

PROCESSING AND INTERPRETATION OF GARDEN-PATH SENTENCES IN
L2 SPEAKERS OF ENGLISH

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF SOCIAL SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

SÜLEYMAN YAMAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF ARTS
IN
THE DEPARTMENT OF ENGLISH LANGUAGE TEACHING

AUGUST 2023

Approval of the thesis:

**PROCESSING AND INTERPRETATION OF GARDEN-PATH SENTENCES
IN L2 SPEAKERS OF ENGLISH**

submitted by **SÜLEYMAN YAMAN** in partial fulfillment of the requirements for
the degree of **Master of Arts in English Language Teaching, the Graduate School
of Social Sciences of Middle East Technical University** by,

Prof. Dr. Sadettin KIRAZCI
Dean
Graduate School of Social Sciences

Prof. Dr. Nurten BİRLİK
Head of Department
Department of Foreign Language Education

Assoc. Prof. Dr. Duygu SARISOY
Supervisor
Department of Foreign Language Education

Examining Committee Members:

Prof. Dr. Çiğdem SAĞIN ŞİMŞEK (Head of the Examining Committee)
Middle East Technical University
Department of Foreign Language Education

Assoc. Prof. Dr. Duygu SARISOY (Supervisor)
Middle East Technical University
Department of Foreign Language Education

Assist. Prof. Dr. Taylan AKAL
Hacettepe University
Department of English Linguistics

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Süleyman YAMAN

Signature:

ABSTRACT

PROCESSING AND INTERPRETATION OF GARDEN-PATH SENTENCES IN L2 SPEAKERS OF ENGLISH

YAMAN, Süleyman

M.A., The Department of English Language Teaching

Supervisor: Assoc. Prof. Dr. Duygu SARISOY

August 2023, 91 pages

The current thesis aims to examine whether L1 and L2 speakers of English differ from each other in terms of parsing and comprehension of temporarily ambiguous or garden-path sentences. In a self-paced experiment, the ambiguity of the sentences and the type of comprehension questions (i.e. main vs. subordinate clause) were manipulated. The analysis of reading times revealed that both groups experienced garden-path effects at or following the disambiguation, with these effects being smaller and more volatile for the L2 group. With respect to the analysis of off-line comprehension accuracy, the well-known phenomenon of lingering misinterpretations was broadly replicated in both groups, as evidenced by the low accuracy rates to the subordinate clause questions following the ambiguous condition; however, significant ambiguity effects were also found in main clause questions. The L2 group was generally more likely to misinterpret the experimental sentences than the L1 group irrespective of ambiguity and question type. Finally, individual differences in (self-rated) English proficiency did not modulate either the online garden-path effects or comprehension accuracy among the L2 speakers. The observed pattern of results is discussed with respect to the accounts of L2 processing.

Keywords: Psycholinguistics, L2 processing, garden path sentences, good-enough comprehension

ÖZ

İKİNCİ DİL İNGİLİZCE KONUŞUCULARINDA GEÇİCİ OLARAK SÖZDİZİMSEL BELİRSİZLİK İÇEREN TÜMCELERİN İŞLEMLENMESİ VE YORUMLANMASI

