicit method of instruction. The training was given over a period of approximately 10 days in three sessions. An ERP experiment on morpho-syntactic (agreement) violations was then administered at the end of the training. It was found that while the subjects at both low and high proficiency levels didn't differ in performance across training type groups, they did

demonstrate qualitatively different neural activity. The explicit training group with low proficiency didn't show any effect, and the explicit group with high proficiency produced only a P600 accompanied by an anterior positivity. In contrast, the implicit training group showed a transition from non-native to native-like ERP patterns as proficiency improved: the low proficiency group yielded an N400, and the implicit group with high proficiency produced the native-like LAN+P600. Such results suggest that training type could shape L2 neuro-cognition. While it is still unclear how training type interacts with the syntactic properties of the target structure, it is nevertheless a promising future research direction to pursue. The subjects used in this thesis are mostly undergraduate and graduate students who have had naturalistic exposure but were mostly trained with a traditional, explicit teaching style for their first several years of learning English. Such a learning experience could have reduced their sensitivity to purely structural processing. For future research, it would be beneficial to test late learners who received implicit training from very early on to explore the role of training type in L2 processing. The results will not only inform the research on L2 processing overall but also that of L2 acquisition and pedagogy.

A couple other issues can also be better addressed in future research. Both proficiency score and online behavioral data (accuracy rate) were found to be positively correlated with the Working Memory test used in Experiment I, suggesting the validity of that measure. Nevertheless, it would be ideal if the learners' WM capacity could be tested in their native language as well, to minimize any language transfer effects. Developing a reliable WM test in Chinese is thus necessary. In

216

addition, although the vocabulary used in the two studies is relatively easy and the subjects in Experiment I received training on less common animal names, the variability in their word recognition abilities was not carefully examined in this work. Previous L2 research (e.g., Harrington, 2006; Muljani, Koda & Moates (1998) has revealed that L2 learners' lexical access can be inefficient, which could negatively impact their processing. Including a word recognition measure in future experiments and correlating the results with the neurophysiological data would help to clarify the role of L2 lexical access in online processing.

5.4   Concluding Remarks

Overall, this thesis has provided novel neurophysiological evidence demonstrating that while L2 learners are able to effectively use syntactic information in a native-like way for some structures, they resort to a fundamentally non-native parsing strategy guided mostly by semantics for other structures. These empirical findings are largely in line with the views of the Shallow Structure Hypothesis and other accounts arguing for a fundamental difference between the L1 and L2 online processing of complex syntax. However, based on the one-system view of grammar and parser (the view that grammatical theories and language processing models describe a single cognitive system, as argued in Lewis and Phillips (2013)), this work reached a different theoretical conclusion from that of the SSH. It is stressed here that L2 shallow processing is structure dependent, of limited scope, and persists only when it does not interfere with successful meaning computation.

217

REFERENCES

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs?. *Psychological bulletin*, *131*(1), 30.

Allen, M., Badecker, W., & Osterhout, L. (2003). Morphological analysis in sentence processing: An ERP study. *Language and Cognitive Processes,* 18(4), 405-430.

Amenta, S., & Crepaldi, D. (2012). Morphological processing as we know it: an analytical review of morphological effects in visual word identification. *Frontiers in psychology*, *3*.

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

Ardal, S., Donald, M. W., Meuter, R., Muldrew, S., & Luce, M. (1990). Brain responses to semantic incongruity in bilinguals. *Brain and language*, *39*(2), 187-205.

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language*, *37*(1), 94-117.

Baddeley, A. D., Hitch, G. J., & Bower, G. H. (1974). The psychology of learning and motivation. *The psychology of learning and motivation*, *8*, 47-90.

Baddeley, A. (1998). The central executive: A concept and some misconceptions. *Journal of the International Neuropsychological Society*, *4*(5), 523-526.

Baddeley, A. (2000). The episodic buffer: a new component of working memory?. *Trends in cognitive sciences*, *4*(11), 417-423.

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience*, *17*(1), 137-153.

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157-194). Hillsdale, NJ: Lawrence Erlbaum Associates.

Batterink, L., & Neville, H. J. (2013). The Human Brain Processes Syntax in the Absence of Conscious Awareness. *The Journal of Neuroscience*, *33*(19), 8528-8533.

Baztán, A. M. (2008). La evaluación oral: una equivalencia entre las guidelines de ACTFL y algunas escalas del MCER (doctoral thesis). Universidad de Granada. p. 469–70.

Bernstein, J. & De Jong, J. H.A.L. (2001). An experiment in predicting proficiency within the Common Europe Framework Level Descriptors. In Y.N. Leung et al. (Eds.), *Selected Papers from the Tenth International Symposium on English Teaching* (pp. 8-14). Taipei, ROC: The Crane Publishing.

Bernstein, J. & Cheng, J. (2007). Logic and validation of fully automatic spoken English test. In M. Holland & F.P. Fisher (Eds.), *The path of speech technologies in computer assisted language learning: From research toward practice* (pp. 174-194). Florence, KY: Routledge.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In R. Haynes (Ed.), *Cognition and language development* (pp. 279-362). New York: Wiley and Sons.

Bialystok, E., & Hakuta, K. (1999). Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 161–181). Mahwah, NJ: Lawrence Erlbaum Associates.

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning. *Linguistic perspectives on second language acquisition*, *4*, 1-68.

Boland, J. (1993). The role of verb argument structure in sentence processing: distinguishing between syntactic and semantic effects. *Journal of Psycholinguistic Research 22*:133-152.

Boland, J. (1997). The relationship between syntactic and semantic processes in

sentence comprehension. *Language and Cognitive Processes 12*: 423-484.

Bowey, J. A. (2001). Nonword repetition and young children's receptive vocabulary: A longitudinal study. *Applied Psycholinguistics*, *22*(3), 441-469.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and cognitive processes*, *10*(5), 425-455.

Caplan, D., Alpert, N., & Waters, G. (1999). PET studies of syntactic processing with auditory sentence presentation. *Neuroimage*, *9*(3), 343-351.

Carpenter, P., & Just, M. (1989). The role of working memory in language comprehension. In D. Klahr & K. Kotovsky (Eds.), *Complex Information Processing: The Impact of Herbert A. Simon* (pp. 31-68). Hillsdale: Lawrence Erlbaum Associates.

Carnie, A. (2012). *Syntax: A generative introduction* (Vol. 19). John Wiley & Sons.

Carreiras, M., & Clifton, C., Jr. (1999) Another word on parsing relative clauses: Eyetracking evidence from Spanish and English. *Memory & Cognition*, 27, 826-833.

Carreiras, M., Pattamadilok, C., Meseguer, E., Barber, H., & Devlin, J. T. (2012). Broca's area plays a causal role in morphosyntactic processing. *Neuropsychologia*, 50(5), 816-820.

Carroll, S. (2001). *Input and Evidence: The Raw Material of Second Language* Acquisition.Amsterdam: John Benjamins.

Chen, L., Shu, H. U. A., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject–verb agreement in L2 learning. *Bilingualism: Language and Cognition*, *10*(02), 161-174.

Chomsky, N. (1965). *Aspects of the Theory of Syntax* (Vol. 11). The MIT press.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, N. (1986). *Knowledge of language: Its nature, origins, and use*. Greenwood Publishing Group.

Chomsky, N. (2002). *Syntactic structures*. Walter de Gruyter.

Chun, C. W. (2008). Comments on "Evaluation of the Usefulness of the Versant for English Test: A Response": The Author Responds. *Language Assessment Quarterly*, *5*(2), 168-172.

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

Clahsen, H., & Felser, C. (2006)b. How native-like is non-native language processing? *Trends in Cognitive Sciences*, 10, 564–570.

Clahsen, H. & K. Neubauer (2010). Morphology, frequency, and the processing of derived words in native and non-native speakers. *Lingua* 120: 2627–263.

Clahsen, H., Balkhair, L., Schutter, J. S., & Cunnings, I. (2013). The time course of morphological processing in a second language. *Second Language Research*, *29*(1), 7-31.

Clifton, C., & Frazier, L. (1989). Comprehending sentences with long distance dependencies. In G. M. Carlson & M. K. Tanenhaus (Eds.), *Linguistic structure in language processing* (pp. 273-317). Dordrecht, The Netherlands: Kluwer.

Clifton, C., Traxler, M. J., Mohamed, M. T, Williams, Rihana S., Morris, R. K., & Raynera, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language 49*:317-334.

Colom, R., Shih, P. C., Flores-Mendoza, C., & Quiroga, M. Á. (2006). The real relationship between short-term memory and working memory. *Memory, 14*(7), 804-813.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic bulletin & review*, *12*(5), 769-786.

Cunnings, I., Batterham, C., Felser, C., & Clahsen, H. (2009). Essex research reports in linguistics.

Dallas, A., & Kaan, E. (2008). Second language processing of Filler-Gap dependencies by late learners. *Language and Linguistics Compass*, 2(3), 372-388.

Daneman,M. & P. A. Carpenter (1980). Individual differences in workingmemory and reading. *Journal of Verbal Learning and Verbal Behavior 19.4*, 450–466.

Daneman, M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research, 20(6),* 445-464.

Daneman, M., & Merikle, P. (1996). Working memory and language comprehension: Ameta-analysis. *Psychonomic Bulletin & Review, 3(4),* 422-433.

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 (GASLA 2006)* (pp. 33-40).

Dekydtspotter, L., Donaldson, B., Edmonds, A. C., Fultz, A. L., & Petrush, R. A. (2008). Syntactic and prosodic computations in the resolution of relative clause attachment ambiguity by English-French learners. *Studies in Second Language Acquisition*, *30*(04), 453-480.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature neuroscience*, *8*(8), 1117-1121.

Delong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology*, *48*(9), 1203-1207.

Desmette, D., Hupet, M., Schelstraete, M.A., & Van der Linden, M. (1995). Adaptation en langue française du "Reading Span Test" de Daneman et Carpenter. [A French adaptation of Daneman and Carpenter's "Reading Span Test"] *L'Annee Psychologique*, *95*, 459–482.

Dewaele, J-M. and Veronique, D. (2001) Gender assignment and gender agreement in advanced French interlanguage: A cross-sectional study. *Bilingualism: Language and Cognition 4*, 275_297.

Dien, J. (2010). The ERP PCA Toolkit: An open source program for advanced statistical analysis of event-related potential data. *Journal of neuroscience methods*, *187*(1), 138-145.

