SECOND LANGUAGE SENTENCE PROCESSING: IS IT FUNDAMENTALLY DIFFERENT?

by

Guillermo A. Rodríguez

B.A. in English Language and Literature, Universidad Nacional de Cuyo, 1998
M.A. in Applied Linguistics, University of Pittsburgh, 2002

Submitted to the Graduate Faculty of

Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2008

UNIVERSITY OF PITTSBURGH

ARTS AND SCIENCES

This dissertation was presented

by

Guillermo A. Rodríguez

It was defended on

August 25th, 2008

and approved by

Dr. David Mortensen, Assistant Professor, Department of Linguistics

Dr. Yasuhiro Shirai, Professor, Department of Linguistics

Dr. Tessa Warren, Assistant Professor, Department of Psychology

Dissertation Advisor: Dr. Alan Juffs, Associate Professor, Department of Linguistics

SECOND LANGUAGE SENTENCE PROCESSING: IS IT FUNDAMENTALLY DIFFERENT?

Guillermo A. Rodríguez, PhD

University of Pittsburgh, 2008

In this dissertation, the main assumptions in the Shallow Structure Hypothesis, developed by Clahsen & Felser (2006), are evaluated to determine whether the performance of second language (L2) learners when parsing sentences in the target language is fundamentally different. First, the claim that L2 learners do not employ phrase structure heuristics is assessed with stimuli made up of transitively- and intransitively-biased verbs followed by a noun phrase (Traxler, 2005). The second claim evaluated is that L2 learners do not use structurally defined gaps. This hypothesis is tested by comparing the learners' reading performance of intermediate gaps, stimuli with garden path effects and genitive nominalizations. The third assumption tested involves the use of configurational (binding) principles (Chomsky, 1981) in the parsing of cataphoric reference. The performance of L2 learners of English from Spanish and Chinese backgrounds is compared to that of native English speakers using the moving window paradigm. The relative influence of WM on the processing of these structures was also measured. Results show that both native and non-native speakers present similar parsing profiles and do make use of parsing heuristics. At the same time, both native speakers and L2 learners present difficulties accessing other kinds of structural information and resort instead to other clues that may render 'good-enough' representations (Ferreira et al., 2002). A pervasive finding as regards the WM capacity in L2 learners is the relationship found between the ability to store words and grammatical proficiency in a version of the reading span task (Daneman & Carpenter, 1980).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

This dissertation would not have been possible without the unconditional support, guidance and encouragement of my advisor, Dr. Alan Juffs. From the beginning, he managed to help me focus my academic interests without forcing his own vision, in a wholly open-minded way. Alan was always available to provide feedback on work in progress or advice with things non-academic, no matter what other more pressing issues were on his plate at the time. Dr. Juffs is also responsible for helping me discover new and more fulfilling ways to apply my research to the 'real' world. And he has always managed to do all this with the most positive attitude. I hope I am able to replicate his example with future students of mine. Alan's insight, knowledge and experience have made this dissertation a much better product. I would also like to thank the other members of my committee, David Mortensen, Yasuhiro Shirai and Tessa Warren for their patience and willingness to provide their knowledge and expertise in the improvement of this thesis. They shared extremely useful advice and different perspectives that helped better focus the outcome of my work. Of course, all errors in this thesis remain my own.

Another word of gratitude should go to three people who helped me achieve my goals and provided examples to follow. I would like to thank Dr. Robert DeKeyser who gave me the opportunity to start working with him and profit from his extensive expertise, and who has remained involved in my progress throughout the years. Dr. Pascual Masullo also contributed to

my academic enrichment with his almost encyclopedic knowledge about linguistics and beyond, and I am grateful for the friendship we have developed. I am also extremely thankful to Claire Siskin who was a kind boss and showed me what a relentless drive for improvement can achieve.

My fellow graduate students during my time at Pitt also deserve many thanks. Brian Brubaker, Wendy Martelle, Greg Mizera and Patti Spinner shared the ups and downs of graduate school with a smile. The department's Student Lounge was witness to our most interesting chats about academic and life issues.

I could never have reached the completion of this work without the support of my 'families'. Diana, Ian and Tom accepted me as one of their own and opened their world to me, never complaining when they heard "But Guillermo has to work…." Christine and Dawn (with Ceilidgh, Humo and Tsala) were always there to help, give precious advice and share everything with an open heart. Thanks also to Boston and Mia, the youngest members of the family, for making life seem so simple and fun.

A mi querida familia que aceptó la ausencia y la distancia como requisito ineludible de mis aspiraciones intelectuales. Papi, Mami y Loli, les agradezco tanto el apoyo, los sacrificios y el amor incondicional que hicieron posible este logro.

Finally, I would like to thank my wife Maeve, who stood by me at every step of the way. Her love, support and trust in me mean everything to me and have helped me achieve my dreams.

1.0 INTRODUCTION

A great deal of the research in Second Language Acquisition (SLA) is concerned with the role of input (see Carroll, 2001) in the development of a second language (L2) and how the already present first language (L1) affects its apperception (Odlin, 2005), whether in instructed or naturalistic contexts. Although countless studies have looked at the factors that may affect the noticing of L2 input (e.g., Robinson, 2002) and the type of input enhancement that may trigger a change in interlanguages (see VanPatten, 2004), little has been done as regards the actual processing of target language input online (Harrington, 2001). This dearth of research on the second-to-second processing of L2 input has been identified as highly problematic by many (Juffs, 2001; Papadopoulou, 2005). Since the construction of a grammatical system necessarily entails the analysis of input in order to form generalizations (implicitly or explicitly), the study of how learners comprehend sentences online should help us better understand the development of L2 grammars (Klein & Martohardjono, 1999). In particular, a more detailed picture of how L2 comprehenders integrate words into a sentence and how they assign a structural/semantic organization to the incoming strings at different levels of proficiency can shed light on a needed transition theory for SLA, i.e., how it is that the learner acquires the L2 system (Gregg, 2003; Juffs, 2001).

Another issue that may contribute to the development of a transition theory in SLA which has not been clearly elucidated in the literature so far, is the role of cognitive constraints such as memory. Working Memory (WM) has been described as the capacity to store and process information during the performance of a particular task (Baddeley, 1986). This construct represents an individual cognitive measure that has been attributed a preponderant role in the acquisition of an L2 (Baddeley, Gathercole & Papagno, 1998; Ellis, 2001; Ellis & Sinclair, 1996; Service & Kohonen, 1995). The claim for its importance stems from studies in which different subcomponents of WM, mainly the phonological loop and the central executive (Baddeley, 2003), have been found to correlate with and predict L2 proficiency. The phonological loop is thought of as a temporary buffer where phonological information is stored for further analysis or storage into long-term memory. On the other hand, the central executive is said to be responsible for the assignment of memory resources and for some information processing capacity of its own. Studies that claim that these constructs are related to the acquisition of an L2 have found that individuals with greater WM seem to be better at acquiring a target language (Gathercole, 2006). This advantage granted by a larger memory capacity is argued to be due to the better allocation of memory resources and the greater storage capacity that these learners are believed to have, which allow them to build better and longer-lasting representations of the grammar of an L2. In spite of the academic rigor and exacting experimental designs employed in these studies, there has not been a systematic use of measuring instruments (Conway et al., 2005; Friedman and Miyake, 2004); moreover the scoring procedures used may confound memory capacity with actual L2 proficiency (Rodríguez, 2006).

Juffs (2001) has identified the urgent need to carry out more research that brings together these two aspects, sentence processing and WM constraints, using psycholinguistic methodologies in order to advance our knowledge of how it is that learners process input in order to modify their interlanguage (Selinker, 1972). The study of these phenomena may, for example, help us uncover a further source of transfer (Odlin, 2005). Since the L2 learner approaches the task of acquiring the target language with a fully-fledged set of parsing strategies from their L1, these procedures may indeed be pervasive in their online analysis of input in the L2 (Juffs, 1998). At the same time, further research on L2 sentence processing may help us understand the reasons behind language parsing breakdown (Fodor, 1998). Failure at parsing and ensuing assignment of an accurate semantic and/or structural representation to a particular language string may raise the learner's awareness of a difference between their developing system and the appropriate target language configuration, allowing for reanalysis and modification of their interlanguage. Given that language breakdown may lead to a change in the L2 learner's grammar, being able to predict or identify the particular structures or input that may trigger this failure of the system may help us create materials that foster interlanguage changes based on primed language breakdown episodes (VanPatten, 1996; 2004). Research on sentence processing may also contribute to a better understanding of both the level of variance in individual performance in an L2 as well as the inability of most learners to achieve native-like levels of proficiency. It may be the case that L2 learners do acquire the grammatical knowledge necessary to understand sentences in their target language, but the lack of native-like parsing procedures may still prevent them from building appropriate L2 representations. Thus, L2 learners may show a parsing deficit instead of a competence deficiency (Fernández, 1999; Juffs, 2001; Juffs & Harrington, 1996); or couched in a different terminology, L2 learners may access only declarative information instead of tapping into procedural knowledge when functioning in the target language (Ullman, 2001). Additionally, exploring the interaction between L2 parsing and WM capacity may also help us explain some of the pervasive variation in individual performance (Juffs, 2005).

In spite of the lack of research on L2 sentence parsing procedures pointed out before, a group of researchers from the University of Essex led by Harald Clahsen has recently developed a proposal that attempts to explain the differences between native and learner parsing performance observed in the few studies available. Clahsen and Felser (2006; C&F henceforth) have thus developed the Shallow Structure Hypothesis (SSH) that envisions L2 sentence parsing as a fundamentally different procedure from processing of sentences online in a native tongue. The SSH states that second language learners do not make use of syntactic/structural information when parsing sentences in the target language and that, instead, they resort to all kinds of lexical, semantic and pragmatic information in order to assign an interpretation to a particular L2 string. According to C&F, the full parsing route, shaped by a grammar made up of symbolic rules and principles, is unavailable to L2 comprehenders who have to attempt the assignment of an interpretation to a sentence based only on lexical-semantic, pragmatic and world-knowledge information. C&F base this assertion mostly on studies dealing with the L2 processing of ambiguous relative clauses (RCs) and the parsing of structural gaps in populations with an advanced level of proficiency. C&F observe that possible sources of this difference may be an incomplete interlanguage system that may not provide detailed grammatical information for the L2 parser to act on the incoming string of words, the interference of the L1 parsing routines during L2 processing, and the depletion of cognitive resources triggered by having to access words and grammatical information in the foreign language that may overwhelm the working memory capacity of the learner online. However, the research this group has carried out so far does not seem to help isolate one of these characteristics of L2 speakers' performance as the culprit for the proposed fundamental difference (see Felser et al., 2007).

In this dissertation, I contribute data on the online processing of sentences by L2 learners with the aim of testing the assumptions recently put forth by Clahsen & Felser (2006) in their Shallow Structure Hypothesis (SSH). In particular, I assess C&F's hypothesis on the main assumptions that support the fundamental difference idea. First, I test the claim that L2 learners lack phrase structure heuristics that exploit word category (Papadopolou & Clahsen 2003; Felser et al, 2007) as found for native speakers (Traxler, 2002; 2005) in the resolution of local ambiguities. The second SSH claim evaluated here is the idea that second language learners do not make use of structurally defined gaps when processing sentences in their L2 (Felser et al., 2007; Marinis et al., 2005) and thus cannot parse long distance dependencies in a native-like fashion. Finally, I aim to determine whether L2 parsing is also guided by purely configurational (or structural) principles such as the binding principles advanced by Chomskyan generative grammar (Chomsky, 1981). At the same time, given one of C&F's probable explanations for the fundamental difference, the relative influence of WM on the processing of these structures is measured for both native and non-native speakers.

1.1 OVERVIEW OF THE DISSERTATION

This thesis is divided into six chapters. The first is the Introduction, in which the issues that motivate the three studies included in this work and the aims of the dissertation are laid out. In chapter two, a discussion on the use of phrase structure heuristics constitutes the beginning of the section, followed by the description, design and analysis of Experiment I, meant to test the use of these heuristics in L2 parsing. Chapter three presents the issue of filler-gap long distance dependencies and how they have been studied in L1 and L2 sentence processing. Included in this chapter is Experiment II, which constitutes an attempt at partially replicating previous research on this subject. The parsing of referential relationships with cataphoric pronouns is dealt with in chapter four, where Experiment III tests this parsing issue with non-native speakers. All WM discussion and data are confined to chapter five. In this section, the design of the WM tests used in the three experiments in this thesis, as well as the results stemming from these measures, are reviewed and discussed. Finally, chapter six presents the general summary of results and conclusions, together with the identifiable limitations in this work and possible avenues for future research.

2.0 PHRASE STRUCTURE HEURISTICS IN SENTENCE PARSING

Sentence parsing is the process comprehenders (readers/listeners) engage in when trying to incorporate each new word incrementally into the speech stream or the written discourse (Fodor, 1995). It involves the assignment of categories to words and hierarchical structure to strings of words, taking place very rapidly and usually without conscious awareness. Native speakers become aware of processing difficulties when parsing fails in some serious way, as in garden path sentences (Bever, 1970; Pritchett, 1992).

One of the most controversial issues in the study of sentence parsing is what kind of information the human parsing mechanism first uses in order to integrate a new word into a sentence. The camps are divided as to whether the process is a modular one, with only syntactic information being considered first (Frazier, 1987; Frazier & Clifton, 1989) or whether the many different syntactic, lexical and contextual cues interact simultaneously in order to arrive at a single interpretation of a particular utterance (e.g., MacDonald et al., 1994).

Principle-based parsing theories claim that the parser utilizes only syntactic or structural information in the first pass over a particular sentence (Frazier, 1987). This syntactic information is represented by the word class of each of the lexical items processed (noun, verb, determiner, etc.). Furthermore, the principles that operate over these word categories are believed to be universal and to foster the evaluation of a single structural alternative at a time. If the structural representation achieved during the first parse turns out to be inappropriate (due to semantic or plausibility anomalies) the parser is forced to reanalyze the string in order to determine a better configuration. The most representative model within this type of approach is the Garden Path Model (Frazier, 1987), which focuses primarily on two principles: minimal attachment and late closure. Minimal attachment refers to the need to join a new word without generating unnecessary structural nodes (or clauses). The second principle, Late Closure, posits that when finding a new word, the parser tries to integrate it within the current clause being processed. The two principles are meant to achieve efficiency and economy during parsing with only one structural alternative being considered at a time for both native and non-native sentence processing.

On the other hand, constraint-based parsing accounts envision the processing of sentences as a competition between different structural alternatives that are activated and considered simultaneously. Under the assumptions of this type of model, each new word encodes a plethora of cues, e.g. lexical, semantic, pragmatic, frequency counts, that are also joined by contextual cues stemming from the previous discourse encountered. All of these cues are activated at the same time and compete until one combination surfaces beyond a threshold that designates it as the winner over the other cue combination. In this framework there are no principles that determine a single interpretation at a time; instead, all possible configurations are considered in parallel. One of the best representatives of this type of approach is the Interactive Activation Model (MacDonald et al., 1994), which questions the existence of modular (syntactically encapsulated) sources of information in the parsing process.

Although there have been many studies on this topic using diverse psycholinguistic methodologies (see Mitchell, 1994 for a comprehensive review), the issue of what kind of information is accessed first when parsing remains a question with no definitive answer.

In the sections that follow I review studies from L1 parsing research that highlight evidence for the initial application of word category (phrase structure) heuristics in L1 parsing of local ambiguities. After that, I focus on the available studies on the processing of ambiguous stimuli in L2. In particular, I evaluate the evidence marshaled by C&F as proof of shallow processing of ambiguous stimuli by L2 learners.

2.1 PHRASE STRUCTURE HEURISTICS AND AMBIGUITY IN L1

The most widely used strategy to unveil the workings of the human parser entails exposing it to ambiguous stimuli that force the application of pervasive biases behind the normal functioning of the system. This is the data elicitation technique used in most of the studies that have attempted to tap into the processing of local ambiguities and, in particular, studies that have dealt with the role of subcategorization information in the resolution of ambiguity online.

Mitchell's (1987) work on subject/object ambiguity represents a good example of the type of study described above. Using the self-paced reading paradigm (Just, Carpenter & Wooley, 1982), in which participants are able to read sentences one word at a time in a non-cumulative fashion (moving window), Mitchell studied the heuristics used by native speakers to assign structure to sentences like (1) and (2).

> (1) After the child had sneezed (during surgery) / the doctor // prescribed a course of injections.
>
> (2) After the child had visited (during surgery) / the doctor // prescribed a course of injections.

In these stimuli, the ambiguity arises in (2) where the noun phrase (NP) *the doctor* may function as both the direct object (DO) of *visit* or the subject of the main predicate *prescribed*, which is ultimately the correct assignment for (2). Mitchell was also interested in whether the inclusion of an adverbial, *during surgery*, would sway readers' preferences as to one or the other structural assignment for the ambiguous NP in a similar fashion to the use of good or bad cues in Juffs' (1998) work on reduced relatives. Each sentence was divided in two sections for presentation, at the (/) marker for one version and at the (//) for another list. In spite of the clear difference in transitivity value for predicates such as *sneeze* and *visit,* participants in Mitchell's study encountered difficulty, meaning longer reading times (RTs) when parsing the ambiguous noun, *the doctor*, in both conditions. Even though these results provide evidence of a tendency in native speakers to follow the heuristics included in the Garden Path model described before, Mitchell's findings have been criticized on the basis of the artificial segmentation used for the presentation of the stimuli. This segmentation, according to some of the critics (Boland et al., 1995; Fodor, 1989), may have already biased the adjunction of the ambiguous NP as a DO of the verb, in spite of its subcategorization preferences (cf. Adams et al., 1998 for eye movement data that corroborates this assumption about Mitchell's materials).

In a similar study to Mitchell's, van Gompel & Pickering (1998) used materials that also included a subject/object local ambiguity, but this time the methodology used was eye-tracking. Van Gompel & Pickering were interested in determining whether the length of the ambiguous NP in (3) and (4) played any role in the resolution of its ambiguous role.

(3) After the dog struggled the vet and his new assistant took off the muzzle.

(4) After the dog scratched the vet and his new assistant took off the muzzle.

The researchers monitored eye fixations for each word in these stimuli and found that participants had more difficulty processing the ambiguous phrase, *the vet and his new assistant*, in (3) where the subordinate verb was clearly intransitive than when the predicate allowed for a transitive subcategorization frame as in (4). Another interesting finding stemming from this study is the fact that the transitivity value of the subordinate verb seemed to ease the reanalysis process in order to reassign a role to the ambiguous NP. When the subordinate verb was intransitive, participants were able to recover and reassign the role of subject of the main clause to the ambiguous NP, *the vet and his new assistant* (in (3)), more easily than when the subordinate verb presented a transitive bias (in (4)). The strength of the plausible misanalysis was very difficult to overcome once the disambiguating material in the shape of the main verb, *took*, was found.

The study to be replicated with L2 learners in the following section, Traxler (2005), is another attempt at determining the role of subcategorization information during the parsing of ambiguous stimuli in an L1. Hence, a more detailed review of this work follows.

Traxler's (2005) main motivation was to assess whether native speakers used transitivity as a lexical cue to avoid misanalyzing an NP V NP string. At the same time, his work was focused on the role of subcategorization and plausibility in the reanalysis needed for the kind of sentences used in Mitchell (1987) and van Gompel & Pickering (1998). In order to study this phenomenon, Traxler used the moving window paradigm to test the performance of 60 native speakers of English with 26 sets of stimuli like the sentences in (5) through (7).

(5) When Susan fell (,) the policeman stopped and picked her up.

(6) When Susan tripped (,) the table crashed to the ground.

(7) When Susan tripped (,) the policeman stopped and picked her up.

Each item in a set has two versions: one with a disambiguating comma and the other disambiguated once the reader reached the main verb (*stopped, crashed*). All materials in Traxler's stimuli sets were normed for subcategorization preferences (with a sentence generation task) and for the plausibility of the NP V NP strings (e.g. Susan fell the policeman; Susan tripped the table; Susan tripped the policeman) to avoid intrinsic biases in the materials. Consequently, sentences like (5) presented an intransitive verb, *fell*, followed by an NP that stood as an implausible DO for the subordinate verb. The second type of sentence, (6), incorporated a verb with transitive bias and an implausible DO and, finally, the third type (7), included a verb with transitive subcategorization preference and a plausible NP as DO. The sentences with commas included in the experimental sets were meant to act as a baseline for comparison with the sentences lexically disambiguated by the main verb of each sentence. The predictions advanced by Traxler involved, as mentioned before, the role of subcategorization information as well as the plausibility fit between the subordinate verb and the NP following it. Given the difference in subcategorization preference for sentences (5) and (7), if the reading profiles for these two conditions are similar (in the without-commas conditions), that would mean that comprehenders do not make use of this type of lexical cue, and apply some sort of heuristics based on word category instead. Even if this is the case, Traxler predicted that the degree of subcategorization bias (as normed) of the subordinate verb, together with the plausibility of the combination, should gravitate on the ease of recovery from misanalysis - implausible analyses are hard to adopt and plausible analyses are hard to abandon (Pickering et al., 2000). A comprehension question followed each target item and participants were equally accurate in answering these questions with an accuracy rate of 98% for items like (5), 98% for (6) and 94% for (7).

Traxler's results as regards RTs for the relevant regions of the stimuli are presented in Table 1.

Table 1. Raw reaction times from Traxler (2005)

Condtion	Subordinate Verb	Determiner	Noun	Matrix Verb	Spillover 1
(5) Ambiguous	541	509	488	472	450
Unambiguous	750	492	416	441	459
(6) Ambiguous	518	508	520	561	524
Unambiguous	652	498	460	476	508
(7) Ambiguous	487	458	510	573	533
Unambiguous	657	464	468	471	495

As mentioned above, Traxler's predictions indicated that if participants were using the subcategorization preference of a verb like *fell* in (5), they should be able to close the clause headed by this verb and open a new phrase when encountering the following NP, *the policeman*, in order to treat the latter as the subject of the incoming main verb. If, on the other hand, these comprehenders were not resorting to lexical cues initially, they should try to incorporate the NP as the DO of the intransitive verb, following structural principles such as minimal attachment and late closure. The ensuing implausible attachment would prompt reanalysis at a later stage in the processing of this same NP and would translate into significantly longer RTs for *policeman* in (5). The raw RTs shown in Table 1 corroborate Traxler's predictions, since participants were processing a subordinate verb like *fell* without considering its subcategorization preference as evidenced by the increment in reading times in the ambiguous condition (488ms in (5)) when compared to the comma condition (416ms). At the same time, the anticipated interaction between the plausibility of the misanalysis of the NP as DO of the subordinate verb modulated the difficulty found when overcoming the local ambiguity. In conditions (6) and (7), RTs are significantly higher than those found in condition (5), where the implausibility of the wrong attachment is found earlier and then eases the reassignment of sentence function to the ambiguous NP. When trying to explain these results, Traxler resorts to Frazier's (1987) ideas on the architecture of the human parsing mechanism. The parser needs to perform under temporal and cognitive constraints that force it to adopt the first analysis available with all of the lexical items processed so far. Thus it allocates a particular configuration to the string based on plain word category information that is afterwards revised (or filtered out) by using further lexical information.

Traxler (2002) also found similar parsing behavior in English-speaking children who performed using only structural information in the processing of the same kind of stimuli. Overall, Traxler's findings seem to lend support to the principle-based parsing approaches that claim it is only structural information that the human parser uses in the first pass over a string of words.

What is relevant about these findings for the experiment to be described below is the fact that native speakers of English present a tendency to analyze NP V NP strings as a unit, following the purported economical and efficient principles posited by frameworks like the Garden Path Model (or the Distance/Locality framework, Gibson, 1998). In the following section, I review a study of L2 sentence parsing that found similar results to Traxler's. In this same section, I assess research carried out by Clahsen and colleagues on this topic which constitutes the basis for their claim about a lack of phrase-based structure building in L2 parsing. The last section of this chapter introduces the study I carried out in an attempt to partially replicate Traxler's findings with L2 learners and the results obtained.

2.2 PHRASE STRUCTURE HEURISTICS AND AMBIGUITY IN L2

2.2.1 Subject/Object Ambiguity Resolution in L2 parsing.

Even though the focus of their paper was on the online resolution of ambiguity by L2 learners, Juffs & Harrington (1996) found similar results to those presented in Traxler (2005) with Chinese-speaking learners of English. Their emphasis was placed on the possible dichotomy between knowledge of grammar and parsing performance in L2 comprehension. Juffs & Harrington were working under the assumption that L2 learners may have the necessary grammatical competence in their L2, but they lack the ability to deploy this knowledge during parsing (comprehension). In order to test this idea, they presented sentences like (8) and (9) to 26 Chinese learners of English and 25 participants in the native speaker baseline group during an online grammaticality judgment task using the moving window paradigm explained before.

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

(9) After Sam arrived the guests began to eat and drink.

These are similar materials to those employed by Traxler (2005); however, the focus was not on the interaction between subcategorization and plausibility, but on the actual resolution of the local ambiguity once learners reach the disambiguating verb, *proved*, in (8). Juffs & Harrington were also interested in grammaticality judgment accuracy on these sentences, in order to assess whether dissociation between competence and parsing performance obtained. Results showed that both groups, native English and Chinese speakers, were very inaccurate in judging garden path sentences such as (8) as ungrammatical (58% and 47% accurate respectively), which indicates that both groups may have guessed at this stage of the task due to their inability to recover from the initial misanalysis of the words parsed (Warren, p.c.). On the other hand, sentences like (9), with an intransitive subordinate verb, were processed with significantly higher accuracy (92% for native speakers and 82% for the Chinese group). This difference in accuracy between conditions may be related to the hypothesis of Good Enough Representations (GER) first introduced to explain inaccuracy with this type of sentence in native speakers' performance by Ferreira and colleagues (2002), which will be further examined later on in this thesis.

The parsing profiles of both native and non-native speakers in Juffs & Harrington (1996) showed that all participants were surprised when encountering the main verb in sentences such as (8), even though Chinese speakers tended to pause longer on all predicates (probably a consequence of a difference in automatized processes in L2 (DeKeyser, 2001; Segalowitz, 2001)). A very interesting finding in Juffs & Harrington (1996) is relevant to the question of phrase-structure based parsing which, C&F claim, is not part of the L2 learner's parsing repertoire. Figure 1 shows the parsing profile for sentences like (9) in Juffs & Harrington's results. This profile demonstrates a significant spike in the RTs of non-native speakers when they processed the ambiguous NP, *the man*, in (9).

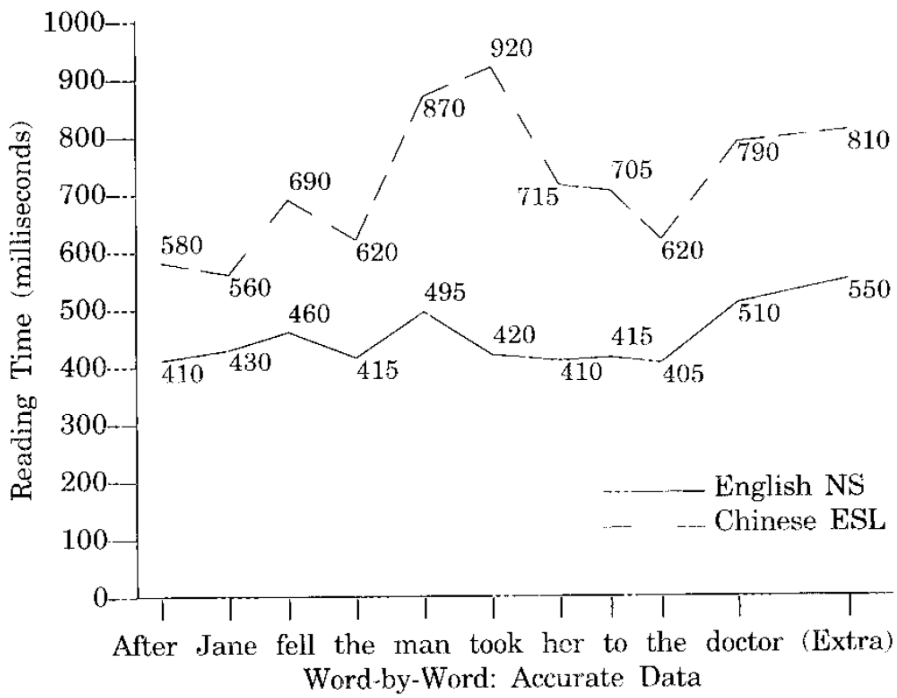


Figure 1. Parsing profiles from Juffs & Harrington (1996).

This finding did not represent the focus of Juffs & Harrington's investigation and they did not dwell on an explanation for this difference, except to say that in spite of the similarities in accuracy rates, the non-native speakers in their study seemed to present a parsing deficit/difference that was masked by their grammatical accuracy. It is also noteworthy that the pattern found in the non-native profiles followed the general shape of the native speakers' profile, in spite of significant differences in raw RTs between the groups. This stands then as preliminary evidence that the effects found in Traxler's (2005) data for native speakers may also be present in the parsing performance of non-native speakers, contrary to what has been suggested by C&F (2006).

Juffs (1998) went on to replicate the findings in Juffs & Harrington (1996), but included different L1 groups of language learners in order to assess the influence of the participants' mother tongue in the processing of this subject/object local ambiguity. This study also included the processing of the causative/inchoative alternation that is not the focus of the present work, but that justified Juffs' choices as regards L1 groups. Juffs tested the parsing performance of 4 different L1 groups: Chinese, Korean, Japanese and Romance L1 speakers. The items tested with these groups were similar to Juffs & Harrington's materials and are repeated in (10) and (11).

(10) Before Mary ate the pizza arrived from the local restaurant.

(11) After Mary died her husband married a woman from Texas.

The sentences appeared on a computer screen one word at a time, following a self-paced presentation methodology, and once the participant had finished reading the sentence, an extra button press prompted the question: "Was this a possible sentence in English?", to which a YES/NO reply was expected. Juffs (1998) reports no difference between groups when it comes to accuracy on this grammaticality judgment, lending support to the idea that non-native speakers are able to reanalyze these target items and come to the same accurate interpretation that native speakers extract from the sentences.

RTs for the regions considered in Juffs (1998) are shown in Figure 2.

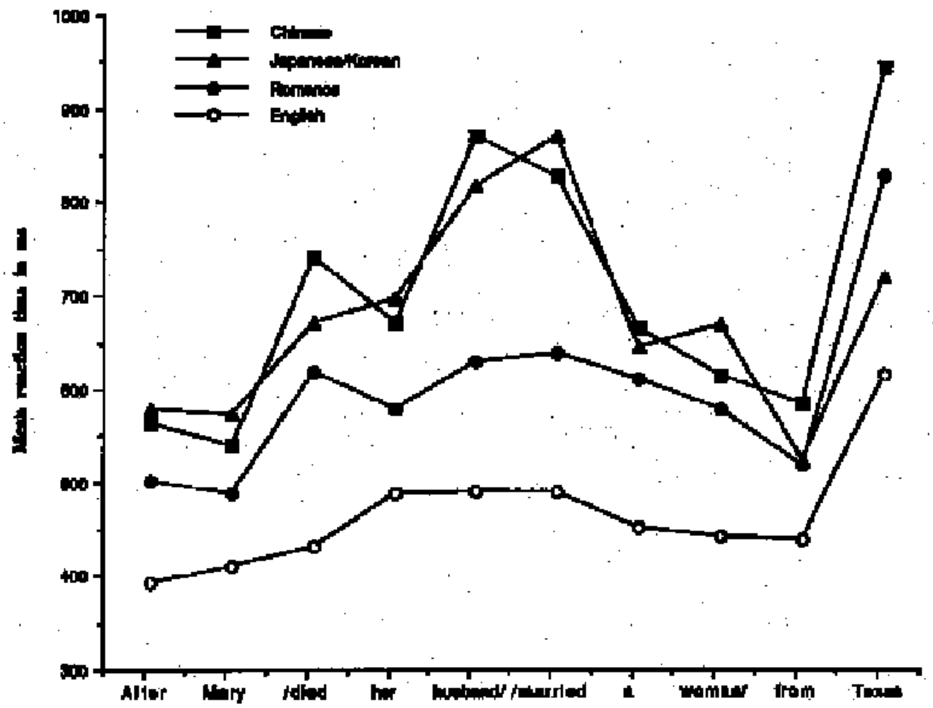


Figure 2. Parsing profiles from Juffs (1998)

Juffs decided to collapse the Japanese and Korean learners into a single group due to these languages' head-final nature, relevant to the discussion on the causative/inchoative alternation that was another aim of his study.

As can be seen in the profiles in Figure 2, the raw RTs for the NP following an intransitive verb in (11) replicate the findings in Juffs & Harrington (1996) with Chinese speakers. This time Korean, Japanese and Romance L1 speakers all show a similar tendency to the one presented in the 1996 study. The native speakers in Juffs' sample also seemed to have some difficulty when encountering the possessive, *her*, as the onset of an incoming NP, but this "surprise" effect seems to vanish and level off more rapidly than for non-native speakers. However, the biases found in this replication seem to point to some sort of universal tendency in parsing the NP V NP string in these stimuli that is followed not only by non-native speakers but also by L1 speakers of the language as in Traxler (2005).

The conclusion that can be drawn from the findings of the two studies as regards the use of phrase structure information seems to contradict the assumption of the SSH, namely that non-native speakers do not have access to the full parsing route in their model. It is evident from the profiles in Juffs' work that L2 learners of different L1 backgrounds do seem to apply the heuristics advocated in principle-based models of parsing, making use of word category information as soon as it becomes available and then having to reanalyze the structure based on the later plausibility effects.

2.2.2 Relative Clause Ambiguity Resolution in L2 parsing.

The evidence marshaled by C&F to justify their claim as to the use of phrase-structure heuristics comes from a different kind of structure altogether and, at this point, I shift focus to review the studies that have dealt with this phenomenon.

Clahsen & Felser (2006) state that L2 learners are guided by frequency, lexical, semantic and pragmatic cues encoded in the content words they parse, and also by cues from the previous context they have encountered. Most of their evidence for this claim stems from studies of relative clause (RC) ambiguity resolution. Two cases in point are Papadopoulou & Clahsen (2003) and Felser et al. (2003), who studied the performance of L2 learners when reading sentences such as (12).

(12) The reporter interviewed the daughter of the colonel who had an accident.

In (12) the RC *who had an accident* is globally ambiguous because it may modify the first noun phrase (NP1), *the daughter*, or it may attach to the second noun phrase (NP2), *the colonel*. Several studies have identified an attachment bias in native speakers of English, who prefer to incorporate the RC as a modifier of the NP2 (Cuetos & Mitchell, 1988), a phenomenon known in the literature as Low Attachment (making reference to the height of the attachment site in a traditional syntactic tree). At the same time, native speakers of other languages, such as Spanish, Dutch and Greek, present a different preference for the attachment site of the RC in this same structural configuration: in these languages the bias is towards attaching the RC to NP1, or High Attachment.

Papadopoulou & Clahsen (2003) tested very advanced L2 learners of Greek from Spanish, German, and Russian L1 backgrounds. Greek has been identified as having a High Attachment preference with the kind of stimuli presented in (12); a similar tendency has been documented for Spanish (Cuetos & Mitchell, 1988) and German (Hemforth et al., 2000). The researchers' rationale for including these groups with the same attachment bias was to investigate the influence of L1 proclivities in the parsing of a very similar structure in their L2. The ambiguous materials in the lists were disambiguated on the basis of gender either towards high or low attachment. Papadopoulou & Clahsen not only included a condition like the one in (12), but also another set of sentences that presented the equivalent of the thematic preposition *with* in Greek. This additional condition was meant to determine whether L2 learners were sensitive to the lexical cue encoded in this preposition, which has been found to trigger a Low Attachment preference crosslinguistically (Gilboy et al., 1995; Frenck-Mestre, 2005; Frenck-Mestre & Pynte, 1997) in tokens such as (13).

(13) Someone shot the servant *with* the actress who was on the balcony.

This preference for low attachment has been explained under the Construal Theory (Frazier & Clifton, 1996), which stands as a re-elaboration of the Garden Path Model (Frazier, 1987). The main principles in this model, Late Closure & Minimal Attachment, are believed to apply universally. However, Clifton & Frazier further qualified this claim by differentiating between primary and non-primary phrases. The universal principles would apply to primary phrases (obligatory constituents), but lexical and other biases would determine the attachment of non-primary phrases such as RC adjuncts. Thus the thematic preposition in (13) assigns a thematic role to its NP2 complement, and this consequently results in the low attachment of the RC to NP2.

The predictions stated in Papadopoulou & Clahsen (2003) anticipated that if L2 learners were able to access the full parsing route in C&F's model, they would follow their L2 (Greek) and their L1 parsing preferences for items like (12) – what they termed as a genitive NP – that favored NP1 attachment. However, the preference should shift for strings like (13) – prepositional NP – if L2 learners are able to access lexical cues in their target language; the latter being the main source of information for their parsing mechanism according to the SSH. Papadopoulou and Clahsen also included an offline task in this study in order to determine what the attachment preferences in this kind of modality were, since it has been claimed that offline tasks allow more time for attachment decisions to be based on all possible sources of information stemming from the stimuli.