YAMAN, Süleyman

Yüksek Lisans, İngiliz Dili Öğretimi Bölümü

Tez Yöneticisi: Doç. Dr. Duygu SARISOY

Ağustos 2023, 91 sayfa

Mevcut çalışma, anadil (D1) ve ikinci (D2) dil İngilizce konuşucusu bireylerin geçici sözdizimsel muğlaklık barındıran tümceleri işleme ve anlamlandırma konusunda birbirlerinden farklılık gösterip göstermediklerini incelemektedir. Buna yönelik uygulanan bir kendi hızında okuma deneyinde, tümcelerin belirsizlik içerip içermediği ve sonrasında sorulan anlam sorularının tümcenin hangi kısmını hedeflediği (yan ve ana tümcecik) manipüle edilmiştir. Okuma sürelerinin analizi her iki grubun da belirsizliğin giderildiği bölgede işleme güçlüğü yaşadığını, ancak bu güçlüğü D2 grubunda daha zayıf ve değişken olduğunu göstermiştir. Çevrimdışı anlam sorularına verilen yanıtların doğruluğuna ilişkin yapılan analizler kapsamında, muğlak koşulda yöneltilen yan tümcecik sorularındaki düşük doğruluk oranları literatürde iyi bilinen yanlış yorumlama olgusunun her iki grupta da replike edildiğini göstermekle birlikte ana tümcecik sorularında da muğlaklığın anlamlı bir etkisi gözlemlenmiştir. D2 grubu, deneysel tümceleri muğlaklık ve soru tipinden bağımsız olarak D1 grubuna göre daha sık yanlış yorumlamıştır. Son olarak, İngilizce (öz) yeterliğindeki bireysel farklılıklar, D2 grubu içinde muğlaklığın giderildiği bölgede yaşanan çevrimiçi işleme

güçlüğünü ya da anlam sorularına verilen yanıtların doğruluğunu etkilememiştir. Elde edilen sonuçlar, D2 işleme teorileri çerçevesinde tartışılmaktadır.

Anahtar Kelimeler: Psikodilbilim, ikinci dil işleme, geçici sözdizimsel muğlaklığa sahip tümceler, yüzeysel anlamlandırma

To every person who accounts for a significant variance in my life

ACKNOWLEDGMENTS

First of all, I would like to thank my advisor Assoc. Prof. Dr. Duygu Sarısoy for supporting me in this tough journey all the time. As a professor from whom I took the majority of the linguistics courses during my undergraduate education, she has been a great role model for me. I am truly grateful to her for encouraging the linguist in me from day one, for always believing in me and for her guidance for me to make the current thesis “good-enough” (pun intended). It is owing to her unconditional support and unique guidance that this 3-years of arduous process was more enjoyable for me. I once again thank you for everything wholeheartedly, Hocam!

I thank my committee members Prof. Dr. Çiğdem Sağın-Şimşek and Assist. Prof. Dr. Taylan Akal for their constructive comments and feedback on my thesis. Even though she was not able to take part in my committee, I am also indebted to Prof. Dr. Martina Gračanin-Yüksek for inspiring and encouraging the (semi) syntactician in me. I feel very lucky to have had a chance to be her student, and will always remember those puzzles from Croatian data in her exams ☺. Besides, I am grateful to Prof. Dr. Bilal Kırkıcı for the perspectives I gained from his courses. They enabled me to become a more critical researcher and take a further step towards being able to ask the right questions as a psycholinguist and scientist.

I express my profound gratitude to all my friends from METU Language and Cognitive Development Lab and the linguistics circle of FLE. Their friendship and support have always been invaluable for me.

I owe a lot to my friend and former roommate Meriç Öztürk. He put in a great deal of effort to find native speaker participants for me in Ames, Iowa even though he had already too much on his plate at the time. I am thankful for his genuine friendship and our conversations on Twitter.

Many heartfelt thanks to my participants—including some fellow Redditors whose names I will never know. Without their help, this thesis wouldn't have made it into the existence.

My last but greatest thanks go to my beloved parents and my dear sister Nazlıcan for all their endless love, support and encouragement. They have always believed in me and stood by my side in every challenge I had to tackle. I am eternally grateful to every single one of them.

TABLE OF CONTENTS

PLAGIARISM	iii
ABSTRACT	iv
ÖZ	vi
DEDICATION	viii
ACKNOWLEDGMENTS	ix
TABLE OF CONTENTS	xi
LIST OF TABLES	xiv
LIST OF FIGURES	xv
LIST OF ABBREVIATIONS	xvi
CHAPTERS	
1. INTRODUCTION	1
1.1. Background to the Study	1
1.2. Significance of the Study	4
2. THEORETICAL BACKGROUND	7
2.1. Sentence Processing and Good-Enough Comprehension	7
2.2. Comprehension of GP Sentences in L1 Speakers	9
2.3. Parsing and Reanalysis in Second Language	20
2.3.1. Overview	20
2.3.2. Second Language Processing Accounts	21
2.3.2.1. Capacity Limitation Accounts	21
2.3.2.2. Declarative/Procedural Model	22
2.3.2.3. Shallow Structure Hypothesis	23
2.3.2.4. Interference Account	27