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.

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*(7), 1651-1659.

Downey, R., Farhady, H., Present-Thomas, R., Suzuki, M., & Van Moere, A. (2008). Evaluation of the usefulness of the Versant for English test: A response. *Language Assessment Quarterly*, 5, 160-167.

Dussias, P. E. (2003). Syntactic ambiguity resolution in L2 learners. *Studies in Second Language Acquisition*, *25*(4), 529-557.

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.

Ellis , N. C. , & Sinclair , S. G . (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order . *Quarterly Journal of Experimental Psychology,49* , 234 – 250 .

Embick, D., & Halle, M. (2005). On the status of stems in morphological theory. *Amsterdam Studies in the Theory and Histroy of Linguistic Science Series 4*, *270*, 37.

Epstein, B., Hestvik, A. (2013). ERPs reveal atypical processing of subject vs. object Wh-questions in children with Specific Language Impairment. To appear in International Journal of Language & Communication Disorders.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309.

Federmeier, K. D., & Kutas, M. (2005). Aging in context: Age‐related changes in context use during language comprehension. *Psychophysiology*, *42*(2), 133-141.

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

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.

Felser, C., L. Roberts, R. Gross, & T. Marinis (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, 24, 453–89.

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

Fender, M. (2001). A review of L1 and L2/ESL word integration skills and the nature of L2/ESLword integration development involved in lower-level text processing. *Language Learning, 51(2),* 319-396.

Ferreira, F., & Henderson, J.M. (1991) Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, 31, 725-745.

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

Ferreira, F., & Patson, N. (2007). The good enough approach to language comprehension. *Language and Linguistics Compass*, 1, pp 71-83.

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

Flege, J. E. (2009). Give input a chance. *Input matters in SLA*, 175-190.

Franceschina, F. (2005). *Fossilized second language grammars: The acquisition of grammatical gender* (Vol. 38). John Benjamins.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension:Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*,*14,* 178-210.

Frazier, L. (1983). Processing sentence structure. In K. Rayner (Ed.), Eye movements in reading: *Perceptual and language processes* (pp. 215-236). New York: Academic Press.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance 12: The psychology of reading* (559-586). Mahwah, NJ: Erlbaum.

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

Frazier, L. (1999). *On sentence interpretation*. Dordrecht: Kluwer.

French, L. M., & O'BRIEN, I. R. E. N. A. (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics*, *29*(03), 463-487.

Frenck-Mestre, C., & Pynte, J. (1997). Syntactic ambiguity resolution with reading in second and native language. *The Quarterly Journal of Experimental Psychology, 50A*(1), 119-148.

Frenck-Mestre, C. (2002). An on-line look at sentence processing in the second language. In R. R. Heredia & J. Altarriba (Eds.), *Bilingual sentence processing* (pp. 217-235). Amsterdam: Benjamins.

Frenck-Mestre, C., Foucart, A., Carrasco-Ortiz, H., & Herschensohn, J. (2009). Processing of grammatical gender in French as a first and second language: Evidence from ERPs. *Eurosla yearbook*, *9*(1), 76-106.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological, and syntactic violations. *Cognitive Brain Research, 1*, 183–192.

Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and language*, *50*(3), 259-281.

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

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

Friederici, A. D., & Weissenborn, J. (2007). "Mapping sentence form onto meaning: The syntax-semantic interface". *Brain Research*, 1146: 52.

Friederici, A. D. (2011). The brain basis of language processing: From structure to function.*Physiological Reviews*, 91, 1357-1392.

Frisch, S., Hahne, A., & Friederici, A. D. (2004). Word category and verb–argument structure information in the dynamics of parsing. *Cognition*, *91*(3), 191-219.

Fodor, J. D. (1995). Comprehending sentence structure. In L. R. Gleitman & M. Liberman (Eds.), *An invitation to cognitive science, Volume 1: Language, 2*[nd] *Edition* (pp. 209-246). Cambridge, MA: MIT Press.

Fodor, J. D., & Inoue, A. (1998). Attach anyway. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 101-141). Dordrecht: Kluwer.

Fortkamp , M. B. M . (1999). Working memory capacity and aspects of L2 speech production. *Communication & Cognition*, 32, 259 – 296 .

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contribution of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language, 37,* 58-93.

Gass, S. M., & Selinker, L. (1992). *Language transfer in language learning*. John Benjamins Publishing Co.

Gibson, E., & Warren, T. (2004). Reading-time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax*, 7, 55-78.

Gor, K. (2010). Beyond the obvious: Do second language learners process inflectional morphology? Language Learning, 60.1, 1-20.

Gor, K., & Cook, S. (2010). Nonnative processing of verbal morphology: In search of regularity. *Language Learning*, *60*(1), 88-126.

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

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24(2), 95-112.

Gregg, K. (1996). The logical and developmental problems of second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 49-81). San Diego, CA: Academic Press.

Gregg, K. (2003). SLA theory construction and assessment. In C. Doughty & M. Long (Eds.). *Handbook of second language acquisition* (pp. 831-865). Oxford: Blackwell

Gross, M., Say, T., Kleingers, M., Clahsen, H., & Münte, T. F. (1998). Human brain potentials to violations in morphologically complex Italian words. *Neuroscience Letters*, 241(2-3), 83-86.

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.

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

Hagoort, P. (2003). How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. *Neuroimage, 20*, S18-S29.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*(5669), 438-441.

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

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*(02), 123-141.

Hahne, A., Mueller, J., & Clahsen, H. (2006). Morphological processing in a second language: Behavioral and event-related brain potential evidence for storage and decomposition. *Journal of Cognitive Neuroscience, 18*, 121–134.

Halle, M., & Marantz, A. (1994). Some key features of Distributed Morphology. In A. Carnie & H. Harley (Eds.), *MIT Working Papers in Linguistics 21: Papers on phonology and morphology*. Cambridge, Massachusetts: MIT Department of Linguistics and Philosophy.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition*, *14*(1), 25-38

Harrington, M. (2006). The Yes/No test as a measure of receptive vocabulary knowledge. *Language Testing*, *23*(1), 73-98.

Hashimoto, K. I. (2009). First language influence on second language syntactic processing of English relative clauses.

Hawkins, R., & Chan, C. Y. (1997). The partial availability of universal grammar in second language acquisition: The "Failed functional features hypothesis." *Second Language Research*, 13, 187–226.

Hestvik, A. (1995). Reflexives and ellipsis. *Natural Language Semantics*, *3*(2), 211-237.

Hestvik, A., Maxfield, N., Schwartz, R. G., & Shafer, V. L. (2007). Brain responses to filled gaps. *Brain and Language*, 100(3), 301-316.

Hestvik, A., Bradley, E., & Bradley, C. (2012). Working Memory Effects of Gap-Predictions in Normal Adults: An Event-Related Potentials Study. *Journal of Psycholinguistic Research*, *41*(6), 425-438.

Hestvik, A., Shafer, V., Schwartz, R. G., Ullman, M., Neumann, Y., & Rinker, T. (Submitted). Brain responses to contextually ungrammatical verb inflection.

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. Second *Language Research*, *22(3)*, 369-397.

Hsu, C.-C. N. (2006), *Issues in head-final relative clauses: derivation, processing, and acquisition*, Unpublished Ph.D. Dissertation, University of Delaware.

Hsu, C.-C. N. (2008), Revisit relative clause islands in Chinese, *Language and Linguistics*, 9 (1), 23-48.

Huang, C.-T. J. (1982). Logical relations in Chinese and the theory of grammar, Ph.D thesis, MIT.

Huang, J., Li, Y. A., & Li, Y. (2009). The Syntax of Chinese. Cambridge, Cambridge University Press.

Huang, Y. C., & Kaiser, E. (2008, April). Investigating filler-gap dependencies in Chinese topicalization. In *Proceedings of The 20th North American Conference on Chinese Linguistics* (Vol. 2, p. 927).

Isel, F. (2007): Syntactic and referential processes in second-language learners: event-related brain potential evidence. *Neuroreport 18*, 1885–89.

Jackson, C. N., & van Hell, J. G. (2011). The effects of L2 proficiency level on the processing of wh-questions among Dutch second language speakers of English. *IRAL-International Review of Applied Linguistics in Language Teaching*, *49*(3), 195-219.

Jegerski, J. (2010). *Ultimate attainment in second language acquisition: Near-native sentence processing in Spanish*. ERIC.

Jeong, H., Sugiura, M., Sassa, Y., Haji, T., Usui, N., Taira, M., et al. (2007). Effect of syntactic similarity on cortical activation during second language processing: A comparison of English and Japanese among native Korean trilinguals. *Human Brain Mapping*, 28, 194–204.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning*, *57*(1), 1-33.

Jiang, X., & Zhou, Z. (2009). Processing different levels of syntactic hierarchy: An ERP study on Chinese. *Neuropsychologia*, 47(5), 1282–1293.

Juffs, A., & Harrington, M. (1995). Parsing effects in second language sentence processing: Subject and object asymmetries in wh-extraction. *Studies in Second Language Acquisition*, 17, 483-516.

Juffs, A. (1998). Main verb vs. reduced relative clause ambiguity resolution in second language sentence processing. *Language Learning*, *48*, 107-147.

Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society*, *102*, 199-225.

Juffs, A. (2005). The influence of first language on the processing of wh-movement in English as a second language. *Second Language Research*, 21, 121-151.

Juffs, A. (2006). Grammar and parsing and a transition theory. *Applied Psycholinguistics*, *27*, 66-69.

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, *44*(02), 137-166.

Just, M., & Carpenter, P. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99(1)*, 122-149.

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

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. I. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, 15(2), 159-201.

Kanno, K., Ariel, M., Bates, E., Macwhinney, B., Bates, E., Macwhinney, B., ... & King, J. (2007). Factors affecting the processing of Japanese relative clauses by L2 learners. *Studies in Second Language Acquisition*, *29*(2), 197.

Kielar, A., Joanisse, M. F., & Hare, M. L. (2008). Priming English past tense verbs: Rules or statistics?. *Journal of Memory and Language*, *58*(2), 327-346.

Kielar, A., & Joanisse, M. F. (2010). Graded effects of regularity in language revealed by N400 indices of morphological priming. *Journal of cognitive neuroscience*, *22*(7), 1373-1398.

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

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

Kluender, R., & Kutas, M. (1993)a. Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, *5*(2), 196-214.