Table 2. Raw reading times from Papadopolou & Clahsen (2003)

Participants	Gen-High	Gen-Low	PP-High	PP-Low
Greek native speakers	882.64	1222.12	938.38	864.32
Spanish learners	1915.85	1821.26	2035.71	1818.23
German learners	2648.49	2849.40	3225.31	2654.04
Russian learners	2285.79	2484.87	2649.23	2223.62

Raw RT results from the disambiguation region in the self-paced reading task are summarized in Table 2. On- and offline, L2 learners strayed from the performance of Greek native speakers with the genitive NP condition. While native speakers presented a high attachment preference for sentences such as (12), a genitive NP, L2 learners seemed to perform at random without a definitive bias in either presentation modality. However, all groups appeared to be sensitive to the thematic domain created by the preposition *with* in Greek, in (13), and were able to attach the RC to NP2 in this condition, following the Construal Model predictions.

In a similar study to Papadopoulou & Clahsen's, Felser et al. (2003) tested advanced learners of English from Greek and German L1 backgrounds. Felser et al. followed the same rationale employed in Papadopoulou & Clahsen, but, since English lacks grammatical gender marking on most nouns, they provided morphological disambiguation cues represented by number morphemes on the critical main verb that followed the genitive NP, as shown in (14) and (15).

> (14) The dean liked / the secretary of the professors / who / was (were) / reading a letter.
>
> (15) The dean liked / the professors with the secretary / who / were (was) / reading a letter.

The presentation modality chosen by Felser and colleagues for these stimuli was region-by-region instead of word-by-word, and the segmentation imposed on the sentences is shown with (/). In separate experiments, 28 German-speaking learners of English and 39 Greek-speaking learners were tested in an offline grammaticality judgment task that included sentences like (14) and (15) in grammatical and ungrammatical versions. The same participants performed in a self-paced reading task with items that were all grammatical and similar to the items in the grammaticality task. Felser et al. found that both groups of learners were very accurate in their knowledge of the ambiguous RC structure, with correct rejections in 92.8% of instances for the German group and 87.7% for the Greek-speaking participants. Table 3 shows raw RTs from the self-paced reading task on the critical region, segment 4 in (14) and (15), for both groups of learners.

Table 3. Raw reading times from Felser et al. (2003)

Segment 4 - (morphological disambiguation)	German	Greek	Native
Genitive (of) disambiguated to NP1	435 (141)	508 (224)	581 (471)
Genitive (of) disambiguated to NP2	439 (139)	533 (256)	495 (229)
Preposition (with) disambiguated to NP1	502 (176)	661 (333)	648 (513)
Preposition (with) disambiguated to NP2	428 (119)	532 (190)	512 (272)

The comparison between native speakers and L2 learners' RTs in these biased materials shows that native English speakers found it difficult (elevated RTs) to attach the RC to NP1 in the genitive condition. A similar kind of delay in the attachment of RCs disambiguated towards NP1 in the thematic preposition condition (*with*) was characteristic of the performance of native speakers. In contrast, Felser et al. report the lack of a similar tendency in the RTs of the learners. Neither the German group nor the Greek participants' RTs showed a clear and significant preference for either attachment site in the morphologically disambiguated materials used in the genitive condition in (14). However, in the thematic preposition condition, in (15), both groups of learners reacted similarly to the native speaker control group, by taking a longer time to process the critical region when these sentences were disambiguated towards NP1 attachment. Felser and colleagues interpreted these results as further evidence that L2 learners are able to access lexical cues such as the thematic nature of the preposition *with* in order to guide their parsing decisions. However, they claim that the results from the genitive condition support the SSH claim that non-native comprehenders are unable to apply the phrase structure heuristics that native speakers are capable of resorting to when there are no lexical biases, semantic or pragmatic information available to inform the L2 parser.

In spite of the experimental rigor demonstrated by both studies reviewed here, Fernández (2006) has recently criticized these results, stating that the stimuli used may have been biased towards one of the attachment sites (see Rodríguez, 2004, for a similar objection). This stands as a plausible flaw of the research carried out by Papadopoulou & Clahsen (2003) and Felser and colleagues (2003), since the materials were not normed. What is more, Fernández claims that it may not be a lack of syntactic information being applied by the learners that is the cause of their random behavior, but instead the misassignment of the correct prosodic structure to these sentences with ambiguous RCs that may have triggered the chance performance. This represents another sensible challenge to the results in both studies, taking into account the Implicit Prosody Hypothesis, developed by Fodor (1998), and given the artificial segmentation imposed on the target materials in these studies. Additionally, both studies provided as evidence for a lack of attachment preferences in L2 report that this was the case even in offline measures of the learners' attachment decisions. This is particularly puzzling, since it is not the canonical finding in the literature on RC attachment ambiguities offline and may reinforce Fernández's claim of inherent biases in the stimuli.

Furthermore, recent research on the processing of morphology by L2 learners (Jiang, 2004; 2007) seems to point to the fact that non-native speakers are more sensitive to violations of argument structure requirements than to morphologically inaccurate stimuli in the target language. Jiang (2007) presented L2 learners with grammatical and ungrammatical versions of the same items including violations of argument structure and number agreement mistakes in their foreign language as shown in (16)-(19).

(16) The visitor took several of the rare coins in the cabinet.

(17) * The visitor took several of the rare coin in the cabinet.

(18) The teacher wanted the student to start all over again.

(19) * The teacher insisted the student to start all over again.

In this study, Jiang recorded word-by-word RTs to both types of stimuli (grammatical and ungrammatical) and predicted that L2 learners would take longer to read the sentence region that presented an argument structure violation, but that the same would not obtain for sentences in which the mistake was located in a region involving a number agreement error. Results, reproduced in Table 4, show that native speakers' RTs were sensitive to both types of violations, argument structure and number agreement.

Table 4. Raw reading times from Jiang (2007)

		Plurality				Subcategorization			
	Test position	1	2	3	4	1	2	3	4
Native Speakers *(N = 26)*	Grammatical	411 (136)	456 (150)	441 (127)	408 (112)	443 (140)	442 (136)	435 (116)	450 (132)
	Ungrammatical	393 (113)	446 (150)	478 (134)	442 (111)	439 (129)	445 (130)	478 (123)	478 (136)
	Difference	−18	−10	37**	34**	−4	3	43**	28**
Nonnative Speakers *(N = 26)*	Grammatical	332 (76)	424 (120)	422 (61)	371 (61)	408 (73)	393 (79)	418 (80)	411 (69)
	Ungrammatical	342 (81)	445 (131)	415 (71)	388 (90)	424 (112)	416 (113)	450 (104)	451 (100)
	Difference	10	21	−7	17	16	23	32*	30**

*Significant at .05 level in participant analysis only.
**Significant at .05 level in both participant and item analyses.

As predicted though, non-native comprehenders showed sensitivity to argument structure errors in test positions 3 (the onset of the error) and 4 (the region right after the mistake) in Table 4, but no difference was found between the grammatical and ungrammatical versions of sentences including number agreement mistakes (see also Silva & Clahsen, 2008). This kind of finding calls into question the claims advanced by Felser et al. (2003) about the lack of a marked preference for RC attachment in the L2 learners they tested in their study. The logic here is that, according to Jiang's work, L2 comprehenders may not be able to incorporate morphological information when parsing online as native speakers do. Insensitivity to the kind of disambiguation used by Felser and colleagues, number agreement on the verb, may have caused the differences, attributed to a lack of phrase structure, between native speakers and L2 learners in the genitive NP conditions.

What is more, Swets et al. (2007) have recently reshaped the discussion of RC attachment ambiguity in L1 parsing by providing evidence that the tendencies found for speakers of English (and also native speakers of Dutch) may be the result of WM constraints. Swets' study found that individuals with high WM capacity (measured by means of reading span tasks in two modalities, language-based and visually-based) had a greater tendency to attach the RC low, whereas comprehenders with low WM capacity more commonly attached the RC high in the structure, contrary to what has been traditionally claimed the interaction between WM resources and attachment preferences should be like (Gibson et al., 2005) and providing further evidence in favor of chunking and silent prosody (Fodor, 1998) as the driving forces behind these attachment preferences.

2.2.3 Experiment I – PHRASE STRUCTURE PARSING

As mentioned before and exemplified by the RC studies analyzed in the preceding section, one of the most controversial claims of the SSH is that L2 learners do not have access to the full parsing route, since they are unable to exploit word category information in order to build initial syntactic representations automatically. However, some of the previously raised criticism involving RC attachment preferences as a basis for the SSH claim leaves room for further experimentation to assess the validity of this claim. Given the many variables that have been shown to affect RC attachment preferences (plausibility, length of the RC, intonation patterns, modification, WM capacity, L1 attrition, etc.), this particular structure does not seem to be the best candidate to study the application of syntactic information in L2 parsing. Consequently, the experiment described below attempts a replication of Traxler's (2005) findings with non-native speakers of English from Spanish L1 background. Even though the use of word category information was identified in the review of some of the studies that dealt with subject/object ambiguity resolution (Juffs, 1998; Juffs & Harrington, 1996) in preceding sections, this phenomenon was not the focus of those studies.

Thus, in trying to partially replicate Traxler's study with this group of L2 learners, I attempt to answer the following research questions:

(a) Do L2 learners parse sentences with local ambiguity in a native-like way in their target language?

(b) What kind of information do L2 learners use in order to initially integrate words into a sentence? Do they make use of subcategorization information in their first parse as Clahsen & Felser would argue, or do they apply a structural pattern based on word category alone in the initial stage like Traxler's native speaker participants?

(c) How does WM capacity relate to their ability to process this kind of stimuli in their L2? This last question will be dealt with in the chapters on WM and L2 parsing performance.

2.2.3.2 Participants

In order to answer the research questions stated above, 20 Spanish-speaking learners of English at the University of Pittsburgh, who were all pursuing or had already obtained graduate degrees, were tested individually. Each Spanish speaker received $30 in compensation for their time. The L2 proficiency of this group of learners was high-intermediate to advanced, as measured by an adapted version of the English Placement Test (EPT) included in the Michigan battery (Corrigan et al., 1978). All of the non-native speakers of English were citizens of Latin American countries who had come to the US in order to further their studies (see Appendix A for detailed biographical information). A control group of 27 native speakers of English provided the baseline necessary for comparison. These participants were undergraduate students at the

University of Pittsburgh and participated in this study in order to comply with a course requirement for Introduction to Psychology.

2.2.3.3 Tasks

The Spanish-speaking learners performed in three different tasks: two of them meant to test their WM capacity and one to test their parsing proficiency, in the form of a self-paced reading (SPR) comprehension task. This SPR task was aimed at determining whether learners parsed sentences in a similar way to the native speakers, without taking into account word category (phrase structure cues) as described in C&F's model of L2 parsing.

The SPR task was administered and controlled using a PC station running the experimental design software LINGER (Rhode, 2001). The presentation of items to be read followed the moving window paradigm (Just, Carpenter & Wooley, 1982), in which sentences are presented one word at a time in a non-cumulative fashion, so that the participants never see the complete sentence on the screen at one time. The words in the sentence are initially covered with a row of dashes and each time a participant presses the space bar on the computer keyboard a new word appears and the previous one simultaneously disappears from the screen. After each of the sentences was read in this task, participants were presented with a yes/no comprehension question. They were told to answer this question as fast and accurately as possible. This is done to keep the readers focused on the meaning of the sentences read, to avoid a very quick pass over the words in order to finish the task. It is also intended to divert the participants' attention from the actual grammatical characteristics of the target stimuli being used. There were 20 target items which were similar to materials used by Traxler (2005) and by Juffs (1998) and Juffs & Harrington (1996). Additionally, participants read 96 distracter items, all of which were also followed by a comprehension question and were unrelated to the aims of this particular study.

There was a 5-minute break after participants had finished reading half of the stimuli in the test. The nature and predictions related to the target stimuli are discussed in the following section.

2.2.3.4 Materials

In using similar stimuli to the one included in Traxler's (2005) study, the purpose was to replicate his study with L2 learners to test the hypothesis that L2 comprehenders are not guided by structural principles or syntactic information (Clahsen & Felser, 2006). Thus, the target items were constructed with adverbial subordinate clauses that included a verb biased either intransitively, *appear*, (20) or transitively, *clean*, (21) followed by an NP, *the bird* or *the stove*, which should be integrated as the subject of the incoming main verb, either *fly* or *heat*.

(20) When the tiger appeared the bird flew away.

(21) After the maid cleaned the stove began to heat up.

Only two of the original conditions tested by Traxler were included in this experiment, since the focus was not on the role of plausibility in reanalysis, but on the role of subcategorization.

The sentence in (21) is locally ambiguous because the NP2 may initially be attached as the DO of the subordinate adjunct verb, *clean*, but when the main verb, *begin*, is found, that initial erroneous allocation has to be reanalyzed and the NP2 has to be integrated as the subject of the main verb. This condition was included in order to compare it to the parsing of sentences like (20), where a principle-based parsing approach predicts that if L2 learners make use of syntactic principles (like minimal attachment and late closure) when integrating the NP2 in the initial parse, the plausibility assessment that takes place after that initial attachment should trigger a reanalysis process that would entail an increase in reading times over the region of the

NP2. If, on the contrary, L2 learners incorporate all cues simultaneously when a new word is parsed, they should be able to close the clause being built for the verb *appear* (in (20)) without expecting any extra material to be integrated as the DO of that intransitive verb.

2.2.3.5 Results

Figure 3 shows the parsing profile of the English native speakers in the relevant regions within the transitively (TBSV) and intransitively-biased (IBSV) target stimuli included in the SPR comprehension task.

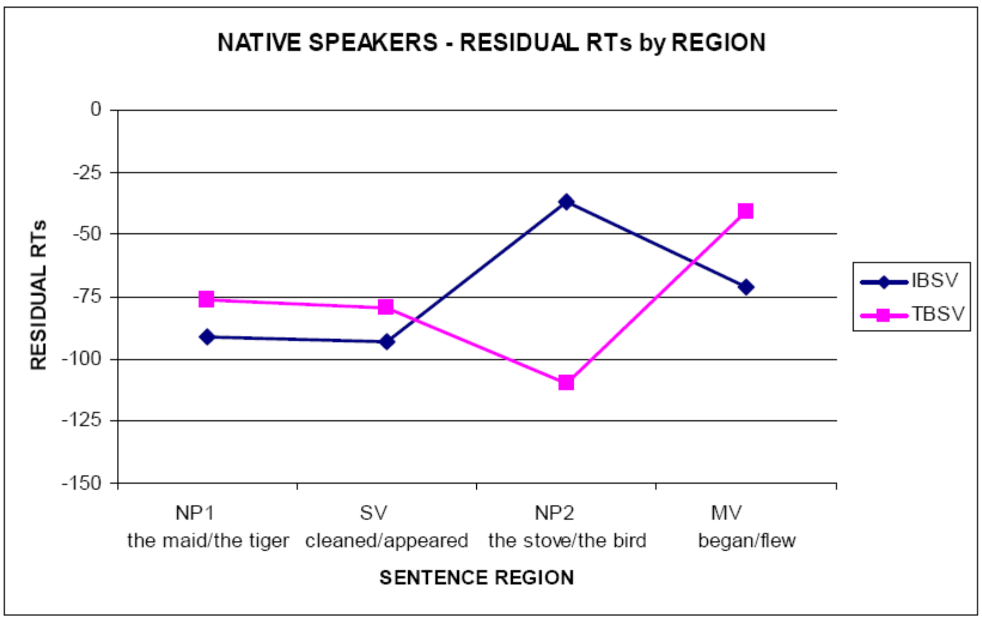


Figure 3. Native speakers' parsing profiles

The reading times referred to in these graphs (Y axis) reflect a transformation on the raw scores obtained for each of the participants, based on a regression analysis that takes into account the number of characters in each word and the average reading time for each of the participants. The result of this regression analysis is then subtracted from the actual raw reading times

(Ferreira & Clifton, 1986). The resulting values are presented in the graph and were used in the statistical analysis included in Tables 5 and 6 below. As is evident from the parsing profiles in Figure 3, the native speakers tested in this study behave similarly to Traxler's participants as regards the integration of the NP2 as the DO of an intransitive verb (IBSV). When these readers encountered the NP, *the bird*, they attempted to attach it to the preceding verb, disregarding its transitivity cue and trying to apply the parsing principles mentioned before. Conversely, when the subordinate verb was transitive (TBSV), the reading times on that NP2 actually decreased, due to the ease of incorporating the NP as the DO of the verb *clean* without a plausibility mismatch that would trigger reanalysis. Paired sample t-tests showed that the difference in reading times between the SV region and the NP2 region in the IBSV stimuli was significant (shown in Table 5).

Table 5. Native speakers' residual reading times by region

English Native Speakers

Type	CP	NP1	SV	NP2	MV	MV1	MV2	MV3
IBSV	-59.92	-91.03	-93.31[a]	-37.1[a]	-70.94	-78.77	-28.07	-101.43
TBSV	-46.38	-76.48	-79.49	-109.63[b]	-40.59[b]	-59.51	-64.63	-55.16

a significantly different $p < .0005$; b significantly different $p < .01$

As the reader may infer from Figure 4, the L2 parsing profile is very similar to the native speakers in reading time fluctuations (peaks and valleys). In the transitive stimuli (TBSV), L2 learners did not show any difficulty when trying to incorporate the NP2 into the parse of transitive verbs, although the increase in RTs when the main verb is encountered is not as steep as the one shown in the natives speakers' profile.

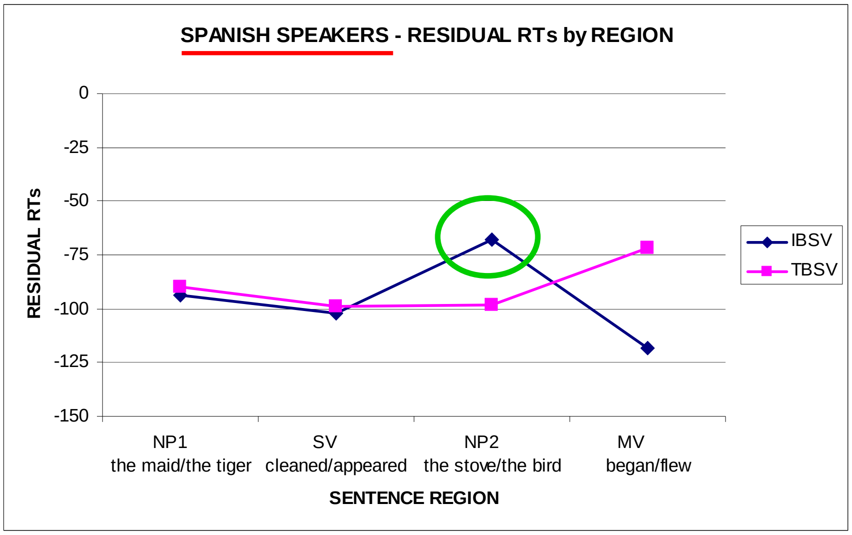


Figure 4. Spanish speakers' parsing profiles

The most important finding in these parsing profiles is the increase in RTs that is triggered by the presence of the NP2 following intransitive verbs in the IBSV stimuli.

Table 6. Spanish speakers' residual reading times by region

Spanish Native Speakers

Type	CP	NP1	SV	NP2	MV	MV1	MV2	MV3
IBSV	-162.01	-94.01	-102.47	-67.68[a]	-118.37[a;b]	-35.59[b]	-55.13	33.85
TBSV	-168.57	-89.96	-98.63	-98.38	-71.77	-30.37	-43.49	-82.89

a approaches significance p < .077; b significantly different p < .001

This increase in RTs for NP2, although not significant when subjected to paired samples t-tests comparing regions within the different sentence types (see Table 6), seems to indicate that the learners tried to deploy the structural principles that would force the attachment of that NP to the intransitive verb. However, the attachment failed after the initial parse, once plausibility constraints were considered.

Another interesting set of results as regards this task has to do with the native speakers' and learners' accuracy when answering the comprehension questions that accompanied the target stimuli. The percentage correct for each sentence type, IBSV and TBSV, for each group appears in Figure 5. The percentages both groups showed for the IBSV stimuli are very high, with 94.5% correct for the Spanish-speaking learners and 93.3% for the native speakers. On the other hand, the significant decrease in the accuracy levels for the TBSV may be the result of the application of what Ferreira et al. (2002) and Christianson et al. (2001) have termed Good Enough Representations (GER); i.e., sometimes the human parser does not compute complete detailed representations of sentences due to time, capacity or ambiguity constraints in the input being processed. Thus, the resulting representations contain a distorted assignment of thematic roles that are stored in this erroneous state for further processing, in this case answering a comprehension question. This finding is not crucial to the current aim of this section, but it may constitute an interesting counterpoint to C&F's theory, and promises to be a fruitful avenue for future research in comparing the GER of native and non-native speakers.

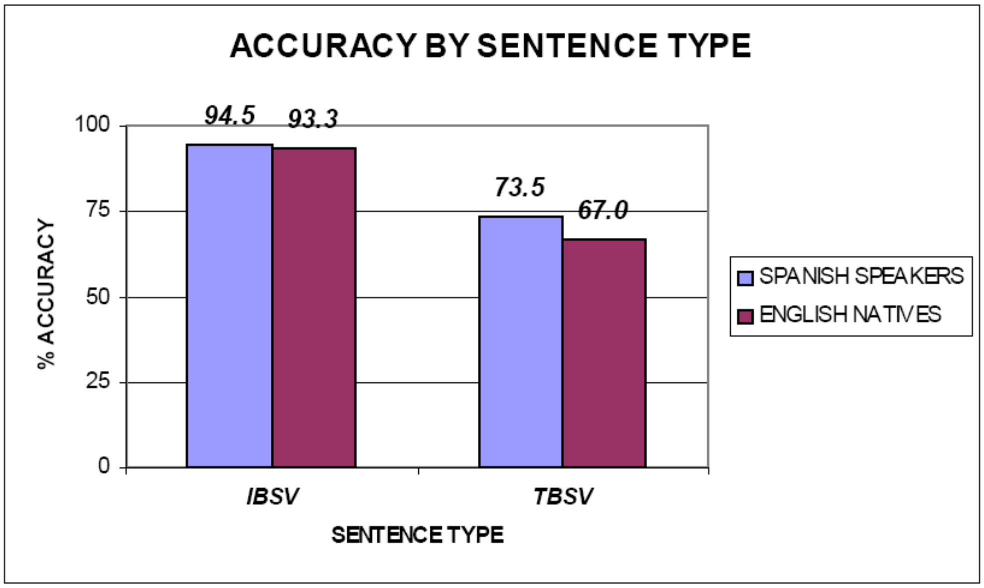


Figure 5. Comprehension accuracy by group and stimuli type

Some explanation should be given to the fact that Spanish-speaking learners were more accurate than native speakers in both conditions. Even though the differences in accuracy within conditions did not prove to be significant (IBSV: $F = .227$; $p = .636$; TBSV: $F = 3.263$; $p = .078$), it may have been the case that the undergraduate students were more eager than the Spanish-speaking learners to finish with the task as quickly as possible, as this test represented another class requirement.

2.2.3.6 Discussion

In order to summarize and attempt an explanation of the findings, I would like to come back to the three research questions stated at the beginning of this section. The first question dealt with a general assessment of the parsing performance of L2 learners, asking whether these participants would parse sentences in a similar fashion to the native speakers in an SPR task. As shown by the parsing profiles in Figures 3 and 4, the response to this question should be in the affirmative, since non-native speakers had very similar peaks and valleys when presented with the same stimuli parsed by native speakers of English. There may be a significant difference though in the parsing of the disambiguating region, the MV, in sentences that contained a transitively biased subordinate verb. If the groups are compared on this particular region, it is evident that the L2 learners' peak on the MV that disambiguates the role of the NP2 is not as sharp as the one presented in the parsing profile of the native speakers. Even though this was not the main object of analysis for this paper, this difference may be motivated by the availability of null subjects in Spanish, which, in certain cases, allows for two readings of this kind of transitive stimuli as in (22).

(22) Si Pedro conduce su coche ganará la carrera.
If Pedro drives his car (he/the car) will win the race.

In (22) the availability of a null-subject and the optionally transitive predicate, *conducir*, may make it still plausible for the NP2, *su coche*, to remain the object of the SV if Spanish speakers were applying their L1 grammatical competence in order to parse the target language. This seems to be the strategy some of the learners resorted to when parsing the target TBSV stimuli included in this task. But, again, the most important conclusion in relation to reading times in both groups is that in the regions of interest, the profiles were very similar, suggesting that both groups are behaving in accordance when integrating new words incrementally into the parse.

The second research question dealt with the type of information used in the initial pass over the string of words parsed by both native and non-native speakers. As was shown in Figures 3 and 4, both groups presented a peak on the NP2 region that could be interpreted as difficulty in integrating this word into the clause being processed. This is predicted to obtain only if the comprehenders are using structural or syntactic principles in order to incorporate that NP without taking into account lexical cues, such as transitivity. This would provide evidence against a strong version of the SSH advocated by Clahsen & Felser (2006). Instead of using lexical cues, both groups resorted to some kind of template in which the configuration NP V NP is initially parsed as a unit. Again this may be explained following the precepts established by the Garden Path Model (Frazier, 1987). The need to generate the fewest new structural nodes and to process each new incoming word inside the clause currently being parsed guides the incorporation of lexical items in the sentence in what seems to be a modular process unaffected by other possible cues. A possible counterargument to the explanation of the data, particularly in the IBSV stimuli (for both native and non-native speakers' profiles), may be to suggest that it represents a surprise effect at finding an unexpected NP after an intransitive. However, in Traxler's original stimuli, there were both ambiguous and unambiguous (with a comma after the subordinate verb) versions of IBSV items, the latter used as a baseline for effect comparisons. A closer look at the regions immediately after the intransitive subordinate verb in the unambiguous (comma) condition shows that in sentences such as (20), there was no statistical difference in RTs (509ms for the ambiguous version vs. 492ms for the comma condition) between conditions for the determiner of the crucial NP. The fact that there were no significant differences between ambiguous and unambiguous versions of IBSV stimuli immediately after the subordinate verb in Traxler's study does not seem to provide support for a surprise effect explanation of the pattern found for both native and L2 speakers.

Traxler (2005) argues that the only way to reconcile similar findings with a constraint-based approach is to claim that a greater number of sentences in the world's languages are of a transitive nature and thus the parser uses this coarse frequency count in order to process sentences like (20) and (21). However, Traxler notes, within a constraint-based approach one would then be forced to explain why it is that this type of frequency invariably wins over all other sources of information, if, as claimed by the Interactive Activation Model (MacDonald, 1994), these cues are all activated and considered at the same time, in parallel, in a non-modular fashion.

Having shown that the non-native speakers tested in this partial replication of Traxler's study show a tendency to use the same phrase-structure heuristics that native speakers employ with adverbial subordinate clauses (like in 20), in the following section I turn to another controversial claim made in the SSH: the idea that L2 learners do not pose structural gaps in the representation of long distance dependencies.

3.0 INTERMEDIATE GAPS IN SENTENCE PROCESSING

The mental representation and psychological reality of syntactic gaps has intrigued both linguists and psycholinguists since the early days of the Derivational Theory of Complexity (Foss & Hakes, 1978). This theory followed tenets from the early version of Transformational Grammar (Chomsky, 1973) and stated that difficulty in comprehension was directly related to the number of transformations (or derivations) that a deep structure (23) required in order to render a particular surface representation like the one in (24).

(23) The nurse knew the doctor saw *who* in his office?
(24) *Who* does the nurse know the doctor saw ____ in his office?

This early theory was disfavored due to evidence from psycholinguistic studies that falsified its main claim of a relationship between the number of transformations and the difficulty readers found when processing sentences (Ferreira, 2007). However, empirical interest in the phenomenon of gap parsing has not waned, as the following studies reviewed below demonstrate. It is the need to understand whether the grammar informs the parser, or how accurately the parser employs the proposed model of the grammar, that remains of interest to linguists and psycholinguists as well.

3.1 INTERMEDIATE GAPS IN L1 PARSING

The literature on the processing of intermediate structure (gaps) in L1 is truly extensive (see Mitchell, 1994; Phillips, 2006, for comprehensive reviews) and it involves mostly local dependencies in the processing of WH-movement and NP-movement (Gibson & Warren, 2004). However, the focus of this review will be on theoretically proposed intermediate gaps in long distance dependencies (Chomsky, 1981), since this kind of structure seems to have rendered evidence supporting the claim in the SSH of an absence of structurally-defined empty categories in the L2 parsing repertoire.

In the only prior study identified by Gibson & Warren (2004), Frazier & Clifton (1989) investigated the extraction of a filler (or the posing of gaps) across a clause and also across two consecutive clauses. The phenomenon Frazier & Clifton were trying to investigate involves the psychological reality of what linguists have termed as subjacency (or successive cyclicity). Subjacency (Chomsky, 1973) is a constraint on movement that prevents a moved constituent from crossing more than one bounding node at a time. Under this theoretical construct, bounding nodes are represented by noun phrases (NPs) and inflectional phrases (IP).

(25) Who$_i$ did the consultant (IP) claim t_i that (IP) the proposal had pleased t_i?

In a sentence like (25), the wh-filler *who* must cross two bounding nodes (IPs). Chomsky proposed that this movement should take place in two steps, with only one bounding node being crossed at a time. Moreover, the wh-filler should leave a coindexed trace once the first bounding node has been crossed at the specifier position of the CP for the predicate *had pleased* and at its originating node.

The two-consecutive-clauses condition in Frazier & Clifton (1989) proved to be more complex than conditions that did not include extraction phenomena. Frazier & Clifton characterized their results as evidence suggesting the existence of intermediate structure, and in particular, of successive cyclicity in the stimuli they used.

In spite of the evidence in favor of successive cyclicity in Frazier & Clifton's work, there are some inherent flaws identified by Gibson & Warren (2004) that may call into question the claim that there is evidence for intermediate structure stemming from this particular study. Gibson & Warren (2004) stress the fact that there was no norming of the materials used in Frazier & Clifton's stimuli; as a result, inherent lexical and plausibility biases may have made the intermediate structure sentences more complicated to process than those items that were included as baseline. At the same time, Gibson & Warren notice that all of the items used in the experiment were locally ambiguous. This poses the most serious objection to using Frazier & Clifton's evidence to support the existence of intermediate structure in these sentences, since the results obtained could potentially be explained by resorting to principles of ambiguity resolution, such as the Minimal Chain Principle (De Vincenzi, 1991), which states that the parser's aim is to reduce the distance between the filler (or constituent moved) and the actual trace or gap it should be associated with. Gibson & Warren (2004) also argue that this phenomenon could be equally accounted for by the Dependency Locality Theory (Gibson, 1998), which states that the difficulty in interpreting a gap increases with the distance between the filler and the gap position, where the filler should be reactivated.

It is precisely the Dependency Locality Theory (DLT) that Gibson and Warren (2004) resort to in order to explain their own experimental results as regards the parsing of intermediate gaps in L1 comprehenders. Based on the ideas in the DLT and their own Intermediate Structure

Hypothesis (ISH), Gibson & Warren predict that the presence of an intermediate gap (IG) in sentences such as (26) should facilitate the processing of the filler's subcategorizer, *pleased*, since this filler should have been reactivated at the stipulated gap that occurs before the head of the CP, *that*. On the other hand, an item like (27) presents a structure where filler reactivation is blocked by the lack of a clause boundary at which to pose an intermediate gap. Thus, in (27), the distance between the filler and its subcategorizer increases and, consequently, Gibson & Warren predict the RTs on the verb phrase, *had pleased*, which assigns a thematic role to the filler, should be longer. Since there has been no chance for that constituent to be reactivated at an intermediate site before reaching the role-assigning predicate, it should take longer to achieve filler integration in (27).

(26) Extraction across a VP (intermediate structure):
The manager who$_{i/}$ the consultant claimed/ ***IG*** that/ the new proposal/ had pleased/ t_i will hire/ five workers tomorrow.

(27) Extraction across an NP (no intermediate structure):
The manager who$_{i/}$ the consultant's claim/ about/ the new proposal/ had pleased/ t_i will hire/ five workers tomorrow.

Gibson & Warren added two more conditions to their target sets, justified by a confound they identified in items (26) and (27). Since the subject of the embedded predicate *pleased* in (26) is *the new proposal,* whereas the same function for *pleased* in (27) is performed by *the consultant's claim*, Gibson & Warren argued that the difference in distance between the subjects and the predicate may also help us predict that reading times over the embedded verb will be shorter in (26) than in (27). Hence, the additional conditions presented in (28) and (29) were derived from the target items in (26) and (27), but did not include extraction phenomena. As a result, both regions presented the same subject-verb integration distance than their extraction counterparts.

> (28) No Extraction - local subject-verb integration (VP):
> The consultant claimed that the new proposal had pleased the manager who will hire five workers tomorrow.
>
> (29) No Extraction - non-local subject-verb integration (NP):
> The consultant's claim about the new proposal had pleased the manager who will hire five workers tomorrow.

These additional items were meant to test the effect of extraction without the confounding factor of subject/predicate distance mentioned before. Gibson & Warren claimed that the ISH predicts longer RTs for the region *had pleased* in the extraction conditions (26 & 27) than in the stimuli where no filler gap dependency is present (28 & 29). Another prediction stemming from the application of the ISH to these stimuli involves longer RTs where the intermediate gaps are first posed. In (26), the site of the structurally-defined gap would be between the predicate *claim* and the head of CP, *that* (defined in syntactic terms as the specifier of CP). Thus longer RTs would be expected in this region in items like (26), but not in the NP extraction condition (27), since no gap is predicted to occur at this point in its structural representation. However, Gibson & Warren caution against the interpretation of their findings for this region, because, as the reader may infer, there are lexical differences between the conditions, with (26) taking *that* in the region following the stipulated gap, and the preposition *about* as its counterpart in (27). The verbs used as intermediate predicates, i.e. the verbs before the positing of a gap in (26), were all "bridge" verbs that are highly biased towards a sentence complement, according to published norms (Garnsey et al., 1997). Gibson & Warren also performed a separate norming procedure for these items to make sure that there were no intrinsic plausibility or lexical biases that may influence the participants' RTs for the crucial regions.

A group of 98 English native speakers participated in Gibson & Warren's self-paced reading task with the stimuli presented above, which was divided into four lists containing 20 target items and 70 fillers. All stimuli were followed by a comprehension question. The sentences were presented one word at a time, following the moving window paradigm, but RTs were later collapsed into seven different regions for analysis (as indicated in (26) and (27) with /). Their results showed that participants were less accurate when dealing with the extracted conditions (26 & 27) than with the items that did not present extraction. Another interesting outcome related to accuracy was that participants were more accurate with extractions from VP than from NP (this point will be revisited later on in the discussion section of Experiment II). Gibson & Warren used the regression equation described in the results section of Experiment I to transform raw RTs into residual RTs, following Ferreira & Clifton (1986).

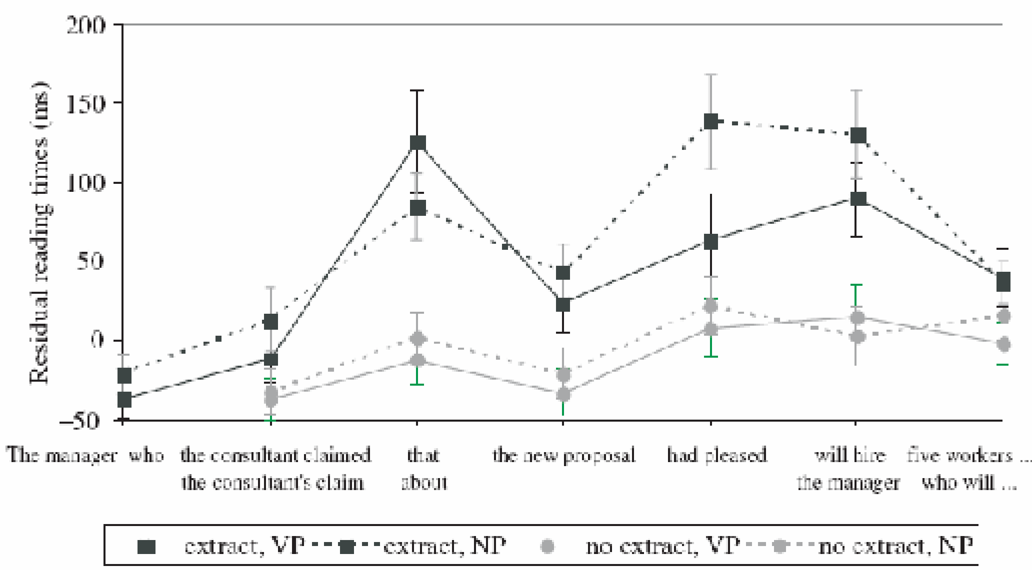


Figure 6. Parsing profiles from Gibson & Warren (2004)

Residual reading times for the seven regions in Gibson & Warren's (2004) four conditions are shown in Figure 6. The main regions of interest were 5, *had pleased*, and 6, *will hire*, in the extraction conditions. As predicted by the ISH, RTs were shorter for VP extraction stimuli than for NP extraction items, providing support for the claim that the difference is motivated by the existence of an intermediate landing site in the VP condition that allows the filler to be reactivated and thus easier to integrate with its subcategorizer in region 5. Furthermore, the additional non-extraction conditions that were included to test whether the difference in distance between the subject and verb in (26) and (27) could be a factor affecting reading times on the embedded predicate did not present the same effect and were consequently discarded as a possible explanation for the findings with extraction conditions. Larger residual RTs on region 3 including the complementizer *that* (for VP extraction items) could be understood to support the idea that an intermediate gap had been postulated before this region, when compared with lower RTs on the same region with the preposition *about* (for NP extraction items). However, as pointed out by Gibson & Warren, this particular finding should not be taken as definitive, since the lexical items in this region could be responsible for the difference in RTs. Additionally, they further report that this difference was not significant in the items analysis.