2.3.3. Reanalysis Processes in L2 Speakers	30
3. THE PRESENT STUDY	37
3.1. Aim of the Study	37
3.2. Research Questions and Predictions	38
3.3. Method	39
3.3.1. Participants	39
3.3.2. Materials and Design.....	40
3.3.3. Procedure.....	43
3.4. Data Analysis	44
4. RESULTS.....	46
4.1. Between-Groups Analyses	46
4.1.1. RT Analyses	46
4.1.2. Accuracy Analyses.....	48
4.2. Individual Differences in L2 Proficiency.....	49
4.2.1. RT Analyses	49
4.2.2. Accuracy Analyses.....	50
5. DISCUSSION	52
5.1. The Nature of Lingering Misinterpretations in GP Sentences	53
5.2. Reanalysis and Mechanisms of L2 Sentence Processing	54
5.3. Conclusion.....	57
5.4. Limitations and Suggestions for Further Research	58
REFERENCES.....	60
APPENDICES	
A. APPROVAL OF THE METU HUMAN SUBJECTS ETHICS COMMITTEE	75
B. CONSENT FORM FOR THE L1 GROUP.....	76
C. CONSENT FORM FOR THE L2 GROUP.....	77
D. BACKGROUND QUESTIONNAIRE FOR THE L1 GROUP.....	78

E. BACKGROUND QUESTIONNAIRE FOR THE L2 GROUP	79
F. TURKISH SUMMARY / TÜR KÇE ÖZET	80
G. THESIS PERMISSION FORM / TEZ İZİN FORMU	91

LIST OF TABLES

Table 1. Background of the L2 participants	40
Table 2. Summary of the RT analyses across the groups.....	47
Table 3. Summary of accuracy analyses between groups	49
Table 4. Summary of the proficiency analyses for RTs.....	50
Table 5. Summary of the proficiency analyses for accuracy rates.....	50

LIST OF FIGURES

Figure 1. The architecture of L2 parser under SSH (Clahsen & Felser, 2006b, p. 118)	24
Figure 2. Mean reading times (ms) across groups and conditions	46
Figure 3. Mean accuracy rates across question type and group	48

LIST OF ABBREVIATIONS

CP: Complementizer Phrase

DO: Direct Object

ERP: Event Related Potentials

GP: Garden Path

L1: First Language

L2: Second Language

GMM: Gender Mismatch

Min: Minimum

Max: Maximum

ms: Milisecond(s)

NP: Noun Phrase

OPT: Optionally Transitive

RAT: Reflexive Absolute Transitive

RC: Relative Clause

RT: Reading Time

SC: Sentential Complement

SD: Standard Deviation

SSH: Shallow Structure Hypothesis

CHAPTER 1

INTRODUCTION

1.1. Background to the Study

Humans have a distinguished capability to communicate through comprehending and producing language. Central to this capability is parsing during which the linguistic input is momentarily converted into the structural representations. In the course of parsing, the human comprehension system chunks the sentences into syntactic constituents and assign these constituents relevant thematic roles, while also integrating cues from different sources of information such as context or prosody. One striking property of parsing is its incremental nature which allows the comprehenders to process each incoming linguistic unit automatically and integrate it with the sentential representation that is being built as soon as it is detected across the bottom-up input (Frazier & Rayner, 1982; Altmann & Kamide, 1999; Kamide et al., 2003; Altmann & Mirković, 2009; Omaki et al., 2014; see also Özge, 2020 for a recent review). It is thanks to this property of the human processor that we mostly succeed in everyday communication by being able to comprehend utterances around us almost with no or little delay, without having to wait for the end of these utterances.

As advantageous and time-efficient as it may be, incremental processing could however sometimes come with certain costs when the unfolding input presents a linguistic adversity as in temporarily ambiguous or garden-path sentences where an initial interpretation later turns out to be incorrect and thus comprehenders are assigned with an extra task to undo this interpretation:

- (1) While Mary dressed the baby spit up on the bed.