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

Kluender, R. (2004). Are subjects islands subject to a processing account? In A. Rodriguez, V. Chand, A. Kelleher, & B. Scheiser (Eds.), *Proceedings of WCCFL* 23 (pp. 475–499). Somerville, MA: Cascadilla Press.

Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. Cambridge University Press.

Kondo, H., & Osaka, N. (2004). Effect of concreteness of target word on verbal working memory: An evaluation using Japanese version of reading span test. *Japanese Journal of Psychology*, *71*, 51–56.

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition*, *11*(02), 261-271.

Kotz, S. A., Holcomb, P. J., & Osterhout, L. (2008). ERPs reveal comparable syntactic sentence processing in native and non-native readers of English. *Acta Psychologica*, 128(3), 514–527.

Kotz, S. A. (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain and Language*, *109*(2), 68-74.

Krashen, S. (1982). *Principles and practice in second language acquisition* (pp. 65-78). Pergamon: Oxford.

Kutas, M., & Hillyard, S. A. (1980)a. Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology*,11, 99–116.

Kutas, M., & Hillyard, S. A. (1980)b. Reading between the lines: event-related brain potentials during natural sentence processing. *Brain and Language*,11, 354–73.

Kutas, M., & Hillyard, S. A. (1980)c. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203-208.

Kutas, M., Federmeier, K.D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences* 4, 463–470.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, *62*, 621-647.

LaPointe, L.B., & Engle, R.W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 1118–1133.

Larson-Hall, J. (2006). What does more time buy you? Another look at the effects of long-term residence on production accuracy of English /r/and/l/ by Japanese speakers. *Language and Speech*, 521–548.

Lau, E., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain and Language, 98*, 74–88.

Lau, E.F., Rozanova, K., & Phillips, C. (2007). Syntactic prediction and lexical frequency effects in sentence processing. *University of Maryland Working Papers 16,* 163-200.

Lauro, L. J. R., Reis, J., Cohen, L. G., Cecchetto, C., & Papagno, C. (2010). A case for the involvement of phonological loop in sentence comprehension. *Neuropsychologia*, *48*(14), 4003-4011.

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

Leeser , M. J . (2007). Learner-based factors in L2 reading comprehension and processing grammatical form: Topic familiarity and working memory . *Language Learning, 57*, 229 – 270.

Lewis, S. & Phillips, C. (2013). Aligning grammatical theories and language processing models. *Journal of Psycholinguistic Research*.

Li, C., & Thompson, S. (1981). A functional reference grammar of Mandarin Chinese.

Lin, Y., & Garnsey, S. M. (2011). *Animacy and the resolution of temporary ambiguity in relative clause comprehension in Mandarin* (pp. 241-275). Springer Netherlands.

Love, T. E. (2007). The processing of non-canonically ordered constituents in long distance dependencies by pre-school children: A real-time investigation. *Journal of Psycholinguistic research, 36,* 191-206.

Luck, S. J. (2005). An introduction to the event-related potential technique (cognitive neuroscience).

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

MacDonald, M. C. (1999). Production, and Acquisition: Three Puzzles and a Moral. *The emergence of language*, 177.

MacWhinney, B. (2008). A Unified Model. In P. Robinson & N. Ellis (Eds.), *Handbooks of Cognitive Linguistics and Second Language Acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B., & Bates, E. (1989). The crosslinguistic study of sentence processing.

MacWhinney, B. (2006). Emergentism—use often and with care. *Applied Linguistics*, *27*(4), 729-740.

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

Marslen-Wilson, W. D., & Tyler, L. K. (2007). Morphology, language and the brain: the decompositional substrate for language comprehension. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 823-836.

Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition*, *34*(03), 379-413.

McClelland, J. L., & Patterson, K. (2002). Rules or connections in past-tense inflections:what does the evidence rule out? *Trends in Cognitive Sciences*, 6(11), 465-472.

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.

McLaughlin, J., Tanner, D., Pitkänen, I., Frenck‐Mestre, C., Inoue, K., Valentine, G., & Osterhout, L. (2010). Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning*, *60*(s2), 123-150.

Miller, A. K. (2011). Rethinking Nonnative Processing Constraints: Evidence from L2 French.

Mitchell, D. C. (1994). Sentence parsing. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 375-409). San Diego, CA: Academic Press.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge University Press.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive psychology*, *41*(1), 49-100.

Molinaro, N., Vespignani, F., & Job, R. (2008). A deeper reanalysis of a superficial feature: An ERP study on agreement violations. *Brain research*, *1228*, 161-176.

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

Molinaro, N., Barraza, P., & Carreiras, M. (2013). Long-range neural synchronization supports fast and efficient reading: EEG correlates of processing expected words in sentences. *NeuroImage*.

Montrul, S. A. (2004). *The acquisition of Spanish: Morphosyntactic development in monolingual and bilingual L1 acquisition and adult L2 acquisition* (Vol. 37). John Benjamins Publishing.

Montrul, S., Foote, R., & Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning*, *58*(3), 503-553.

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second Language Acquisition of Gender Agreement in Explicit and Implicit Training Conditions: An Event-Related Potential Study. *Language learning*, *60*(1), 154-193.

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.

Morris-Florack, J. & Holcomb, P.J. (2005). ERPs to violations of inflectional verb morphology.*Cognitive Brain Research*, 25, 963-981.

Muljani, D., Koda, K., & Moates, D. (1998). The development of word recognition in a second language. Applied Psycholinguistics, 19, 99-113.

Net, A. H. G. S. Electrical Geodesics, Inc.

234

Neubauer, K., & Clahsen, H. (2009). Decomposition of inflected words in a second language: An experimental study of German participles. *Studies in Second Language Acquisition*, 31(3), 403–435.

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

Newman, A. J., Ullman, M. T., Pancheva, R., Waligura, D. L., & Neville, H. J. (2007). An ERP study of regular and irregular English past tense inflection. *NeuroImage,*34(1), 435-445.

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

Nicol, J. L. (1993). Reconsidering reactivation. In G. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing: The Second Sperlonga Meeting* (pp. 321-347). Mahwah, NJ: Erlbaum.

Nieuwland, M. S., & Van Berkum, J. J. A. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research, 24,* 691-701.

Norris, J. M., & Ortega, L. (2001). Does Type of Instruction Make a Difference? Substantive Findings From a Meta‐analytic Review. *Language Learning*, *51*(s1), 157-213.

Oberauer, K., Süß, H. M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, *31*(2), 167-193.

O'Brien, A. R., Chiaravalloti, N., Goverover, Y., & DeLuca, J. (2008). Evidenced-based cognitive rehabilitation for persons with multiple sclerosis: a review of the literature. *Archives of physical medicine and rehabilitation*, 89(4), 761-769.

O'BRIEN, I. R. E. N. A., Segalowitz, N., Collentine, J., & Freed, B. (2006). Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics*, *27*(03), 377-402.

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience*, 17, 1212–1228.

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.

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

Osterhout, L., & Holcomb, P.J. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes* 8, 413–437.

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice Learners, Longitudinal Designs, and Event Related Potentials: A Means for Exploring the Neurocognition of Second Language Processing. *Language Learning*, 56(s1), 199-230.

Öztekin, I., & McElree, B. (2007). Proactive interference slows recognition by eliminating fast assessments of familiarity. *Journal of Memory and Language*, 57(1), 126-149.

Packard, J. L., Ye, Z., & Zhou, X. (2011). Filler-gap processing in Mandarin relative clauses: Evidence from event-related potentials. In *Processing and producing head-final structures* (pp. 219-240). Springer Netherlands.

Papadopoulou, D., & H. Clahsen (2003). Parsing strategies in L2 and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition*, 25, 501-528.

Papagno, C., Valentine, T., & Baddeley, A. D. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language*, 30, 331–347.

Pakulak, E., & Neville, H. J. (2011). Maturational constraints on the recruitment of early processes for syntactic processing. *Journal of cognitive neuroscience*, 23(10), 2752-2765.

Paradis, M. (2004). A Neurolinguistic Theory of Bilingualism. Amsterdam: Benjamins.

236

Pastizzo, M. J., & Feldman, L. B. (2002). Discrepancies between orthographic and unrelated baselines in masked priming undermine a decompositional account of morphological facilitation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(1), 244.

Penfield, W., & Roberts, L. (1959). *Speech and brain mechanisms*. Princeton: Princeton University Press.

Penke, M., Weyerts, H., Gross, M., Zander, E., Munte, T.F., Clahsen, H. (1997). How the brain processes complex words: an event-related potential study of German verb inflections. *Cognitive Brain Research* 6, 37–52.

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

Phillips, C., & Wagers, M. (2007). Relating structure and time in linguistics and psycholinguistics. *Oxford handbook of psycholinguistics*, 739-756. Phillips, C., & Wagers, M. (2007). Relating structure and time in linguistics and psycholinguistics. *Oxford handbook of psycholinguistics*, 739-756.

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

Pinker, S. & Prince, A. (1994). Regular and irregular morphology and the psychological status of rules of grammar. In S. D. Lima, R. L., Corrigan, & G. K. Iverson (Eds.), *The reality of linguistic rules*. Philadelphia: John Benjamins.

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends Cogn Sci, 6*(11), 456-463.

Pliatsikas, C., & Marinis, T. (2011). Processing of regular and irregular past tense morphology in highly proficient L2 learners of English: A self-paced reading study. *Applied Psycholinguistics*.

Portin, M., Lehtonen, M., Harrer, G., Wande, E., Niemi, J., & Laine, M. (2007). L1 effects on the processing of inflected nouns in L2. *Acta Psychologica*, 128(3),452–465.

Portin, M., Lehtonen, M., & Laine, M. (2007). Processing of inflected nouns in late bilinguals. *Applied Psycholinguistics*, 28(1), 135–156.

Pulvermüller, F., Shtyrov, Y., Hasting, A.S., & Carlyton, R.P. (2008). "Syntax as a reflex: Neurophysiological evidence for the early automaticity of syntactic processing". *Brain and Language* 104 (3): 245.

Rah, A., & Adone, D. (2010). Processing of the reduced relative clause versus main verb ambiguity in L2 learners at different proficiency levels. *Studies in Second Language Acquisition*, 32(1), 79-109.

Reichle, R. V., Tremblay, A., & Coughlin, C. E. (2013). Working-Memory Capacity Effects in the Processing of Non-adjacent Subject-Verb Agreement: An Event-Related Brain Potentials Study. In *Selected Proceedings of the 2011 Second Language Research Forum. Somerville, MA: Cascadilla Proceedings Project*.