Summing up, Gibson & Warren's results can be taken to support the existence of intermediate structures in the processing of long distance dependencies, following the successive cyclicity constraint for filler-gap dependencies proposed by current theories of syntactic structure (Chomsky, 1981; Gazdar et al., 1985). In the section that follows, a review of studies dealing with the behavior of L2 learners when parsing similar stimuli in their target language is provided, and the second study in this dissertation is presented and discussed.

3.2 INTERMEDIATE GAPS IN L2 PARSING

In the introduction to this thesis, the lack of empirical data on the parsing of sentences in an L2 was identified as a problem that SLA should address in order to fully understand the role of input and to help develop a needed transition theory (Juffs, 2001; Papadopoulou, 2005). Given the small number of studies dealing with sentence parsing in SLA, and even scarcer data focusing on the L2 processing of long distance dependencies, it is somewhat surprising that C&F clearly identified this as one of the features that make L2 parsing fundamentally different from native comprehension. In this section, I review the few available studies that have gathered evidence supporting the existence of intermediate gaps in L2 performance and I also analyze the issues that Clahsen and colleagues marshal against these findings, including the work of their own research group, purportedly showing a lack of intermediate gaps in the representation of long distance dependencies in L2 parsing.

Juffs & Harrington (1995) tested a group of Chinese-speaking learners of English and, in addition to the adverbial subordinate clauses previously described here, included stimuli to assess their online processing of subject and object wh-extractions from finite clauses as exemplified in (30) and (31).

(30) *Who*$_i$ did Ann say e_i likes her friend?

(31) *Which man*$_i$ did Jane say her friends like e_i ?

The aim of this study was to confirm a previous claim that subject extraction from an embedded clause was more difficult to process than object extraction. The study sought to assess whether this difficulty was the result of competence (lack of grammatical knowledge) or a parsing deficit (inability to deploy grammatical knowledge online) in the learners' performance.

The methodology used in this study for the first time in SLA research was the self-paced reading paradigm with non-cumulative presentation and, since the stimuli included ungrammatical tokens, the software used recorded grammaticality judgments after each sentence the participants read. Juffs & Harrington (1995) found that the L2 learners presented more difficulties with items like (30), with subject extraction from finite clauses, a finding not replicated with the native speaker control group in their study. Table 7 shows the raw reading times for the pre and post-gap regions in both subject and object extraction tokens.

Table 7. Raw reading times for subject extraction in Juffs & Harrington (1995)

	Pre-gap 1	Post-gap 1	Post-gap 2
Chinese			
subject extr. (31)	744	1,035	825
object extr. (32)	675	797	759
English			
subject extr. (31)	480	491	452
object extr. (32)	455	442	429

A three way ANOVA with group, structure and position as factors showed significant effects for all factors. Additionally, the 300ms increment for subject extractions in the Post-gap 1 region for the Chinese learners (in Table 7) was significantly different from the other measures, whereas the English natives did not show this effect. Nevertheless, a similar increment on the Post-gap 1 region, at a minor scale, was found in the native speaker controls. Juffs & Harrington (1995) attributed the differences in reading times on this region to possible transfer from the L1 of these learners. The transferred characteristic of the learners' L1 being that Chinese is a 'wh-in situ' language (Aoun & Li, 1993), i.e. it does not involve movement of constituents in this type of constructions. Juffs & Harrington (1995) arrive at this conclusion based on the plausibility judgment data and accuracy rates, which were similar for both groups when parsing subject extractions, and which led them to discard the possibility of a competence difference.

In a similar study conducted to replicate these results with speakers of Japanese and Spanish as L1 as well as a group of Chinese-speaking L2 learners, Juffs (2005) assessed the influence of the L1 characteristics on the parsing of grammatical and ungrammatical wh-long-distance dependencies in English (as in 32 & 33 below). The need for replication arose from the lack of certainty as to whether the differences found in Juffs & Harrington (1995) stemmed from the wh-in-situ nature of Chinese (Aoun and Li, 1993; Watanabe, 2001) or whether there were other performance factors that prompted these L2 learners to behave differently from native speakers. Juffs (2005) also modified the stimuli used in the original studies, because the placement of the gap for object extractions coincided with the last word of the sentences used, which already triggers higher RTs for wrap-up effects. Furthermore, in trying to determine whether individual cognitive differences affected the parsing of these structures in English, Juffs (2005) added WM measures to the design. The relevant finding from Juffs' RT and accuracy data in this study is that differences between the Spanish-speaking and the Chinese/Japanese-speaking learners when judging grammatical and ungrammatical wh-extractions were not significant; i.e., all learner groups presented the same asymmetry in accuracy and raw RTs in their parsing profiles between subject and object extractions. Consequently, no matter whether the L1 presented wh-movement or wh-in-situ characteristics, subject-extractions proved more difficult to parse and judge than object extractions.

(32) Who does the boss expect ___ to meet the customers next Monday?
(33) Who does the boss expect to meet ___ next Monday?

However, drawing on data from sentences including extraction from nonfinite clauses in subject (32) and object position (33), which should trigger the same kind of reanalysis as extractions from finite clauses, Juffs (2005) arrives at a different conclusion. He suggests that, since nonfinite extractions proved easier for the learners, what may really be the problem for these non-native speakers is the juxtaposition of two finite verbs for the finite stimuli and not the revision of structural parsing assignments made on-line. These results with non-finite clauses put into question the claims made as regards the influence of the L1 for the Chinese learners in Juffs & Harrington (1995).

The findings reviewed above may lend support to the idea that non-native speakers do have access to complex structural representations, the full parsing route in C&F's model. Nevertheless, objections to the interpretations of these results raised from the SSH camp will be discussed after the description of similar results from Williams and colleagues' work (2001) are reviewed below.

Williams, Mobius & Kim (2001) focused on the use of plausibility information in L2 sentence processing by learners of English from Chinese, Korean and German L1 backgrounds. In order to test the availability of this kind of information in the parsing repertoire of L2 learners, they chose to study the filled-gap effect in prepositional adjunct phrases. This effect involves a filler whose subcategorizer is a preposition, but which is initially taken to satisfy the argument requirements of the main verb in the wh-construction. Williams et al. manipulated the plausibility fit between the filler NP and the main verb in the wh-questions that constituted their stimuli as exemplified in (34) and (35).

(34) Plausible at V
Which girl$_i$ did the man push the bike into e_i late last night?

(35) Implausible at V
Which river$_i$ did the man push the bike into e_i late last night?

Available empirical evidence has shown that native speakers attempt to associate the filler, *which girl*, as the object of the main verb, *push*, following the Active Filler Strategy (Stowe, 1986) and they later have to reanalyze this NP's role as the object of the preposition, *into*, when the actual object of the main verb, *the bike*, is found. This filled-gap effect triggers an increase in RTs on the NP, *the bike*, in (34). In order to test whether L2 learners employed the same heuristics native speakers apply, Williams and colleagues used a stop-making-sense task that required the participants to press a button when they detected the sentence they were reading, presented with a moving window display, would end up being implausible. As predicted, both native and non-native speakers interrupted their reading with a stop-making-sense decision in the implausible at V condition when they reach the verb or immediately after this region. In sentences like (35), when learners reached the main verb, they expected it to attach to the wh-phrase, which seems implausible.

Interesting results also arose from the reading time data on the stop-making-sense task in Williams et al.'s study (the Korean learners' profiles are shown in Figure 7 to exemplify L2 behavior, together with the native speakers controls in Figures 8).

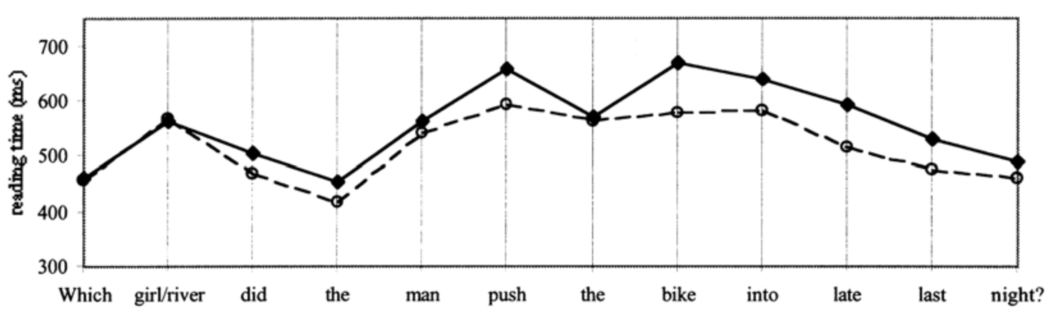


Figure 7. Native speakers' parsing profiles from Williams et al. (2001)

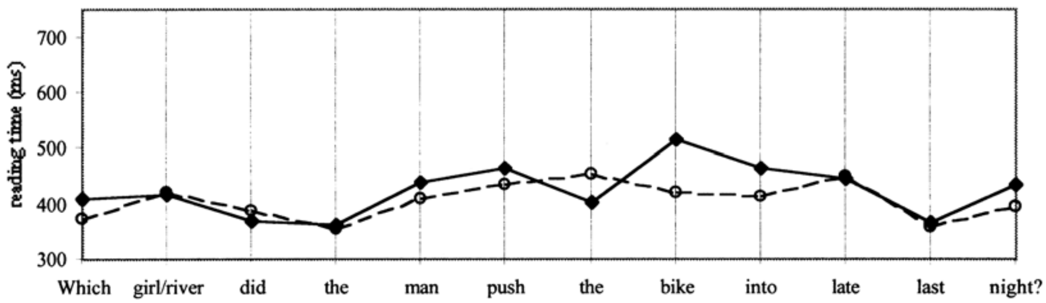


Figure 8. Korean speakers' parsing profiles from Williams et al. (2001)

The expected filled-gap effect obtained and was corroborated by elevated RTs on the NP following the main verb in the wh-questions for all groups, even though the onset of this filled-gap effect took place on the NP's determiner for native speakers and on the actual noun for the learners. These results replicate the findings in Traxler (2005) for native speakers, as well as results found in this dissertation with L2 learners, as to the application of phrase structure heuristics in L1 and L2 parsing.

The reviewed findings from Juffs & Harrington (1995) and Williams et al. (2001) could be taken to indicate that L2 learners do indeed process wh-gaps in a native-like fashion and that they are able to achieve accurate representations in order to answer the comprehension questions included in the experimental designs of these studies. Nevertheless, Clahsen & Felser (2006) and Marinis et al. (2003) do not accept that these studies provide conclusive evidence for the existence of gaps in L2 parsing of local and long-distance dependencies. They argue that due to the location of the gap, next to its subcategorizer, in sentences such as (34) and (35), it is not feasible to discard the possibility that the non-native comprehenders were using a direct association strategy based on lexical (thematic) and plausibility information, instead of structurally informed parsing decisions purportedly present in the native speakers' performance. Instead, Clahsen and colleagues provide empirical support in favor of their SSH with studies involving the processing of filler-gap relationships in long-distance dependencies in which the subcategorizer and the gap are not adjacent. A review of the most representative of these studies follows.

In a very interesting experimental design, Felser & Roberts (2007) used a crossmodal picture priming task (Swinney et al., 1979) in order to investigate whether L2 learners showed any signs of reactivation of fillers at intermediate gaps in long distance dependencies. Felser & Roberts tested Greek-speaking learners of English that were classified as having advanced proficiency (according to Oxford Placement Test scores) and compared their performance to that of 54 native speakers (from a different study: Roberts et al., 2007). The experimental task consisted of listening to sentences involving intermediate gaps with indirect prepositional objects (as shown in conditions (36)-(39)) and answering comprehension questions for some of the items. At the same time, the participants were asked to judge whether an entity being shown in a picture (SQUIRREL, TOOTHBRUSH) at gap or pre-gap positions was alive or not.

(36) Identical – Gap position
Fred chased the squirrel to which the nice monkey explained the game's difficult rules [SQUIRREL] in the class last Wednesday.

(37) Identical – Pre-gap position
Fred chased the squirrel to which the nice monkey explained the game's [SQUIRREL] difficult rules in the class last Wednesday.

(38) Unrelated - Gap position
Fred chased the squirrel to which the nice monkey explained the game's difficult rules [TOOTHBRUSH] in the class last Wednesday.

(39) Unrelated – Pre-gap position
Fred chased the squirrel to which the nice monkey explained the game's [TOOTHBRUSH] difficult rules in the class last Wednesday.

The predictions advanced by Felser & Roberts anticipated that if native speakers and L2 learners behaved similarly, both groups would experience antecedent priming effects, evidenced by faster reaction times to the alive/not alive decision, when the picture presented was identical to the antecedent and appeared at structural gap positions. Since the antecedent (*the squirrel*) should be reactivated at this site (after *the game's difficult rules*), it would be easier to make a judgment there than in a pre-gap position which did not entail recovering the antecedent from memory. The researchers also included parallel conditions in which the pictures shown were unrelated to the antecedent (e.g. TOOTHBRUSH) to control for mere lexical priming effects.

Comprehension accuracy was high for the non-native speakers, with an average of 96% for end-of-trial questions. The L2 learners also performed very accurately in the alive/not alive picture decision task with 94% accurate ratings. Felser & Roberts draw a comparison between the performance of these L2 learners and native speakers' data stemming from a separate study by Roberts et al (2007). In the latter study, Roberts and colleagues found an effect for identical picture priming at gap positions with alive/not alive decisions taking a significantly shorter time than the same judgment realized at pre-gap positions. However, this effect was significant only for those native speakers who presented high scores in a WM measure following Daneman & Carpenter's (1980) reading span task. The relevant WM results for this study will be discussed later in this thesis, but the important finding from the SSH perspective here is that the non-native speakers behaved differently from the English natives, as shown in Figure 9.

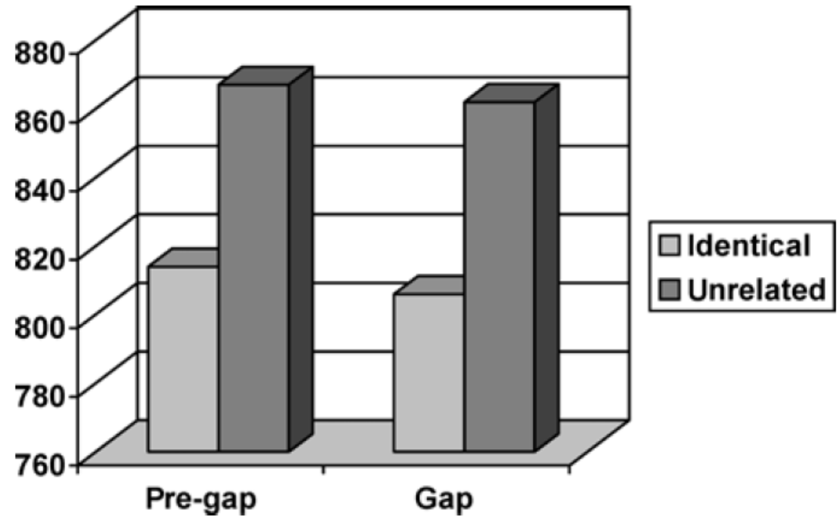


Figure 9. Non-native speakers' reaction times to probes from Felser & Roberts (2007)

Figure 9 shows average reaction times on the alive/not alive judgment required of L2 learners when presented with an identical or unrelated (to the antecedent) picture at pre-gap and gap positions. The lack of a significant difference between the pre-gap and gap positions with identical picture primes seems to indicate that the L2 comprehenders were not sensitive to the presence of a structural gap that required the reactivation of the antecedent, unlike high-WM native speakers in Roberts et al. (2007). This result is one of the pieces of evidence Clahsen & Felser's model takes into account to claim that L2 learners do not pose structural gaps in long distance dependencies, and instead join fillers with their licenser based on a direct association procedure that is informed by semantic and lexical cues encoded in the lexical items.

However, given that Felser & Roberts (2007) compared the performance of the L2 learners to the native speakers in Roberts et al. (2007), it would serve our purpose to look at the raw data gathered in that study. Since Felser & Roberts did not find an interaction between WM, measured with a version of Harrington & Sawyers' RST (1992), they collapsed results for the non-native speakers, as shown in Figure 9. Thus, if we performed the same analysis, averaging across native speakers' RTs in Roberts et al. (2007), instead of separating them according to RST measures, the picture that arises from the data is very similar for both learners and native speakers. Figure 10 shows the averaging of high and low WM participants in Roberts et al. (2007).

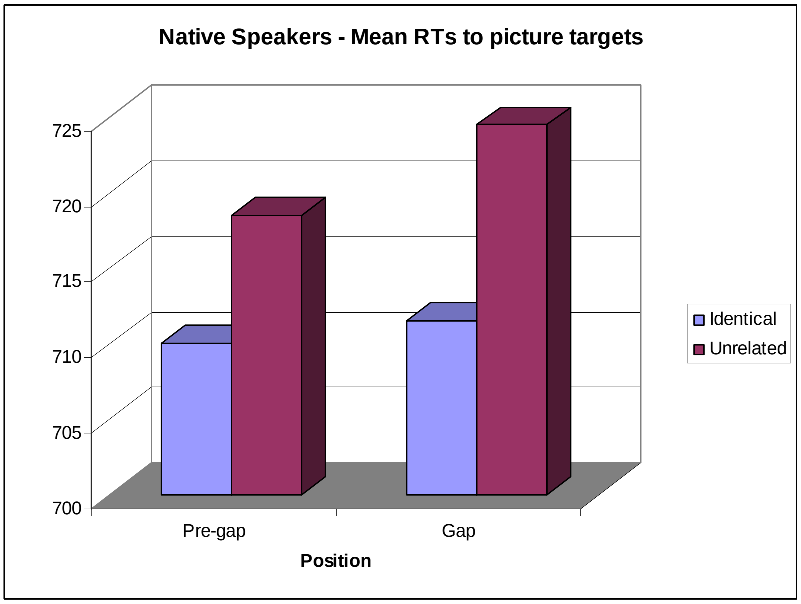


Figure 10. Native speakers' reaction times to probes adapted from Roberts et al. (2007)

Even though we have not tested this assertion statistically, native speakers processed the identical picture prime very similarly in both conditions, pre-gap and gap. This lack of a difference is not predicted by Felser & Roberts (2007) or Roberts et al. (2007) and goes clearly against their claim of a differential performance between native and non-native speakers in this respect. Hence, the results shown in Figure 10 could be evidence of a mere lexical priming effect instead of true antecedent reactivation at intermediate gaps.

What is more, Felser & Roberts (2007) report that 38 items in their stimuli were followed by an auditory comprehension question, but unfortunately, they do not include those questions in their article. This may be relevant, because, having to answer a comprehension question as part of the experimental treatment, both the native speakers and L2 learners may have engaged in the application of a strategy to answer the probe correctly. For example, they may have tried to remember all animate entities at the beginning of the sentences read, because they knew the comprehension question would be related to those entities. However, without knowing what the questions were like in Felser & Roberts (2007), this objection stands as tentative.

Finally, included in Felser & Roberts' (2007) own review of the literature is a study that shows that reactivation effects are not found only in purported intermediate gap sites. Balogugh et al. (1998) have found that antecedent reactivation effects have also been observed at the end of sentences, in the wrap-up region of target items. This may suggest that the effect is not triggered only by structural processes; semantic or plausibility evaluations may also cause the reactivation of entities previously found in a string of words. Thus, the data analyzed by Felser & Roberts (2007) in favor of one of the SSH's main assumptions does not seem to provide conclusive evidence as to the presence of intermediate structure in native speakers, and therefore cannot be evidence for its absence in L2 learners' performance. We turn our attention next to another study that has been claimed to show the lack of intermediate representational gaps in L2 comprehension.

In a study meant to test the same assumption of the SSH, Marinis et al. (2005) attempted a replication of Gibson & Warren's findings with intermediate gaps (IG) for native (n=24) and also non-native speakers of English from four different L1 backgrounds: Chinese (n=34), Japanese (n=26), German (n=24) and Greek (n=30). The phenomenon investigated by Gibson and Warren (2004) and Marinis and colleagues (2005) involved filler-gap relationships in long distance dependencies and a sample of the materials used is included in (40) and (41) to remind the reader of the structures involved in this type of stimuli.

(40) Extraction across a VP – intermediate structure
The nurse who / the doctor argued __IG / that / the rude patient / had angered
R1 R2 R3 R4 R5
__G2 / is refusing to work late.
R6

(41) Extraction across an NP - no intermediate structure
The nurse who / the doctor's argument / about / the rude patient / had angered ___
R1 R2 R3 R4 R5
/ is refusing to work late.
R6

In order for the reader to extract the correct representation from this sentence, which includes center-embedded clauses, they should activate the fronted argument, *the nurse*, at the IG after the verb *argued* and also after the predicate *angered*. According to Gibson & Warren (2004), the availability of this IG should make it easier for the reader to reactivate the antecedent *the nurse* at the second gap in (40). This ease of activation should be reflected in lower RTs for *angered* in (40) when compared to the RTs for the same lexical item in (41), a type that lacks the additional gap mediating between the filler and the subcategorizer. As mentioned before, in (41) there is no available IG and the distance between the antecedent, *the nurse*, and the gap where it should be reactivated is much longer.

Marinis et al. (2005) included two additional conditions without wh-movement, following Gibson & Warren (2004), in order to rule out a possible confound involving the distance between the subject and the verb preceding the main predicate where the effect of intermediate gaps was expected. However, Marinis et al. (2005) introduced changes to these additional stimuli, because they wanted to keep the same number of words in the four conditions up to the embedded verb. Thus, they added a further level of embedding to these non-extraction conditions, examples of which are listed in (42) and (43).

(42) No Extraction / local subject-verb integration (VP):
The nurse thought / the doctor argued / that / the rude patient / had angered / the
R1 R2 R3 R4 R5
staff at the hospital.
R6

(43) No Extraction / non-local subject-verb integration (NP):
The nurse thought / the doctor's argument / about / the rude patient / had angered /
R1 R2 R3 R4 R5
the staff at the hospital.
R6

These stimuli were presented in a region-by-region modality, instead of the word-by-word presentation used in Gibson & Warren (2004), and the regions chosen are separated by (/) in (42) and (43). Following Gibson & Warren's (2004) predictions as regards the regions where effects for the different conditions would be observed, Marinis et al. hypothesized that if non-native speakers were able to pose intermediate gaps, they would present elevated RTs on region 5 in the extraction condition involving a genitive NP (in (41) above). Conversely, the same region in the VP extraction condition should render lower RTs due to the previous reactivation of the filler in the intermediate gap site. Another prediction stemming from Gibson & Warren's work is related to region 3, where Marinis and colleagues anticipated longer RTs for this segment in the VP extraction condition, as opposed to the same region in the VP non-extraction stimuli, since in the former an intermediate gap should be posed before *that*, but no such gap should be present in the latter.

Even though they do not report the statistical results of the test performed, Marinis and colleagues claim there were no significant differences as regards accuracy when answering comprehension questions after the target stimuli among the groups in this experiment: native speakers, 79.5%; Chinese learners, 79%; Japanese learners, 74.5%; German learners, 84.75% and Greek learners, 79.75%. The crucial finding needed to corroborate the assumptions of the SSH and Marinis et al.'s prediction involved a difference in raw RTs on region 5 between the VP and NP extraction conditions. The charts in Figures 11 and 12 show the reading times obtained for regions 3 and 5 for all groups.

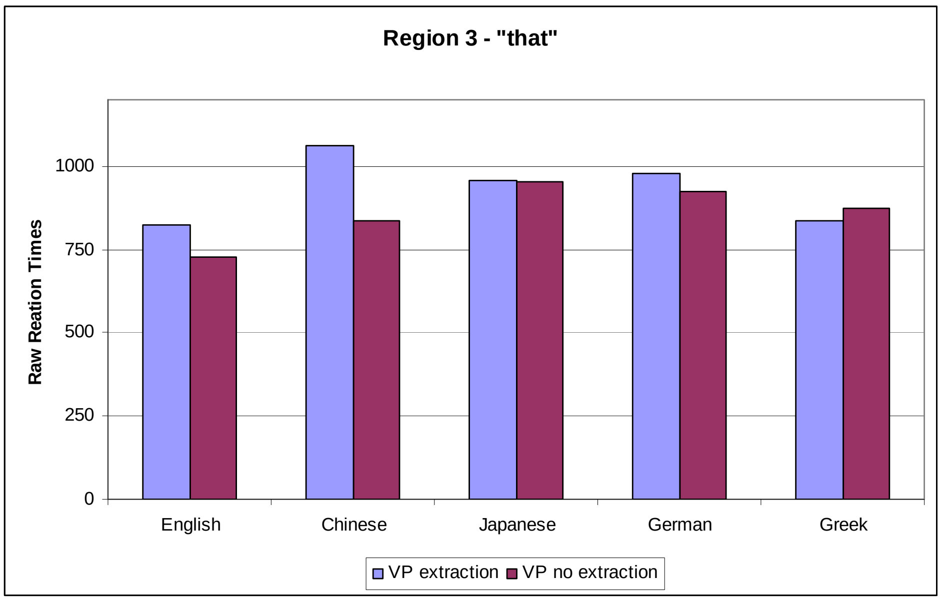


Figure 11. Raw reading times for region 3 adapted from Marinis et al. (2005)

The difference between VP extraction and VP no-extraction was significant and in the expected direction for native speakers on region 3, with longer RTs for the complementizer that follows the intermediate gap position in VP extractions, replicating Gibson & Warren's findings (although, as stated before, Marinis et al., like Gibson & Warren, stress the fact that due to lexical differences between the regions, this result should be taken with caution). On the learners' performance, Marinis et al. report that there was no significant effect of extraction for the non-native groups, even though the raw RTs seem to show a trend in the right direction for all learner groups except the Greek L2 speakers. The authors conclude that the lack of a significant effect on this region shows that learners do not seem to pose the intermediate gap that is apparent from the RTs of native speakers on this particular region. Figure 12 shows results for the other relevant segment, region 5, for all groups.

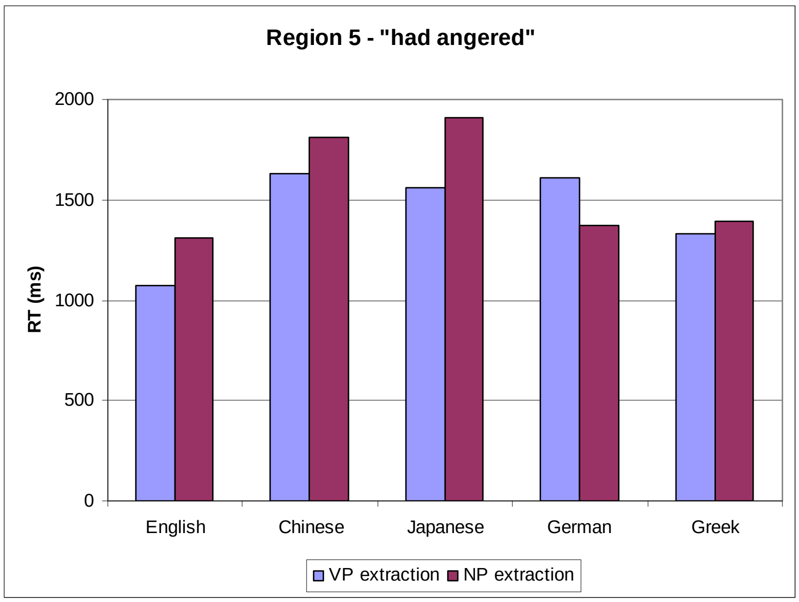


Figure 12. Raw reading times for region 5 adapted from Marinis et al. (2005)

The difference between the VP extraction (40) and NP extraction (41) conditions for the native speakers on region 5 was significant by participants and items. Further statistical tests showed that extraction conditions were read overall more slowly than non-extraction conditions, but there was no effect for extraction/no extraction on region 5, thus discarding the probable subject/verb distance confound problem. These findings, coupled with what was discussed for region 3, seem to show the predicted effect for native speakers, replicating Gibson & Warren's

(2004) results. Native speakers do seem to have the filler readily available in the VP extraction condition, which helps them later during filler integration processes, once they reach region 5.

In the non-native speaker groups, a different picture emerges. In spite of a similar tendency to the one found for native speakers in the raw RTs shown in Figure 12 for most learner groups, Marinis and colleagues report no significant interaction for extraction and phrase type for the non-native speakers, with one group, the German learners, showing the opposite tendency to the one predicted on region 5. There was, however, a significant effect for extraction/non-extraction conditions, with the former being read more slowly than the non-extraction conditions by all learner groups.

Marinis et al. (2005) conclude from these results that L2 learners were able to process these sentences in a manner that allowed them similar accuracy rates to those of native speakers, but with a parsing algorithm that does not generate the same kind of structural representation. They thus claim that L2 parsing is guided by a direct association mechanism that exploits thematic relationships between the words being integrated into a sentence.

There are a few aspects of Marinis et al.'s (2005) design and results that cast doubt on the strong claim of L2 structure-less parsing in long distance dependencies. First, even though the authors used a very similar design to Gibson & Warren's (2004), they presented their stimuli already divided into regions, instead of using a word-by-word presentation like Gibson and Warren. This artificial segmentation may have prompted participants to respond differently to the materials, since, for example, silent prosody has been found to influence structural assignments in L1 parsing (Fodor, 1998). Second, Gibson & Warren analyzed their results based on residual reading times (Ferreira & Clifton, 1986) due to the lexical differences between conditions as regards word and sentence length. In contrast, Marinis and colleagues decided to analyze raw reading times instead, leaving these preexisting differences unaccounted for. Third, Marinis et al. used 3 animate role names in their target stimuli, instead of two animate and one inanimate, as in Gibson & Warren's study. This choice can further complicate comprehension for non-native speakers, since they have to keep track of three names of the same kind of entity, a characteristic that has been found to distort even native speaker comprehension (Gordon et al., 2004). Fourth, it may be the case that the greater degree of variability in L2 learners' performance in these tasks coupled with a low sample size may help obscure pervasive patterns, which are easily identifiable in native speakers' more stable performance, due to lack of statistical power (T. Warren, p.c.). Finally, as mentioned before, the authors report different accuracy values for the participant groups and state that there were no significant differences between the groups. Nevertheless, details of the pertinent statistical test to show this absence of a significant difference are not reported.

Finally, having reanalyzed the raw reading time profiles presented in Marinis et al. (2005), there is a noticeable difference in reading times between the second region in the extraction conditions, which the authors report as significant. This increase in RTs over the genitive construction may have a role in the later differences found on region 5 for both native speakers and L2 learners. Figure 13 shows that Region 2, *the doctor's argument* (in (41)), took longer to read than *the doctor argued* __ in (40). It could be the case that the complexity of the genitive NP generates an effect that lingers in the NP conditions and makes the filler gap integration at region 5 more difficult. As Gibson & Warren (2004) themselves point out, according to the Distance Locality Theory (Gibson, 1998, p. 59), "the difficulty of reactivating previous elements in the structure depends on how far back in the input they have occurred, and what kind of elements have occurred in the interim, and how much they interfere with each other and the head to be connected."

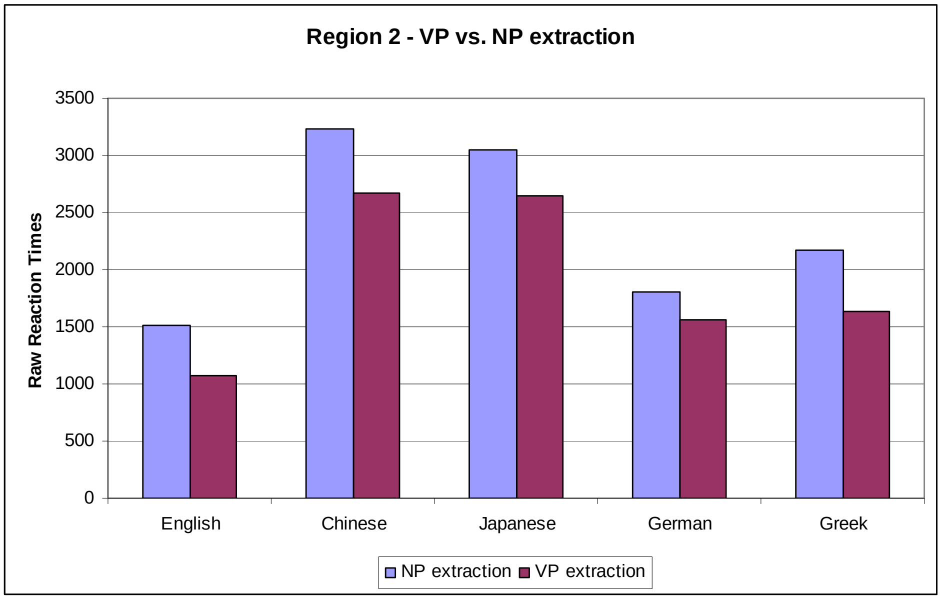


Figure 13. Raw reading times for region 2 adapted from Marinis et al. (2005)

In the section that follows, I describe the design and results of Experiment II, meant to test the SSH claim that L2 learners do not pose intermediate gaps when processing long distance dependencies and included similar materials to the ones employed in Marinis et al. (2005) and Gibson & Warren (2004).

3.2.1 Experiment II – INTERMEDIATE GAPS

In this second experiment, an attempt at replicating Gibson & Warren's, and Felser et al.'s findings with learners of English from Chinese and Spanish L1 backgrounds is made. This partial replication takes into account most of the criticism raised against Felser and colleagues' study in the preceding section. The relevant research questions to be answered with the data presented here are:

(a) Do L2 learners parse sentences that include intermediate gaps in a native-like way in their target language?
(b) Could gaps with NP-complement-biased verbs, i.e. those that trigger a garden path effect, pose the same difficulty that the genitive NP condition does (with no intermediate gaps) for both native and L2 learners?
(c) How does WM capacity relate to the L2 learners' ability to process this kind of stimuli?[1]

The tasks included in this experiment were completed by 20 Spanish-speaking and 21 Chinese-speaking learners of English, whose scores and RTs were compared to a native speaker control group of 20 individuals.

3.2.1.2 Participants

This experiment, as well as Experiment III, involved a comparison of two groups of non-native speakers of English: a Spanish L1 group and a group of Chinese native speakers. These two language groups were chosen in order to assess whether influence from the L1 affects the parsing of sentences that include wh-movement in English, since Spanish shares this feature of the grammar with English, whereas Chinese is considered to be a language with wh-in situ (Aoun & Li, 1993).

These L2 learners' proficiency in English was measured by an adapted version of the English Placement Test (EPT) included in the Michigan battery (Cordigan et al., 1978). This test consisted of 70 multiple choice items that assessed the participants' knowledge of English

[1] The answer to this last question will be developed in the chapter dealing with WM and L2 parsing performance.

grammar and vocabulary. This proficiency test was completed during a twenty-minute window and, due to the nature of the stimuli used in the SPR task, the proficiency threshold chosen for this test was 85% correct. These participants' accuracy when answering comprehension questions in the SPR was also taken into account as a criterion for inclusion in the final analyses, with a requirement of 80% correct for all of the items presented in the SPR task, since it has been found that even native speakers' accuracy decreases in online tasks when compared to paper-and-pencil tests (Juffs, 2005). These high proficiency thresholds decreased the number of participants in each group not only for the non-native speakers, but also for the native speaker control group. All groups were finally composed of 15 participants after triage based on the above criteria. Detailed biographical information for the non-native speakers appears in Appendix B.[2]

A one way ANOVA on the resulting proficiency scores indicated that there were no significant differences in the EPT proficiency between the learner groups in this test ($F(1)=2.498$, $p=.126$), with an average of 95% for Spanish speakers and 93.07% for the Chinese learners (see APPENDIX C). This lack of a difference between the groups was taken to mean that both presented high-intermediate to advanced proficiency in English and should be able to process the stimuli included in Experiments II and III. This was also confirmed with a comparison of the results in accuracy for the SPR task described in the following section.

3.2.1.3 Tasks

All participants in this experiment performed in a self-paced reading comprehension task in English that was meant to assess their parsing preferences when processing materials that

[2] The same two groups of learners participated in Experiment III in this thesis and thus this information will not be provided again in the design of the third experiment.

included intermediate gaps (IG) and garden path sentences that are described in the section that follows. The SPR task was administered with a personal computer running LINGER (Rohde, 2001), an experiment design software suite. The administration procedure was similar to the one employed in Experiment I, and is the same used in Experiment III. Sentences were presented one word at a time in a non-cumulative fashion and participants were able to advance through the sentence by pressing the space bar on the computer keyboard. Once they had finished reading the last word in a sentence they pressed the space bar once again and a comprehension question in the form of a true/false statement appeared on the screen. They were then asked to respond to this probe by pressing keys marked for YES or NO on the keyboard. The software recorded the time taken to read each word and the accuracy and response time for each of the probes answered.

The Spanish speakers and the native speakers also performed in a WM test, a reading span task in their respective L1s, to be described in the WM section below.