In the sentence (1), the noun phrase (NP) *the baby* is first understood as the object of the verb *dressed*. At the next phrase *spit up* further downstream, however, the parser undergoes a processing difficulty regarding how to integrate this unexpected verb into the current representation, usually in the shape of elevated reaction times. For the parser, the only way this conflict can be grammatically resolved is to revise the NP *the baby* as the subject of *spit up* and the next verb *dress* as reflexive-intransitive verb, which yields the globally correct interpretation that “it was during which Mary dressed (herself) that the baby spit up”.

In psycholinguistics, garden-path sentences have played a substantial role in motivating various theories of sentence processing regarding how the human parser generates real-time analyses of the sentential structure from the bottom-up input (Bever, 1970; Frazier & Rayner, 1982; Ferreira & Clifton, 1986; Altmann et al., 1992; Ferreira & Henderson, 1991; MacDonald et al., 1994; McRae et al., 1998). Although these theories often differ in explaining the reasons behind how or why the comprehenders are lured into adopting a garden-path (GP) interpretation in specific and the architectures of sentence comprehension in general, they have long converged in their implicit assumption that the parser’s ultimate task is always to create a fully-fledged and accurate representation of the linguistic input and the sentence meaning is consequently built in a compositional fashion. In fact, the extent to which this assumption has governed these models can be seen in MacDonald et al.’s (1994) view that there could be circumstances where “the communicative goals of the reader or listener can be achieved with only a partial analysis of a sentence, but we view these as degenerate cases” (p. 686). In accordance with this rationale, one may in turn expect reanalysis and repair processes to be mostly successful for the linguistic environments of the sort exemplified in (1), as has been largely assumed in the earlier years of psycholinguistic research.

Nevertheless, the last two decades have witnessed a serious bulk of work challenging this idea especially since the seminal findings of Christianson et al. (2001). In their study, Christianson and his colleagues presented native English speaker participants with the GP sentences (1) and asked comprehension questions probing the direct object interpretation (i.e. *Did Mary dress the baby?*). The striking result was that the participants gave incorrect *Yes* responses endorsing the incorrect interpretation,

although one could normally expect the disambiguating verb *spit up* to serve as a reliable cue to override this reading. Against the aforementioned traditional models of parsing, Christianson et al. (2001) interpreted their findings as evidence that reanalysis processes may not run to completion as previously assumed and thus the parser may be sometimes inclined to construct “good-enough” representations under certain circumstances. Even though Christianson et al.’s (2001) findings initially received a substantial amount of criticism mainly due to an alternative explanation that the explicit comprehension questions may have reactivated the direct object reading even after a successful reanalysis (Tabor et al., 2004), similar findings were later reported using more implicit measures such as paraphrasing (Karsenti & Meltzer-Asscher, 2022; Patson et al., 2009), structural priming (van Gompel et al., 2006), downstream manipulation of semantic consistency (Sturt, 2007; Şafak & Hopp, 2022) and nested texts (Slattery et al., 2013; Fujita & Cunnings, 2020, 2021).

Among evidence from other phenomena, the documented instances of persistent misinterpretations of GP sentences have constituted an important driving force in motivating dual-pathways or “good-enough” models of sentence processing (Ferreira et al., 2002; Ferreira, 2003; Kuperberg, 2007; Karimi & Ferreira, 2016). Under these models, it is maintained that the human parser essentially makes of two processing pathways: one corresponds to an algorithmically-driven route that creates detailed structural representations of the input, and the other to a heuristic route that largely operates on semantics and yields “good-enough” representations. The core idea is that the comprehenders can sometimes resort to the latter route in an aim to alleviate the relative burden of constructing detailed syntactic representations, especially when this option turns out to be too costly to employ. In these models, the lingering misinterpretations of GP sentences are thus captured with the heuristic route dominating the algorithmic one in a way that the parser either fails to derive a complete syntactic representation (e.g., Christianson et al., 2001; Chromý, 2022) or may retain the initially assigned misinterpretation despite a fully complete syntactic repair (e.g., Kaschak & Glenberg, 2004; Sturt, 2007; Slattery et al., 2013).