Roberts, L., Marinis, T., Felser, C. & Clahsen, H. (2007). Antecedent priming at gap positions in children's sentence processing. *Journal of Psycholinguistic Research, 36*, 175-188.

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics 25*, 45–73.

Rodriguez-Fornells, A., Clahsen, H., Lleo, C., Zaake, W., & Münte, T. F. (2001). Event- related brain responses to morphological violations in Catalan. *Brain Research Cognitive Brain Research*, 11(1), 47-58.

Rodríguez, G. A. (2008). *Second Language Sentence Processing: Is it fundamentally different?*. ProQuest.

Rösler, F., Putz, P., Friederici, A. D., & Hahne, A. (1993). Event-Related Brain Potentials While Encountering Semantic and Syntactic Constraint Violations. *Journal of Cognitive Neuroscience, 5*(3), 345-362.

Rossi, S., Pasqualetti, P., Zito, G., Vecchio, F., Cappa, S. F., Miniussi, C., ... & Rossini, P. M. (2006). Prefrontal and parietal cortex in human episodic memory: an interference study by repetitive transcranial magnetic stimulation. *European Journal of Neuroscience, 23*(3), 793-800.

Sabourin, L., & Stowe, L. A. (2004). Memory effects in syntactic ERP tasks. *Brain and Cognition*, 55, 392–395.

Sabourin, L., Stowe, L. A., & de Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research*, 22, 1–29.

Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research*, 24, 397–430.

Sag, I. A. (1976). "Deletion and Logical Form". *PhD Thesis, Massachusetts Institute of Technology* (New York: Garland Publishing).

Sag, I. A., & Fodor, J. D. (1994). Extraction without traces. In *Proceedings of the 13th West Coast Conference on Formal Linguistics* (pp. 365-384). Stanford, CA: CSLI Publications.

Sagarra, N., & Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua*, *120*(8), 2022-2039.

Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in cognitive sciences*, *6*(9), 382-386.

Sawyer, M. & Ranta. L.(2001). Aptitude, individual differences, and instructional design. *Cognition and second language instruction*, 310-353.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime Reference Guide*. Pittsburgh: Psychology Software Tools, Inc.

Schremm, A. (2012). *Processing filler-gap dependencies in an L2: An Event-Related Potential study*. (Unpublished MA thesis). Lund University, Lund, Sweden

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

Selinker, L. (1969). Language transfer. *General linguistics*, *9*(2), 67-92.

Selinker, L. (1972). Interlanguage. *IRAL-International Review of Applied Linguistics in Language Teaching*, *10*(1-4), 209-232.

Service, E. (1992). Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology, 45*, 21 – 50.

Shyu, S. I., & Sybesma, R. (1996). The syntax of focus and topic in Mandarin Chinese. *Letters & Replies*, *2*(4), 11.

Silva, R., & Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition*, 11, 245–260.

Solomyak, O., & Marantz, A. (2010). Evidence for early morphological decomposition in visual word recognition. *Journal of Cognitive Neuroscience*, *22*(9), 2042-2057.

Sorace, A. (2004). Native language attrition and developmental instability at the syntax-discourse interface: Data, interpretations and methods. *Bilingualism Language and Cognition*, *7*(2), 143-145.

Sorace, A., & Serratrice, L. (2009). Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism*, *13*(2), 195-210.

Sprouse, J., & Almeida, D. (2012). Assessing the reliability of textbook data in syntax: Adger's Core Syntax. *Journal of Linguistics*, *48*(3), 609-652.

Stanners, R. F., Neiser, J. J., Hernon, W. P., & Hall, R. (1979). Memory representation for morphologically related words. *Journal of Verbal Learning and Verbal Behavior*, *18*(4), 399-412.

Steinhauer, K., & Ullman, M. T. (2002). Consecutive ERP effects of morpho-phonology and morpho-syntax. *Brain and Language*, 83, 62-65.

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

Stockall, L., & Marantz, A. (2006). A single route, full decomposition model of morphological complexity: MEG evidence. *The Mental Lexicon*, *1*(1), 85-123.

Swets, B., Desmet, T., Hambrick, D.Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General, 136*, 64-81.

Swets, B., Desmet, T., Clifton, C., Jr, & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition, 36*, 201-216.

Swinney, D, Ford, M., & Bresnan, J. (1989). Coreference assignment during sentenceprocessing. In *Language Structure and Processing*, eds. B. Grosz, R.

Kaplan, M. Mackenand I. Sag. Stanford: CSLI.

Szmalec, A., Verbruggen, F., Vandierendonck, A., & Kemps, E. (2011). Control of interference during working memory updating. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 137.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of verbal learning and verbal behavior*, *14*(6), 638-647.

Tannenbaum, R. & Caroline E. W. (2005). Research Reports: Mapping English Language Proficiency Test Scores onto the Common European Framework. Educational Testing Services.

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, *83*(3), 382-392.

Tolentino, L. C., & Tokowicz, N. (2011). Across Language, space and time. *Studies in Second Language Acquisition*, *33*(01), 91-125.

Tokowicz, N., & MacWhinney, B. (2005). Implicit vs. explicit measures of sensitivity to violations in L2 grammar: An event-related potential investigation. *Studies in Second Language Acquisition*, 27, 173–204.

Townsend, D., & Bever, T. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.

Traxler, M., & Pickering, M. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, 35, 542-562.

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

Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language 35*:566-585.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent?. *Journal of Memory and Language*, *28*(2), 127-154.

Ueno, M., & Kluender, R. (2003). Event-related brain indices of Japanese scrambling. *Brain and Language*, *86*(2), 243-271.

Ueno, M., & Garnsey, S. M. (2008). An ERP study of the processing of subject and object relative clauses in Japanese. *Language and Cognitive Processes*, *23*(5), 646-688.

Ullman, M. T. (2001)a. The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguist Research*, 30(1), 37-69.

Ullman, M. T. (2001)b. A neurocognitive perspective on language: The declarative/procedural model. *Nature reviews neuroscience*, *2*(10), 717-726.

Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, 92(1-2), 231-270.

Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. *Mind and context in adult second language acquisition*, 141-178.

Ullman, M. T. (2006). The declarative/procedural model and the shallow structure hypothesis. *Applied Psycholinguistics*, 27(01), 97-105.

Ullman, M. T., Miranda, R. A., & Travers, M. L. (2008). Sex differences in the neurocognition of language. In J. B. Becker, K. J. Berkley & N. Geary (Eds.), *Sex on the Brain: From Genes to Behavior* (pp. 291-309). New York, NY: Oxford University Press.

Vallar, G. & Baddeley, A.D. (1984). Fractionation of working memory. Neuropsychological evidence for a phonological short-term store. *Journal of Verbal Learning and Verbal Behaviour*, *23*, 151–161.

van den Noort, M. W., Bosch, P., & Hugdahl, K. (2006). Foreign language proficiency and working memory capacity. *European Psychologist*, *11*(4), 289-296.

Van Hell, J. G., & Tokowicz, N. (2010). Event-related brain potentials and second language learning: Syntactic processing in late L2 learners at different L2 proficiency levels. *Second Language Research*, *26*(1), 43-74.

Van Heurten, M., Chwilla, D. J., & Kolk, H. H. J. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience*, *18*, 1181-1197.

Van Herten, M., Kolk, H. & Chwilla, D. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22,241–55.

Van Patten, B. (1996). *Input Processing and Grammar Instruction: Theory and Research*. Norwood:Ablex Publishing Corporation.

Van Patten, B. (2006). Input processing in adult second language acquisition. In B. VanPatten & J.Williams (Eds.), *Theories in Second Language Acquisition* (pp. 115-136). Mahwah:Lawrence Erlbaum.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., Parks,M. (1999). Timecourse of word identification and semantic integration in spoken language. *Journal of Experimental Psychology. Learning, Memory, and Cognition* 25, 394–417.

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

Vos, S. H., Gunter, T. C., Kolk, H. H. J., & Mulder, G. (2001). Working memory constraints on syntactic processing: An electrophysiological investigation. *Psychophysiology*, 38, 41–63.

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

Waters, G. S. & D. Caplan (1996)a. Processing resource capacity and the comprehension of garden path sentences. *Memory and Cognition 24*, 342–355.

Waters, G. S. & D. Caplan (1996)b. The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology-Human Experimental Psychology 49.A*, 51–79.

Waters, G. S. & D. Caplan (1996)c. The capacity theory of sentence comprehension: Critique of Just & Carpenter (1992). *Psychological Review 103.4*, 761–772.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50(1), 1-25.

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

Weyerts, H., Penke, M., Dohrn, U., Clahsen, H., Munte, T.F. (1997). Brain potentials indicate differences between regular and irregular German plurals. *Neuroreport* 8, 957–962.

White, L. (2003). *Second language acquisition and Universal Grammar*. Cambridge University Press.

White, L. (2011). Second language acquisition at the interfaces. *Lingua*, *121*(4), 577-590.

White, L. (2012). Universal Grammar, crosslinguistic variation and second language acquisition. *Language Teaching*, *45*(03), 309-328.

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

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

Xu, L., & Langendoen, D. T. (1985). Topic structures in Chinese. *Language*, 1-27.

Ye, Z., Luo, Y., Friederici, A. D., & Zhou, X. (2006). Semantic and syntactic processing in Chinese sentence comprehension: Evidence from event-related potentials. *Brain Research*, 1071, 186–196.