3.2.1.4 Materials

As previously mentioned, this experiment attempted to replicate findings in Gibson & Warren (2004) and Marinis et al. (2005) with second language learners of different L1 backgrounds in order to assess claims put forth by Clahsen & Felser (2006) as regards their SSH for L2 parsing. The sentences included in the SPR task involved the VP (44) and NP (46) extraction conditions taken from the studies to be replicated. A decision was made not to include the non-extraction conditions in Gibson & Warren (2004) and Marinis et al. (2005), since both studies clearly demonstrated that subject/predicate distance was not a confound responsible for the differences in RTs on the region where reactivation of the filler should occur. However, based on the ideas in Gibson's (1998, 2000) framework to explain parsing difficulty and on the rationale that Gibson, Warren, and Marinis et al. used to explain the lack of a filler-reactivation effect in (46), we decided to include a third type of stimuli involving garden path sentences, as in (45).

(44) The man who the lawyer <u>determined</u> that the illegal contract *had confused* will not go to prison.

(45) The man who the lawyer <u>understood</u> that the illegal contract *had confused* will not go to prison.

(46) The man who the <u>lawyer's determination</u> about the illegal contract *had confused* will not go to prison.

If the reactivation of the filler, *the man*, in (44) after the predicate *determined* is what triggers lower reading times when the second predicate, *had confused* – the site of filler gap integration – is encountered, a very similar effect should obtain for sentences like (45). In (45) there should be temporary filler reactivation after the predicate *understood*, since *the man* could be taken as the direct object of the verb *understand* and the parser may end the active search for a filler at this point (following the alleged universal parsing principles discussed in chapter 2). Obviously, the consequence of this temporary integration could be a garden path effect that is conscious (Juffs, 2004), and might trigger elevated reading times on the following words (or regions). At the same time, following Gibson & Warren and Marinis and colleagues, the availability of this intermediate gap and the conscious reactivation of the filler at this point should make access to the filler at a later point easier than in condition (46), in which there are no intermediate gaps where the reactivation should take place. However, if what causes longer RTs on the region *had confused* is not the availability of intermediate gaps between the filler and the reactivation site, but the difficulty in processing a region such as the one containing the genitive NP, *the lawyer's determination*, then a different outcome should be expected. Were the difficulty of the garden path effect in (45) similar to the difficulty of tracking participants once the genitive NP in (46) has been processed, we would expect similar RTs to the crucial region, *had confused*, for these two conditions. Thus these three sentence types were included in the stimuli in order to tease apart the interaction of availability of gaps and parsing difficulty in the SPR task.

The verbs before the purported intermediate gap in (44) and before the garden path effect site in (45) were selected based on their preference for an NP complement as normed in Kennison (1999). Verbs in (44) were *determine, argue, assume, admit* and *confess*, and their averaged NP complement preference was of 13.5 %. On the other hand, their counterparts in (45) were *hear, understand, trust, fear* and *find* with an average NP preference of 82.4%. This difference between conditions as regards NP complement preference increased the chances that the garden path effect expected in (45) actually obtained and could, thus, be compared to the difficulty involved in parsing a genitive NP in (46).

The length of all target items was controlled so that all sentences were composed of 17 words that would later be collapsed into 7 regions (following Gibson & Warren, 2004). There were 10 tokens per condition making 30 target stimuli sentences, which were interspersed with 120 fillers of a different nature unrelated to the aims of this experiment. All of these sentences were pseudo-randomized and counterbalanced in two lists, so that no two target items appeared one after the other. Half of the target items were followed by the type of comprehension probe that was discussed in the previous section. The large number of distractors used in this experiment was meant to divert participants' attention as to the nature of the target stimuli, since all participants read all of the target sentences in this experiment.

3.2.1.5 Results

The reading time results to be discussed here correspond solely to items whose comprehension probe was answered correctly. Table 8 shows accuracy percentages for the relevant conditions for the three participant groups.

Table 8. SPR percentage accuracy by condition by group

SPR ACCURACY	Intermediate Gaps (44) IG	Garden Paths (45) GP	Nominalizations (46) NOM
Native Speakers	85.33_{a}	85.33_{b}	69.33_{ab}
n = 15	(14.07)	(20.65)	(18.30)
Spanish Speakers	89.23_{cd}	75.38_{d}	66.15_{c}
n = 15	(15.52)	(20.25)	(25.01)
Chinese Speakers	69.33_{e}	84_{ef}	69.33_{f}
n = 15	(14.86)	(18.82)	(21.20)

d = approaches significance at p = .069; f = approaches significance at p = .068

A multivariate ANOVA procedure on these accuracy scores showed that there was a significant effect for group (F=2.487, Hdf=6, p=.03) and an interaction between group and stimuli type (F=7.55, Hdf=3, p<.0005). One way ANOVAs on these accuracy rates further qualified this overall result, demonstrating that there were no significant differences among the groups when comprehending sentences in the garden path (F(2)=1.002, p=.376) and the nominalization conditions (F(2)=.099, p=.906), (45) and (46) in Table 8. However, the ANOVA performed on the scores in the intermediate gaps condition (44) showed a significant difference (F(2)=7.316, p=.002). Tukey posthoc tests provided evidence to confirm that this significant difference stemmed from lower accuracy rates in the IG condition for the Chinese group that made it different from both the native and the Spanish groups (see APPENDIX D for statistical summary). Furthermore, the accuracy rates of native speakers were not different from the Spanish speakers' percentages in this condition. In Table 8, means that share subscripts were significantly different ($p \leq .05$) from each other in paired samples t-tests conducted on the accuracy rates for these SPR conditions.

As evident in Table 8, all groups experienced significant difficulty answering the probe for the nominalization condition, which involved processing a genitive NP in English. Chinese speakers also had similar difficulty when parsing the intermediate gap condition, with an accuracy rate that did not differ from that presented for the nominalizations in the same group.

Following Kazanina et al.'s (2007) trimming procedure, individual participants' mean RTs for all stimuli were calculated and any RT that exceeded 2.5 standard deviations from this mean was replaced with the threshold value (mean + 2.5 SD). There were then no missing values in this dataset.

As mentioned before, the presentation of stimuli for the SPR followed a word-by-word non-cumulative modality. However, since both Gibson & Warren (2004) and Marinis et al. (2005) analyzed their results taking into account regions, the word-by-word data obtained was later collapsed into 7 regions as shown in Table 9 (and repeated under each parsing profile for better understanding)[3].

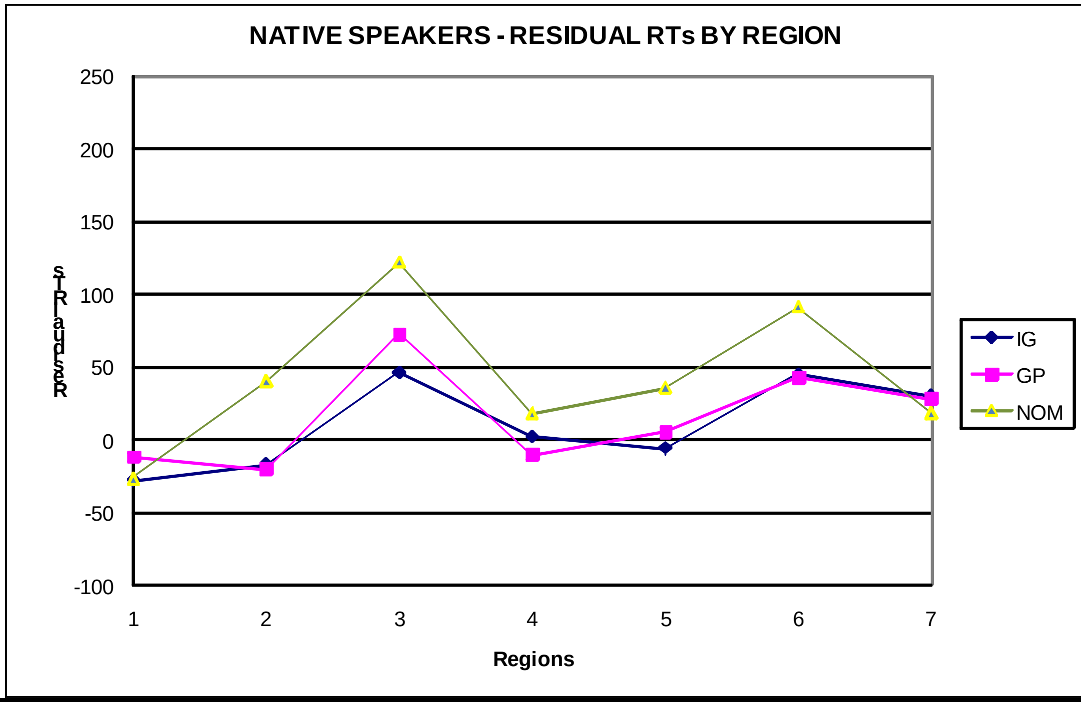
Table 9. Segmentation into regions for analyses following Gibson & Warren (2004)

Regions	1	2	3	4	5	6	7
IG	The man who	the lawyer determined	that	the illegal contract	had confused	will not	go to prison.
GP	The man who	the lawyer understood	that	the illegal contract	had confused	will not	go to prison.
NOM	The man who	the lawyer's determination	about	the illegal contract	had confused	will not	go to prison.

Another detail about the parsing profiles to be described below is that all of the data used for analyses and correlation corresponded to residual reading times instead of raw (actual) reading times collected by the experimental software. This transformation is meant to normalize

[3] Marinis et al. (2005) presented their stimuli directly segmented into similar regions.

results based on the reading rate of each participant and also on the length of word and utterances employed in the stimuli (Ferreira & Clifton, 1986). The parsing profiles for each group will be described and discussed separately, before reaching the overall discussion at the end of this section.



Regions	1	2	3	4	5	6	7
IG	The man who	the lawyer determined	that	the illegal contract	had confused	will not	go to prison.
GP	The man who	the lawyer understood	that	the illegal contract	had confused	will not	go to prison.
NOM	The man who	the lawyer's determination	about	the illegal contract	had confused	will not	go to prison.

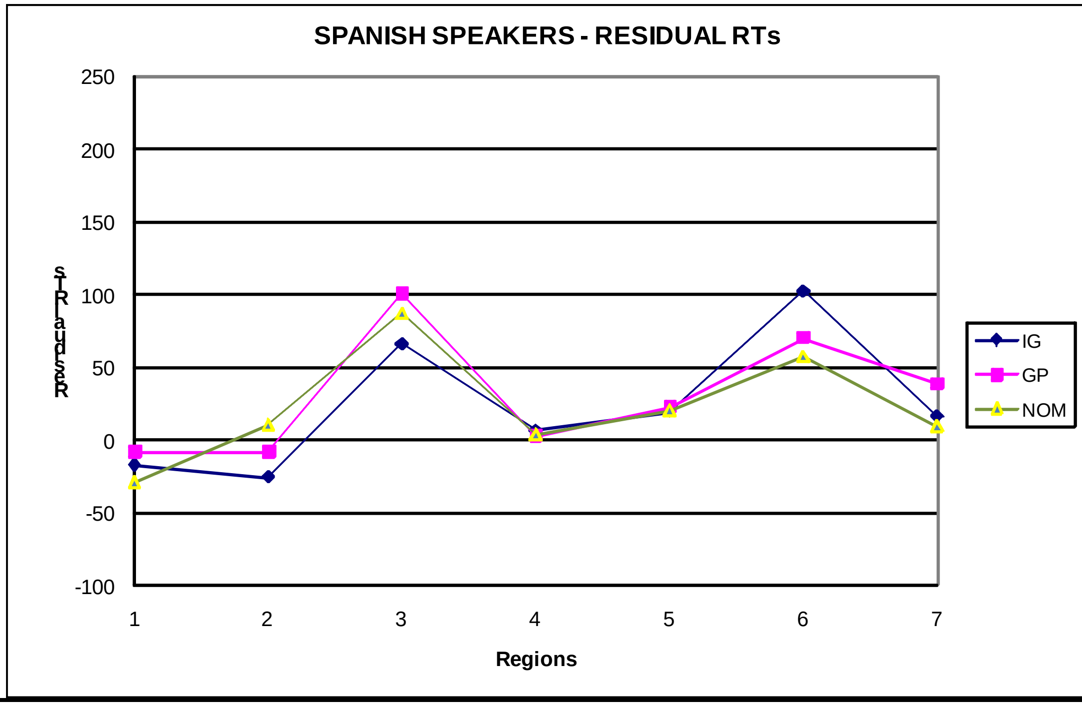
Figure 14. Native speakers - Residual RTs by region in Experiment II

The relevant regions for the purposes of our analyses are 2, 3, 5 and 6 in Figure 14. Region 2 is pertinent because we are trying to determine whether reading a genitive NP generates processing difficulty that is later associated with longer RTs on regions 5 and 6, where gap-filler integration should occur and disambiguation of the garden path effect should also take place. Region 3 is also important because it is the region after reactivation of the filler in conditions (44) and (45). The reading times on region 5, and possibly 6, should reflect whether the filler-integration process has been facilitated or not by the presence of a gap (stipulated or not) between the preceding regions 2 and 3.

Figure 14 shows the performance of the native speaker group for the three conditions. Parsing performance appears similar throughout the parsing profiles for the IG and the GP conditions, with the exception of region 3, where native speakers took longer to read the complementizer *that* in the GP condition. On the other hand, it is evident that the nominalization condition (NOM) poses a greater challenge for native speakers starting on region 2, the actual locus of the genitive NP. This effect appears even more pronounced on region 3, containing the lexical item *about*, and the reading times difference for NOM continues all along the remaining regions up until the last one (region 7) where all conditions coincide. A repeated measures ANOVA with condition and region as factors showed a significant effect for condition ($F(2,14)=10.258$, $p<.0005$) and a significant effect for region ($F(3,14)=9.952$, $p<.0005$). No interaction between condition and region for the native speaker group was found ($F(6,14)=.361$, $p=.902$). The Bonferroni procedure showed that the NOM condition was significantly different from the other two. Additionally, paired sample t-tests were conducted on the 4 relevant regions between conditions. These tests showed that region 2, the genitive NP, in the NOM condition was significantly different from the same region of the other two conditions. At the same time, there was a significant difference between the IG and the NOM conditions on region 3, where the complementizer *that* and the preposition *about* appear respectively. Finally, the last two regions presented significant differences only between the IG and the NOM conditions for 5 and between the GP and the NOM conditions for 6 (please see APPENDIX F for statistical summary of anovas performed for this experiment).

The most interesting result from the parsing profiles of the native speakers in this study is the difficulty found to parse region 2, the genitive NP, in the NOM condition. This issue of the difficulty in processing a genitive construction was not taken up by Gibson & Warren (2004) or Marinis et al. (2005) in their discussions of results, even though close examination of the RTs in their experiments shows that this region was also problematic for their participants (both native and non-native speakers). It is particularly interesting to notice that the level of difficulty (evidenced by RT differences) involved in the reanalysis of a garden path sentence in the GP condition was significantly lower than that found for the parsing of the NOM condition. Moreover, the lack of a significant difference between the IG and the GP conditions for Region 2 (and throughout these parsing profiles) may indicate that the effect found for the NOM condition in later regions, like 5 and 6, is a lingering effect of the difficulty in tracking discourse entities once the genitive NP has been processed. As described before, native speakers' problems with the NOM condition were not limited only to RTs, but were also evident in their comprehension question accuracy for this condition, which reached only 69.33% and was significantly different from the scores obtained for the IG and GP conditions (which did not differ from one another).

Figure 15 presents the parsing profiles for the Spanish-speaking learners of English tested in this experiment. The reader should be reminded that these results are based on the performance of only 15 participants, who scored at least 85% accuracy in the proficiency test and 80% overall accuracy in the comprehension probe of the SPR task. The RTs shown here present similar spikes to the ones found for native speakers on regions 3 and 6 for all conditions. However, the most difficult condition on region 3, the complementizer, appears to be GP (instead of NOM) and it is the IG condition (again instead of the genitive NP condition) on region 5 that takes the longest to read for these learners.



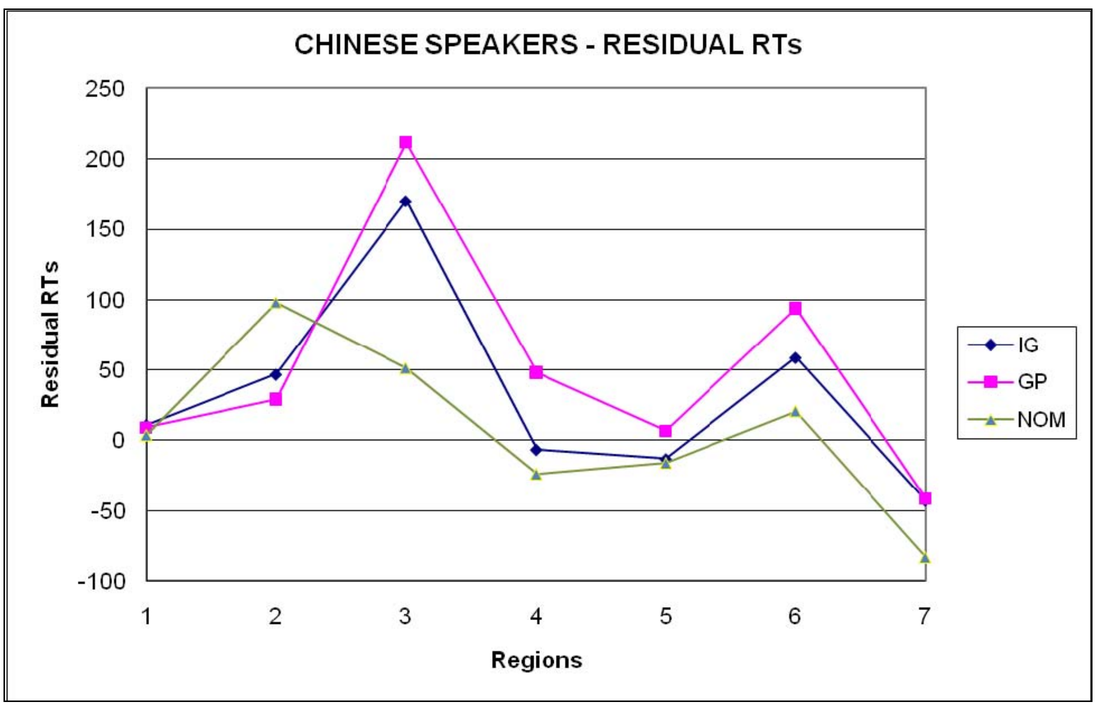
Regions	1	2	3	4	5	6	7
IG	The man who	the lawyer determined	that	the illegal contract	had confused	will not	go to prison.
GP	The man who	the lawyer understood	that	the illegal contract	had confused	will not	go to prison.
NOM	The man who	the lawyer's determination	about	the illegal contract	had confused	will not	go to prison.

Figure 15. Spanish Speakers - Residual RTs by Region in Experiment II

A repeated measures ANOVA with condition and region as factors showed no significant effect for condition ($F(2,14)=.807$, $p=.458$), a significant effect for region ($F(3,14)=10.698$, $p<.0005$) and no interaction between region and condition ($F(6,14)=1.005$, $p=.429$). Paired sample t-tests were conducted on the residual RTs for regions 2, 3, 5 and 6 of these profiles to assess differences between conditions. The only significant difference was found between conditions IG and NOM on region 2, the genitive NP.

As mentioned before, the parsing profiles of the Spanish group are similar to those of the native speakers in that they present similar increases in RTs on regions 3 and 6. However, the statistical tests employed showed no difference between conditions on these regions with elevated RTs, suggesting that these learners encounter the same kind of difficulty when reactivating a filler in the IG and the GP conditions (evident on region 3), and that, contrary to what was found for native speakers, the difficulty after having parsed a genitive NP in NOM is also similar to the other two conditions. Nevertheless, data from region 2 reinforces the idea that parsing a genitive NP is a difficult task, since these learners also showed greater RTs on this region compared to the IG condition, a difference that proved to be the only significant result here. This effect cannot be attributed to lexical differences or a different reading rate for the learners, since the regions in the three conditions included the same number of words and the statistical analyses were performed on residual RTs, which normalize the data taking into account these variables.

The Chinese-speaking learners' performance on this task differed significantly from the profiles obtained from both native and Spanish-speaking participants.



Regions	1	2	3	4	5	6	7
IG	The man who	the lawyer determined	that	the illegal contract	had confused	will not	go to prison.
GP	The man who	the lawyer understood	that	the illegal contract	had confused	will not	go to prison.
NOM	The man who	the lawyer's determination	about	the illegal contract	had confused	will not	go to prison.

Figure 16. Chinese Speakers - Residual RTs by region in Experiment II

Figure 16 shows the parsing profiles for the Chinese group, where the first evident difference involves the NOM condition, which presented the same increase in RTs on region 2, but differed on region 3, the locus of *about* and *that*, with a significant drop in RTs. The remaining regions after 3 present a similar profile to that of the Spanish speakers and the native control group. A repeated measures ANOVA with condition and region as factors indicated a lack of an effect for condition ($F(2,14)=2.594$, $p=.093$), a significant effect for region ($F(3,14)=8.693$, $p<.0005$) and an interaction between condition and region ($F(6,14)=4.034$, $p=.001$). Further paired sample t-tests pointed to significant differences between conditions GP and NOM on region 2. On region 3, both the IG and the GP conditions proved different from the genitive NP condition. Finally there was another significant difference between the NOM condition and the GP condition on region 6.

The performance of these Chinese learners replicates the finding in the two previous groups as regards a significant increase in RTs on the second region of the stimuli, where the genitive NP is located. Again this points to the fact that parsing a genitive construction appears to be difficult for both native speakers and learners alike. Contrary to what obtained for both the native speakers and the Spanish speakers though, when reaching region 3, the Chinese comprehenders show a decrease in RTs for NOM that was significantly different from the spikes in their RTs for IG and GP. What is also interesting about this region for the learners is that, even though the differences were not statistically significant, Spanish- and Chinese-speaking participants presented the longest RTs for the GP condition, following a trend that identifies the GP reanalysis as more difficult than the IG reactivation also in the native speaker group (in spite of the longer RTs for the NOM condition in this latter group). It should also be pointed out that the accuracy rates for conditions IG and NOM in the Chinese were low, with 69.33% correct for both stimuli types. What this accuracy finding reveals is that overall, based on the percentage correct for the three groups, the stimuli that were parsed the most accurately were the sentences included in the GP condition with 85.33%, 75.38% and 84% respectively for the native, Spanish and Chinese speaking groups. Thus, this difference in accuracy may constitute further evidence in favor of the claim that not all sentences that require the reactivation of a filler in a site preceding the actual gap, like the GP sentences included here, facilitate its later activation from memory (Gibson & Warren, 2004), making it easier to build an accurate representation.

3.2.1.6 Discussion

We started this section by asking whether L2 learners parsed sentences with purported intermediate gaps in a way similar to the native speakers in previous research (Gibson & Warren, 2004; Marinis et al., 2005). An examination of the residual RTs for these sentences in each group shows that the peaks and valleys for both learners and native speakers are similar. All participants present elevated RTs on region 3, at the complementizer or the preposition, and also on region 6, after the gap site for all sentences. A further prediction about these parsing profiles stipulated that if the IG in sentences such as (44) motivated a reduction in RTs at the licenser site, the reactivation of a filler caused by a GP effect in (45) should trigger a similar decrease in RTs, when compared to a gapless condition with a nominalization (46). This seemed to obtain only for the native speakers who took part in this experiment.

However, a more detailed look at the results in Marinis et al. (2005) and also Gibson and Warren (2004), together with the data obtained for the three groups tested here seem to suggest a different explanation. In the gapless condition (46), the parsing of the genitive NP posed the most difficulty in comprehension when compared to the same region in the other conditions. This pervasive finding across groups and experiments may suggest that it is not the presence/absence of an intermediate gap that generates longer RTs when the filler must be associated with its corresponding gap, but instead the lingering difficulty of processing a genitive construction that makes the tracking of who did what to whom more challenging not only for L2 learners but also for native speakers of English in this case. These data on the processing of a genitive NP may, in fact, stand as a counterexample to the claims of a fundamental difference advocated by C&F. This may be particularly relevant for Marinis et al.'s study with non-native speakers, since the stimuli included three animate entities in one sentence, which represents exactly the kind of entity-tracking in discourse that has proven problematic even for native speakers (Gordon et al., 2004). It may be the case then that L2 learners are behaving like native speakers in that their parsers find the same kind of difficulty when processing a genitive construction that forces a non-canonical assignment of roles to discourse entities.

In the following section, the focus shifts towards another controversial claim of the SSH as regards the use of configurational constraints in L2 parsing.

4.0 PARSING CATAPHORIC PRONOUNS & BINDING CONSTRAINTS

Cataphoric pronouns engage in coreference relations with an antecedent that follows them (47), as opposed to forward anaphora, in which the antecedent precedes the pronoun it corefers with (48).

(47) After he_i found the club, the $robber_i$ hit the window with might.

(48) $Mike_i$ didn't let anybody tell him_i what to do.

These referential dependencies are similar to the wh-dependencies which were discussed in the previous chapter. Similar to wh-dependencies, cataphoric coreference may involve many clauses in between the pronoun and its antecedent, and these phenomena also share the fact that the dependent element comes before its antecedent. Additionally, the relationship between pronouns and antecedents are governed by structural principles. In particular, Binding Principles (Chomsky, 1981) regulate the relationship between pronouns and their coreferents. These principles state, for example, that a pronoun cannot c-command its antecedent, and are exemplified with sentences (49) to (52). Principle C says that an R expression must be free everywhere, i.e. not c-commanded or co-referenced with anything else, including anaphors and pronouns.

(49) *He$_i$ likes John$_i$.

(50) *He$_i$ said that John$_i$ likes wine.

(51) *He$_i$ drank beer while John$_i$ watched a soccer game.

(52) His$_i$ friends drank beer while John$_i$ watched a soccer game.

Given the similarities between wh-dependencies and referential dependencies, it may be the case that the parser computes the structure for these relationships with a similar active search mechanism (Crain & Fodor, 1985; Frazier & Clifton, 1989; Stowe, 1986) that attempts to link an unanchored pronoun with the antecedent that follows it, guided by structural principles (Stowe 1986; Traxler & Pickering, 1996) like Binding Principle C (Kazanina et al., 2007). Recent results from L1 sentence processing research on the parsing of cataphoric pronouns seem to confirm this claim, thus providing a possible testing ground for the SSH tenet of a lack of structural input to inform the L2 sentence parsing process.

Kazanina et al. (2007) tested English native speakers' knowledge of binding principles using the self-paced reading paradigm. Based on previous findings from van Gompel & Liversedge (2003), Kazanina and colleagues (2007) hypothesized that when presented with an initial pronoun in subject position, the parser engages in a lookup strategy for the co-referent which corresponds to that pronoun. The aim of Kazanina et al.'s study was to further explore the characteristics of the active search mechanism that Van Gompel & Liversedge (2003) had previously identified for cataphoric relationships in English. In their eye-tracking study, Van Gompel & Liversedge looked at adverbial subordinate clauses that included a cataphoric pronoun. Examples are provided in (53) to (55).

(53) When he was at the party, the boy cruelly teased the girl during the games.

(54) When he was at the party, the girl cruelly teased the boy during the games.

(55) When I was at the party, the boy cruelly teased the girl during the games.

Van Gompel & Liversedge attempted to determine the type of information first used by the parser when looking for a pronominal antecedent. First-pass reading times on the adverb *cruelly* were longer for (54) where the pronoun and the noun can structurally co-refer (like in (53)), but gender features on these items differ (*he*; *the girl*), prompting a reanalysis of the dependency formed between the two. Furthermore, Van Gompel & Liversedge included an additional condition, (55), in order to test whether the longer RTs on the region after the critical noun were due to the introduction of a new discourse entity (Gordon et al., 2004), and not triggered by a gender mismatch. Results showed no difference on this region between (53) and (55), suggesting the increase in reading times was dependent upon the gender mismatch in condition (54). The researchers concluded that this finding could be taken as evidence of access to syntactic information about the restrictions on the linking of pronouns and antecedents before any kind of bottom up (semantic) information, such as gender, could be used by the parser.

Nevertheless, Kazanina et al. (2007) put forth a different version of this active search mechanism, which does not prioritize syntactic information when parsing. Instead, the predictive nature of this search may be such that it foresees a coreferential relationship once a *while*-clause has been identified. The parser processes the subordinator *while* and predicts that the subject of the *while*-clause will be followed by a subject for the main clause, and that these two may be related. In this way, there is no need to posit a preferential role for syntactic information as proposed by Van Gompel & Liversedge (Kazanina et al., 2007). However, the relevant characteristic of this search mechanism in native speakers for our purposes is that the parser only looks for possible coreferents for the cataphoric pronoun in positions that are allowed by Principle C (Kazanina et al., 2007; Van Gompel & Liversedge, 2003). These positions are those that are not c-commanded by the pronoun.

In order to test knowledge of this strictly structural/configurational restriction on the coreference properties of noun phrases, Kazanina and colleagues used sentences that included cataphoric relationships between a pronoun and a following antecedent, as shown in (56)-(59). They avoided the initial subordinate *while*-clause used in Van Gompel and Liversedge (2003) and they placed the cataphoric pronoun as the first lexical item of the main clause, followed instead by a subordinate *while*-clause.

(56) He$_i$ chatted amiably with some fans while the talented, young quarterback signed autographs for the kids, but Steve$_i$ wished the children's charity event would end soon so he could go home. – PRINCIPLE C / MATCH

(57) She$_i$ chatted amiably with some fans while the talented, young quarterback signed autographs for the kids, but Carol$_i$ wished the children's charity event would end soon so she could go home. – PRINCIPLE C / MISMATCH

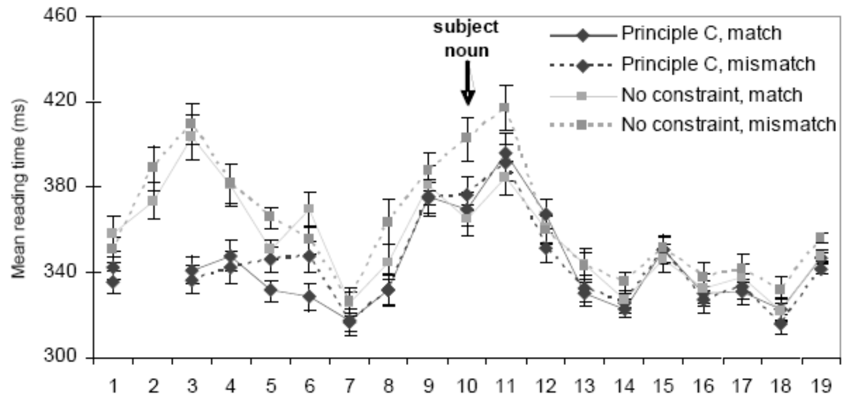
(58) His$_i$ managers chatted amiably with some fans while the talented, young quarterback$_i$ signed autographs for the kids, but Carol wished the children's charity event would end soon so she could go home. – NO CONSTRAINT / MATCH

(59) Her$_i$ managers chatted amiably with some fans while the talented, young quarterback signed autographs for the kids, but Carol$_i$ wished the children's charity event would end soon so she could go home. – NO CONSTRAINT / MISMATCH

All of the items began with a main clause followed by a subordinate adverbial clause headed by a subordinating conjunction such as *while*. The critical region in these sentences was the embedded adverbial clause, *while the talented, young quarterback signed autographs for the kids*, since the noun that stands as the subject of this clause could not be a candidate for coreference with the initial pronoun (in (56) and (57)), due to the restriction posed by Principle C. The further manipulation of gender match/mismatch between the cataphoric pronoun and the embedded noun phrase was meant to test whether the search for an antecedent in native speakers is actually ruled by the binding principle in question.

Consequently, Kazanina and colleagues' predictions were that, if native speakers follow Principle C in their search process, there should be no effect of gender match/mismatch in sentences (56) and (57). This means we should not expect differences in reading times on the embedded noun, because these sentences represent samples of domains where Principle C should apply and disallow coreference between the cataphoric pronoun and the embedded noun. Instead, the mismatch in gender in (59) should trigger longer reading times (contrary to what is expected for items like (58)) on the critical noun, *quarterback*, since Principle C does not pose a constraint on coreference here, allowing the possessive pronoun to corefer with the embedded noun. However, gender features clash and thus should be reflected in increased reading times in (59).

The predictions in Kazanina et al. (2007) did obtain, and raw reading times for each of the conditions are shown in Figure 17 (from Kazanina et al., 2007).



He (she)$_1$ / [Principle C] *or* His (her)$_1$ / managers$_2$ / [no-constraint] / chatted$_3$ / amiably$_4$ / with some fans$_5$ / while$_6$ / the$_7$ / talented,$_8$ / young$_9$ / quarterback$_{10}$ / signed$_{11}$ / autographs$_{12}$ / for$_{13}$ /the kids,$_{14}$ / but$_{15}$ / Steve (Carol)$_{16}$ / wished$_{17}$ / the$_{18}$ / children's charity event would end soon so he could go home$_{19}$.

Figure 17. Parsing profiles for Experiment 3 in Kazanina et al. (2007)

As shown in Figure 17, the critical region in the four conditions, the embedded subject noun, *quarterback*, showed increased reading times only for (59), the NO CONSTRAINT/MISMATCH condition. On the other hand, no significant difference between the other conditions was found, reinforcing the idea that native speakers engage in an active search for an antecedent for the cataphoric pronoun from the moment this item is found in the parse. However, this search is only active for structurally-defined positions that are allowed by the binding principle that determines the behavior of referring expressions, Principle C.

This kind of stimuli presents a feasible test of Clahsen & Felser's arguments as regards the type of information that guides parsing decisions in an L2, and may help refine the SSH by testing whether L2 speakers follow binding principles, which are strictly structural or configurational, are not affected by subcategorization preferences, require the checking of formal features (like gender), and most importantly, involve non-local relationships (Kazanina et al., 2007). The reader should be reminded at this point that C&F (2006) and Marinis et al. (2005) have dismissed previous results on the use of syntactic information in L2 (Juffs & Harrington, 1995; Juffs, 1998; Williams et al. 2001, 2006) based on the immediacy between subcategorizer and gap. Furthermore, these sentences do not seem to provide any kind of semantic/lexical cue as to whether the noun encountered in the input may be an antecedent for the cataphoric pronoun (except from gender features), which also stands as one of the main criticisms of previous research marshaled by C&F.

In what follows I describe the details and findings of Experiment III, which was meant to replicate Kazanina et al.'s findings with non-native speakers of English.

4.1.1 Experiment III

In order to further specify or narrow down the claims brought about by the Shallow Structure Hypothesis (Clahsen & Felser, 2006), this study seeks answers for the following research questions:

(a) Do L2 learners from different L1 backgrounds parse sentences with coreferential relationships governed by syntactic/structural principles in a native-like manner?

(b) Do characteristics from the L1 grammar, such as the gender system, influence the parsing of these sentences in an L2?

The design of the study involving these questions is presented in the following sections together with a description of results and corresponding discussion.

4.1.1.1 Participants

The same groups that took part in Experiment II performed in the tasks involved in this study. This constituted an appropriate choice, since the crucial information to be used in the resolution of coreference between the cataphoric pronoun and its antecedent are gender features.

This grammatical construct has proved problematic for Chinese learners of English, but should not be so for the Spanish-speaking group, since their L1 encodes gender grammatically in a more widespread manner than for English nouns and pronouns.

4.1.1.2 Tasks

As in Experiment II, all participants read sentences on a computer screen following a word-by-word non-cumulative presentation in an SPR task with similar stimuli to those used by Kazanina and colleagues (2007). The native speakers' and the Spanish-speaking learners' WM capacity was also tested with a Reading Span Task, to be discussed within the WM section in chapter 5.

4.1.1.3 Materials

The sentences presented in the SPR task followed the stimuli developed by Kazanina et al (2007) closely. However, some modifications were introduced to make the task more feasible for second language learners and also to reduce the great number of sentences that had to be read in this task (Warren, p.c.). The changes involved omitting adjectives that qualified the critical noun and replacing some lexical items with more frequent counterparts that were more likely considered to be part of the repertoire of L2 speakers, as the examples in Table 10 show.

Table 10. Target stimuli for Experiment III, adapted from Kazanina et al. (2007)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
(60)	He	quickly	opened	the	window	while	the	fireman	entered	the	building,	but	Jeffrey	was	afraid	of	the	black	smoke.
(61)	She	quickly	opened	the	window	while	the	fireman	entered	the	building,	but	Allison	was	afraid	of	the	black	smoke.
(62)	His	children	opened	the	window	while	the	fireman	entered	the	building,	but	Allison	couldn't	shout	or	wave	for	help.
(63)	Her	children	opened	the	window	while	the	fireman	entered	the	building,	but	Allison	couldn't	shout	or	wave	for	help.

Another difference between this experiment and Kazanina et al.'s is that all of the target items had a length of 19-regions/words, so that each region corresponded to a single word and no word string was collapsed into a single region. Following the original design, the stimuli were counterbalanced for the gender of the crucial noun, with lexical items that were either morphologically or stereotypically marked for gender in English. Another factor that was also counterbalanced in the stimuli was the gender of the starting pronouns in region 1 and the proper name used after the adverbial subordinate clause in region 13. The possibility for the cataphoric pronouns to find antecedents co-referents within the target stimuli sentences was always allowed by either coreference with the crucial noun or with the proper name used in all sentences.

These target sentences were accompanied by 110 distractors of a different nature and all target stimuli were pseudo-randomized and counterbalanced in two lists, so that no two target items appeared consecutively. Consequently, all participants saw all target items in this experiment. As in Experiment II, half of the target stimuli were followed by a comprehension probe in the form of true/false statement that the participants had to respond to after they finished reading the sentence.