Appendix A

SUBJECT BACKGROUND QUESTIONNAIRE

Subject identifier: [ ]

<u>**SUBJECT QUESTIONNAIRE**</u>

A. Basic Information   Date: _____

- Gender: M   /   F
- Date of Birth: _____   Age: _____   Level of Education: _____

    Family

- Mother's highest education level:_____   Father's highest education level: _____

B. Language Exposure

- What language(s) are spoken at home? _____

- What is your primary language?_____ _____

- What other languages do you speak? _____

- What languages can you understand (although may not speak)? _____

- Father's primary language: _____ Other languages the father speaks fluently: _____

- Mother's primary language: _____ Other languages the mother speaks fluently: _____

B2. Additional Language Exposure

- Places (exclusive of USA) in which you have lived for more than one year:

    City/State/Country   Years

    _____from _____ to _____

    _____from _____ to _____

    _____from _____ to _____

    _____from _____ to _____

    If you have lived in more places please check here and continue on the back

- At what age did you start learning English? _____ years old
- For how many years did you study English in school and college? _____ years
- Your most recent TOEFL Score: Overall_____   Listening comprehension _____

- GRE/GMAT /Score (optional)_____
- Are you working as TA/RA? Yes _____; No _____

    o If the above answer is Yes, How many hours in a typical week do you actively use English to communicate orally at work? _____hours

- How long have you been living in the USA on a full time basis? _____ years _____ months

- Rate your fluency and understanding of English by checking one of the following which best describe your mastery of this language:
    a. Speak/Understand like a native speaker
    b. Speak with a mild accent and understand native speakers with little or no difficulty
    c. Speak with an accent and understand native speakers, but with some effort
    d. Speak and understand, but with effort
    e. Can not speak or understand this language at all

- What languages did you study as a foreign language in school?
    Circle all applicable: elementary, junior high, high school, college

    Number of semesters:_____

    Rate your fluency and understanding of this language by checking one of the following which best describe your mastery of this language:

    f. Speak/Understand like a native speaker
    g. Speak with a mild accent and understand native speakers with little or no difficulty
    h. Speak with an accent and understand native speakers, but with some effort
    i. Speak and understand, but with effort
    j. Can not speak or understand this language at all

246

C. Speech, Language, & Hearing History

- Is there any history of the following in your family? (check all that apply and state relationship of family):

  Speech or language disorder_____   Hearing impairment_____   Learning Disorder_____

- Did you ever exhibit a language delay as a child?   No   Yes   (explain) _____

- If yes, when was the language delay first apparent?

  _____

- Have you ever been evaluated by or worked with any of the following? (check all that apply and please explain)

	Evaluation Only	Sessions		Evaluation Only	Sessions
Ear Nose and Throat (ENT) Doctor	_____	_____	Reading Specialist	_____	_____
Neurologist	_____	_____	Speech Language Pathologist	_____	_____
Psychologist	_____	_____	Other:		
Audiologist	_____	_____	_____	_____	_____

Explanations: _____

_____

- Have you ever worn hearing aid(s)?   No   Yes

- If yes, at what age did you begin wearing the hearing aid(s)? _____

- If you wear a hearing aid(s), at what time(s) during the day and for what activities?

  _____

247

D. Medical History/ Development

- Have you been diagnosed with…

  PDD?   No   Yes   Asperger's Syndome?   No   Yes

  Autism? No   Yes   ADD/ADHD?   No   Yes

- Do you take any medication?   No   Yes   (explain)_____

  _____

- Which hand do you use most?   Left   Right   Both equally

Is there any information you would like to share with us to help us understand you better?

_____

We are committed to including subjects from all backgrounds in research and therefore collect the following information. You may choose not to provide this information.

(please check one in <u>both</u> categories)

<u>**Ethnic Category (please check one)**</u>   <u>**Racial Category (please check one)**</u>

☐   Hispanic or Latino   ☐   American Indian/Alaska Native

☐   Not Hispanic or Latino   ☐   Asian

☐   Do not wish to respond   ☐   Native Hawaiian or Other Pacific Islander

                                          ☐   Black or African American

                                          ☐   White

                                          ☐   Do not wish to respond

Appendix B

LIST OF STIMULI SENTENCES IN THE WORKING MEMORY TEST

Level I

1. He played baseball all day at the park and got a sore arm.
2. The clerk in the department store put the presents in a bag.

3. I saw a child and her father river near the playing ball.
4. His younger brother played guitar in a rock and roll band.

5. Suddenly the taxi opened in its door front of the bank.
6. The last thing he did was to take a nice hot bath.

Level II

7. Her best memory of England was the Tower of London bell.
8. At the very top of the tree sat small tall a bird.
9. She took a deep breath the reached and into rusty box.

10. The state of Wisconsin is famous for its butter and cheese.
11. He overslept and missed economics all the of morning class.
12. The first thing he does every a is swing morning golf club.

13. Popular foods in the summer are watermelon and sweet corn.
14. The boy was surprised to learn that from milk a came cow.
15. The only thing left in the kitchen cupboard was a broken cup.

Level III

16. The birthday party began in the all and morning lasted day.
17. The young woman and her boyfriend thought they saw a dog.
18. There was nothing left to do the leave except lock and door.
19. In order to attend the dinner she buy to needed a dress.

249

20. The woman screamed and slapped the old in man the face.
21. She leaned over the candle and her hair caught on fire.
22. The drinks were all gone and all that remained was the food.
23. He quickly drank some of the milk and then washed the glass

24. He looked across the room and saw a person holding a gun.
25. The hunting knife was so sharp that it cut his right hand.
26. She soon realized that the man forgot room to leave the key.
27. The saw that he brought was not strong the for enough lock.

Level IV.

28. The first driver out in the morning always picks up the mail.
29. All that remained in the lunch salted was box one nut.
30. The boat engine would not run of because was out it oil.
31. The letter said to come to the market to claim the prize.
32. It was a very simple meal of salted fish and boiled rice.

33. They decided to take an afternoon break by the large rock.
34. He wanted to leave his bags and hotel in jacket the room.
35. There were so many people that I couldn't find a seat.
36. He opened the bottom drawer a and out pulled shirt.
37. The skiing was so wonderful that he didn't mind the snow.

38. They knew that it was impolite to the spaghetti with eat a spoon.
39. The season that people often is with associate love spring.
40. The letter was lost because it not did a postage have stamp.
41. The people in northern Europe always like to travel by train.
42. All morning the two children sat and under a talked tree.

Appendix C

PAPER-AND-PENCIL ACCEPTABILITY JUDGMENT TEST

Filled Gap China   Paper and Pencil Test

On a scare of 1-7, please rate the following sentences based on how acceptable they are. 1 indicates "totally unacceptable", and 7 means the sentence is perfectly acceptable.

1. I was surprised when I found out that Terry has never driven a car.

   1   2   3   4   5   6   7

2. The night that the policeman caught the thief, it was extremely cold.

   1   2   3   4   5   6   7

3. What did Sandy give to whom after the conference last Wednesday?

   1   2   3   4   5   6   7

4. I believed the claim that Philip would visit the city of Athens.

   1   2   3   4   5   6   7

5. The ball that the boy kicked the girl rolled into the gutter.

   1   2   3   4   5   6   7

6. The car that I test-drove last week has some minor scratches.

   1   2   3   4   5   6   7

251

7. Who did Mike give what to at the party in Meredith's apartment?

1   2   3   4   5   6   7

8. The actress that the director hired the photographer won a major award recently.

1   2   3   4   5   6   7

9. Julie became fond of the book after the discussion at her book club meeting.

1   2   3   4   5   6   7

10. Who did Jerry love seem to be known by everyone in the department?

1   2   3   4   5   6   7

11. The professor announced that there would be a test soon and then dismissed the class.

1   2   3   4   5   6   7

12. It rained last night in Baltimore.

1   2   3   4   5   6   7

13. Elliot quickly may free the sick and chained dog in the backyard of his cruel neighbor.

1   2   3   4   5   6   7

14. What Fred wondered whether was the assistant had the tools ready.

1   2   3   4   5   6   7

15. The customer that the waitress greeted the gentleman only comes on Thursday evenings.

1   2   3   4   5   6   7

16. The company would go bankrupt was claimed that by the stockholders.

    1   2   3   4   5   6   7

17. What she thought was that the poison was neutralized.

    1   2   3   4   5   6   7

18. The executive that John met his partner last week was in New York on a business trip.

    1   2   3   4   5   6   7

19. That whether the world is round is unknown upset Helen.

    1   2   3   4   5   6   7

20. My sister thought that I would be home for the rest of the day so she left without her keys.

    1   2   3   4   5   6   7

21. Jeff always thinks that parents of students want to annoy teachers.

    1   2   3   4   5   6   7

22. The afternoon that I had the car accident, I had a few glasses of beer.

    1   2   3   4   5   6   7

23. What Kelsey wondered was whether the store had the DVD in stock.

    1   2   3   4   5   6   7

24. I expected there to be a problem soon.

    1   2   3   4   5   6   7

25. The desk my roommate bought the lamp in IKEA matches the furniture in our room beautifully.

    1   2   3   4   5   6   7

26. The test Jane took last week had more essay questions than she expected.

    1   2   3   4   5   6   7

27. I persuaded there to be a problem of not having enough water.

    1   2   3   4   5   6   7

28. The doctor that the patient talked to the nurse outside the operation room was very cold.

    1   2   3   4   5   6   7

29. That Peter loved Amber seemed to be known by everyone in class.

    1   2   3   4   5   6   7

30. The therapist's analysis of Monica was flawed.

    1   2   3   4   5   6   7

LIST OF STIMULI SENTENCES IN EXPERIMENT I

Conditions	Set	Sentences
adjunct	1	The afternoon that the leopard bumped the cheetah from behind, a fight started.
adjunct	1	The weekend that the frog bumped the duck from behind, there was a holiday.
object	1	The duck heard that the leopard bumped the cheetah from behind and then moved through the weeds.
object	1	The turtle thought that the frog bumped the duck from behind and then moved through the weeds.
trace	1	The turtle that the frog bumped from behind moved through the weeds.
trace	1	The duck that the leopard bumped from behind moved through the weeds.
ungram	1	The duck that the leopard bumped the cheetah from behind moved through the weeds.
ungram	1	The turtle that the frog bumped the duck from behind moved through the weeds.
adjunct	2	The morning that the chicken called the lamb in the meadow, it was thundering.
adjunct	2	The month that the mouse called the turtle in the meadow, they danced.
object	2	The calf knew that the chicken called the lamb in the meadow and then danced in a puddle.
object	2	The duckling heard that the mouse called the turtle in the meadow and then danced in a puddle.
trace	2	The duckling that the mouse called in the meadow danced in a puddle.
trace	2	The calf that the chicken called in the meadow danced in a puddle.