The predictions for these stimuli are the same as those advanced by Kazanina et al. (2007) for native and non-native speakers. If these readers are sensitive to the constraints imposed on coreference by Binding Principle C, we should observe elevated RTs in (63) on region 9 or the subsequent regions, since coreference between the possessive pronoun *her* and the crucial noun, *fireman*, is structurally allowed, but should be avoided due to a gender mismatch effect between these lexical items. If, on the other hand, these comprehenders do not restrict coreference relations following structural principles, no difference should be noticeable between condition (63) and its counterpart in (61). Both learners and native speakers would consider the noun on region 9 as a possible antecedent for either pronoun if Principle C were not to apply for these target items.

4.1.1.4 Results

For the analyses to be developed here, the target conditions will be labeled in the following way: condition (60), A-PrinC-Ma; condition (61), B-PrinC-Mi; condition (62), C-NoConstr-Ma; and condition (63), D-NoConstr-Mi. Let us look first at the accuracy rates for the comprehension probe included after half of the target stimuli. Table 11 shows accuracy rates by condition and group, and provides information as to which scores are significantly different from each other with subscripts.

Table 11. SPR accuracy rates by condition and group in Experiment III

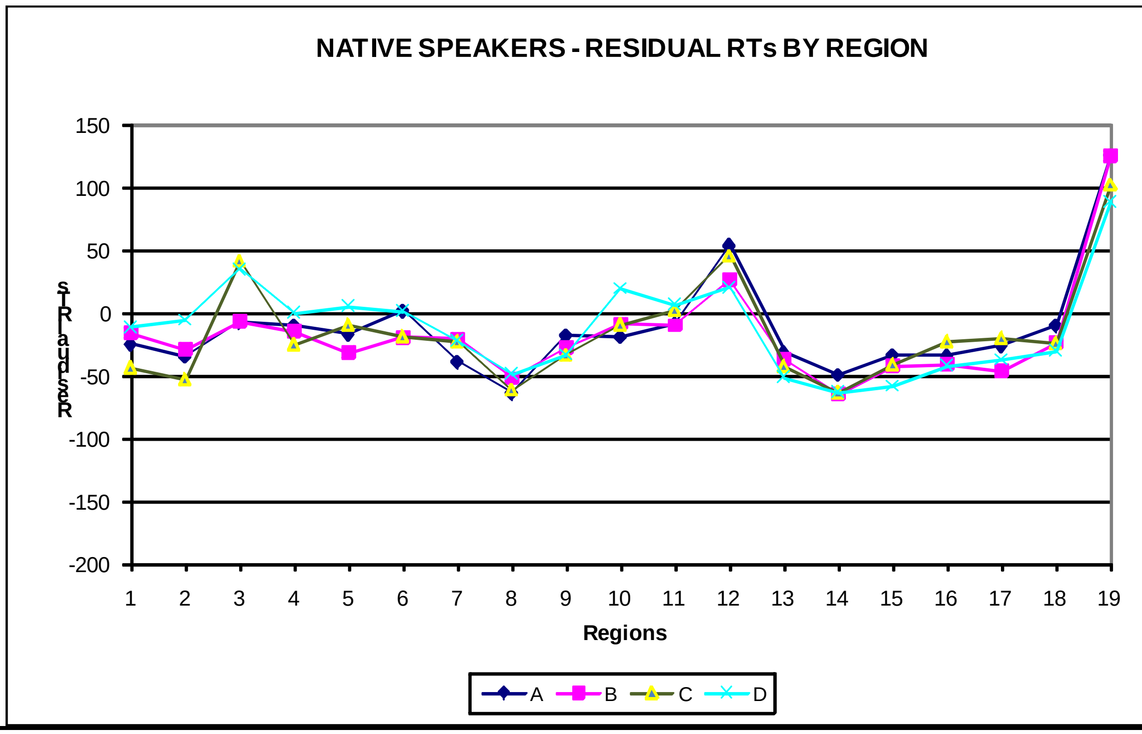
Condition	A-PrinC-Match	B-PrinC-MisM	C-NoConstr-Match	D-NoConstr-MisM
Native	$89.33\%_{a}$	$98.67\%_{b}$	86.67%	85.33%
	(10.32)	(5.15)	(9.76)	(20.65)
Spanish	$80\%_{a}$	$98.46\%_{c}$	90.77%	89.23%
	(8.16)	(5.54)	(10.37)	(15.52)
Chinese	81.33%	$89.33\%_{bc}$	93.33%	89.33%
	(11.87)	(14.86)	(9.76)	(14.86)

.

A multivariate ANOVA procedure was conducted on these accuracy percentages, which showed a significant effect for language group (F=2.55, Hdf=8, p=.016) and an interaction between group and stimuli condition (F=2.009, Hdf=4, p<.005). Further ANOVAs conducted for each stimuli condition showed that there were significant differences for condition A (F(2,44)=3.459, p=.041) and condition B (F(2,44)=4.349, p=.02), but no significant differences were found among the groups' accuracy for conditions C (F(2,44)=1.711, p=.194) and D (F(2,44)=.255, p=.776). Tukey post-hoc tests qualified these results by showing that in condition A, the difference between native and Spanish speakers approached significance (p=.056) and that the Chinese group was different from the other two groups in condition B. Finally, there were no significant differences in accuracy on conditions C and D.

These results suggest that the easiest condition to comprehend was B, followed by C. Furthermore, based on the greater SDs for condition D, this is the stimuli that seemed to pose the most difficulty in understanding for both natives and learners, even though the overall average is greater than for condition A.

The reading time data to be described below provides residual RTs as described for Experiment II, following the normalization procedure in Ferreira & Clifton (1986).



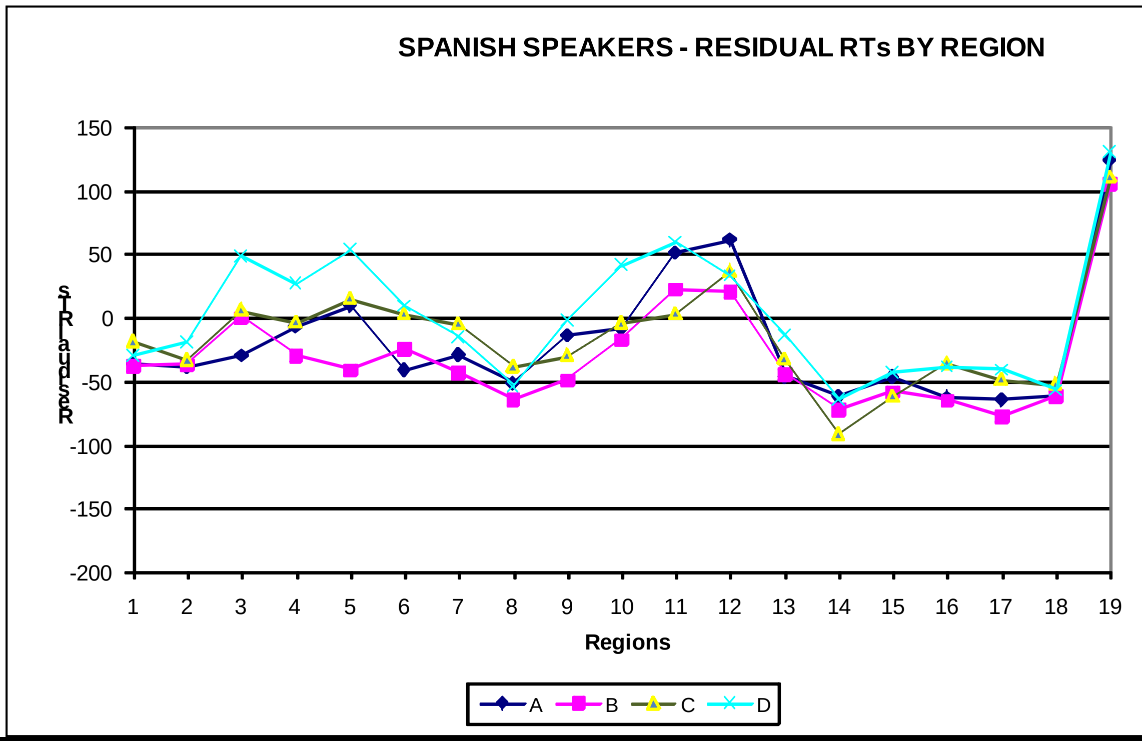
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A	He	quickly	opened	the	window	while	the	fireman	entered	the	building,	but	Jeffrey	was	afraid	of	the	black	smoke.
B	She	quickly	opened	the	window	while	the	fireman	entered	the	building,	but	Allison	was	afraid	of	the	black	smoke.
C	His	children	opened	the	window	while	the	fireman	entered	the	building,	but	Allison	couldn't	shout	or	wave	for	help.
D	Her	children	opened	the	window	while	the	fireman	entered	the	building,	but	Allison	couldn't	shout	or	wave	for	help.

Figure 18. Native Speakers - Residual RTs for Experiment III

Based on the findings in Kazanina et al. (2007), the regions input into the statistical analyses were 7, 8, 9, 10 and 11. The effects of constraints on coreference could first be noticed at the determiner of the crucial noun, region 7, on the noun itself, region 8, or as spill-over effects in the subsequent regions.

Figure 18 shows the parsing profiles for the native speaker group, whose residual RTs for each word were submitted to a 2 x 2 repeated measures ANOVA with constraint (PrinC or NoConstr) and gender congruency (Match or Mismatch) as factors for each of the relevant regions mentioned above. The ANOVA procedures rendered no significant effects or interactions for regions 7, 8, 9 and 11. However, there was a tendency approaching significance for the factor constraint on region 10 ($F(1,14)=3.339$, $p=.089$). Additional paired sampled t-tests between the conditions for this region showed that this tendency was motivated by significant differences between the means of conditions A and B (and a tendency for C) with respect to D, the mismatch condition where coreference between the cataphoric pronoun and the crucial noun is allowed. Even though the locus of this effect takes place after the mismatch region (8) in Kazanina and colleagues' findings, it could be the case that the native speakers in this study became aware of the mismatch only after having processed the whole noun phrase and some of the words that follow it. This remains a plausible explanation, since Kazanina et al. (2007) found similar spill-over effects for regions beyond the crucial NP in their stimuli. The reader should also be reminded of the fact that in the original study (Kazanina et al., 2007), the statistical analyses were performed on raw reading times as opposed to the residual RTs that were used for this experiment. In spite of the statistically significant effect shown for region 10, another pattern worth mentioning from the analysis of the native speakers' performance involves region 8, the critical word. Even though it did not prove to be statistically significant, region 8 shows a gender-related trend, where the two mismatch conditions (B and C) pattern together in spite of the configurational restrictions for coreference that should separate them (T. Warren, p.c.). This purported gender effect on the critical word may be an indication that the native speakers in this study may not have used the configurational information provided by principle C and instead guided their parsing decisions by resorting to lexical cues such as gender. I will return to this point later on, in the overall discussion of results for this thesis.

We turn now to the results for the Spanish-speaking group of learners, whose residual RTs were used in the same kind of statistical procedure employed for the native speakers.



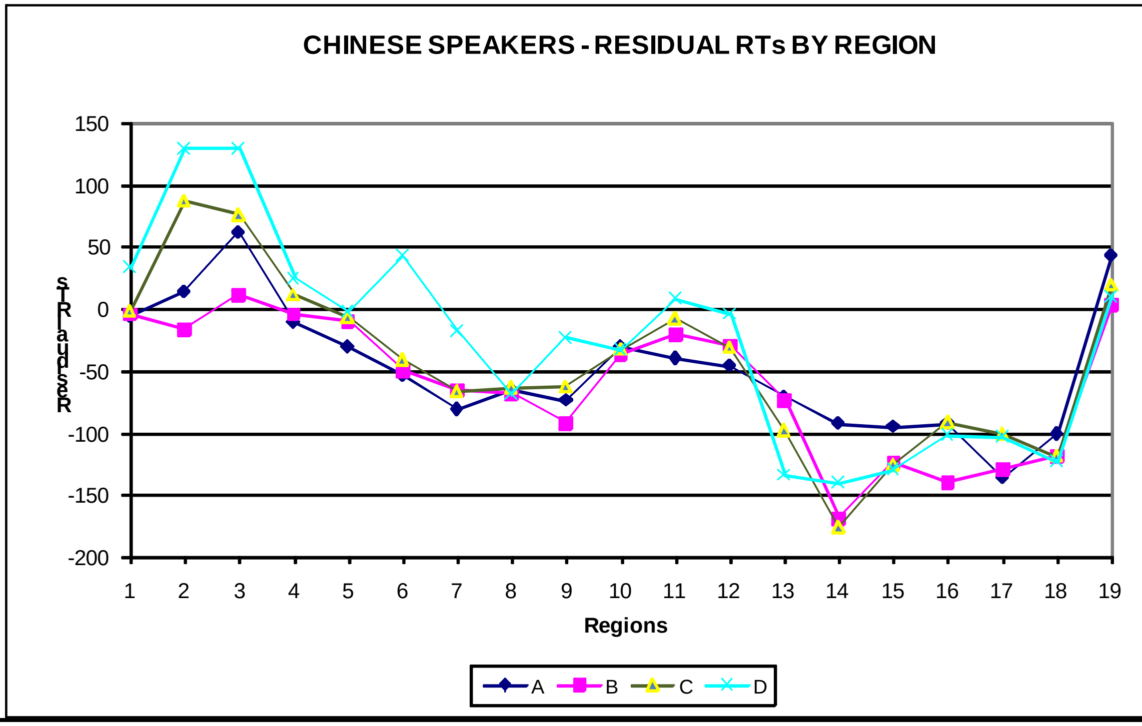
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A	He	quickly	opened	the	window	while	the	fireman	entered	the	building,	but	Jeffrey	was	afraid	of	the	black	smoke.
B	She	quickly	opened	the	window	while	the	fireman	entered	the	building,	but	Allison	was	afraid	of	the	black	smoke.
C	His	children	opened	the	window	while	the	fireman	entered	the	building,	but	Allison	couldn’t	shout	or	wave	for	help.
D	Her	children	opened	the	window	while	the	fireman	entered	the	building,	but	Allison	couldn’t	shout	or	wave	for	help.

Figure 19. Spanish Speakers - Residual RTs for Experiment III

Figure 19 provides residual RTs for each word in the 4 different conditions stemming from the Spanish speakers’ performance. Repeated measures ANOVAs for each region showed that there was a significant effect for constraint on region 7 ($F(1,14)=8.2$, $p=.014$). This effect was motivated, according to paired t-test results, by a significant difference between conditions B and C on this region, and a tendency for the difference between A and C. Regions 8 and 9 presented no significant differences. On the contrary, region 10 rendered a significant effect for constraint ($F(1,14)=5.584$, $p=.036$), a tendency that could prove significant with more participants for gender congruency ($F(1,14)=3.373$, $p=.091$), and an interaction between constraint and congruency ($F(1,14)=6.297$, $p=.027$). Paired t-tests helped determine that these effects stemmed from significant differences on this region between conditions A, B and C with condition D. Finally, region 11 showed an interaction between constraint and congruency ($F(1,12)=7.176$, $p=.02$), which was further qualified by paired samples t-tests, which showed a significant difference between conditions A and B, and trends for significance between conditions A and C, and C and D.

As pointed out in the description of results for the native speaker group, it seems that the mismatch effect expected on region 8, according to Kazanina's findings, takes place in later regions for the Spanish speakers tested in this experiment, too. Region 10 showed significant differences and an interaction between constraint and congruency in the predicted direction, if these participants' performance is influenced by Binding Principle C. Condition D, where the gender mismatch and the possibility for coreference should trigger longer RTs, was significantly different from the remaining conditions on region 10, showing a delay in the locus of the mismatch effect for this condition. Again this delay may be triggered by the fact that participants were exposed to the totality of target items in this experiment, which could have generated a familiarization effect that took longer to overcome when reading the gender mismatch condition in D. Another issue worth considering here is the possibility that these readers were overall slower than the subjects tested by Kazanina et al. (2007) and thus the onset of the mismatch effect took place later on during the parse (Philips, p.c.).

The ensuing description of results corresponds to the Chinese-speaking learner group, whose performance in this experiment was included in order to assess whether characteristics of the L1, in particular the Chinese gender system, may influence the parsing of these stimuli, which depend on gender information in order to be successfully comprehended.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A	He	quickly	opened	the	window	while	the	fireman	entered	the	building,	but	Jeffrey	was	afraid	of	the	black	smoke.
B	She	quickly	opened	the	window	while	the	fireman	entered	the	building,	but	Allison	was	afraid	of	the	black	smoke.
C	His	children	opened	the	window	while	the	fireman	entered	the	building,	but	Allison	couldn't	shout	or	wave	for	help.
D	Her	children	opened	the	window	while	the	fireman	entered	the	building,	but	Allison	couldn't	shout	or	wave	for	help.

Figure 20. Chinese Speakers - Residual RTs for Experiment III

Figure 20 depicts the parsing profiles for Chinese speakers in the four conditions included in Experiment III. Regions 7, 8, 9, 10 and 11 were submitted to 2 x 2 repeated measures ANOVAs. These procedures showed that there were no significant effects for regions 8, 9, 10, and 11. However, region 7 showed an effect for constraint that approached significance ($F(1,14)=4.123$, $p=.062$). This was the only tendency that approached significance and a paired samples t-test provided evidence that the trend towards significance was triggered by a difference between regions A and D ($p=.094$) on this region.

In spite of not reaching statistical significance, an increase in residual RTs for regions 9 and 11 for condition D in the Chinese parsing profiles may indicate a similar tendency to the one found to be significant for native and Spanish speakers. In this experiment it is always after the crucial region 8 that increases in RTs are found, even for the native speakers assessed here.

4.1.1.5 Discussion

Restating the research questions posed at the beginning of this section, our interest in this grammatical phenomenon from English was motivated by the need to determine whether non-native speakers replicate the performance of native speakers (in Kazanina et al., 2007) when it comes to the kind of information employed by the parser. It was predicted that if L2 learners' parsing of cataphoric pronouns was ruled by the same structural constraints that have proved to govern native speakers' parsing, they would not posit coreferential relationships at structural positions disallowed by these grammatical principles. This prediction goes clearly against one of the main tenets of the SSH, which claims that this kind of information is out of reach for L2 learners during parsing in their target language.

Our findings show that the parsing performance of all groups was similarly affected by the coreference restrictions between cataphoric pronouns and their antecedents, though the effect was delayed when compared to the RTs included in Kazanina et al. (2007) for native speakers of English. Elevated residual RTs were found for the condition that allowed coreference relations but presented a gender mismatch effect on regions following the onset of the effect, region 8, in Kazanina et al.'s work. This difference in the loci of the expected effects for all participant groups may have arisen due to a familiarization with the stimuli included in these conditions. Since all participants saw all target items in this experiment, the exposure to the full target stimuli set may have habituated the participants to the words included up to region 8 in these sentences. Additionally, as previously mentioned, it may be the case that other processes involved in parsing (word recognition and integration; Fender (2001)) are slower for these readers and this delay tends to slow down the application of configurational constraints and/or retrieval of gender information necessary to build up the appropriate representation (Philips, p.c.).

However, in condition D, where the gender mismatch should generate a reanalysis of the structure initially assigned to the string, the participants reacted as expected, but in the regions immediately following the locus of the effect for Kazanina et al.'s (2007) native participants. Another variable to take into account here that may have caused differences in the position of the mismatch effect has to do with the use of raw RTs in Kazanina et al. Even though Kazanina's outlier trimming procedure was used for Experiment III, it was decided residual RTs would be used to track reaction times along the parsing profiles of both native speakers and L2 learners. On the other hand, Kazanina et al. (2007) reported and analyzed parsing profiles using raw RTs, in spite of lexical and length differences between the conditions tested. However, this should not be taken as a crucial factor to explain differences between the two experiments, given that in the original study Kazanina and colleagues also found spill-over effects for the gender mismatch cues in the regions following the crucial one, region 8.

Summing up, in this study, the effects of coreference restrictions between cataphoric pronouns and antecedents located at structural positions licensed by Binding Principle C seemed to obtain for L2 learners as well as English native speakers. In spite of the fact that the effect was delayed to later regions in the target stimuli (and that it showed only a tendency for the Chinese speakers), this finding suggests that L2 learners do parse these sentences by resorting to configurational principles, such as binding constraints, contrary to what has been suggested by Clahsen & Felser (2006) in their SSH.

5.0 WORKING MEMORY AND L2 PARSING

The idea that an individual difference such as WM capacity may play a role in the acquisition of an L2 has interested SLA researchers for almost two decades (e.g. Harrington, 1987). The construct has been defined as the cognitive capacity to store and process items, these being segments, syllables, words, or digits (Baddeley, 1993; Just & Carpenter, 1992). The assumption behind most studies of this kind is that learners with higher WM capacity will be able to store and analyze more chunks of language and, hence, be able to acquire phonological and morphosyntactic competence more easily than those whose capacity is lower (Ellis, 2001). The model most commonly employed in this kind of research is Baddeley's WM model (2003), which separates the construct into 4 subcomponents, which are responsible for handling all kinds of sensory stimuli that may be stored, processed and possibly transferred to long-term memory, as shown in Figure 21. These components are the phonological loop, the visuo-spatial sketchpad, the central executive and the episodic buffer, the last element being a more recent addition to the model, and meant to account for individuals with very limited WM capacity but normal language skills (Baddeley, 2000).

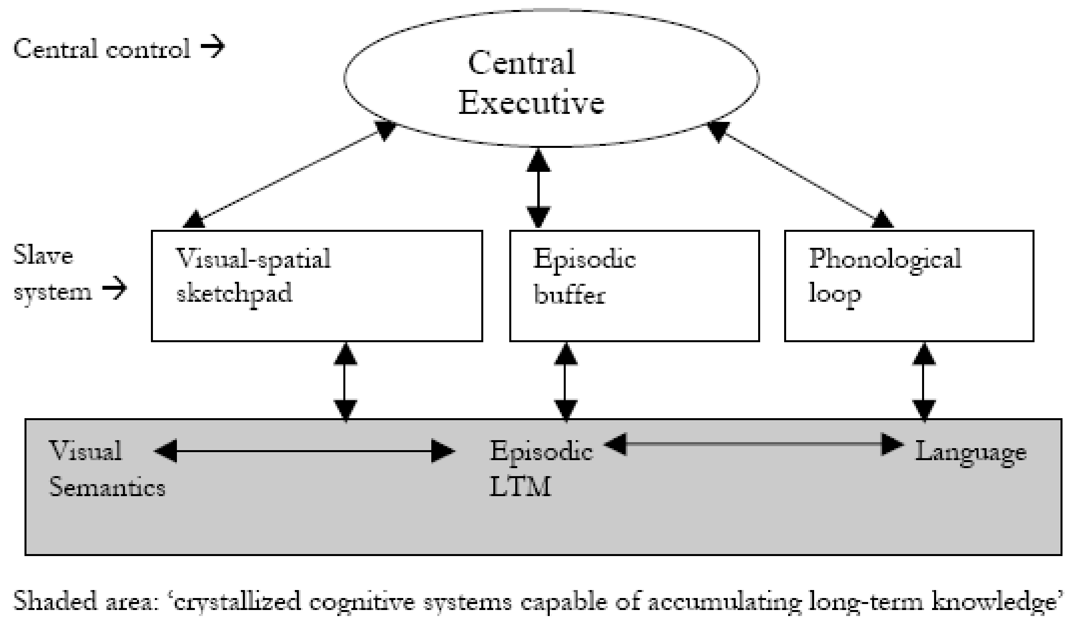


Figure 21. Baddeley's WM model (2003) from Juffs & Rodríguez (2006).

The phonological loop is in charge of storing auditory input and the visuo-spatial sketchpad performs a similar function with visual stimuli. The central executive behaves like a controller, assigning attentional resources to particular stimuli and is also believed to process items that are transferred to long-term memory. The new component in the system, the episodic buffer, is believed to store amalgams of experience into episodes. However, its role and interaction with the other components are at this point not completely clear. The relevant parts of the model for the study of WM influence on the acquisition of language are primarily the central executive and the phonological loop, with a possible secondary role for the episodic buffer in cases where the capacities of the phonological loop and the central executive may be damaged, but there has still been language learning taking place (Baddeley, 2000).

One of the major problems when trying to implement a WM test of either component under the assumptions of Baddeley's model is the variety of methodologies employed in the literature. Some researchers choose to implement tests that require participants to remember digits or words; some scholars use real or made-up words in order to assess the capacity of the phonological loop. The workings of the central executive are believed to require tests that try to measure not only storage of items, but also some kind of processing or computation demand at the same time. Thus, for this kind of test, researchers resort to multiple combinations of tasks within a single test (Ellis & Sinclair, 1996; Just & Carpenter, 1991; Swets et al., 2007), asking participants to solve equations and remember a word appearing after the equation or reading a sentence aloud and then remembering the last word of each sentence read (Daneman & Carpenter, 1980). The variety of methodologies and scoring methods for these tests make it a very difficult task to extract from the literature a coherent picture of previous findings as regards the influence of WM on the development of L2 proficiency. Of course, two camps can be easily identified: those scholars who have found evidence of a preponderant role for WM, in particular for the acquisition of vocabulary in a target language (e.g. Cheung, 1996; Ellis & Sinclair, 1996; Gathercole, 2006; Masoura & Gathercole, 1999, 2005; Service & Kohonen, 1995), and those studies in which WM was not related to either vocabulary knowledge, morphosyntactic competence or parsing performance in the L2 (e.g. Juffs, 2004; Juffs, 2005; Sagarra, 2000).

5.1 WM TESTS

The two WM tests included in this thesis are among the most widely used instruments for the assessment of the phonological loop (involving the storage of segments) and the central executive (incorporating measures of storage and processing). In this thesis, WM capacity was assessed only in the native speaker and Spanish-speaking groups included, since the RST used to measure the central executive function of WM was administered in the L1 of each group and the differences in script and the unavailability of such a test in Chinese made it impossible to test this group's WM capacity in order to compare it with those of the other groups.

5.1.1 Non-word Repetition Test

The phonological WM capacity test used in Experiment I consisted of one of the components of the Comprehensive Test of Phonological Processing (CTOPP; Torgesen et al., 1999), the non-word repetition test included in this battery, which was meant to test phonological processing and awareness. In this test, participants listen to and repeat nonsense words that range from one to six syllables in length and contain some segments from the English phonological repertoire that are not present in the Spanish phonological inventory. The phonological loop task included in Experiment I comprised 18 target items of increasing syllable length that were recorded by a female voice, and each item was separated from the following one by a 5-second inter-stimuli interval. All participants were presented with 5 practice items to familiarize themselves with the task.

A crucial difference between the current implementation of this test and those reported in the literature is the scoring procedure followed after the participants' productions had been digitally recorded. Most of the existing studies that attempted to assess the relationship between this kind of non-word repetition task and actual mastery of an L2 vocabulary followed a scoring procedure that discarded words (gave a 0) that contained a mispronounced segment (Ellis & Sinclair, 1996; Gathercole, 2006; Masoura & Gathercole, 1999, 2005; Service & Kohonen, 1995). Results in studies following this scoring system led researchers to claim that "… learners' ability to repeat total gobbledygook is a remarkably good predictor of their ability to acquire sophisticated language skills in both the L1 and L2" (Ellis, 1996, p.102). The outcome of this kind of scoring represents accuracy as regards manner and place of articulation, and also voicing value, instead of measuring actual segment storage capacity. It comes as no surprise that studies following a phonological accuracy criterion found significant relationships between this measure and vocabulary acquisition. In order to exemplify this kind of procedure here, Masoura & Gathercole's (1999) Phonological Short Term Memory (PSM) task with Greek-speaking learners of English (children) can be taken as a representative study (for similar scoring methods see Gathercole, 1995; Service, 1992; Service & Craik, 1993; Service & Kohonen, 1995).

In Masoura & Gathercole's (1999) experiment the aim was to assess the existence of a link between the acquisition of foreign language vocabulary and phonological short-term memory (PSM) in Greek children aged 8-11. The foreign language these children were learning was English and their PSM capacity was measured using the Children's Test of Nonword Repetition (CNRep; Gathercole & Baddeley, 1996). The participants' receptive and productive native vocabulary repertoires were assessed using adapted measures, and their English vocabulary was also tested with bi-directional translation-equivalent tasks. Other variables, such as nonverbal ability, length of study of the foreign language and age were part of the design of this study as well.

Table 12. Correlation matrix from Masoura & Gathercole (1999)

Measures	*1*	*2*	*3*	*4*	*5*	*6*	*7*
1 Greek nonword repetition total		.47**	.43**	.44**			
2 English nonword repetition total	.48**		.31*	.43*			
3 Greek (native) vocabulary	.50**	.35*		.56**			
4 English (foreign) vocabulary	.36*	.39**	.66**				
5 Nonverbal ability	.24	.24	.40**	.34*			
6 Length of study	.04	.16	.03	.73**	.05		
7 Chronological age (months)	.03	.31*	.36*	.32*	.34*	.43**	

Lower triangle presents simple correlations coefficients, $df = 45$; upper triangle presents partial correlations after nonverbal ability, length of study and age partialled out, $df = 39$.

* $P < .05$; ** $P < .001$; all others nonsignificant.

Table 12 shows the results from correlations in Masoura & Gathercole (1999). The findings point to a significant relationship between the English non-word repetition task, the PSM measure, and knowledge of English vocabulary in the children tested ($r = .39$, $p < .001$). Masoura & Gathercole take this association as evidence of the predictive nature of PSM in success at learning foreign words. However, if attention is paid to the details of the scoring procedure for the foreign non-word repetition task, it is evident that the focus was on phonological accuracy: "Each of the 40 nonwords was scored as either correct or incorrect, with no penalty for the characteristic prosody of a Greek accent provided that the phonemes were correct" (p. 384). This fact renders the significant association between the PSM and foreign vocabulary knowledge meaningless, since phonological accuracy is a measure of proficiency, which should correlate with foreign language vocabulary measures.

If one assesses phonological proficiency, this measure will undoubtedly be related not only to vocabulary knowledge, but also morphosyntactic accuracy in the learners' interlanguage. Hence, in the scoring procedures handed to raters for this kind of test in Experiment I, instructions specified that a repetition should be given a 0 if, and only if, there was an omission or addition of a segment present in the stimuli, and also if the segments were scrambled or the position of word stress had changed. This scoring method ensured that the measure obtained was actually a storage capacity unit and not a covert measure of proficiency.

5.1.2 Reading Span Task

The WM storage and processing test included in this thesis was aimed at measuring the capacity of the central executive within Baddeley's (2000, 2003) model of WM and was a modified version of the Reading Span Task, first used by Daneman and Carpenter (1980).

The RST in Experiment I (as well as in Experiments II & III) was administered in the participants' L1 (Spanish) and consisted of sets of sentences that they had to read outloud at a normal pace. For this particular implementation, participants simultaneously had to try to detect morphosyntactic mistakes included in half of the stimuli presented to them. This monitoring of the morphosyntactic accuracy of the sentences read was meant to tackle the processing capacity of the central executive (in some other implementations it involves a judgment as to the plausibility of the sentence instead of the identification of grammatical mistakes, e.g. Waters & Caplan, 1996). The storage capacity is measured by asking participants to remember the last word of each of the sentences read. Stimuli were presented in sets varying from 2 to 5 sentences, each with a total of 100 sentences and the same number of two-syllable words to be remembered in the participants' L1. Both groups, native speakers and Spanish-speaking learners, performed in this test with stimuli presented in their corresponding L1s.

An important detail pertaining to the administration of this test was that the procedure was experimenter-controlled, following Friedman & Miyake's (2004) recommendations. In previous research, participants have been left alone during the experimental sessions, which Friedman & Miyake (2004) argue allows them to implement a diverse set of strategies (subvocalization, semantic association, targeting the last word first), meant to improve their final score in this test. Consequently, in this implementation, the last word of the sentence, the item to be recalled later, did not appear with the rest of the words at the beginning of the reading phase. Instead, the experimenter followed the reading pace of the participant and when they had reached approximately the middle of the sentence, the researcher pressed a button for the last word to appear. Another modification involved the experimenter controlling the onset of the question asking about the morphosyntactic accuracy of the string (Is this sentence correct?) once the participant had finished reading the sentence aloud. When a participant had finished reading a set of sentences, a message appeared on the computer screen asking them to recall the last word of each of the sentences read.

A further difference with previous research studies that used this kind of test is the scoring method followed to evaluate participants' performance. Based on Conway et al.'s (2005) advice, the score resulting from the storage measure of this test reflected a more accurate picture of their memory capacity, since partial credit was assigned for incomplete sets. In previous research, a missing word in the recalling stage of the test prompted the annulment of that particular set with a score of 0. In this study, the partial credit unit scoring method was applied to the results so that if, for example, 2 out of 3 words were remembered in a set, that production was given a .67 score instead of a null score, thereby more accurately representing the actual storage capacity of the participants.

Another score stemming from the RST is accuracy in detecting morpho-syntactic mistakes in L1 sentences, which was measured as a percentage out of a possible 100 items judged. These two tests then provide us with three different measures of WM capacity: a segment storage-capacity score[4] (non-word repetition test), a word-storage, and a processing accuracy measure for all participants.

Finally, following suggestions in Conway et al. (2005) and Swets et al. (2007), the use of extreme-groups design for the processing of WM data in these experiments was avoided. Conway et al. (2005) suggest that dividing participants into high and low WM groups when conducting correlational analyses for these measures takes away variance that could prove useful in uncovering relationships between WM and other cognitive abilities. Thus, the correlations reported in the WM results and discussion sections for the three experiments reported here include all participants.

The following three sections describe and discuss the outcomes of testing the WM capacity of the native and Spanish-speakers with the tasks outlined above.

5.1.2.1 WM Results from Experiment I

The aim of Experiment I was to determine what kind of information, structural or otherwise, is used by L2 learners when parsing locally ambiguous adverbial subordinate clauses, similar to the stimuli used in Traxler (2002, 2005). The RT findings from the SPR task showed that the Spanish-speaking learners of English used a similar parsing heuristic to the one used by native speakers, which is initially based on the word class of the lexical items parsed. Table 13 shows average scores for the L2 learners in Experiment I.

---

[4] A phonological working memory test was designed and implemented following a model previously used by French (2001) for the English native speakers by using Arabic words. Three native speakers of Arabic, following the same guidelines used for the CTOPP test, judged the performance of these participants, but, unfortunately, there was no inter-rater reliability between the three scorers and thus these data are not included here.

Table 13. Spanish Speakers - Scores for WM tests and proficiency test (EPT)

n = 20	Non-Word Rep	RST-STOR	RST-ACC	EPT	EPT-Vocab	EPT-Grammar
Mean	13.93	17.45	91.25	67.75	23.9	20.85
SD	2.15	3.06	3.93	9.24	3.78	3.29
out of	18	25	100	78	27	25

Out of a possible total of 18 points (or words repeated correctly) the Spanish-speaking learners achieved an average of 13.93 points in the non-word repetition test used here. As discussed before, the RST renders two different measures of WM capacity, a storage measure (RST-STO) and a processing measure (RST-ACC). Just as a reminder to the reader, this test was administered in the participants' L1. The L2 learners in this study were able to remember 17.25 sets of words out of a possible 25. They were also very accurate in their identification of morpho-syntactic mistakes in the sentences read during the RST, being correct on 91.25% of instances on average. Let us now analyze the correlation data between the working memory measures and the L2 proficiency of the participants. These data are presented in Table 14.

Table 14. Spanish Speakers - Correlations between WM and Proficiency (EPT) in Exp. I

n = 20	Non-Word Rep	RST-ACC	RST-STOR	EPT	EPT-Vocab	EPT-Grammar
Non-Word Rep	1					
RST-ACC	-0.13	1				
RST-STOR	0.06	0.16	1			
EPT	0.26	0.05	0.44	1		
EPT-Vocab	0.25	-0.13	0.36	0.93*	1	
EPT-Grammar	0.21	0.2	0.5*	0.94*	0.77**	1

* p<=.05; ** p<=.0005

The first relevant finding shown in Table 14 is the complete lack of a relationship between the segment-storage WM measure represented by the CTOPP/Non-word repetition test and the rest of the WM measures and proficiency scores. Thus, predictions confirmed in other studies as to the validity of the non-word repetition ability of L2 learners being a successful predictor of vocabulary acquisition or L2 proficiency in general (e.g. in Masoura & Gathercole, 1999) were not supported in this study.

In order to provide more evidence in favor of the argument against the traditional way of scoring non-word repetition tests purely on a phonological accuracy basis, another two scorers were asked to judge the repetitions attempted by the Spanish speakers for this task following different guidelines. Their instructions were to give a null score to all repetitions that included mistakes in place and/or manner of articulation, and errors in voicing and stress placement. This is the strictly phonological scoring method (NW Rep Phon) followed in most of the studies that have claimed a relationship between non-word repetition scores and proficiency in a L2. The results of the correlations between this WM measure (NW Rep Phon) and the remaining WM tests and proficiency scores for the Spanish-speaking learners appear in Table 15.