ungram	2	The chick that the eagle chased the seagull around the yard climbed on a rock.
ungram	2	The hen that the rooster chased the turkey around the yard climbed on a rock.
adjunct	3	The year that the eagle chased the seagull around the yard, it was hot.
adjunct	3	The evening that the rooster chased the turkey around the yard, there was a fight.
object	3	The chick dreamt that the eagle chased the seagull around the yard and then climbed on a rock.
object	3	The hen hoped that the rooster chased the turkey around the yard and then climbed on a rock.
trace	3	The hen that the rooster chased around the yard climbed on a rock.
trace	3	The chick that the eagle chased around the yard climbed on a rock.
ungram	3	The owl that the goose frightened the pigeons at night landed in the tree.
ungram	3	The parrot that the cub frightened the woodpecker at night landed in the tree.
adjunct	4	The night that the gorilla followed the tiger in the woods, there was a storm.
adjunct	4	The year that the lion followed the bear in the woods, they were hunting.
object	4	The bear knew that the gorilla followed the tiger in the woods and then hid behind a tree.
object	4	The tiger guessed that the lion followed the bear in the woods and then hid behind a tree.
trace	4	The tiger that the lion followed in the woods hid behind a tree.
trace	4	The bear that the gorilla followed in the woods hid behind a tree.
ungram	4	The flamingo that the alligator heard the elephant in the dark stared at the moon.
ungram	4	The wolf that the buffalo heard the crocodile in the dark stared at the moon.
adjunct	5	The weekend that the goose frightened the pigeon at night, they flew away.
adjunct	5	The month that the cub frightened the woodpecker at night, there was a hurricane.

object	5	The owl pretended that the goose frightened the pigeon at night and then landed in the tree.
object	5	The parrot pretended that the cub frightened the woodpecker at night and then landed in the tree.
trace	5	The parrot that the cub frightened at night landed in the tree.
trace	5	The owl that the goose frightened at night landed in the tree.
ungram	5	The otter that the penguin hugged the walrus at bedtime snored during the night.
ungram	5	The reindeer that the otter hugged the guinea pig at bedtime snored during the night.
adjunct	6	The evening that the monkey greeted the kangaroo at the zoo, they a meeting.
adjunct	6	The minute that the ostrich greeted the giraffe at the zoo, the zookeeper walked in.
object	6	The giraffe hoped that the monkey greeted the kangaroo at the zoo and then jumped over the fence.
object	6	The kangaroo said that the ostrich greeted the giraffe at the zoo and then jumped over the fence.
trace	6	The kangaroo that the ostrich greeted at the zoo jumped over the fence.
trace	6	The giraffe that the monkey greeted at the zoo jumped over the fence.
ungram	6	The camel that the rhino kissed the zebra on the nose ran far away.
ungram	6	The zebra that the hippo kissed the camel on the nose ran far away.
adjunct	7	The second that the alligator heard the elephant in the dark, he jumped.
adjunct	7	The week that the buffalo heard the crocodile in the dark, it was windy.
object	7	The flamingo hoped that the alligator heard the elephant in the dark and then stared at the moon.
object	7	The wolf thought that the buffalo heard the crocodile in the dark and then stared at the moon.
trace	7	The wolf that the buffalo heard in the dark stared at the moon.
trace	7	The flamingo that the alligator heard in the dark stared at the moon.

ungram	7	The chick that the eagle loved the seagull for years sat in the dust.
ungram	7	The hen that the rooster loved the turkey for years sat in the dust.
adjunct	8	The minute that the hamster helped the rat in the afternoon, they shook hands.
adjunct	8	The winter that the snail helped the snake in the afternoon, they chatted.
object	8	The mouse said that the hamster helped the rat in the afternoon and then slept in the sun.
object	8	The toad guessed that the snail helped the snake in the afternoon and then slept in the sun.
trace	8	The toad that the snail helped in the afternoon slept in the sun.
trace	8	The mouse that the hamster helped in the afternoon slept in the sun.
ungram	8	The dog that the cat patted the rabbit on the back napped in the grass.
ungram	8	The rabbit that the bird patted the dog on the back napped in the grass.
adjunct	9	The evening that the penguin hugged the walrus at bedtime, they stayed up late.
adjunct	9	The night that the otter hugged the guinea pig at bedtime, he was snoring.
object	9	The otter said that the penguin hugged the walrus at bedtime and then snored during the night.
object	9	The reindeer guessed that the otter hugged the guinea pig at bedtime and then snored during the night.
trace	9	The reindeer that the otter hugged at bedtime snored during the night.
trace	9	The otter that the penguin hugged at bedtime snored during the night.
ungram	9	The dolphin that the panda poked the ape on the side laughed out loud.
ungram	9	The whale that the seal poked the shark on the side splashed in the waves.
adjunct	10	The winter that the goat kicked the horse in the field, it was snowing.
adjunct	10	The moment that the cow kicked the donkey in the field, the horse walked in.

object	10	The donkey heard that the goat kicked the horse in the field and then trotted into the barn.
object	10	The horse hoped that the cow kicked the doneky in the field and then trotted into the barn.
trace	10	The horse that the cow kicked in the field trotted into the barn.
trace	10	The donkey that the goat kicked in the field trotted into the barn.
ungram	10	The dolphin that the panda raced the ape to the shore blinked in the sun.
ungram	10	The kangaroo that the ostrich raced the giraffe down the road came to the river.
adjunct	11	The year that the rhino kissed the zebra on the nose, they ran away.
adjunct	11	The week that the hippo kissed the camel on the nose, there was a party.
object	11	The camel dreamt that the rhino kissed the zebra on the nose and then ran far away.
object	11	The zebra thought that the hippo kissed the camel on the nose and then ran far away.
trace	11	The zebra that the hippo kissed on the nose ran far away.
trace	11	The camel that the rhino kissed on the nose ran far away.
ungram	11	The duck that the leopard saw the cheetah in the morning floated in the pond.
ungram	11	The turtle that the frog saw the duck in the morning floated in the pond.
adjunct	12	The year that the bat liked the pig so much, he hardly cried.
adjunct	12	The time that the puppy liked the kitten so much, they were playmates.
object	12	The kitten said that the bat liked the pig so much and then played in the garden.
object	12	The pig pretended that the puppy liked the kitten so much and then played in the garden.
trace	12	The pig that the puppy liked so much played in the garden.
trace	12	The kitten that the bat liked so much played in the garden.
ungram	12	The otter that the penguin smelled the walrus in the air slipped on the ice.
ungram	12	The reindeer that the otter smelled the guinea pig in the air slipped on the ice.

259

adjunct	13	The time that the eagle loved the seagull for years, I won't forget.
adjunct	13	The summer that the rooster loved the turkey for years, we all remember.
object	13	The chick dreamt that the eagle loved the seagull for years and then sat in the dust.
object	13	The hen hoped that the rooster loved the turkey for years and then sat in the dust.
trace	13	The hen that the rooster loved for years sat in the dust.
trace	13	The chick that the eagle loved for years sat in the dust.
ungram	13	The chipmunk that the beaver surprised the coyote in the woods rolled down the hill.
ungram	13	The skunk that the raccoon surprised the squirrel in the woods rolled down the hill.
adjunct	14	The minute that the hamster met the rat near the rose bushes, it was raining.
adjunct	14	The weekend that the snail met the snake near the rose bushes, they played.
object	14	The mouse said that the hamster met the rat near the rose bushes and then rested after dinner.
object	14	The toad guessed that the snail met the snake near the rose bushes and then rested after dinner.
trace	14	The toad that the snail met near the rose bushes rested after dinner.
trace	14	The mouse that the hamster met near the rose bushes rested after dinner.
ungram	14	The owl that the goose teased the pigeon all day flew through the air.
ungram	14	The parrot that the cub teased the woodpecker all day flew through the air.
adjunct	15	The second that the cat patted the rabbit on the back, the door opened.
adjunct	15	The moment that the bird patted the dog on the back, it flew away.
object	15	The dog dreamt that the cat patted the rabbit on the back and then napped in the grass.
object	15	The rabbit pretended that the bird patted the dog on the back and then napped in the grass.
trace	15	The rabbit that the bird patted on the back napped in the grass.

trace	15	The dog that the cat patted on the back napped in the grass.
ungram	15	The giraffe that the monkey took the kangaroo down the road came to the river.
ungram	15	The whale that the seal took the shark to the shore blinked in the sun.
adjunct	16	The morning that the ant pinched the butterfly for no reason, it was raining .
adjunct	16	The summer that the spider pinched the bee for no reason, it was hot.
object	16	The bee knew that the ant pinched the butterfly for no reason and then looked for a new home.
object	16	The butterfly knew that the spider pinched the bee for no reason and then looked for a new home.
trace	16	The butterfly that the spider pinched for no reason looked for a new home.
trace	16	The bee that the ant pinched for no reason looked for a new home.
ungram	16	The kitten that the bat visited the pig before breakfast crawled across the floor.
ungram	16	The pig that the puppy visited the kitten before breakfast crawled across the floor.
adjunct	17	The day that the panda poked the ape on the side, they were swimming.
adjunct	17	The winter that the seal poked the shark on the side, it was snowing.
object	17	The dolphin heard that the panda poked the ape on the side and then laughed out loud.
object	17	The whale thought that the seal poked the shark on the side and then splashed in the waves.
trace	17	The whale that the seal poked on the side splashed in the waves.
trace	17	The dolphin that the panda poked on the side laughed out loud.
ungram	17	The duckling that the mouse called the turtle in the meadow danced in a puddle.
ungram	17	The calf that the chicken called the lamb in the meadow danced in a puddle.
adjunct	18	The morning that the goat pushed the horse through the door, there was a visitor.

adjunct	18	The afternoon that the cow pushed the donkey through the door, they got sick.
object	18	The donkey heard that the goat pushed the horse through the door and then fell in the mud.
object	18	The horse hoped that the cow pushed the donkey through the door and then fell in the mud.
trace	18	The horse that the cow pushed through the door fell in the mud.
trace	18	The donkey that the goat pushed through the door fell in the mud.
ungram	18	The tiger that the lion followed the bear in the woods hid behind a tree.
ungram	18	The bear that the gorilla followed the tiger in the woods hid behind a tree.
adjunct	19	The summer that the panda raced the ape to the shore, it was sunny.
adjunct	19	The winter that the ostrich raced the giraffe down the road, it was humid.
object	19	The dolphin heard that the panda raced the ape to the shore and then laughed out loud.
object	19	The kangaroo said that the ostrich raced the giraffe down the road and then came to the river.
trace	19	The kangaroo that the ostrich raced down the road came to the river.
trace	19	The dolphin that the panda raced to the shore blinked in the sun.
ungram	19	The kangaroo that the ostrich greeted the giraffe at the zoo jumped over the fence.
ungram	19	The giraffe that the monkey greeted the kangaroo at the zoo jumped over the fence.
adjunct	20	The afternoon that the ant rubbed the butterfly on the back, we were away.
adjunct	20	The weekend that the spider rubbed the bee on the back, they got lost.
object	20	The bee knew that the ant rubbed the butterfly on the back and then dreamed about flowers.
object	20	The butterfly knew that the spider rubbed the bee on the back and then dreamed about flowers.
trace	20	The butterfly that the spider rubbed on the back dreamed about flowers.

trace	20	The bee that the ant rubbed on the back dreamed about flowers.
ungram	20	The toad that the snail helped the snake in the afternoon slept in the sun.
ungram	20	The mouse that the hamster helped the rat in the afternoon slept in the sun.
adjunct	21	The day that the leopard saw the cheetah in the morning, there was a fire.
adjunct	21	The time that the frog saw the duck in the morning, we escaped.
object	21	The duck heard that the leopard saw the cheetah in the morning and then floated in the pond.
object	21	The turtle thought that the frog saw the duck in the morning and then floated in the pond.
trace	21	The turtle that the frog saw in the morning floated in the pond.
trace	21	The duck that the leopard saw in the morning floated in the pond.
ungram	21	The horse that the cow kicked the donkey in the field trotted into the barn.
ungram	21	The donkey that the goat kicked the horse in the field trotted into the barn.
adjunct	22	The afternoon that the gorilla scared the tiger by accident, he apologized.
adjunct	22	The spring that the lion scared the bear by accident, it was freezing.
object	22	The bear knew that the gorilla scared the tiger by accident and then went swimming in the pool.
object	22	The tiger guessed that the lion scared the bear by accident and then went swimming in the pool.
trace	22	The tiger that the lion scared by accident went swimming in the pool.
trace	22	The bear that the gorilla scared by accident went swimming in the pool.
ungram	22	The pig that the puppy liked the kitten so much played in the garden.
ungram	22	The kitten that the bat liked the pig so much played in the garden.
adjunct	23	The night that the penguin smelled the walrus in the air, it was windy.

adjunct	23	The spring that the otter smelled the guinea pig in the air, it was warm.
object	23	The otter said that the penguin smelled the walrus in the air and then slipped on the ice.
object	23	The reindeer guessed that the otter smelled the guinea pig in the air and then slipped on the ice.
trace	23	The reindeer that the otter smelled in the air slipped on the ice.
trace	23	The otter that the penguin smelled in the air slipped on the ice.
ungram	23	The toad that the snail met the snake near the rose bushes rested after dinner.
ungram	23	The mouse that the hamster met the rat near the rose bushes rested after dinner.
adjunct	24	The moment that the beaver startled the coyote near the log, we screamed.
adjunct	24	The summer that the raccoon startled the squirrel near the log, they were in a fight.
object	24	The chipmunk dreamt that the beaver startled the coyote near the log and then hopped up and down.
object	24	The skunk guessed that the raccoon startled the squirrel near the log and then hopped up and down
trace	24	The skunk that the raccoon startled near the log hopped up and down.
trace	24	The chipmunk that the beaver startled near the log hopped up and down.
ungram	24	The butterfly that the spider pinched the bee for no reason looked for a new home.
ungram	24	The bee that the ant pinched the butterfly for no reason looked for a new home.
adjunct	25	The moment that the beaver surprised the coyote in the woods, it started pouring.
adjunct	25	The minute that the raccoon surprised the squirrel in the woods, we got angry.
object	25	The chipmunk dreamt that the beaver surprised the coyote in the woods and then rolled down the hill.
object	25	The skunk guessed that the raccoon surprised the squirrel in the woods and then rolled down the hill.
trace	25	The skunk that the raccoon surprised in the woods rolled down the hill.

trace	25	The chipmunk that the beaver surprised in the woods rolled down the hill.
ungram	25	The horse that the cow pushed the donkey through the door fell in the mud.
ungram	25	The donkey that the goat pushed the horse through the door fell in the mud.
adjunct	26	The spring that the alligator tapped the elephant on the head, it was warm.
adjunct	26	The second that the buffalo tapped the crocodile on the head, it became quiet.
object	26	The flamingo hoped that the alligator tapped the elephant on the head and then ate breakfast early in the morning.
object	26	The wolf thought that the buffalo tapped the crocodile on the head and then ate breakfast early in the morning.
trace	26	The wolf that the buffalo tapped on the head ate breakfast early in the morning.
trace	26	The flamingo that the alligator tapped on the head ate breakfast early in the morning.
ungram	26	The butterfly that the spider rubbed the bee on the back dreamed about flowers.
ungram	26	The bee that the ant rubbed the butterfly on the back dreamed about flowers.
adjunct	27	The month that the goose teased the pigeon all day, we came to help.
adjunct	27	The time that the cub teased the woodpecker all day, they were arguing.
object	27	The owl pretended that the goose teased the pigeon all day and then flew through the air.
object	27	The parrot pretended that the cub teased the woodpecker all day and then flew through the air.
trace	27	The parrot that the cub teased all day flew through the air.
trace	27	The owl that the goose teased all day flew through the air.
ungram	27	The tiger that the lion scared the bear by accident went swimming in the pool.
ungram	27	The bear that the gorilla scared the tiger by accident went swimming in the pool.
adjunct	28	The spring that the chicken tickled the lamb in the hay, they became friends.
adjunct	28	The week that the mouse tickled the turtle in the hay, they fun.

object	28	The calf knew that the chicken tickled the lamb in the hay and then drank from a bowl.
object	28	The duckling heard that the mouse tickled the turtle in the hay and then drank from a bowl.
trace	28	The duckling that the mouse tickled in the hay drank from a bowl.
trace	28	The calf that the chicken tickled in the hay drank from a bowl.
ungram	28	The skunk that the raccoon startled the squirrel near the log hopped up and down.
ungram	28	The chipmunk that the beaver startled the coyote near the log hopped up and down.
adjunct	29	The week that the monkey took the kangaroo down the road, there was a contest.
adjunct	29	The evening that the seal took the shark to the shore, it was stormy.
object	29	The giraffe hoped that the monkey took the kangaroo down the road and then came to the river.
object	29	The whale thought that the seal took raced the shark to the shore and then blinked in the sun.
trace	29	The whale that the seal took to the shore blinked in the sun.
trace	29	The giraffe that the monkey took down the road came to the river.
ungram	29	The wolf that the buffalo tapped the crocodile on the head ate breakfast early in the morning.
ungram	29	The flamingo that the alligator tapped the elephant on the head ate breakfast early in the morning.
adjunct	30	The day that the cat touched the rabbit very carefully, the farmer appeared.
adjunct	30	The second that the bird touched the dog very carefully, he yelped.
object	30	The dog dreamt that the cat touched the rabbit very carefully and then went to the playground.
object	30	The rabbit pretended that the bird touched the dog very carefully and then went to the playground.
trace	30	The rabbit that the bird touched very carefully went to the playground.
trace	30	The dog that the cat touched very carefully went to the playground.

ungram	30	The duckling that the mouse tickled the turtle in the hay drank from a bowl.
ungram	30	The calf that the chicken tickled the lamb in the hay drank from a bowl.
adjunct	31	The morning that the bat visited the pig before breakfast, he learned to walk.
adjunct	31	The day that the puppy visited the kitten before breakfast, it was cloudy.
object	31	The kitten said that the bat visited the pig before breakfast and then crawled across the floor
object	31	The pig pretended that the puppy visited the kitten before breakfast and then crawled across the floor.
trace	31	The pig that the puppy visited before breakfast crawled across the floor.
trace	31	The kitten that the bat visited before breakfast crawled across the floor.
ungram	31	The rabbit that the bird touched the dog very carefully went to the playground.
ungram	31	The dog that the cat touched the rabbit very carefully went to the playground.
adjunct	32	The month that the rhino watched the zebra from a distance, it was foggy.
adjunct	32	The night that the hippo watched the camel from a distance, it was raining.
object	32	The camel dreamt that the rhino watched the zebra from a distance and then peeked through the leaves.
object	32	The zebra thought that the hippo watched the camel from a distance and then peeked through the leaves.
trace	32	The zebra that the hippo watched from a distance peeked through the leaves.
trace	32	The camel that the rhino watched from a distance peeked through the leaves.
ungram	32	The zebra that the hippo watched the camel from a distance peeked through the leaves.
ungram	32	The camel that the rhino watched the zebra from a distance peeked through the leaves.

Appendix E

LIST OF STIMULI SENTENCES IN EXPERIMENT II

A1. Irregular verb sentences

1. (Yesterday,) I bend a spoon   (Yesterday,) I bent a spoon
2. (Yesterday,) I bleed on it   (Yesterday,) I bled on it
3. (Yesterday,) I break a glass   (Yesterday,) I broke a glass
4. (Yesterday,) I bring an apple   (Yesterday,) I brought an apple
5. (Yesterday,) I build a castle   (Yesterday,) I built a castle
6. (Yesterday,) I buy one shoe   (Yesterday,) I bought one shoe
7. (Yesterday,) I catch a trout   (Yesterday,) I caught a trout
8. (Yesterday,) I choose a shirt   (Yesterday,) I chose a shirt
9. (Yesterday,) I deal a card   (Yesterday,) I dealt a card
10. (Yesterday,) I dig a hole   (Yesterday,) I dug a hole
11. (Yesterday,) I drive around town   (Yesterday,) I drove around town
12. (Yesterday,) I eat a banana   (Yesterday,) I ate a banana
13. (Yesterday,) I feed our eagle   (Yesterday,) I fed our eagle
14. (Yesterday,) I feel an earthquake   (Yesterday,) I felt an earthquake
15. (Yesterday,) I fight with Larry   (Yesterday,) I fought with Larry
16. (Yesterday,) I fly over Disney Land   (Yesterday,) I flew over Disney Land
17. (Yesterday,) I freeze a steak   (Yesterday,) I froze a steak
18. (Yesterday,) I give an answer   (Yesterday,) I gave an answer
19. (Yesterday,) I grow an inch   (Yesterday,) I grew an inch
20. (Yesterday,) I hear a story   (Yesterday,) I heard a story
21. (Yesterday,) I hide a coin   (Yesterday,) I hid a coin
22. (Yesterday,) I hold our baby   (Yesterday,) I held our baby