Table 15. Spanish speakers – Correlations- NW Rep test strict phonological scoring method

n = 20	NW Rep Phon	RST-ACC	RST-STOR	EPT	EPT-Vocab	EPT-Grammar
NW Rep Phon	1					
RST-ACC	.23	1				
RST-STOR	.04	0.16	1			
EPT	.599**	0.05	0.44	1		
EPT-Vocab	.509*	-0.13	0.36	0.93*	1	
EPT-Grammar	.586**	0.2	0.5*	0.94*	0.77**	1

* p <= .05; ** p <= .001

It is evident from Table 15 that the scoring method chosen is the culprit for the resulting significant correlations found between the NW Rep Phon test and the EPT from the Michigan battery. This confirms the hypothesis that previous research using this strict scoring method was not actually tapping onto the *storage* capacity advocated for the phonological loop in Baddeley's model, but for another measure of proficiency of the L2 learners tested. What is more, additional criticism has been raised against this type of measure, based on the many different variables involved in the repetition of a non-word with unfamiliar sounds. Smith (2006, p. 585) states it very clearly when criticizing work carried out by Gathercole and colleagues: "Given that speech output always constitutes the "end product" of this process, it is difficult to know with certainty if output errors that subjects make in an imitation task reflect whether they may have misheard the production of a stimulus to be repeated, it was heard correctly but then stored in short-term memory inaccurately, it was mispronounced for reasons unrelated to hearing or memory factors, or it was potentially due to some combination of these factors." These methodological uncertainties, as well as the correlational data just discussed, motivated the exclusion of this type of test from the WM measures used in the remaining experiments in this thesis. A decision was made to focus on the processing and storage test, the RST, instead.

An interesting result in relation to the RST is the significant relationship found between the proficiency in the grammar section of the EPT and the storage score in the RST. This finding points to a relationship between grammatical competence and the capacity to store lexical items after having processed information contained in a sentence.

It should also be noted that the performance of these learners in the detection of morpho-syntactic mistakes in their L1 (RST-ACC) did not correlate with any of the proficiency measures used in this study. However, there was a significant relationship between this capacity and the learners' speed of reading in the self-paced reading task. Tables 16 and 17 show the relationship between the Residual RTs for native and Spanish speakers in the SPR task and the processing measure of the RST, the accuracy rate at detecting morphosyntactic mistakes in L1, used in this experiment. The regions of interest for this analysis are the second NP (NP2), the main verb (MV) and a spill-over region after the main verb (MV1).

Table 16. Correlations - Residual RTs and RST-ACC for IBSV condition in Experiment I

Participants	IBSV-NP1	IBSV-SV	IBSV-NP2	IBSV-MV	IBSV-MV1
Spanish group	.174	.387	.611**	.422#	.460*
Native group	.096	.160	.130	.110	.070

*** p < .05; ** p < .01; # approaches significance p =.064**

Table 17. Correlations - Residual RTs and RST-ACC for TBSV condition in Experiment I

Participants	TBSV-NP1	TBSV-SV	TBSV-NP2	TBSV-MV	TBSV-MV1
Spanish group	.113	.337	.58**	.446*	.466*
Native group	.175	-.009	-.078	-.133	-.082

As the significant correlations for both types of target stimuli over the relevant regions seem to indicate, it was only in the performance of L2 learners that the time taken to read these regions was related to the accuracy in spotting morpho-syntactic mistakes in the L1 of these non-native speakers. To my knowledge, there is only one other study where this relationship has obtained. Juffs & Rodríguez (2006) found a very similar effect for RST-ACC with sentences containing relative clauses that modified subjects and objects with adjacent and non-adjacent verbs.

5.1.2.2 WM results from Experiment II and III

The RST used to measure the storage and processing WM capacity of participants in Experiments II and III had the same characteristics of the test employed in Experiment I. The averages by group and the EPT results for the Spanish speakers are displayed in Table (18).

Table 18. RST and EPT averages by group for Experiments II & III

	RST-STOR	RST-ACC	EPT
Spanish (n=15)			
Mean	16.69 (1.99)	88.92 (3.09)	95.00% (3.56)
English (n=15)			
Mean	16.19 (3.07)	92.4 (4.88)	NA
Possible Total	25	100	70

ANOVAs for RST-STOR showed no difference between groups on this measure that was administered in the participants' L1 ($F(1,29)=.251$, $p=.621$). A similar procedure for RST-ACC showed that the difference in rates at detecting morphosyntactic mistakes in the L1 was significantly different between these two groups ($F(1,29)=4.88$, $p=.036$).

Correlations between the two kinds of scores from the RST test and the proficiency measure used in these experiments, shown in Table 19, indicated a trend towards a significant association between the storage measure, how many words per set were recalled, and the EPT (see Appendices M and N for the full set of correlations). This tendency replicates findings in Experiment I as regards a relationship between storage and L2 proficiency in the two groups of Spanish-speaking learners tested for this thesis.

Table 19. Spanish speakers - Correlations between RST and EPT

n = 15	RST-ACC	RST-STOR	EPT
RST-ACC	1		
RST-STOR	-.002	1	
EPT	.401	.491#	1

 # tendency towards significance p = .08

However, in spite of this seemingly pervasive finding of a relationship between storage and proficiency, the remaining WM measure did not render any significant results. There were no significant trends that associated the RST-ACC measure with either the proficiency of the learners or the residual RTs for both learners and native speakers with the stimuli used in Experiments II and III (see Appendices O and P for the full set of correlations). This appears to be a surprising finding here, since the very significant trends between RST-ACC and residual RTs in Experiment I led us to suggest that the Spanish-speaking learners may have engaged in an error-detection strategy when reading sentences in their L2. Having used the same experimenter-controlled test for all experiments discussed here, this difference between these experiments and Experiment I seems puzzling. Nevertheless, there was a difference in the order of presentation of the tasks between Experiments I, and II & III that may be responsible for the different results for RST-ACC. In Experiment I, the order of tasks was such that for half of the participants, the RST took place before the SPR task, whereas for Experiments II & III, the SPR task preceded the RST for all subjects. Those participants who were exposed to the RST before having to read sentences in the SPR may have continued to apply a reading strategy that was useful for the RST while they were performing in the SPR that followed it. This kind of error-detection reading strategy while parsing sentences in their L2 may have generated the correlations found between RST-ACC and the residual RTs for Experiment I.

5.1.3 Discussion

There are two findings stemming from the experiments in this thesis that help us further qualify the relationship between WM measures and proficiency in an L2. The first relevant result has to do with previous research that has claimed a relationship between non-word repetition tests, purportedly measuring the phonological loop capacity, and vocabulary and/or grammar competence in the target language. Experiment I clearly demonstrated that if the scoring method used to assess the performance of learners in repetition of L2 non-word items is based on phonological accuracy, the significant correlations obtained should not be taken as an indication of WM influence on L2 competence. The strict phonological scoring method renders another proficiency measure for the L2 which, once correlated with traditional L2 knowledge measures (like vocabulary or grammar tests), is bound to produce a statistically significant association, because they are measuring the same underlying trait.

Another consistent result from these experiments is the relationship found between the capacity to store words, tested in the participants' L1, with the RST and proficiency in the L2. This finding suggests that a greater capacity for remembering words, under cognitive and time pressures, may be beneficial when acquiring an L2. Furthermore, this interaction between storage and proficiency may provide support for theories of SLA that put emphasis on the capacity of learners to chunk target language input into smaller units, which should be stored for later use or analysis in the development of an interlanguage (Ellis, 2001; Myles et al. 1998; Zyzik, 2006). However, it should be noted that no trends or significant relationships were found between the RST-STOR measure and residual RTs for any of the stimuli used in these experiments. This should cast some doubt over the chunking argument in SLA, since a relationship between parsing measures and the capacity to break apart language input into smaller parts could be expected, but again was not found in these data.

Claims made as regards the relationship between RST-ACC and residual RTs in Experiment I should be taken with caution. It was suggested that L2 learners may engage in an error-detection mode when reading sentences in their target language, and that this was reflected in the significant correlations found between the scores in the WM test and the residual RTs. However, as mentioned before, the participants in Experiment I may have engaged in this type of reading strategy due to the order of presentation of tasks in that study, with the RST preceding the SPR for half of the subjects.

6.0 SUMMARY AND CONCLUSIONS

In order to summarize the findings in this thesis, it is necessary to return to the main tenets of the SSH (Clahsen & Felser, 2006) as regards L2 sentence parsing. As the reader may remember, one of the facets of the fundamental difference in this hypothesis was the idea that L2 learners do not utilize syntactic information, in the form of word class, in order to parse sentences in their target language. Experiment I was meant to assess this claim while partially replicating findings in Traxler (2002, 2005) for native speakers. The Spanish-speaking learners who participated in Experiment I showed very similar parsing profiles to the native speakers (in spite of a difference in reading rate) and seemed to be using the same parsing principles that have been described for the performance of native speakers. These parsing principles make exclusive use of structural information, in order to assign an initial representation to a string. Nevertheless, a difference in the way L2 learners resolved the local ambiguity involved in these sentences provided evidence that, although non-native speakers may apply principles that profit from word class in order to assign structure, the grammatical characteristics of their L1 may still play a part in the online deployment of their L2 grammatical knowledge when parsing sentences. This pervasive influence of the L1 grammar during L2 sentence processing has also obtained with different L1 populations in a diverse number of studies (Juffs, 2005; Juffs & Harrington, 1995, Williams et al., 2001). In spite of this difference due to L1 influence, the learners in this study did resort to word class information in order to draw an initial representation for a string of lexical items in their L2.

The second claim tested here as part of the SSH is the issue of gapless parsing in an L2. Clahsen & Felser (2006) claim, based on recent findings (Marinis et al., 2005; Felser et al., 2007), that L2 learners do not pose intermediate structure in their syntactic representations (as native speakers are supposed to do) and that, instead, non-native speakers match fillers and gaps with an associative mechanism that only takes into account semantic or contextual cues in order to assign thematic roles online. The first objection to this claim stems from another line of research in L1 parsing, which has found that native speakers of English do not assign fully detailed structural representations to language strings that present ambiguities, also known as the Good Enough Representations Hypothesis (Ferreira et al., 2002). As the reader may infer, this issue cannot be grounds for a fundamental difference between L1 and L2 parsing given that native speakers do end up with GER sometimes (depending on cognitive and time constraints during the experimental tasks). However, Clahsen & Felser (2006) claim that the difference arises because L2 learners are completely restricted to parsing L2 input without accessing the full parse route in their model. Experiment II was an attempt at not only partially replicating one of the studies that has provided evidence for the Marinis et al.'s (2005) claim, but also testing whether the assumptions behind the design of this study were accurate. The findings from Experiment II paint a more complicated picture where the parsing performance of the three groups tested was fairly similar, without a significant difference between the two conditions that are claimed to include long distance dependencies (IG & GP) of the type tested by Marinis and colleagues. The very interesting finding in this experiment was the consistent significant difference found for a region that included a genitive NP before the filler-gap integration site in one of the conditions compared (NOM). It is claimed here that the effects found in previous research, attributed to the availability of intermediate structure in the performance of both native and non-native speakers (Gibson & Warren, 2004; Marinis et al., 2005), may be motivated by a lingering effect of the genitive construction that makes the assignment of thematic roles more challenging in this kind of stimuli and thus perpetuates the RT difference that surfaces again at the filler-gap integration region. Thus this particular structure may not be the best candidate to assess the reliance on structural information of either native or L2 learners of English. The stimuli in this third experiment posed difficulties to the three groups tested, but again, this may have been generated by other characteristics of the sentences not relevant to the existence of purported intermediate structural gaps in L2 parsing.

The third controversial claim in Clahsen & Felser's SSH deals with the lack of access to configurational principles, i.e. Binding Principles, in order to build syntactic structure in L2 parsing. This assumption would prevent L2 learners from performing in a native-like fashion with input that required, for example, the relationships afforded between pronouns and coreferents with respect to c-command (Chomsky, 1981). The lack of access to these strictly structural principles was tested in Experiment III, following the design and procedures in Kazanina et al. (2007). The effects found by Kazanina and colleagues with native speakers were replicated here, although the locus of the effect appeared later in the parse. The three groups tested presented elevated RTs for regions following a noun that was structurally allowed to corefer with a cataphoric pronoun, but that did not match in gender with the preceding pronoun. This mismatch effect was only expected if the parsing algorithms used in L1 and L2 followed the c-commanding restrictions delineated by Binding Principle C. Thus, the use of Principle C in parsing L2 sentences by the learners in Experiment III goes against one of the main assumptions in the SSH. Some significant differences were found in the parsing profiles of the Chinese-speaking learners, but as suggested before, this may be the result of differences between English and their L1 grammar as regards gender marking. In spite of these spillover effects for the three participant groups, the native speakers' parsing profile presented an interesting tendency that did not reach significance, but may also call into question the idea of a fundamental difference. As the reader may remember, region 8 in Experiment III showed a gender-based grouping of stimuli conditions, with gender mismatch conditions patterning together on the crucial NP. This may indicate that even native speakers here were not resorting to structural information at that stage of the parse and instead guided their parsing decisions based on lexical cues such as gender. If when assessed with a greater number of native speakers, this trend were to prove statistically significant, it would further undermine the idea of a fundamental difference between L2 learners and native speakers as regards the use of configurational information in sentence parsing. It would also lend further support to the ideas that inform the good-enough representations approach to parsing (Christianson et al., 2006; Ferreira et al., 2002; Sanford & Sturt, 2002), in that all comprehenders (native and L2 learners) may feed underspecified structural information to the parser when trying to understand an utterance in order to maximize cognitive resources employed in comprehension.

Finally, the WM data gathered in these three experiments provided evidence against a traditional method of scoring non-word repetition tests which are supposed to measure the storage capacity of the phonological loop in Baddeley's model (2000). The focus on phonological accuracy in previous research has rendered significant correlations between NWRep tests and proficiency measures, which have led some researchers to claim a fundamental role for the phonological loop as a language learning device. The scoring method adopted for the

NWRep test in Experiment I, based solely on segment storage capacity, showed that this measure did not show any relationship with either proficiency or RT measures in this study. Further evidence came from a re-scoring of the results following the traditional phonological accuracy method that resulted, as expected, in significant correlations between the test, vocabulary, and grammar competence in the L2. These results prompted the decision not to use this same measure for Experiments II and III, and to focus instead on the storage and processing test, the RST. The inclusion of the RST was motivated by claims that suggest a preponderant role for the central executive in Baddeley's WM model (2003) during parsing of an L2. The expectations were that this measure would not only be related to L2 proficiency, but also to measures of parsing efficiency, such as RTs and comprehension probe accuracy. Findings from Experiment I seemed to indicate that both predictions would obtain with a significant relationship between the storage section of the test and grammatical competence, and other significant associations between RTs and accuracy in the RST. The former result proved consistent in Experiments II and III, however the relationship between RST-ACC and RTs in the SPR in Experiment I may have been the result of an ordering effect. Thus, the only measure of WM that correlated with L2 proficiency was the storage part of the RST and, as mentioned before, this may be related to the fact that individuals with a higher storage capacity may be able to better store and analyze more chunks of the target language grammar. This, in turn, would prove beneficial to the development of their interlanguage grammar towards the native-like target.

In conclusion, the experiments included in this thesis stand as problematic evidence for the fundamental difference assumption included in the SSH. Experiment I showed that L2 learners do use phrase structure heuristics that are similar to the parsing strategies employed by native speakers. It also provided evidence for the existence of L1 transfer effects, in this case the null subject characteristic of Spanish, which affected the online assignment of structure to an incoming string of words in the L2. At the same time, even though the prediction in Experiment II as regards the difficulty of a GP condition did not obtain, the data stemming from the parsing of a genitive phrase cast doubts on the actual justification for the comparison postulated in previous studies. The third experiment also showed that another major claim in the SSH, the lack of configurational principles in L2 parsing, may also actually be part of the parsing repertoire not only of L2 learners but also of native speakers of English. Consequently, in this dissertation, no empirical evidence supporting Clahsen & Felser's assumptions as to a fundamental difference between the parsing algorithms employed by L1 and L2 comprehenders was found.

The most important finding as regards the role of WM in L2 parsing and competence has to do with the consistent relationship found between the storage measure in the RST and the proficiency measure tapping into grammatical knowledge used here. This relationship should be replicated in future studies given the differences found for the RST-ACC scores between Experiments I and II & III. It would also prove useful to implement tests that tap into different modalities of WM capacity, so that the results of language-specific tests can be compared to those of, for example, visual WM measures (following work with native speakers by Swets et al., 2007). This approach will help better refine the construct of WM when applied to L2 learners.

The data and claims included in this dissertation should be taken with caution, since the number of participants is not ideal for sound statistical testing. Besides, more technologically-advanced measurement instruments, such as eye-tracking technology and ERP measurement, could provide a more fine-grained picture of the relationship between structural information, WM, and parsing in a foreign language. It is also feasible that improvements to the implementation of the WM measures included here may not only improve the reliability of these measures, but also provide a better representation of the capacities being studied. For example, it has been claimed that the RST measures storage and processing capacities; however, the storage measure in this RST is taken after and not while the processing has taken place (Omaki & Ariji, 2005).

In future work it would be interesting to test the opposite L1 relationship with English-speaking learners of Spanish, a language that presents rich agreement morphology, which may be used not only to test claims about the use of structural information, but also to test claims of morphological insensitivity in an L2 (Jiang, 2004, 2007). At the same time, it would be important to find a pedagogical application in order to exploit the phrase structure heuristics that learners were found to apply in these experiments, following efforts in, for example, VanPatten's work (1996; 2004). Another relevant avenue of research that has sparked a great deal of interest in L1 parsing, but that has not been explored in SLA, is the topic of syntactic prediction and how this may hinder or facilitate processes in the acquisition of a foreign language (Lau et al., 2007).

APPENDIX A

NON-NATIVE SPEAKERS - BIOGRAPHICAL INFORMATION – EXPERIMENT I

Table 20. Biographical information for non-native speakers - Experiment I

Participant #	*Gender*	*EPT*	*Origin*	*Age*	*Education*
SPANISH SPEAKERS		%			
1	F	62	Peru	30	MA
2	F	53	Peru	52	BA
3	F	63	Chile	28	BA
4	M	70	Colombia	27	PHD
5	M	68	Colombia	40	MA
6	M	72	Honduras	46	PHD
7	M	75	Peru	44	MBA
8	M	71	Colombia	28	PHD
9	F	60	Colombia	57	MA
10	F	71	Colombia	67	BA
11	F	76	Peru	31	MA
31	M	68	Colombia	37	BA
32	M	41	Chile	26	BA
33	M	76	Venezuela	28	MA
34	F	77	Colombia	28	PHD
35	M	67	Venezuela	34	MA
36	F	74	Honduras	38	MA
37	F	77	Colombia	27	MA
38	F	75	Venezuela	35	MA
39	M	59	Colombia	30	BA
Average/Totals	**10F/10M**	**67.75**		**36.7**	

APPENDIX B

NON-NATIVE SPEAKERS - BIOGRAPHICAL INFORMATION – EXPS. II & III

Table 21. Biographical information for non-native speakers - Experiments II & III

Participant #	*Gender*	*EPT*	*Origin*	*Age*	*Education*
SPANISH SPEAKERS		%			
702	M	97	Colombia	29	Postgraduate
703	M	99	Colombia	34	Postgraduate
705	M	93	Colombia	41	Postgraduate
709	M	96	Chile	51	Postgraduate
710	F	93	Chile	26	Postgraduate
711	M	98	Colombia	30	Postgraduate
751	F	99	Peru	32	Postgraduate
753	F	97	Panama	28	Postgraduate
754	F	96	Colombia	36	Postgraduate
756	F	92	Colombia	24	Postgraduate
757	F	94	Chile	29	Undergraduate
759	F	92	Colombia	21	Undergraduate
760	M	91	Venezuela	27	Postgraduate
761	F	90	Argentina	32	Postgraduate
762	F	98	Chile	30	Postgraduate
Average/Totals	**8-F / 5-M**	**95**		**31.3**	
CHINESE SPEAKERS		%			
801	F	93	Taiwan	26	Postgraduate
802	F	91	Taiwan	27	Postgraduate
803	F	97	Taiwan	27	Postgraduate
805	F	96	Taiwan	30	Postgraduate
806	F	89	China	28	PhD
807	F	93	Taiwan	27	Postgraduate
808	F	94	Taiwan	28	Postgraduate
809	F	90	Taiwan	25	Postgraduate
810	M	98	Taiwan	27	Postgraduate
851	M	90	Taiwan	26	Postgraduate
853	F	90	Taiwan	27	Postgraduate
856	F	91	Taiwan	27	Postgraduate
857	F	94	Taiwan	24	Postgraduate
859	F	97	Taiwan	24	BA
860	F	93	China	30	Postgraduate
Average/Totals	**13-F/2-M**	**93.07**		**26.9**	

APPENDIX C

ANOVA FOR EPT (PROFICIENCY TEST) BETWEEN GROUPS

Table 22. Anova for proficiency - Descriptives

Descriptives

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
spanish	15	95.0000	3.55903	.98710	92.8493	97.1507	89.00	99.00
chinese	15	93.0667	2.91466	.75256	91.4526	94.6808	89.00	98.00
Total	30	93.9643	3.31643	.62675	92.6783	95.2503	89.00	99.00

Table 23. Anova for proficiency

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	26.031	1	26.031	2.498	.126
Within Groups	270.933	28	10.421		
Total	296.964	29			

APPENDIX D

D.1 MULTIVARIATE ANOVA FOR EXP. II CONDITIONS ACCURACY

Table 24. Multivariate anova for Exp. II on accuracy - Descriptives

Descriptive Statistics

	GROUP	Mean	Std. Deviation	N
IGs	native	85.3333	14.07463	15
	spanish	89.2308	15.52500	15
	chinese	69.3333	14.86447	15
	Total	80.9302	16.87686	45
GP	native	85.3333	20.65591	15
	spanish	75.3846	20.25479	15
	chinese	84.0000	18.82248	15
	Total	81.8605	19.91121	45
NOM	native	69.3333	18.30951	15
	spanish	66.1538	25.01282	15
	chinese	69.3333	21.20198	15
	Total	68.3721	21.03628	45

Table 25. Multivariate anova for Exp. II on accuracy

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	AccIG	3203.816[a]	2	1601.908	7.316	.002
	AccGP	794.753[b]	2	397.376	1.002	.376
	AccNOM	91.688[c]	2	45.844	.099	.906
Intercept	AccIG	282921.026	1	282921.026	1292.028	.000
	AccGP	284827.817	1	284827.817	718.518	.000
	AccNOM	199525.153	1	199525.153	431.537	.000
GROUP	AccIG	3203.816	2	1601.908	7.316	.002
	AccGP	794.753	2	397.376	1.002	.376
	AccNOM	91.688	2	45.844	.099	.906
Error	AccIG	8758.974	40	218.974		
	AccGP	15856.410	40	396.410		
	AccNOM	18494.359	40	462.359		

a. R Squared = .268 (Adjusted R Squared = .231)

b. R Squared = .048 (Adjusted R Squared = .000)

c. R Squared = .005 (Adjusted R Squared = -.045)

Table 26. Multivariate anova for Exp. II on accuracy - Comparisons

Multiple Comparisons

Tukey HSD

Dependent Variable	(I) GROUP	(J) GROUP	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
SPRaccE	native	spanish	-3.8974	5.60736	.768	-17.5453	9.7504
		chinese	16.0000*	5.40339	.014	2.8486	29.1514
	spanish	native	3.8974	5.60736	.768	-9.7504	17.5453
		chinese	19.8974*	5.60736	.003	6.2496	33.5453
	chinese	native	-16.0000*	5.40339	.014	-29.1514	-2.8486
		spanish	-19.8974*	5.60736	.003	-33.5453	-6.2496
SPRaccF	native	spanish	9.9487	7.54456	.393	-8.4141	28.3116
		chinese	1.3333	7.27012	.982	-16.3616	19.0282
	spanish	native	-9.9487	7.54456	.393	-28.3116	8.4141
		chinese	-8.6154	7.54456	.494	-26.9782	9.7475
	chinese	native	-1.3333	7.27012	.982	-19.0282	16.3616
		spanish	8.6154	7.54456	.494	-9.7475	26.9782
SPRaccG	native	spanish	3.1795	8.14801	.920	-16.6521	23.0111
		chinese	.0000	7.85162	1.000	-19.1102	19.1102
	spanish	native	-3.1795	8.14801	.920	-23.0111	16.6521
		chinese	-3.1795	8.14801	.920	-23.0111	16.6521
	chinese	native	.0000	7.85162	1.000	-19.1102	19.1102
		spanish	3.1795	8.14801	.920	-16.6521	23.0111

Based on observed means.

The error term is Mean Square(Error) = 462.359.

*. The mean difference is significant at the .05 level.

APPENDIX E

MULTIVARIATE ANOVA FOR EXP. III CONDITIONS ACCURACY

Table 27. Multivariate anova for Exp. III on accuracy

Descriptive Statistics

	GROUP	Mean	Std. Deviation	N
SPRaccA	native	89.3333	10.32796	15
	spanish	80.0000	8.16497	15
	chinese	81.3333	11.87234	15
	Total	83.7209	10.91597	45
SPRaccB	native	98.6667	5.16398	15
	spanish	98.4615	5.54700	15
	chinese	89.3333	14.86447	15
	Total	95.3488	10.54443	45
SPRaccC	native	86.6667	9.75900	15
	spanish	90.7692	10.37749	15
	chinese	93.3333	9.75900	15
	Total	90.2326	10.11561	45
SPRaccD	native	85.3333	20.65591	15
	spanish	89.2308	15.52500	15
	chinese	89.3333	14.86447	15
	Total	87.9070	16.98152	45

Table 28. Multivariate anova for Exp. III

Multivariate Tests

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.995	2.009E3	4.000	37.000	.000
	Wilks' Lambda	.005	2.009E3	4.000	37.000	.000
	Hotelling's Trace	217.200	2.009E3	4.000	37.000	.000
	Roy's Largest Root	217.200	2.009E3	4.000	37.000	.000
GROUP	Pillai's Trace	.423	2.550	8.000	76.000	.016
	Wilks' Lambda	.608	2.608[a]	8.000	74.000	.014
	Hotelling's Trace	.591	2.661	8.000	72.000	.013
	Roy's Largest Root	.483	4.592[b]	4.000	38.000	.004

a. Exact statistic

b. The statistic is an upper bound on F that yields a lower bound on the significance level.

c. Design: Intercept + GROUP

Table 29. Multivariate anova for Exp. III - Tests between-subjects

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	SPRaccA	737.984[a]	2	368.992	3.459	.041
	SPRaccB	833.870[b]	2	416.935	4.348	.020
	SPRaccC	338.700[c]	2	169.350	1.711	.194
	SPRaccD	152.654[d]	2	76.327	.255	.776
Intercept	SPRaccA	298843.577	1	298843.577	2801.659	.000
	SPRaccB	390286.379	1	390286.379	4069.831	.000
	SPRaccC	348697.936	1	348697.936	3523.114	.000
	SPRaccD	331223.465	1	331223.465	1107.866	.000
GROUP	SPRaccA	737.984	2	368.992	3.459	.041
	SPRaccB	833.870	2	416.935	4.348	.020
	SPRaccC	338.700	2	169.350	1.711	.194
	SPRaccD	152.654	2	76.327	.255	.776
Error	SPRaccA	4266.667	40	106.667		
	SPRaccB	3835.897	40	95.897		
	SPRaccC	3958.974	40	98.974		
	SPRaccD	11958.974	40	298.974		

a. R Squared = .147 (Adjusted R Squared = .105)

b. R Squared = .179 (Adjusted R Squared = .137)

c. R Squared = .079 (Adjusted R Squared = .033)

d. R Squared = .013 (Adjusted R Squared = -.037)

Multiple Comparisons

Tukey HSD

Dependent Variable	(I) GROUP	(J) GROUP	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
SPRaccA	native	spanish	9.3333	3.91360	.056	-.1920	18.8587
		chinese	8.0000	3.77124	.098	-1.1789	17.1789
	spanish	native	-9.3333	3.91360	.056	-18.8587	.1920
		chinese	-1.3333	3.91360	.938	-10.8587	8.1920
	chinese	native	-8.0000	3.77124	.098	-17.1789	1.1789
		spanish	1.3333	3.91360	.938	-8.1920	10.8587
SPRaccB	native	spanish	.2051	3.71078	.998	-8.8266	9.2369
		chinese	9.3333*	3.57580	.033	.6301	18.0365
	spanish	native	-.2051	3.71078	.998	-9.2369	8.8266
		chinese	9.1282*	3.71078	.047	.0965	18.1599
	chinese	native	-9.3333*	3.57580	.033	-18.0365	-.6301
		spanish	-9.1282*	3.71078	.047	-18.1599	-.0965
SPRaccC	native	spanish	-4.1026	3.76984	.527	-13.2781	5.0729
		chinese	-6.6667	3.63271	.171	-15.5084	2.1751
	spanish	native	4.1026	3.76984	.527	-5.0729	13.2781
		chinese	-2.5641	3.76984	.776	-11.7396	6.6114
	chinese	native	6.6667	3.63271	.171	-2.1751	15.5084
		spanish	2.5641	3.76984	.776	-6.6114	11.7396
SPRaccD	native	spanish	-3.8974	6.55207	.824	-19.8447	12.0498
		chinese	-4.0000	6.31373	.803	-19.3671	11.3671
	spanish	native	3.8974	6.55207	.824	-12.0498	19.8447
		chinese	-.1026	6.55207	1.000	-16.0498	15.8447
	chinese	native	4.0000	6.31373	.803	-11.3671	19.3671
		spanish	.1026	6.55207	1.000	-15.8447	16.0498

Based on observed means.

The error term is Mean Square(Error) = 298.974.

*. The mean difference is significant at the .05 level.

APPENDIX F

REPEATED MEASURES ANOVA TABLES – EXPERIMENT II

F.1 NATIVE SPEAKERS

Table 30. Repeated Measures ANOVA - Experiment II - NS - Descriptives

Descriptive Statistics

	Mean	Std. Deviation	N
CeRegion2	-16.7617	42.90389	15
CeRegion3	50.4123	126.13194	15
CeRegion5	-6.0660	30.39951	15
CeRegion6	44.7666	70.16279	15
CfRegion2	-20.4619	41.35774	15
CfRegion3	73.8662	105.51956	15
CfRegion5	5.3377	56.60421	15
CfRegion6	42.3443	54.24398	15
CgRegion2	39.9579	71.74515	15
CgRegion3	112.8552	69.96055	15
CgRegion5	35.5210	80.23770	15
CgRegion6	90.5614	116.63098	15

Table 31. Repeated measures ANOVA - Exp. II - NS

Tests of Within-Subjects Effects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
condition	Sphericity Assumed	93877.986	2	46938.993	10.258	.000
	Greenhouse-Geisser	93877.986	1.554	60427.026	10.258	.001
	Huynh-Feldt	93877.986	1.712	54847.203	10.258	.001
	Lower-bound	93877.986	1.000	93877.986	10.258	.006
Error(condition)	Sphericity Assumed	128129.244	28	4576.044		
	Greenhouse-Geisser	128129.244	21.750	5890.982		
	Huynh-Feldt	128129.244	23.963	5347.009		
	Lower-bound	128129.244	14.000	9152.089		
region	Sphericity Assumed	189332.720	3	63110.907	9.952	.000
	Greenhouse-Geisser	189332.720	2.692	70321.612	9.952	.000
	Huynh-Feldt	189332.720	3.000	63110.907	9.952	.000
	Lower-bound	189332.720	1.000	189332.720	9.952	.007
Error(region)	Sphericity Assumed	266348.783	42	6341.638		
	Greenhouse-Geisser	266348.783	37.693	7066.198		
	Huynh-Feldt	266348.783	42.000	6341.638		
	Lower-bound	266348.783	14.000	19024.913		
condition * region	Sphericity Assumed	6367.925	6	1061.321	.361	.902
	Greenhouse-Geisser	6367.925	3.570	1783.734	.361	.815
	Huynh-Feldt	6367.925	4.942	1288.410	.361	.872
	Lower-bound	6367.925	1.000	6367.925	.361	.558
Error(condition*region)	Sphericity Assumed	246983.047	84	2940.274		
	Greenhouse-Geisser	246983.047	49.980	4941.642		
	Huynh-Feldt	246983.047	69.195	3569.400		
	Lower-bound	246983.047	14.000	17641.646		

F.2 SPANISH SPEAKERS

Table 32. Repeated measures ANOVA - Exp. II - Descrptives - SS

Descriptive Statistics

	Mean	Std. Deviation	N
CeRegion2	-31.7436	54.68249	15
CeRegion3	76.7440	112.57023	15
CeRegion5	17.1275	53.91266	15
CeRegion6	106.6802	104.55751	15
CfRegion2	-13.6424	59.72622	15
CfRegion3	92.9005	120.84182	15
CfRegion5	19.1489	49.06169	15
CfRegion6	69.7728	81.56027	15
CgRegion2	15.0336	56.08561	15
CgRegion3	118.3118	172.39610	15
CgRegion5	24.8917	66.48193	15
CgRegion6	66.8631	47.91097	15

Table 33. Repeated measures ANOVA - Exp. II - SS

Tests of Within-Subjects Effects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
condition	Sphericity Assumed	6943.212	2	3471.606	.807	.458
	Greenhouse-Geisser	6943.212	1.693	4101.213	.807	.441
	Huynh-Feldt	6943.212	1.941	3576.667	.807	.455
	Lower-bound	6943.212	1.000	6943.212	.807	.387
Error(condition)	Sphericity Assumed	103213.323	24	4300.555		
	Greenhouse-Geisser	103213.323	20.316	5080.500		
	Huynh-Feldt	103213.323	23.295	4430.703		
	Lower-bound	103213.323	12.000	8601.110		
region	Sphericity Assumed	293793.931	3	97931.310	10.698	.000
	Greenhouse-Geisser	293793.931	2.084	140948.292	10.698	.000
	Huynh-Feldt	293793.931	2.531	116073.758	10.698	.000
	Lower-bound	293793.931	1.000	293793.931	10.698	.007
Error(region)	Sphericity Assumed	329552.356	36	9154.232		
	Greenhouse-Geisser	329552.356	25.013	13175.290		
	Huynh-Feldt	329552.356	30.373	10850.116		
	Lower-bound	329552.356	12.000	27462.696		
condition * region	Sphericity Assumed	32169.783	6	5361.631	1.005	.429
	Greenhouse-Geisser	32169.783	2.395	13432.159	1.005	.391
	Huynh-Feldt	32169.783	3.033	10605.587	1.005	.402
	Lower-bound	32169.783	1.000	32169.783	1.005	.336
Error(condition*region)	Sphericity Assumed	384190.252	72	5335.976		
	Greenhouse-Geisser	384190.252	28.740	13367.888		
	Huynh-Feldt	384190.252	36.399	10554.841		
	Lower-bound	384190.252	12.000	32015.854		

F.3 CHINESE SPEAKERS

Table 34. Repeated measures ANOVA - Exp. II - Descriptives - CS

Descriptive Statistics

	Mean	Std. Deviation	N
CeRegion2	46.6089	102.08402	15
CeRegion3	165.4635	130.46212	15
CeRegion5	-13.5010	91.40761	15
CeRegion6	58.5342	104.36535	15
CfRegion2	29.0821	96.83956	15
CfRegion3	230.0755	151.79351	15
CfRegion5	6.8576	96.77175	15
CfRegion6	93.8968	138.85997	15
CgRegion2	97.8977	130.06276	15
CgRegion3	51.2909	199.21371	15
CgRegion5	-15.8710	129.40367	15
CgRegion6	20.8102	128.60159	15

Table 35. Repeated measures ANOVA - Exp. II - CS

Tests of Within-Subjects Effects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
condition	Sphericity Assumed	79400.909	2	39700.454	2.594	.093
	Greenhouse-Geisser	79400.909	1.864	42599.696	2.594	.097
	Huynh-Feldt	79400.909	2.000	39700.454	2.594	.093
	Lower-bound	79400.909	1.000	79400.909	2.594	.130
Error(condition)	Sphericity Assumed	428603.273	28	15307.260		
	Greenhouse-Geisser	428603.273	26.094	16425.117		
	Huynh-Feldt	428603.273	28.000	15307.260		
	Lower-bound	428603.273	14.000	30614.519		
region	Sphericity Assumed	558215.405	3	186071.802	8.693	.000
	Greenhouse-Geisser	558215.405	2.087	267520.777	8.693	.001
	Huynh-Feldt	558215.405	2.459	226975.199	8.693	.000
	Lower-bound	558215.405	1.000	558215.405	8.693	.011
Error(region)	Sphericity Assumed	898986.796	42	21404.448		
	Greenhouse-Geisser	898986.796	29.213	30773.789		
	Huynh-Feldt	898986.796	34.431	26109.699		
	Lower-bound	898986.796	14.000	64213.343		
condition * region	Sphericity Assumed	249595.706	6	41599.284	4.034	.001
	Greenhouse-Geisser	249595.706	2.858	87336.471	4.034	.015
	Huynh-Feldt	249595.706	3.668	68049.152	4.034	.008
	Lower-bound	249595.706	1.000	249595.706	4.034	.064
Error(condition*region)	Sphericity Assumed	866285.128	84	10312.918		
	Greenhouse-Geisser	866285.128	40.010	21651.668		
	Huynh-Feldt	866285.128	51.350	16870.130		
	Lower-bound	866285.128	14.000	61877.509		

APPENDIX G

REPEATED MEASURES TABLES – EXPERIMENT III – SIGNIFICANT REGIONS

G.1 NATIVE SPEAKERS – REGION 10

Table 36. Repeated measures ANOVA - Exp. III- NS - R10

Tests of Within-Subjects Effects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
constraint	Sphericity Assumed	5038.752	1	5038.752	3.339	.089
	Greenhouse-Geisser	5038.752	1.000	5038.752	3.339	.089
	Huynh-Feldt	5038.752	1.000	5038.752	3.339	.089
	Lower-bound	5038.752	1.000	5038.752	3.339	.089
Error(constraint)	Sphericity Assumed	21124.594	14	1508.900		
	Greenhouse-Geisser	21124.594	14.000	1508.900		
	Huynh-Feldt	21124.594	14.000	1508.900		
	Lower-bound	21124.594	14.000	1508.900		
congruency	Sphericity Assumed	5712.860	1	5712.860	2.150	.165
	Greenhouse-Geisser	5712.860	1.000	5712.860	2.150	.165
	Huynh-Feldt	5712.860	1.000	5712.860	2.150	.165
	Lower-bound	5712.860	1.000	5712.860	2.150	.165
Error(congruency)	Sphericity Assumed	37195.064	14	2656.790		
	Greenhouse-Geisser	37195.064	14.000	2656.790		
	Huynh-Feldt	37195.064	14.000	2656.790		
	Lower-bound	37195.064	14.000	2656.790		
constraint * congruency	Sphericity Assumed	1072.652	1	1072.652	1.278	.277
	Greenhouse-Geisser	1072.652	1.000	1072.652	1.278	.277
	Huynh-Feldt	1072.652	1.000	1072.652	1.278	.277
	Lower-bound	1072.652	1.000	1072.652	1.278	.277
Error(constraint*congruency)	Sphericity Assumed	11754.056	14	839.575		
	Greenhouse-Geisser	11754.056	14.000	839.575		
	Huynh-Feldt	11754.056	14.000	839.575		
	Lower-bound	11754.056	14.000	839.575		

G.2 SPANISH SPEAKERS

G.2.1 REGION 7

Table 37. Repeated measures ANOVA - Exp. III - SS - R7

Tests of Within-Subjects Effects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
constraint	Sphericity Assumed	7495.922	1	7495.922	8.200	.014
	Greenhouse-Geisser	7495.922	1.000	7495.922	8.200	.014
	Huynh-Feldt	7495.922	1.000	7495.922	8.200	.014
	Lower-bound	7495.922	1.000	7495.922	8.200	.014
Error(constraint)	Sphericity Assumed	10969.757	14	914.146		
	Greenhouse-Geisser	10969.757	14.000	914.146		
	Huynh-Feldt	10969.757	14.000	914.146		
	Lower-bound	10969.757	14.000	914.146		
congruency	Sphericity Assumed	2713.652	1	2713.652	1.051	.325
	Greenhouse-Geisser	2713.652	1.000	2713.652	1.051	.325
	Huynh-Feldt	2713.652	1.000	2713.652	1.051	.325
	Lower-bound	2713.652	1.000	2713.652	1.051	.325
Error(congruency)	Sphericity Assumed	30982.857	14	2581.905		
	Greenhouse-Geisser	30982.857	14.000	2581.905		
	Huynh-Feldt	30982.857	14.000	2581.905		
	Lower-bound	30982.857	14.000	2581.905		
constraint * congruency	Sphericity Assumed	130.978	1	130.978	.070	.796
	Greenhouse-Geisser	130.978	1.000	130.978	.070	.796
	Huynh-Feldt	130.978	1.000	130.978	.070	.796
	Lower-bound	130.978	1.000	130.978	.070	.796
Error(constraint*congruency)	Sphericity Assumed	22551.718	14	1879.310		
	Greenhouse-Geisser	22551.718	14.000	1879.310		
	Huynh-Feldt	22551.718	14.000	1879.310		
	Lower-bound	22551.718	14.000	1879.310		

G.2.2 REGION 10

Table 38. Repeated measures ANOVA - Exp. III - SS - R10

Tests of Within-Subjects Effects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
constraint	Sphericity Assumed	12714.007	1	12714.007	5.584	.036
	Greenhouse-Geisser	12714.007	1.000	12714.007	5.584	.036
	Huynh-Feldt	12714.007	1.000	12714.007	5.584	.036
	Lower-bound	12714.007	1.000	12714.007	5.584	.036
Error(constraint)	Sphericity Assumed	27323.888	14	2276.991		
	Greenhouse-Geisser	27323.888	14.000	2276.991		
	Huynh-Feldt	27323.888	14.000	2276.991		
	Lower-bound	27323.888	14.000	2276.991		
congruency	Sphericity Assumed	5288.413	1	5288.413	3.373	.091
	Greenhouse-Geisser	5288.413	1.000	5288.413	3.373	.091
	Huynh-Feldt	5288.413	1.000	5288.413	3.373	.091
	Lower-bound	5288.413	1.000	5288.413	3.373	.091
Error(congruency)	Sphericity Assumed	18814.990	14	1567.916		
	Greenhouse-Geisser	18814.990	14.000	1567.916		
	Huynh-Feldt	18814.990	14.000	1567.916		
	Lower-bound	18814.990	14.000	1567.916		
constraint * congruency	Sphericity Assumed	12979.131	1	12979.131	6.297	.027
	Greenhouse-Geisser	12979.131	1.000	12979.131	6.297	.027
	Huynh-Feldt	12979.131	1.000	12979.131	6.297	.027
	Lower-bound	12979.131	1.000	12979.131	6.297	.027
Error(constraint*congruency)	Sphericity Assumed	24735.594	14	2061.299		
	Greenhouse-Geisser	24735.594	14.000	2061.299		
	Huynh-Feldt	24735.594	14.000	2061.299		
	Lower-bound	24735.594	14.000	2061.299		

G.2.3 REGION 11

Table 39. Repeated measures ANOVA - Exp. III - SS - R11

Tests of Within-Subjects Effects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
constraint	Sphericity Assumed	1669.952	1	1669.952	.166	.691
	Greenhouse-Geisser	1669.952	1.000	1669.952	.166	.691
	Huynh-Feldt	1669.952	1.000	1669.952	.166	.691
	Lower-bound	1669.952	1.000	1669.952	.166	.691
Error(constraint)	Sphericity Assumed	120483.176	14	10040.265		
	Greenhouse-Geisser	120483.176	14.000	10040.265		
	Huynh-Feldt	120483.176	14.000	10040.265		
	Lower-bound	120483.176	14.000	10040.265		
congruency	Sphericity Assumed	1305.786	1	1305.786	.562	.468
	Greenhouse-Geisser	1305.786	1.000	1305.786	.562	.468
	Huynh-Feldt	1305.786	1.000	1305.786	.562	.468
	Lower-bound	1305.786	1.000	1305.786	.562	.468
Error(congruency)	Sphericity Assumed	27880.370	14	2323.364		
	Greenhouse-Geisser	27880.370	14.000	2323.364		
	Huynh-Feldt	27880.370	14.000	2323.364		
	Lower-bound	27880.370	14.000	2323.364		
constraint * congruency	Sphericity Assumed	26609.625	1	26609.625	7.176	.020
	Greenhouse-Geisser	26609.625	1.000	26609.625	7.176	.020
	Huynh-Feldt	26609.625	1.000	26609.625	7.176	.020
	Lower-bound	26609.625	1.000	26609.625	7.176	.020
Error(constraint*congruency)	Sphericity Assumed	44499.218	14	3708.268		
	Greenhouse-Geisser	44499.218	14.000	3708.268		
	Huynh-Feldt	44499.218	14.000	3708.268		
	Lower-bound	44499.218	14.000	3708.268		

G.3 CHINESE SPEAKERS

G.3.1 REGION 7

Table 40. Repeated measures ANOVA - Exp. III - CS - R7

Tests of Within-Subjects Effects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
constraint	Sphericity Assumed	14443.782	1	14443.782	4.123	.062
	Greenhouse-Geisser	14443.782	1.000	14443.782	4.123	.062
	Huynh-Feldt	14443.782	1.000	14443.782	4.123	.062
	Lower-bound	14443.782	1.000	14443.782	4.123	.062
Error(constraint)	Sphericity Assumed	49043.245	14	3503.089		
	Greenhouse-Geisser	49043.245	14.000	3503.089		
	Huynh-Feldt	49043.245	14.000	3503.089		
	Lower-bound	49043.245	14.000	3503.089		
congruency	Sphericity Assumed	15570.379	1	15570.379	2.028	.176
	Greenhouse-Geisser	15570.379	1.000	15570.379	2.028	.176
	Huynh-Feldt	15570.379	1.000	15570.379	2.028	.176
	Lower-bound	15570.379	1.000	15570.379	2.028	.176
Error(congruency)	Sphericity Assumed	107461.341	14	7675.810		
	Greenhouse-Geisser	107461.341	14.000	7675.810		
	Huynh-Feldt	107461.341	14.000	7675.810		
	Lower-bound	107461.341	14.000	7675.810		
constraint * congruency	Sphericity Assumed	4128.587	1	4128.587	.872	.366
	Greenhouse-Geisser	4128.587	1.000	4128.587	.872	.366
	Huynh-Feldt	4128.587	1.000	4128.587	.872	.366
	Lower-bound	4128.587	1.000	4128.587	.872	.366
Error(constraint*congruency)	Sphericity Assumed	66265.265	14	4733.233		
	Greenhouse-Geisser	66265.265	14.000	4733.233		
	Huynh-Feldt	66265.265	14.000	4733.233		
	Lower-bound	66265.265	14.000	4733.233		

APPENDIX H

EXPERIMENT I – TARGET ITEMS

Intransitively-biased Subordinate Verbs (IBSV)
1. When the tiger appeared the bird flew away.
? Did the tiger appear? Y

2. When the boy sneezed the girl walked to the door.
? Did the girl sneeze? N

3. When the girl screamed the boy ran away from the dog.
? Did the boy run away? Y

4. When the child fell the policeman stopped and helped him up.
? Did the child stop? N

5. While the waiter worked the cook prepared more meals.
? Did the cook prepare more meals? Y

6. When the baby smiled the nurse clapped her hands.
? Did the baby clap his hands? N

7. After the doctor laughed the woman jumped to her feet.
? Did the doctor laugh? Y

8. After the girl coughed the doctor gave her a pill.
? Did the doctor cough? N

9. While the dog slept the cat drank its milk.
? Did the cat drink the milk? Y

10. When the customer complained the salesman called the boss.
? Did the customer call the boss? N

Transitively-biased Subordinate Verbs (TBSV)

11. While the boy drank the milk got warm and the food got cold.
? Did the boy drink? Y

12. While the girl ate the ice-cream melted and ran down the side of the bowl.
? Did the girl eat the ice-cream? N

13. After the woman cleaned the stove began to heat up.
? Did the stove heat up? Y

14. While the mailman walked the dog was barking and jumping up and down.
? Did the mailman walk the dog? N

15. After the player hit the ball went out of the field.
? Did the ball go out of the field? Y

16. Before the man parked the car lost power and went on the curb.
? Did the man park the car? N

17. When the boy played the saxophone sounded a lot better.
? Did the boy play? Y

18. While the assistant baked the cookies were tasted by the chef.
? Did the assistant bake the cookies? N

19. As the man drove the truck crashed into the store.
? Did the truck crash into the store? Y

20. When the child asked the teacher answered right away.
? Did the child ask the teacher? N

APPENDIX I

EXPERIMENT II – TARGET ITEMS

INTERMEDIATE GAPS CONDITION

1. The manager who the employee determined that the computer program had helped will hire five workers tomorrow.
? The manager will hire five workers next week. N

2. The student who the teacher argued that the new class had encouraged will study chemistry in college.

3. The patient who the doctor assumed that the imported drug had cured will stay in the hospital.
? The doctor assumed the imported drug had cured the patient. Y

4. The mother who the daughter admitted that the difficult decision had worried will leave home for good.

5. The tourist who the guide confessed that the new schedule had angered will return to camp today.
? The guide will return to camp today. N

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

7. The chef who the waiter argued that the wedding party had exhausted will look for another job.
? The chef worked for the wedding party. Y

8. The girl who the policeman assumed that the loud noises had frightened will stop going to school.

9. The singer who the musician admitted that the broken microphone had bothered will perform one more song
? The singer will not perform anymore. N

10. The actress who the journalist confessed that the Russian play had interested will go on stage tonight.

GARDEN PATH CONDITION

11. The manager who the employee heard that the computer program had helped will hire five workers tomorrow.
? The employee heard something about a computer program. Y

12. The student who the teacher understood that the new class had encouraged will study chemistry in college.

13. The patient who the doctor overheard that the imported drug had cured will stay in the hospital.
? The patient will leave the hospital soon. N

14. The mother who the daughter feared that the difficult decision had worried will leave home for good.

15. The tourist who the guide found that the new schedule had angered will return to camp today.
? The new schedule had angered the tourist. Y

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

17. The chef who the waiter heard that the wedding party had exhausted will look for another job.
? The waiter will look for another job. N

18. The girl who the policeman feared that the loud noises had frightened will stop going to school.

19. The singer who the musician found that the broken microphone had bothered will perform one more song
? The broken microphone had bothered the singer. Y

20. The actress who the journalist overheard that the Russian play had interested will go on stage tonight.

NOMINALIZATION CONDITION

21. The manager who the employee's determination about the computer program had helped will hire five workers tomorrow.
? The manager helped the employee. N

22. The student who the teacher's argument about the new class had encouraged will study chemistry in college.

23. The patient who the doctor's assumption about the imported drug had cured will stay in the hospital.
? The drug the patient took was imported. Y

24. The daughter who the mother's admission about the difficult decision had worried will leave home for good.

25. The tourist who the guide's confession about the new schedule had angered will return to camp today.
? The tourist confessed something about the new schedule. N

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

27. The chef who the waiter's argument about the wedding party had exhausted will look for another job.
? The waiter and the chef argued about the wedding party. Y

28. The girl who the policeman's assumption about the loud noises had frightened will stop going to school.

29. The singer who the musician's admission about the broken microphone had bothered will perform one more song.
? The singer will perform three more songs. N

30. The actress who the journalist's confession about the Russian play had interested will go on stage tonight.

APPENDIX J

EXPERIMENT III – TARGET ITEMS

- Condition (a) = PrinC-Match
- Condition (b) = PrinC-Mismatch
- Condition (c) = NoConstraint-Match
- Condition (d) = NoConstraint-Mismatch

1. a.He slowly drank cheap beer while the bachelor sang bad karaoke, but Jake didn't want to go home late.
? Jake was drinking beer. Y

2. b.She slowly drank cheap beer while the bachelor sang bad karaoke, but Kate didn't want to go home late.

3. c.His friends drank cheap beer while the bachelor sang bad karaoke, but Kate preferred to watch the soccer game.
? Kate preferred to watch rugby. N

4. d.Her friends drank cheap beer while the bachelor sang bad karaoke, but Kate preferred to watch the soccer game.

5. a.He quickly opened the window while the fireman entered the building, but Jeffrey was afraid of the black smoke.

6. b.She quickly opened the window while the fireman entered the building, but Alison was afraid of the black smoke.
? The fireman entered the building. Y

7. c.His children opened the window while the fireman entered the building, but Alison couldn't shout or wave for help.

8. d.Her children opened the window while the fireman entered the building, but Alison couldn’t shout or wave for help.
? Alison opened the window. N

9. a.He carefully wrote the receipts while the mechanic finished the job, but Steve was too tired to keep working.
? Steve finished the job. N

10. b.She carefully wrote the receipts while the mechanic finished the job, but Carol was too tired to keep working.

11. c.His employees wrote the receipts while the mechanic finished the job, but Carol couldn't understand what the problem was.
? Carol could not understand the problem. Y

12. d.Her employees wrote the receipts while the mechanic finished the job, but Carol couldn't understand what the problem was.

13. a.He patiently talked to people while the priest opened the doors, but Paul decided to leave before mass started.

14. b.She patiently talked to people while the priest opened the doors, but Liza decided to leave before mass started.
? Liza stayed for the mass. N

15. c.His neighbors talked to people while the priest opened the doors, but Liza didn’t want to wait any longer.

16. d.Her neighbors talked to people while the priest opened the doors, but Liza didn’t want to wait any longer.
? Her neighbors talked to people. Y

17. a.He happily fed the horses while the prince planned the trip, but Bill was eager to finish the job.
? Bill wanted to be done with the job. Y

18. b.She happily fed the horses while the prince planned the trip, but Mary was eager to finish the job.

19. c.His soldiers fed the horses while the prince planned the trip, but Mary didn't want him to leave yet.
? The soldiers planned the trip. N

20. d.Her soldiers fed the horses while the prince planned the trip, but Mary didn't want him to leave yet.

21. a.She quickly left the theater while the ballerina took a break, but Sara did not want to drive home.

22. b.He quickly left the theater while the ballerina took a break, but Frank did not want to drive home.
? The ballerina took a break. Y

23. c.Her dancers left the theater while the ballerina took a break, but Frank was prepared to control the lights.

24. d.His dancers left the theater while the ballerina took a break, but Frank was prepared to control the lights.
? Frank left the theater. N

25. a.She slowly drank iced coffee while the maid cleaned the rooms, but Sheila decided to complain about the service.
? She slowly drank vodka. N

26. b.He slowly drank iced coffee while the maid cleaned the rooms, but Andrew decided to complain about the service.

27. c.Her parents drank iced coffee while the maid cleaned the rooms, but Andrew was not interested in having breakfast.
? Andrew was not interested in breakfast. Y

28. d.His parents drank iced coffee while the maid cleaned the rooms, but Andrew was not interested in having breakfast.

29. a.She patiently prepared the dessert while the sister entertained the kids, but Wendy decided to buy ice-cream for everybody.

30. b.He patiently prepared the dessert while the sister entertained the kids, but Matt decided to buy ice-cream for everybody.
? Matt decided to buy cake for everybody. N

31. c.Her cousins prepared the dessert while the sister entertained the kids, but Matt couldn’t find the big toy box.

32. d.His cousins prepared the dessert while the sister entertained the kids, but Matt couldn’t find the big toy box.
? The sister was in charge of entertaining the kids. Y

33. a.She carefully fixed the dress while the actress learned the lines, but Karen didn't enjoy the job at all.
? Karen was not enjoying the job. Y

34. b.He carefully fixed the dress while the actress learned the lines, but Adam didn't enjoy the job at all.

35. c.Her assistants fixed the dress while the actress learned the lines, but Adam tried to leave the theater early.
? Adam wanted to stay in the theater. N

36. d.His assistants fixed the dress while the actress learned the lines, but Adam tried to leave the theater early.

37. a.She happily cut the cake while the mother served the soda, but Molly decided to ask for more food.

38. b.He happily cut the cake while the mother served the soda, but Tom decided to ask for more food.
? Tom wanted to eat more. Y

39. c.Her relatives cut the cake while the mother served the soda, but Tom did not want to eat anything.

40. d.His relatives cut the cake while the mother served the soda, but Tom did not want to eat anything.
? Tom served the soda. N

APPENDIX K

EXPERIMENT II – RAW REACTION TIMES

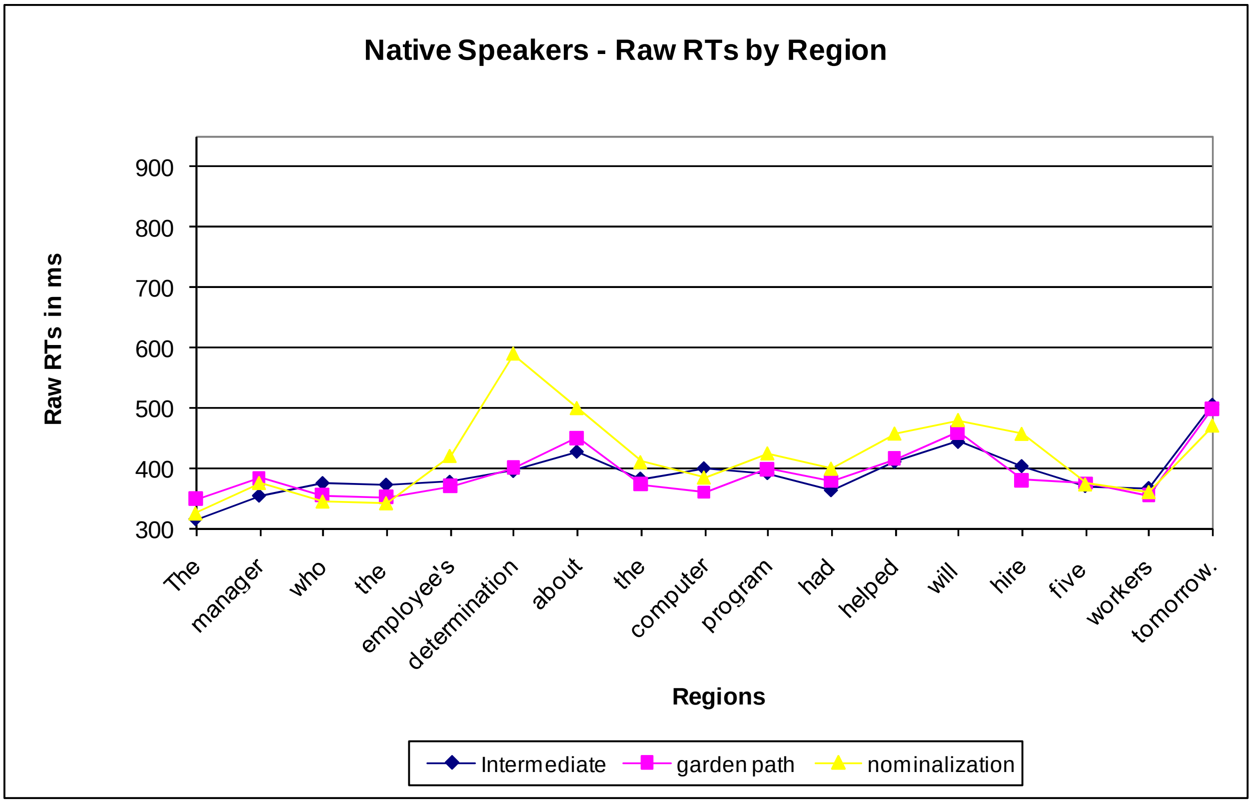


Figure 22. Raw RTs by region - Exp. II - NS

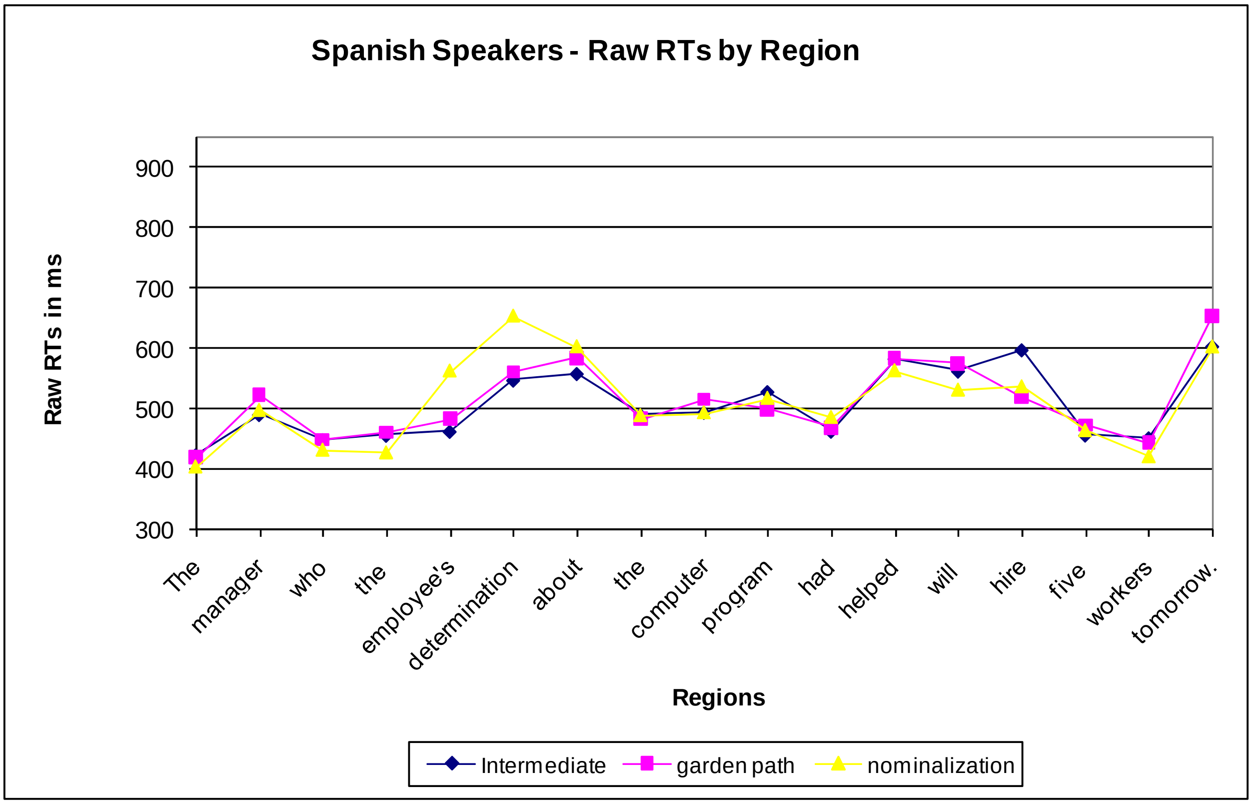


Figure 23. Raw RTs by region - Exp. II - SS

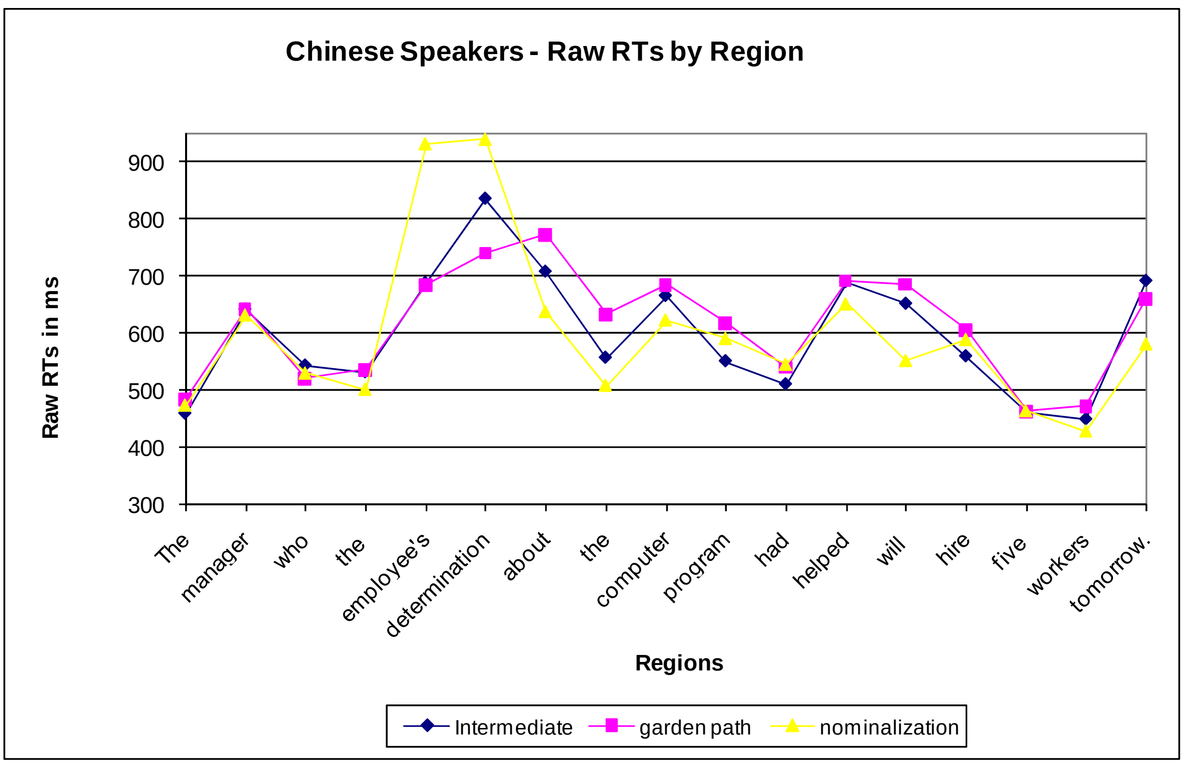


Figure 24. Raw RTs by region - Exp. II - CS

APPENDIX L

EXPERIMENT III – RAW READING TIMES

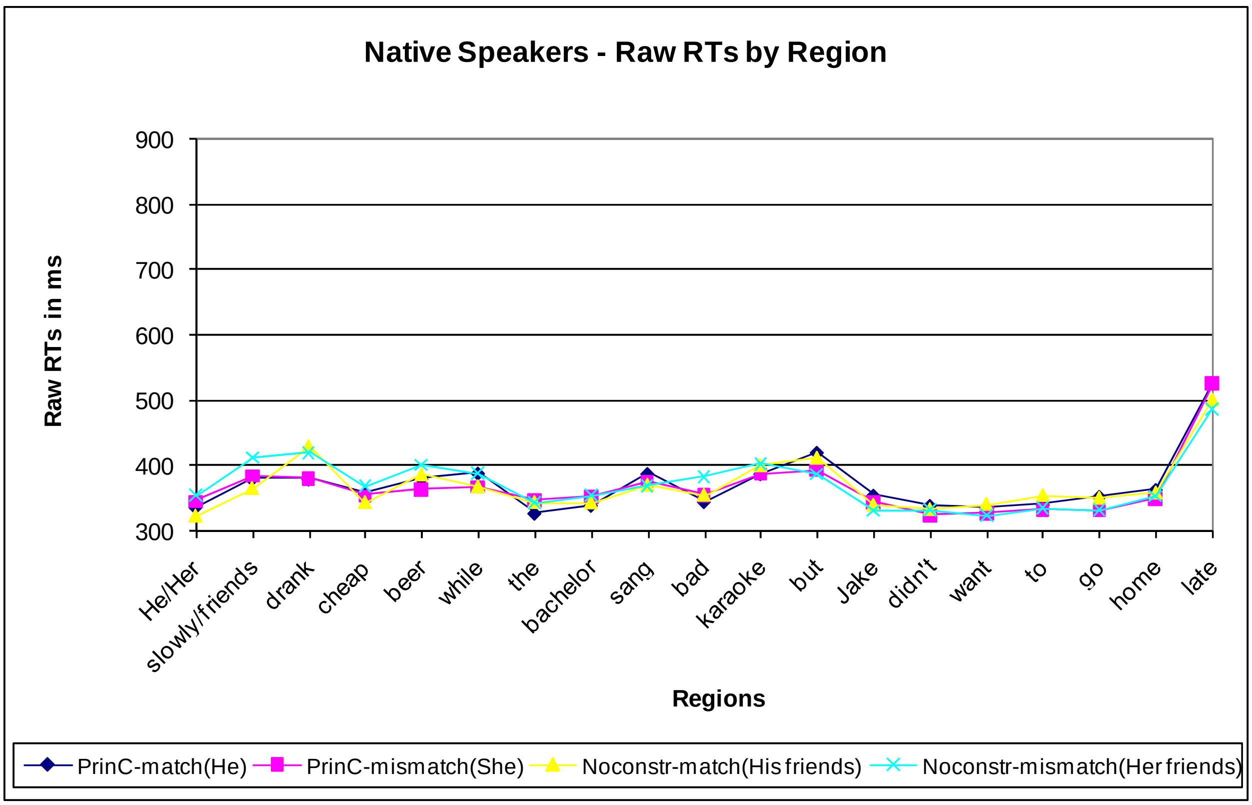


Figure 25. Raw RTs by region - Exp. III - NS

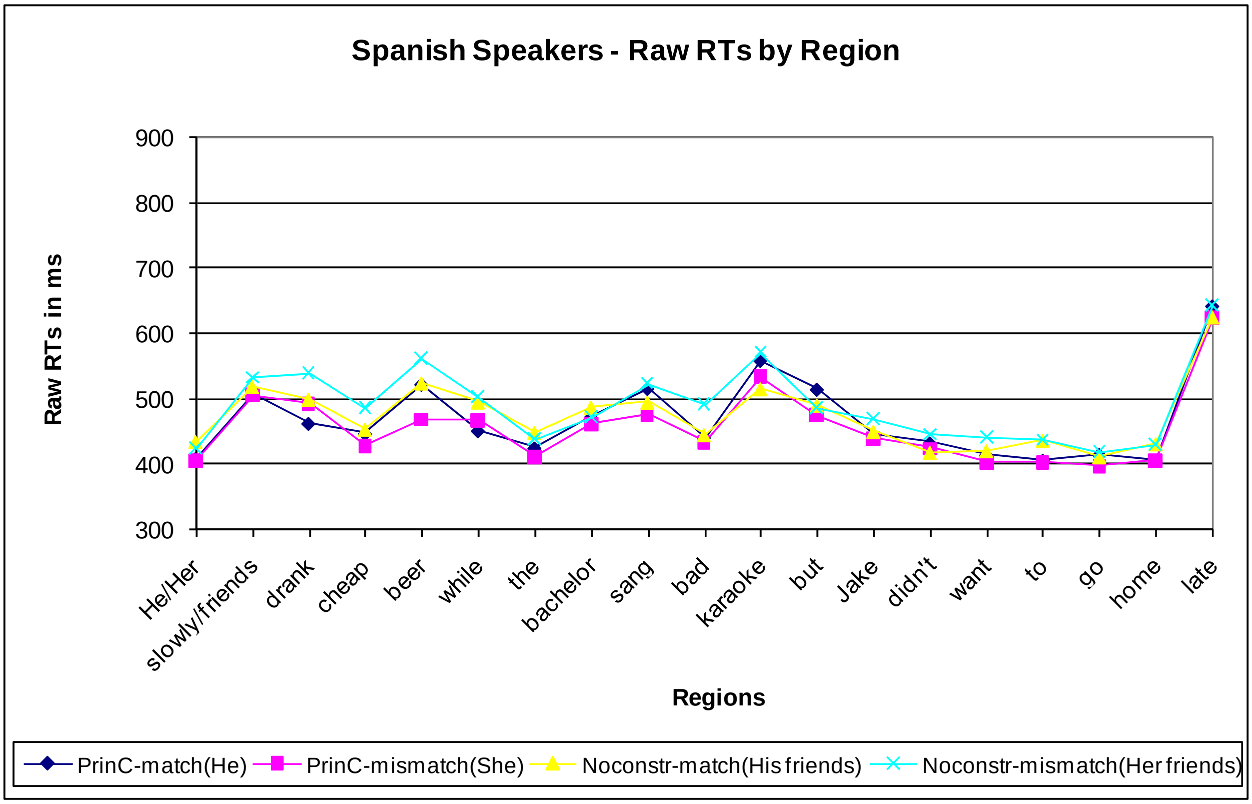


Figure 26. Raw RTs by region - Exp. III - SS

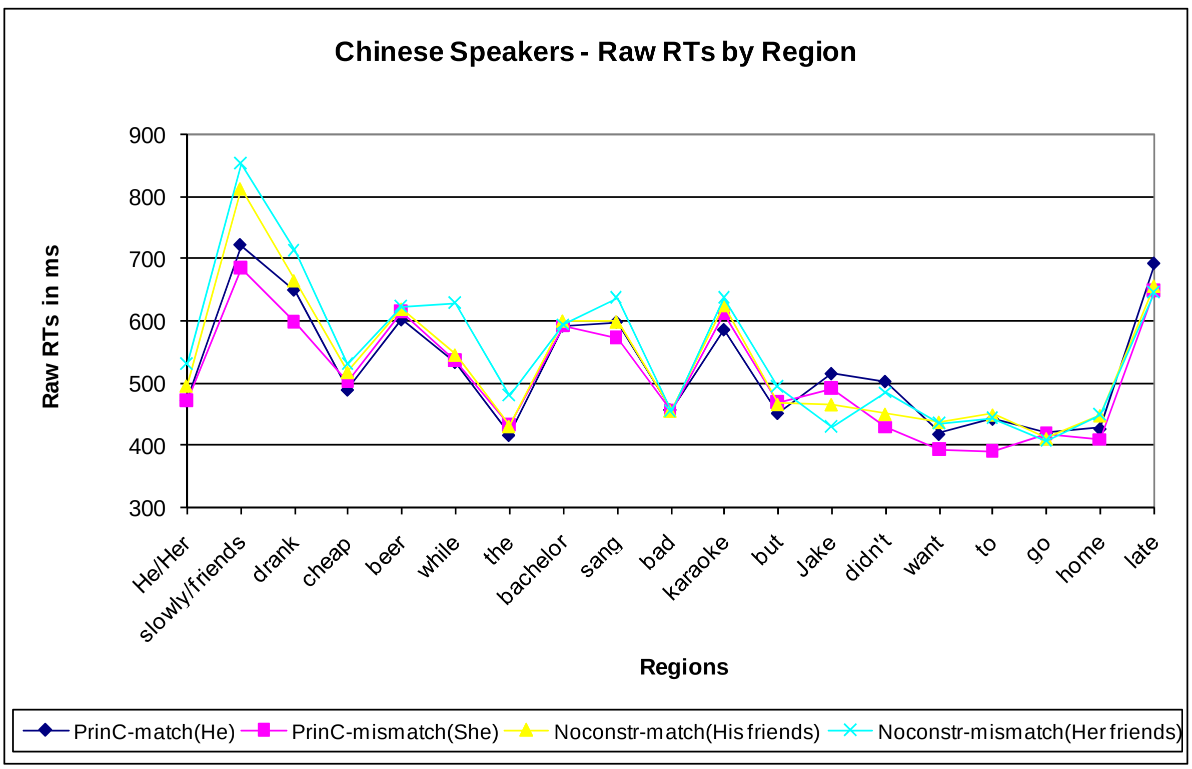


Figure 27. Raw RTs by region - Exp. III - SS

APPENDIX M

NATIVE SPEAKERS – WM AND ACCURACY CORRELATIONS

Table 41. WM and Accuracy Correlations – Exp. 1 – NS - Descriptives

Descriptive Statistics

	Mean	Std. Deviation	N
RSTotal	58.2000	13.77990	15
RSTpc	16.1889	3.07502	15
RSTAcc	92.4000	4.88145	15
SPRacc	90.9260	3.93285	15
SPRaccA	89.3333	10.32796	15
SPRaccB	98.6667	5.16398	15
SPRaccC	86.6667	9.75900	15
SPRaccD	85.3333	20.65591	15
SPRaccIG	85.3333	14.07463	15
SPRaccGP	85.3333	20.65591	15
SPRaccNOM	69.3333	18.30951	15