23. (Yesterday,) I keep a dime	(Yesterday,) I kept a dime
24. (Yesterday,) I lose a key	(Yesterday,) I lost a key
25. (Yesterday,) I make a cake	(Yesterday,) I made a cake
26. (Yesterday,) I meet a friend	(Yesterday,) I met a friend
27. (Yesterday,) I read a story	(Yesterday,) I read a story
28. (Yesterday,) I ride a horse	(Yesterday,) I rode a horse
29. (Yesterday,) I ring our bell	(Yesterday,) I rang our bell
30. (Yesterday,) I run a mile	(Yesterday,) I ran a mile
31. (Yesterday,) I sell a car	(Yesterday,) I sold a car
32. (Yesterday,) I send a letter	(Yesterday,) I sent a letter
33. (Yesterday,) I shoot an arrow	(Yesterday,) I shot an arrow
34. (Yesterday,) I sing in bed	(Yesterday,) I sang in bed
35. (Yesterday,) I sink a ship	(Yesterday,) I sank a ship
36. (Yesterday,) I sit in bed	(Yesterday,) I sat in bed
37. (Yesterday,) I sleep in bed	(Yesterday,) I slept in bed
38. (Yesterday,) I slide on ice	(Yesterday,) I slid on ice
39. (Yesterday,) I speak with Betty	(Yesterday,) I spoke with Betty
40. (Yesterday,) I spend a dollar	(Yesterday,) I spent a dollar
41. (Yesterday,) I spin on ice	(Yesterday,) I spun on ice
42. (Yesterday,) I steal a pie	(Yesterday,) I stole a pie
43. (Yesterday,) I stick around him	(Yesterday,) I stuck around him
44. (Yesterday,) I sting an eye	(Yesterday,) I stung an eye
45. (Yesterday,) I strike a nail	(Yesterday,) I struck a nail
46. (Yesterday,) I swear at school	(Yesterday,) I swore at school
47. (Yesterday,) I sweep our floor	(Yesterday,) I swept our floor
48. (Yesterday,) I swim a mile	(Yesterday,) I swam a mile
49. (Yesterday,) I swing a bat	(Yesterday,) I swung a bat
50. (Yesterday,) I take a penny	(Yesterday,) I took a penny
51. (Yesterday,) I teach a class	(Yesterday,) I taught a class
52. (Yesterday,) I tell a story	(Yesterday,) I told a story

53. (Yesterday,) I think about Mary   (Yesterday,) I thought about Mary
54. (Yesterday,) I weep with joy   (Yesterday,) I wept with joy
55. (Yesterday,) I win a prize   (Yesterday,) I won a prize
56. (Yesterday,) I write you poetry   (Yesterday,) I wrote you poetry

A2. Regular verb sentences

57. (Yesterday,) I ask a question.   (Yesterday,) I asked a question
58. (Yesterday,) I beg in town   (Yesterday,) I begged in town
59. (Yesterday,) I call a friend   (Yesterday,) I called a friend
60. (Yesterday,) I cause a riot   (Yesterday,) I caused a riot
61. (Yesterday,) I change a diaper   (Yesterday,) I changed a diaper
62.(Yesterday,) I clear a debt   (Yesterday,) I cleared a debt
63. (Yesterday,) I crawl into bed   (Yesterday,) I crawled into bed
64. (Yesterday,) I cry with joy   (Yesterday,) I cried with joy
65. (Yesterday,) I drop a plate   (Yesterday,) I dropped a plate
66. (Yesterday,) I dry a flower   (Yesterday,) I dried a flower
67. (Yesterday,) I fail an exam   (Yesterday,) I failed an exam
68. (Yesterday,) I fan our king   (Yesterday,) I fanned our king
69. (Yesterday,) I file a lawsuit   (Yesterday,) I filed a lawsuit
70. (Yesterday,) I fire a rifle   (Yesterday,) I fired a rifle
71. (Yesterday,) I gain a pound   (Yesterday,) I gained a pound
72. (Yesterday,) I glue one stamp   (Yesterday,) I glued one stamp
73. (Yesterday,) I help a stranger   (Yesterday,) I helped a stranger
74. (Yesterday,) I hire a nanny   (Yesterday,) I hired a nanny
75. (Yesterday,) I look after Sue   (Yesterday,) I looked after Sue
76. (Yesterday,) I move a chair   (Yesterday,) I moved a chair
77. (Yesterday,) I owe a dollar   (Yesterday,) I owed a dollar
78. (Yesterday,) I pass one test   (Yesterday,) I passed one test

79. (Yesterday,) I pay a fine　　　　　　(Yesterday,) I paid a fine

80. (Yesterday,) I plan a party　　　　　(Yesterday,) I planned a party

81. (Yesterday,) I play an instrument　　(Yesterday,) I played an instrument

82. (Yesterday,) I pour one gallon　　　(Yesterday,) I poured one gallon

83. (Yesterday,) I pray in bed　　　　　(Yesterday,) I prayed in bed

84. (Yesterday,) I prove a point　　　　(Yesterday,) I proved a point

85. (Yesterday,) I pull a tooth　　　　　(Yesterday,) I pulled a tooth

86. (Yesterday,) I raise a hand　　　　　(Yesterday,) I raised a hand

87. (Yesterday,) I reach a conclusion　　(Yesterday,) I reached a conclusion

88. (Yesterday,) I roar with laughter　　(Yesterday,) I roared with laughter

89. (Yesterday,) I roll a marble　　　　(Yesterday,) I rolled a marble

90. (Yesterday,) I sail a ship　　　　　(Yesterday,) I sailed a ship

91. (Yesterday,) I save a quarter　　　　(Yesterday,) I saved a quarter

92. (Yesterday,) I score a point　　　　(Yesterday,) I scored a point

93. (Yesterday,) I scrape our floor　　　(Yesterday,) I scraped our floor

94. (Yesterday,) I share a cake　　　　　(Yesterday,) I shared a cake

95. (Yesterday,) I sign a letter　　　　　(Yesterday,) I signed a letter

96. (Yesterday,) I slip on ice　　　　　(Yesterday,) I slipped on ice

97. (Yesterday,) I spy on Chris　　　　(Yesterday,) I spied on Chris

98. (Yesterday,) I stare around me　　　(Yesterday,) I stared around me

99. (Yesterday,) I stay after school　　　(Yesterday,) I stayed after school

100. (Yesterday,) I step on gum　　　　(Yesterday,) I stepped on gum

101. (Yesterday,) I stir our soup　　　　(Yesterday,) I stirred our soup

102. (Yesterday,) I stop a cab　　　　　(Yesterday,) I stopped a cab

103. (Yesterday,) I talk with Elbert　　　(Yesterday,) I talked with Elbert

104. (Yesterday,) I tie a ribbon　　　　　(Yesterday,) I tied a ribbon

105. (Yesterday,) I try her soup　　　　(Yesterday,) I tried her soup

106. (Yesterday,) I use a map　　　　　(Yesterday,) I used a map

107. (Yesterday,) I view a movie　　　　(Yesterday,) I viewed a movie

108. (Yesterday,) I walk after lunch　　　(Yesterday,) I walked after lunch

109. (Yesterday,) I weigh a package   (Yesterday,) I weighed a package

110. (Yesterday,) I whip an egg   (Yesterday,) I whipped an egg

111. (Yesterday,) I wish you joy   (Yesterday,) I wished you joy

112. (Yesterday,) I work with Fred   (Yesterday,) I worked with Fred

Appendix F

INSTITUTIONAL REVIEW BORAD (IRB) APPROVAL DOCUMENTS

UNIVERSITY OF DELAWARE   RESEARCH OFFICE

210 Hullihen Hall
University of Delaware
Newark, Delaware 19716-1551
Ph: 302/831-2136
Fax: 302/831-2828

DATE: September 11, 2013

TO: Arild Hestvik, PhD
FROM: University of Delaware IRB

STUDY TITLE: [273017-5] Past Tense

SUBMISSION TYPE: Continuing Review/Progress Report

ACTION: APPROVED
APPROVAL DATE: September 11, 2013
EXPIRATION DATE: September 23, 2014
REVIEW TYPE: Expedited Review

REVIEW CATEGORY: Expedited review category # 4, 7

Thank you for your submission of Continuing Review/Progress Report materials for this research study. The University of Delaware IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a study design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

This submission has received Expedited Review based on the applicable federal regulation.

Please remember that informed consent is a process beginning with a description of the study and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the study via a dialogue between the researcher and research participant. Federal regulations require each participant receive a copy of the signed consent document.

Please note that any revision to previously approved materials must be approved by this office prior to initiation. Please use the appropriate revision forms for this procedure.

All SERIOUS and UNEXPECTED adverse events must be reported to this office. Please use the appropriate adverse event forms for this procedure. All sponsor reporting requirements should also be followed.

Please report all NON-COMPLIANCE issues or COMPLAINTS regarding this study to this office.

Please note that all research records must be retained for a minimum of three years.

Based on the risks, this project requires Continuing Review by this office on an annual basis. Please use the appropriate renewal forms for this procedure.

- 1 -   Generated on IRBNet

273

If you have any questions, please contact Nicole Farnese-McFarlane at (302) 831-1119 or nicolefm@udel.edu. Please include your study title and reference number in all correspondence with this office.

DATE: May 16, 2013

TO: Zhiyin (Renee) Dong
FROM: University of Delaware IRB

STUDY TITLE: [437058-2] Filled Gap China

SUBMISSION TYPE: Amendment/Modification

ACTION: APPROVED
APPROVAL DATE: May 16, 2013
EXPIRATION DATE: February 27, 2014
REVIEW TYPE: Expedited Review

REVIEW CATEGORY: Expedited review category # 4, 7

Thank you for your submission of Amendment/Modification materials for this research study. The University of Delaware IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a study design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

This submission has received Expedited Review based on the applicable federal regulation.

Please remember that informed consent is a process beginning with a description of the study and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the study via a dialogue between the researcher and research participant. Federal regulations require each participant receive a copy of the signed consent document.

Please note that any revision to previously approved materials must be approved by this office prior to initiation. Please use the appropriate revision forms for this procedure.

All SERIOUS and UNEXPECTED adverse events must be reported to this office. Please use the appropriate adverse event forms for this procedure. All sponsor reporting requirements should also be followed.

Please report all NON-COMPLIANCE issues or COMPLAINTS regarding this study to this office.

Please note that all research records must be retained for a minimum of three years.

Based on the risks, this project requires Continuing Review by this office on an annual basis. Please use the appropriate renewal forms for this procedure.

If you have any questions, please contact Jody-Lynn Berg at (302) 831-1119 or jlberg@udel.edu. Please include your study title and reference number in all correspondence with this office.