Table 42. WM and accuracy correlations - Exp. I - NS

Correlations

		RSTotal	RSTpc	RSTAcc	SPRacc	SPRaccA	SPRaccB	SPRaccC	SPRaccD	SPRaccIG	SPRaccGP	SPRaccNOM
RSTotal	Pearson Correlation	1	.996**	-.024	.290	.046	.165	.011	.222	.421	-.200	-.382
	Sig. (2-tailed)		.000	.934	.295	.870	.558	.970	.427	.118	.475	.160
	N	15	15	15	15	15	15	15	15	15	15	15
RSTpartial credit	Pearson Correlation	.996**	1	-.002	.283	.027	.185	-.001	.177	.433	-.211	-.366
	Sig. (2-tailed)	.000		.995	.307	.925	.509	.996	.528	.107	.451	.179
	N	15	15	15	15	15	15	15	15	15	15	15
RSTAccuracy	Pearson Correlation	-.024	-.002	1	-.054	-.023	-.261	.330	-.164	-.012	.544	.019
	Sig. (2-tailed)	.934	.995		.848	.936	.348	.230	.558	.965	.360	.946
	N	15	15	15	15	15	15	15	15	15	15	15
SPRoverallaccuracy	Pearson Correlation	.290	.283	-.054	1	.047	.544*	.188	.520*	.113	.313	.126
	Sig. (2-tailed)	.295	.307	.848		.867	.036	.503	.047	.689	.255	.654
	N	15	15	15	15	15	15	15	15	15	15	15

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

APPENDIX N

SPANISH SPEAKERS – WM AND ACCURACY CORRELATIONS

Table 43. WM and Accuracy Correlations – Exps. II & III – SS - Descriptives

Descriptive Statistics

	Mean	Std. Deviation	N
RSTotal	62.0000	9.14695	15
RSTpc	16.6885	1.99608	15
RSTAcc	88.9231	3.09466	15
PROFIC90	95.0000	3.55903	15
SPRacc	90.6285	4.05821	15
SPRaccA	80.0000	8.16497	15
SPRaccB	98.4615	5.54700	15
SPRaccC	90.7692	10.37749	15
SPRaccD	89.2308	15.52500	15
SPRaccIG	89.2308	15.52500	15
SPRaccGP	75.3846	20.25479	15
SPRaccNOM	66.1538	25.01282	15

Table 44. WM and Accuracy Correlations – Exp. 1I & III – SS

Correlations

		RSTotal	RSTpc	RSTAcc	PROFIC	SPRacc	SPRaccA	SPRaccB	SPRaccC	SPRaccD	SPRaccIG	SPRaccGP	SPRaccNOM
RSTotal	Pearson Correlation	1.000	.983**	.118	.438	-.062	.201	-.394	.070	-.047	-.575	.198	.204
	Sig. (2-tailed)		.000	.702	.135	.840	.511	.183	.820	.879	.140	.517	.504
	N	15	15	15	15	15	15	15	15	15	15	15	15
RSTpartial credit	Pearson Correlation	.983**	1.000	.033	.491	-.070	.131	-.408	-.008	-.067	-.570	.260	.165
	Sig. (2-tailed)	.000		.915	.088	.820	.669	.166	.979	.828	.142	.390	.590
	N	15	15	15	15	15	15	15	15	15	15	15	15
RSTaccuracy	Pearson Correlation	.118	.033	1.000	.401	.273	.792	-.105	.547	.502	-.088	.419	-.036
	Sig. (2-tailed)	.702	.915		.174	.367	.321	.734	.053	.081	.775	.154	.906
	N	15	15	15	15	15	15	15	15	15	15	15	15
PROFICIENCY	Pearson Correlation	.438	.491	.401	1.000	.626*	.459	-.338	.226	.664*	.060	.763**	.243
	Sig. (2-tailed)	.135	.088	.174		.022	.115	.259	.459	.013	.845	.002	.423
	N	15	15	15	15	15	15	15	15	15	15	15	15
SPRoverallaccuracy	Pearson Correlation	-.062	-.070	.273	.626*	1.000	.254	-.010	.302	.732**	.554*	.674*	.574*
	Sig. (2-tailed)	.840	.820	.367	.022		.402	.975	.317	.004	.049	.012	.040
	N	15	15	15	15	15	15	15	15	15	15	15	15

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

APPENDIX O

NATIVE SPEAKERS – WM AND SPR REGIONS CORRELATIONS

Table 45. WM and RTs correlations - Exps. II & III – NS - Descriptives

Descriptive Statistics

	Mean	Std. Deviation	N
RSTotal	58.2000	13.77990	15
RSTpc	16.1889	3.07502	15
RSTAcc	92.4000	4.88145	15
CaRe07	-39.1593	33.27766	15
CaRe08	-64.0809	43.39365	15
CaRe09	-18.2443	57.80967	15
CaRe10	-19.2220	24.69267	15
CaRe11	-8.4287	75.02656	15
CbRe07	-20.3608	32.71232	15
CbRe08	-51.6702	38.78034	15
CbRe09	-27.7015	66.94445	15
CbRe10	-8.1628	58.34540	15
CbRe11	-9.6673	66.99619	15
CcRe07	-22.8445	31.59747	15
CcRe08	-61.5193	55.58491	15
CcRe09	-33.6843	42.13103	15
CcRe10	-9.3503	41.77247	15
CcRe11	2.6385	76.28260	15
CdRe07	-23.1644	37.14842	15
CdRe08	-49.6695	41.23695	15
CdRe09	-34.0915	38.14169	15
CdRe10	18.6216	51.13238	15
CdRe11	5.9266	74.01531	15
CIGRegion2	-16.7617	42.90389	15
CIGRegion3	50.4123	126.13194	15
CIGRegion5	-6.0660	30.39951	15
CIGRegion6	44.7666	70.16279	15
CGPRegion2	-20.4619	41.35774	15
CGPRegion3	73.8662	105.51956	15
CGPRegion5	5.3377	56.60421	15
CGPRegion6	42.3443	54.24398	15
CNOMRegion2	39.9579	71.74515	15
CNOMRegion3	112.8552	69.96055	15
CNOMRegion5	35.5210	80.23770	15
CNOMRegion6	90.5614	116.63098	15

Table 46. WM & RTs correlations – Exps. II & III - NS

Correlations

		RSTotal	RSTpc	RSTAcc
RSTotal	Pearson Correlation	1	.996**	-.024
	Sig. (2-tailed)		.000	.934
	N	15	15	15
RSTpc	Pearson Correlation	.996**	1	-.002
	Sig. (2-tailed)	.000		.995
	N	15	15	15
RSTAcc	Pearson Correlation	-.024	-.002	1
	Sig. (2-tailed)	.934	.995	
	N	15	15	15
CaRe07	Pearson Correlation	-.026	-.051	.099
	Sig. (2-tailed)	.928	.857	.727
	N	15	15	15
CaRe08	Pearson Correlation	-.104	-.126	-.072
	Sig. (2-tailed)	.711	.654	.798
	N	15	15	15
CaRe09	Pearson Correlation	.025	.033	.247
	Sig. (2-tailed)	.931	.908	.375
	N	15	15	15
CaRe10	Pearson Correlation	-.362	-.335	.209
	Sig. (2-tailed)	.185	.223	.454
	N	15	15	15
CaRe11	Pearson Correlation	.458	.448	.279
	Sig. (2-tailed)	.086	.094	.313
	N	15	15	15
CbRe07	Pearson Correlation	-.412	-.400	.174
	Sig. (2-tailed)	.127	.139	.536
	N	15	15	15
CbRe08	Pearson Correlation	-.536	-.529	.099
	Sig. (2-tailed)	.400	.420	.725
	N	15	15	15
CbRe09	Pearson Correlation	-.290	-.306	-.075
	Sig. (2-tailed)	.294	.267	.791
	N	15	15	15
CbRe10	Pearson Correlation	-.258	-.231	.239
	Sig. (2-tailed)	.352	.408	.392
	N	15	15	15
CbRe11	Pearson Correlation	.330	.300	-.002
	Sig. (2-tailed)	.230	.278	.995
	N	15	15	15
CcRe07	Pearson Correlation	-.201	-.189	.591
	Sig. (2-tailed)	.473	.501	.200
	N	15	15	15
CcRe08	Pearson Correlation	-.276	-.296	-.077
	Sig. (2-tailed)	.320	.284	.784
	N	15	15	15
CcRe09	Pearson Correlation	-.469	-.453	-.228
	Sig. (2-tailed)	.078	.090	.414
	N	15	15	15
CcRe10	Pearson Correlation	-.295	-.283	.095
	Sig. (2-tailed)	.286	.306	.736
	N	15	15	15
CcRe11	Pearson Correlation	.057	.069	.470
	Sig. (2-tailed)	.841	.808	.077
	N	15	15	15
CdRe07	Pearson Correlation	-.257	-.291	-.148
	Sig. (2-tailed)	.356	.292	.600
	N	15	15	15
CdRe08	Pearson Correlation	-.146	-.160	-.176
	Sig. (2-tailed)	.602	.569	.531
	N	15	15	15
CdRe09	Pearson Correlation	.077	.045	-.030
	Sig. (2-tailed)	.784	.872	.916
	N	15	15	15
CdRe10	Pearson Correlation	-.094	-.074	-.048
	Sig. (2-tailed)	.740	.794	.866
	N	15	15	15
CdRe11	Pearson Correlation	.801	.780	-.015
	Sig. (2-tailed)	.316	.425	.958
	N	15	15	15
CIGRegion2	Pearson Correlation	-.130	-.127	.409
	Sig. (2-tailed)	.645	.652	.130
	N	15	15	15
CIGRegion3	Pearson Correlation	.167	.178	.035
	Sig. (2-tailed)	.552	.526	.903
	N	15	15	15
CIGRegion5	Pearson Correlation	.134	.127	-.605
	Sig. (2-tailed)	.635	.651	.170
	N	15	15	15
CIGRegion6	Pearson Correlation	.564	.575	.307
	Sig. (2-tailed)	.280	.200	.265
	N	15	15	15
CGPRegion2	Pearson Correlation	-.134	-.119	.446
	Sig. (2-tailed)	.633	.673	.095
	N	15	15	15
CGPRegion3	Pearson Correlation	.301	.308	.021
	Sig. (2-tailed)	.276	.264	.941
	N	15	15	15
CGPRegion5	Pearson Correlation	.157	.153	-.381
	Sig. (2-tailed)	.576	.586	.161
	N	15	15	15
CGPRegion6	Pearson Correlation	.433	.441	-.195
	Sig. (2-tailed)	.107	.100	.487
	N	15	15	15
CNOMRegion2	Pearson Correlation	.008	.015	-.320
	Sig. (2-tailed)	.978	.957	.244
	N	15	15	15
CNOMRegion3	Pearson Correlation	-.099	-.103	-.478
	Sig. (2-tailed)	.725	.715	.072
	N	15	15	15
CNOMRegion5	Pearson Correlation	.331	.322	-.154
	Sig. (2-tailed)	.228	.241	.584
	N	15	15	15
CNOMRegion6	Pearson Correlation	.318	.323	-.315
	Sig. (2-tailed)	.248	.241	.253
	N	15	15	15

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

APPENDIX P

SPANISH SPEAKERS – WM AND SPR REGIONS CORRELATIONS

Table 47. WM and RTs correlations - Exps. II & III – SP - Descriptives

Descriptive Statistics

	Mean	Std. Deviation	N
RSTotal	62.0000	9.14695	15
RSTpc	16.6885	1.99608	15
RSTAcc	88.9231	3.09466	15
PROFIC90	95.0000	3.55903	15
CaRe07	-36.3969	30.36069	15
CaRe08	-46.3432	40.34303	15
CaRe09	-23.4703	61.57511	15
CaRe10	-13.3176	52.13678	15
CaRe11	47.4106	127.06793	15
CbRe07	-47.6707	48.43364	15
CbRe08	-70.8980	68.36636	15
CbRe09	-59.8928	78.78556	15
CbRe10	-24.7457	30.66914	15
CbRe11	12.1902	104.82262	15
CcRe07	-9.2101	60.52198	15
CcRe08	-62.0332	55.86735	15
CcRe09	-42.2030	50.89769	15
CcRe10	-13.6420	50.30814	15
CcRe11	-9.1659	49.02067	15
CdRe07	-26.8322	70.11296	15
CdRe08	-55.6943	47.38127	15
CdRe09	-12.6155	83.44040	15
CdRe10	38.1247	47.06233	15
CdRe11	46.0989	107.51219	15
CIGRegion2	-31.7436	54.68249	15
CIGRegion3	76.7440	112.57023	15
CIGRegion5	17.1275	53.91266	15
CIGRegion6	106.6802	104.55751	15
CGPRegion2	-13.6424	59.72622	15
CGPRegion3	92.9005	120.84182	15
CGPRegion5	19.1489	49.06169	15
CGPRegion6	69.7728	81.56027	15
CNOMRegion2	15.0336	56.08561	15
CNOMRegion3	118.3118	172.39610	15
CNOMRegion5	24.8917	66.48193	15
CNOMRegion6	66.8631	47.91097	15

Table 48. WM and RTs correlations - Exps. II & III - SS

Correlations

		RSTotal	RSTpc	RSTAcc	PROFIC
RSTotal	Pearson Correlation	1.000	.983**	.118	.438
	Sig. (2-tailed)		.000	.702	.135
	N	15	15	15	15
RSTpartialcredit	Pearson Correlation	.983**	1.000	.033	.491
	Sig. (2-tailed)	.000		.915	.088
	N	15	15	15	15
RSTAccuracy	Pearson Correlation	.118	.033	1.000	.401
	Sig. (2-tailed)	.702	.915		.174
	N	15	15	15	15
PROFICIENCY	Pearson Correlation	.438	.491	.401	1.000
	Sig. (2-tailed)	.135	.088	.174	
	N	15	15	15	15
CaRe07	Pearson Correlation	-.175	-.102	-.523	-.404
	Sig. (2-tailed)	.567	.741	.067	.171
	N	15	15	15	15
CaRe08	Pearson Correlation	-.299	-.329	-.304	-.514
	Sig. (2-tailed)	.321	.273	.313	.072
	N	15	15	15	15
CaRe09	Pearson Correlation	.289	.255	.053	.326
	Sig. (2-tailed)	.337	.400	.864	.277
	N	15	15	15	15
CaRe10	Pearson Correlation	-.449	-.336	-.590	-.139
	Sig. (2-tailed)	.124	.262	.340	.650
	N	15	15	15	15
CaRe11	Pearson Correlation	.054	.012	.147	.457
	Sig. (2-tailed)	.861	.969	.631	.116
	N	15	15	15	15
CbRe07	Pearson Correlation	.051	.159	-.478	-.041
	Sig. (2-tailed)	.868	.603	.099	.895
	N	13	13	13	13
CbRe08	Pearson Correlation	.274	.332	-.126	.022
	Sig. (2-tailed)	.365	.267	.681	.944
	N	15	15	15	15
CbRe09	Pearson Correlation	-.269	-.248	-.363	-.316
	Sig. (2-tailed)	.374	.413	.222	.293
	N	15	15	15	15
CbRe10	Pearson Correlation	-.176	-.128	-.307	.098
	Sig. (2-tailed)	.565	.677	.307	.750
	N	15	15	15	15
CbRe11	Pearson Correlation	.140	.119	.010	.462
	Sig. (2-tailed)	.648	.698	.974	.112
	N	15	15	15	15
CcRe07	Pearson Correlation	-.115	-.015	-.467	-.006
	Sig. (2-tailed)	.708	.961	.108	.984
	N	15	15	15	15
CcRe08	Pearson Correlation	.539	.607	-.229	.366
	Sig. (2-tailed)	.058	.280	.451	.218
	N	15	15	15	15
CcRe09	Pearson Correlation	.456	.524	-.056	.384
	Sig. (2-tailed)	.118	.066	.856	.195
	N	15	15	15	15
CcRe10	Pearson Correlation	.498	.565	-.095	.548
	Sig. (2-tailed)	.083	.440	.757	.052
	N	15	15	15	15
CcRe11	Pearson Correlation	.158	.216	-.006	.553*
	Sig. (2-tailed)	.605	.479	.985	.050
	N	15	15	15	15
CdRe07	Pearson Correlation	-.056	.028	-.133	-.075
	Sig. (2-tailed)	.856	.928	.665	.808
	N	15	15	15	15
CdRe08	Pearson Correlation	.313	.229	.286	-.095
	Sig. (2-tailed)	.298	.452	.344	.756
	N	15	15	15	15
CdRe09	Pearson Correlation	.400	.331	.258	.218
	Sig. (2-tailed)	.176	.269	.395	.475
	N	15	15	15	15
CdRe10	Pearson Correlation	-.323	-.317	-.037	-.388
	Sig. (2-tailed)	.282	.291	.906	.190
	N	15	15	15	15
CdRe11	Pearson Correlation	.089	.039	.525	.178
	Sig. (2-tailed)	.773	.899	.065	.560
	N	15	15	15	15
CIGRegion2	Pearson Correlation	.045	.005	.013	-.328
	Sig. (2-tailed)	.884	.987	.966	.274
	N	15	15	15	15
CIGRegion3	Pearson Correlation	.195	.112	.338	-.061
	Sig. (2-tailed)	.523	.716	.258	.843
	N	15	15	15	15
CIGRegion5	Pearson Correlation	-.294	-.391	.090	-.170
	Sig. (2-tailed)	.330	.186	.770	.580
	N	15	15	15	15
CIGRegion6	Pearson Correlation	.102	.044	.328	.018
	Sig. (2-tailed)	.740	.887	.273	.954
	N	15	15	15	15
CGPRegion2	Pearson Correlation	.423	.399	.272	.212
	Sig. (2-tailed)	.150	.177	.368	.487
	N	15	15	15	15
CGPRegion3	Pearson Correlation	-.031	-.040	-.370	-.319
	Sig. (2-tailed)	.920	.897	.213	.287
	N	15	15	15	15
CGPRegion5	Pearson Correlation	.322	.327	.197	.078
	Sig. (2-tailed)	.283	.275	.519	.800
	N	15	15	15	15
CGPRegion6	Pearson Correlation	-.057	.030	-.210	.298
	Sig. (2-tailed)	.854	.921	.491	.323
	N	15	15	15	15
CNOMRegion2	Pearson Correlation	-.174	-.197	-.484	-.355
	Sig. (2-tailed)	.570	.520	.094	.233
	N	15	15	15	15
CNOMRegion3	Pearson Correlation	.377	.252	.342	-.024
	Sig. (2-tailed)	.205	.406	.252	.937
	N	15	15	15	15
CNOMRegion5	Pearson Correlation	-.130	-.054	.011	.262
	Sig. (2-tailed)	.673	.860	.971	.388
	N	15	15	15	15
CNOMRegion6	Pearson Correlation	-.235	-.260	.035	-.123
	Sig. (2-tailed)	.440	.391	.909	.688
	N	15	15	15	15

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

BIBLIOGRAPHY

Adams, B.C., Clifton,C., Jr.,& Mitchell, D.C. (1998). Lexical guidance in sentence processing? *Psychonomic Bulletin and Review, 5*, 265–270.

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

Baddeley, A. D. (1986). *Working memory.* London: Oxford University Press.

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders 36*(3), 189-208.

Baddeley, A. (2000). Short-term and working memory. In Tulving, E. and Craik, F. (Eds). *The Oxford handbook of memory*. New York: Oxford University Press, 77–92.

Baddeley, A., Gathercole, S. and Papagno, C. (1998). The phonological loop as a language learning device. *The Psychological Review 105*, 158–73.

Balogh, J., Zurif, E., Prather, P., Swinney, D. and Finkel, L. 1998: Gapfilling and end-of sentence effects in real-time language processing: Implications for modeling sentence comprehension in aphasia. *Brain and Language 61,* 169–82.

Bever, T.G. (1970). The cognitive basis for linguistic structures. In Hayes, J.R. (Ed). *Cognition and the development of language.* New York: Wiley, 279-362.

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

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

Carroll, S. 2001: *Input and evidence: the raw material of second language acquisition*. Amsterdam: Benjamins.

Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary learning. *Developmental Psychology, 32*, 867–873.

Chomsky, N. (1973) Conditions on transformations. In Anderson, S.R. & Kiparsky, P. (Eds). *A festschrift for Morris Halle*. New York: Holt, Rinehart, & Winston, 232-286.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

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

Clifton, C., & Frazier, L. (1989). Comprehending sentences with long-distance dependencies. In G. M. Carlson & M. K. Tanenhaus (Eds.). *Linguistic structure in language processing.* Dordrecht: Kluwer, 273–317.

Conway, A. R. A., 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 and Review 12*, 769-786.

Corrigan, A., Dobson, B., Kellman, E., Palmer, A., Peterson, J., Spaan, M., Strowe, L., Tyma, S. and Upshur, J. (1978). *English placement test*. Ann Arbor, MI: University of Michigan English Language Institute.

Crain, S. and Fodor, J.D. (1985). How can grammars help parsers? In Dowty, D., Karttunen, L. and Zwicky, A. (Eds). *Natural language parsing: psychological, computational and theoretical perspectives.* Cambridge: Cambridge University Press, 94–128.

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

Christianson, K., Williams, C., Zacks, R. & Ferreira, F. (2006). Younger and older adults' "good-enough" interpretations of garden-path sentences. *Discourse Processes 42*(2), 205-238.

Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition 30,* 73–105.

Daneman, M. and Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior 19*, 450–66.

De Vincenzi, M. (1991). Filler-gap dependencies in a null subject language: Referential and nonreferential WHs. *Journal of Psycholinguistic Research, 20*(3), 197-213.

DeKeyser, R. (2001). Automaticity and automatization. In Robinson, P. (Ed). *Cognition and Second Language Instruction*. New York: Cambridge University Press, 125–151.

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition, 18*(1), 91-126.

Ellis, N. (2001). Memory for language. In Robinson, P. (Ed). *Cognition and second language instruction*, New York: Cambridge University Press, 33–68.

Ellis, N. C., & Sinclair, S. G. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology, 49*(1), 234-250.

Felser, C, & Roberts, L (2007). Processing wh-dependencies in English as a second language: A crossmodal priming study. *Second Language Research*, *23*, 1–28.

Felser, C., Roberts, L. and Marinis, T. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics 24*, 453–89.

Fender, M. (2001). A review of L1 and L2/ESL word integration skills and the nature of L2/ESL word integration development involved in lower-level text processing. *Language Learning, 51*(2), 319-396.

Fernández, E. M. (2006). How do language learners build syntactic structure on-line? *Applied Psycholinguistics, 27*(1), 59-64.

Fernández, E. 1999: Processing strategies in second language acquisition: some preliminary results. In Klein, E. and Martohardjono G. (Eds). *The development of second language grammars: a generative approach*. Amsterdam: Benjamins, 217–39.

Ferreira, F. and Clifton, C. Jr. (1986). The Independence of syntactic processing. *Journal of Memory and Language 25*, 348–68.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*(1), 11-15.

Fiebach, C.J., Schlesewsky, M. and Friederici, A.D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: the processing of German *wh*-questions. *Journal of Memory and Language 47*, 250–72.

Fodor, J. D. (1989). Empty categories in sentence processing. *Language and Cognitive Processes, 4*, 155–209.

Fodor, J.D. 1998: Learning to parse? *Journal of Psycholinguistic Research 27*, 285–319.

Foss, D. & Hakes, D. (1978). *Psycholinguistics: An introduction to the psychology of language.* Englewood Cliffs, N.J.: Prentice-Hall.

Frazier, L. (1987). Theories of sentence processing. In Garfield, J. (Ed). *Modularity in knowledge representation and natural language understanding*. Cambridge, MA: MIT Press, 291–307.

Frazier, L. and Clifton, C. (1996). *Construal*. Cambridge, MA: MIT Press.

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

Frenck-Mestre, C. (2005). Ambiguities and anomalies: What can eye movements and event-related potentials reveal about second language processing? In J. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism*. Oxford: Oxford University Press, 268-281.

Frenck-Mestre, C. and Pynte, J. (1997) Syntactic ambiguity resolution while reading in second and native languages. *Quarterly Journal of Experimental Psychology 50A*, 119–48.

Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language, 51*(1), 136-158.

Garnsey, S., Pearlmutter, N., Myers, E., & Lotocky, M. (1997). The contribution of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, *37*, 58–93.

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition, 23*(1), 83-94.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences, 3*(11), 410-419.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*(4), 513-613.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language.* Hove: Erlbaum.

Gibson, E. (1998). Syntactic complexity: Locality of syntactic dependencies. *Cognition*, *68*, 1–75.

Gibson, E. and Hickok, G. (1993). Sentence processing with empty categories. *Language and Cognitive Processes* 8, 147–61.

Gibson, E. and Schütze, C.T. (1999). Disambiguation preferences in noun phrase conjunction do not mirror corpus frequency. *Journal of Memory and Language* 40, 263–79.

Gibson, E., & Warren, T. (2004). Reading-time evidence for intermediate linguistic structure in long distance dependencies. *Syntax*, *7*, 55–78.

Gibson, E., Desmet, T., Watson, D., Grodner, D., & Ko, K. (2005). Reading relative clauses in English. *Cognitive Linguistics, 16,* 313–353.

Gilboy, E., Sopena, J.M., Clifton, C. and Frazier, L. (1995). Argument structure and association preferences in Spanish and English compound NPs. *Cognition 54,* 131–67.

Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language, 51*(1), 97-114.

Gregg, K. R. (2003). The state of emergentism in second language acquisition. *Second Language Research, 19*(2), 95-128.

Harrington, M. (1987). Processing transfer: Language-specific processing strategies as a source of interlanguage variation. *Applied Psycholinguistics, 8*(4), 351-377.

Harrington, M. (2001). Sentence processing. In Robinson, P. (Ed.), *Cognition and second language instruction*. Cambridge: Cambridge University Press, 91-124.

Harrington, M. and Sawyer, M. (1992). L2 working memory capacity and L2 reading skills. *Studies in Second Language Acquisition 14*, 25–38.

Hemforth, B., Konieczny, L., & Scheepers, C. (2000). Syntactic attachment and anaphor resolution: The two sides of relative clause attachment. In Crocker, M. W., M. Pickering & C. Clifton Jr. (Eds.). *Architectures and mechanisms for language processing*. Cambridge: Cambridge University Press, 259-281.

Hsiao, F. and Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition 90*, 3–27.

Huang, C.-T.J. (1982). Move *wh* in a language without *wh* movement. *Linguistic Review* 1, 369–416.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning, 57*(1), 1-33.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics, 25*(4), 603-634.

Juffs, A. (1998). Main verb vs. reduced relative clause ambiguity resolution in second language sentence processing. *Language Learning* 48, 107–47.

Juffs, A. (2001). Psycholinguistically-oriented second language research. *Annual Review of Applied Linguistics* 21, 207–23.

Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society, 102*, 199–226.

Juffs, A. (2005). The influence of first language on the processing of wh-movement in English as a second language. *Second Language Research* 21 (2), 121-151.

Juffs, A. and Harrington, M. (1995). Parsing effects in L2 sentence processing: subject and object asymmetries in *wh*-extraction. *Studies in Second Language Acquisition 17*, 483–516.

Juffs, A., & Harrington, M. (1996). Garden path sentences and error data in second language sentence processing. *Language Learning*, *46*, 283–326.

Juffs, A. and Rodríguez, G. (2006). Second language sentence processing and working memory in college-educated and low-educated learners of English as a second language. Paper presented at EuroSLA, Antalya, Turkey.

Just, M.A. and Carpenter, P. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review 99*, 122–49.

Just, M.A., Carpenter, P. and Keller, T. (1996). The capacity theory of comprehension: new frontiers of evidence and arguments. *The Psychological Review 103*, 773–80.

Just, M.A., Carpenter, P. and Woolley, J.D. (1982). Paradigms and processes and in reading comprehension. *Journal of Experimental Psychology: General* 3, 228–238.

Just, M.A. and Varma, S. (2002). A hybrid architecture for working memory: reply to MacDonald and Christianson 2002. *Psychological Review 109*, 55–65.

Kazanina, N., Lau, E., Lieberman, M., Yoshida, M. & Phillips, C. (2007). The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language 56*, 384–409.

Kennison, Sheila M. (1999). American English usage frequencies for noun phrase and tensed sentence complement-taking verbs. *Journal of Psycholinguistic Research 28*, 165–77.

Kilborn, K. (1989). Sentence processing in a second language: The timing of transfer. *Language and Speech 32,* 1–23.

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

Klein, E. and Martohardjono, R. (1999). Investigating second language grammars: some conceptual and methodological issues in generative SLA research. In Klein, E. and Martohardjono, G. (Eds), *The development of second language grammars: a generative approach*. Amsterdam: John Benjamins, 3–34.

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

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.

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

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

MacDonald, M.C. and Christiansen, M.H. (2002). Reassessing working memory: comment on Just and Carpenter 1992 and Waters and Caplan 1996. *Psychological Review 109*, 35–54.

Mackey, A., Philp, J., Egi, T., Fujii, A. and Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. In Robinson, P. (Ed). *Individual differences and instructed language learning*. Philadelphia: John Benjamins, 181–209.

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

Masoura, E., & Gathercole, S. E. (1999). Phonological short-term memory and foreign vocabulary learning. *International Journal of Psychology 34*, 383–388.

Masoura, E. V., & Gathercole, S. E. (2005). Phonological short-term memory skills and new word learning in young Greek children. *Memory, 13*, 422–429.

Mitchell, D. C. (1987). Lexical guidance in human parsing: Locus and processing characteristics. In M. Coltheart (Ed.). *Attention and Performance XII*. Hillsdale, NJ: Erlbaum, 601-618.

Mitchell, D. (1994). Sentence parsing. In Gernsbacher, M.A. (Ed). *Handbook of psycholinguistics*. New York: Academic Press, 375-409.

Myles, F., Hooper, J. and Mitchell, R. (1998). Rote or rule? Exploring the role of formulaic language in the foreign language classroom. *Language Learning 48*, 323–364.

Myles, F., Mitchell, R. and Hooper, J. (1999). Interrogative chunks in French L2: a basis for creative construction? *Studies in Second Language Acquisition 21*, 49–80.

Nakano, Y., Felser, C. and Clahsen, H. (2002). Antecedent priming at trace positions in Japanese long-distance scrambling. *Journal of Psycholinguistic Research 31*, 531–71.

Odlin, T. (2005). Crosslinguistic influence and conceptual transfer: What are the concepts? *Annual Review of Applied Linguistics 25*, 3-25.

Omaki, A., & Ariji, K. (2005). Testing and attesting the use of structural information in L2 sentence processing. In L. Dekydtspotter, R. A. Sprouse, A. Liljestrand (Eds.).

Proceedings of the 7th Generative Approaches to Second Language Acquisition Conference. Somerville, MA: Cascadilla Proceedings Project, 205-218.

Osaka, M. & Osaka, N. (1992). Language independent working memory as measured by Japanese and English reading span tests. *Bulletin of the Psychonomic Society 30*, 287–89.

Papadopoulou, D. (2005). Reading-time studies of second language ambiguity resolution. *Second Language Research 21,* 98–120.

Papadopoulou D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition 25*, 501–528.

Phillips, C. (2006). The real-time status of island phenomena. *Language, 82*(4), 795-823.

Pickering, M. J., Traxler, M. J., & Crocker, M. W. (2000). Ambiguity resolution in sentence processing: Evidence against frequency-based accounts. *Journal of Memory and Language 43*, 447–475.

Pritchett, B.L. (1992a). *Grammatical competence and parsing performance*. Chicago: Chicago University Press.

Pritchett, B.L. (1992b). Parsing with grammar: Islands, heads, and garden paths. In Goodluck, H. and Rochemont, M. (Eds). *Island constraints: theory, acquisition, and processing*. Dordrecht: Kluwer, 321–50.

Roberts, L., Marinis, T., Felser, C., & Clahsen, H. (2007). Antecedent priming at trace positions in children's sentence processing. *Journal of Psycholinguistic Research 36*(2), Mar-188.

Roberts, R. and Gibson, E. (2002). Individual differences in sentence memory. *Journal of Psycholinguistic Research 31*, 573–598.

Robinson, P. (2002). Effects of individual differences in intelligence, aptitude and working memory on incidental SLA. In Robinson, P. (Ed). *Individual differences and instructed language learning*. Philadelphia, PA: John Benjamins, 211–51.

Rodríguez, G. (2004). Relative clause attachment preferences in second language learners' parsing performance. *University of Pennsylvania Working Papers in Linguistics 10*(1), 157-169.

Rodríguez, G. (2006). Using only word class: Evidence against shallow parsing in second language sentence process. Manuscript.

Saah, K.K. and Goodluck, H. (1995). Island effects in parsing and grammar: evidence from Akan. *The Linguistic Review 12*, 381–409.

Sagarra, N. (2000). The longitudinal role of working memory on adult acquisition of L2 grammar. Unpublished PhD Dissertation, University of Illinois at Urbana-Champaign.

Sanford, A., & Sturt, P. (2002). Depth of processing in language comprehension: not noticing the evidence. *Trends in Cognitive Science 6*(9), 382-386.

Scheepers, C., Hemforth, B., & Konieczny, L. (2000). Linking syntactic functions with thematic roles: Psych-verbs and the resolution of subject-object ambiguity. In B. Hemforth & L. Konieczny (Eds.). *German sentence processing*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 95-137.

Segalowitz, N. (2001). On the evolving connections between psychology and linguistics. *Annual Review of Applied Linguistics 21*, 3-22.

Selinker, L. (1972). Interlanguage. *IRAL, International Review of Applied Linguistics in Language Teaching 10*(3), 209-231.

Service, E. (1992). Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology 45A*, 21–50.

Service, E., & Craik, F. I.M. (1993). Differences between young and older adults in learning a foreign vocabulary. *Journal of Memory and Language 32*, 608–623.

Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition. *Applied Psycholinguistics 16*, 155–172.

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.

Smith, B. (2006). Precautions regarding nonword repetition tasks. *Journal of Applied Psycholinguistics 27*(4), 584-587.

Stowe, L. (1986). Parsing *wh*-constructions: evidence for on-line gap location. *Language and Cognitive Processes 2*, 227–46.

Swets, B., Desmet, T., Hambrick, D. & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology 136*, 64–81.

Swinney, D. A., Onifer, W., Prather, P., & Hirshkowitz, M. (1979). Semantic facilitation across sensory modalities in the processing of individual words and sentences. *Memory & Cognition 7*(3), 159-165.

Tanenhaus, M.K., Carlson, G. and Trueswell, J.C. (1989). The role of thematic structures in interpretation and parsing. *Language and Cognitive Processes 4*, 211–34.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Lindamood, P., Rose, E., Conway, T., et al. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology 91*(4), 579-593.

Traxler, M. (2002). Plausibility and subcategorization preference in children's processing of temporarily ambiguous sentences: Evidence from selfpaced reading. *The Quarterly Journal of Experimental Psychology 55A*, 75–96.

Traxler, M. (2005). Plausibility and verb subcategorization in temporarily ambiguous sentences: Evidence from self-paced reading. *Journal of Psycholinguistic Research 34*(1), 1-30.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The Declarative/Procedural model. *Bilingualism: Language and Cognition 4*(2), 105-122.

Van Gompel, R. P. G., & Liversedge, S. P. (2003). The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition 29*(1), 128-139.

Van Gompel, R. P. G., & Pickering, M. J. (1998). Lexical guidance in sentence processing: A note on Adams, Clifton, & Mitchell. *Psychonomic Bulletin and Review 8*, 851–857.

VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition*. Norwood, NJ: Ablex.

VanPatten, B. (2004). Input processing in SLA. In B. VanPatten (Ed.). *Processing instruction: Theory, research, and commentary*. Mahwah, NJ: Lawrence Erlbaum, 5-32.

Watanabe, A. (2001). *Wh*-in-situ languages. In M. Baltin & C. Collins (Eds.), *The handbook of contemporary syntactic theory*. Oxford: Blackwell, 203-225.

Waters, G. S., & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review 103,* 761–772.

Williams, J.N., Möbius, P. and Kim, C. (2001). Native and non-native processing of English *wh*-questions: parsing strategies and plausibility constraints. *Applied Psycholinguistics 22*, 509–40.

Zyzik, E. (2006). Transitivity alternations and sequence learning: Insights from L2 Spanish production data. *Studies in Second Language Acquisition 28*(3), 449-485.