The degree to which good-enough processing route operates can vary between and within populations. As fundamental to the present thesis, Shallow Structure Hypothesis (SSH) proposed by Clahsen and Felser (2006a, 2006b, 2018) argues that

second language speakers (L2) compute shallow representations of the syntax and rely more on pragmatic, semantic and surface cues during real-time sentence processing. This account builds upon the existing models of dual-pathways of language processing mentioned above that incorporate two routes of parsing, one of which corresponds to a full parsing route where detailed and hierarchical syntactic representations are computed for a given utterance and the other to a shallow or heuristic parsing route which provides less detailed representations in which non-syntactic cues are assigned a greater amount of weight to guide the interpretation. Based on this assumption, SSH maintains that the non-native processing is more likely to be dominated by the latter pathway such that L2 speakers may not construct complex hierarchical structures in a similar fashion to the L1 speakers and resort more to the other sources of information. Critically, SSH emphasizes that the processing differences between the native and non-native speakers cannot be simply attributed to the influence of L1 and incomplete acquisition of grammar since the studies taken to lend support to SSH suggest that even highly proficient speakers of L2 may apply distinct parsing routines different from that of L1 speakers and use non-syntactic information more reliably irrespective of their L1 background. With regard to the GP phenomena, it could be thus expected that L2 speakers would potentially experience more difficulty in recovering from the initial misinterpretations of these sentences due to a greater tendency to base their processing of the input on good-enough or semantically motivated representations compared to the L1 speakers, by committing to an initially plausible analysis more strongly (Roberts & Felser, 2011; Jacob & Felser, 2016; cf. Cunnings, 2017), which will constitute a core examination of the current thesis.

1.2. Significance of the Study

The present study aims to add to the existing body of literature on second language processing as well as good-enough comprehension in terms of several aspects. First of all, most of the available studies in L2 processing research have taken advantage of GP phenomena as a diagnostic of whether L2 speakers can use various sources of information such as plausibility (Roberts & Felser, 2011), verb bias (Dussias & Cramer-Scaltz, 2008) and case marking (Jackson, 2008) incrementally in a similar manner to the L1 speakers. Such studies have usually manipulated the relevant cues that precede the disambiguating segment, and systematically compared the processing

difficulty at or following the disambiguation across conditions as an index of whether these cues are integrated reliably during on-line comprehension. Although such line of investigation has yielded a vast body of findings that could inform the ongoing debates regarding the degree with which L1 and L2 parsing could be similar, there are relatively fewer studies that explored the potential differences between L1 and L2 speakers as to the success with which misinterpretations are successfully abandoned (c.f. Jacob & Felser, 2016; Fujita & Cunnings, 2020, 2021). As the examination of the reanalysis success among the two populations has a potential to inform the debates regarding which of the two concomitant routes of processing is more likely to guide L2 comprehension, the present study aims to address the research gap in question and add to the current literature by combining on-line and off-line measures of reanalysis in the same experimental design. Secondly, the bulk of research that investigated the interpretation of GP sentences has largely utilized comprehension probes that exclusively target the initial misanalysis and neglected to directly test other parts of the sentence. Although there are some exceptions to this trend (e.g., Christianson et al., 2001; Chromý, 2022; Fujita & Cunnings, 2021), the number of such studies is still rather limited to reach a more fine-grained picture of GP recovery. Against this backdrop, the experiment reported in the current study examines both the initially adopted misinterpretation and the nature of the internal syntactic representation by making use of comprehension probes that target the respective parts of the experimental sentences. Thirdly and finally, the present study will also explore the role of L2 proficiency on the success with which reanalysis is conducted. As the investigation of proficiency can have some theoretical implications regarding the properties of the developing L2 parser, the current investigation is also expected to cast light on the degree to which GP recovery can become more native-like over time as well as on the inter-individual variability of L2 processing.

Against this backdrop, the present study aims to seek answer to the following research questions:

1. Do L2 speakers of English differ from L1 speakers of English in terms of GP recovery?

2. Does the level of proficiency modulate the degree to which GP recovery is successful within L2 speakers?

CHAPTER 2

THEORETICAL BACKGROUND

2.1. Sentence Processing and Good-Enough Comprehension

As one of the most basic linguistic units, the way sentences are processed in real-time comprehension garnered a good deal of interest since the dawn of psycholinguistics. At the center of this interest has often been the questions of what kind of linguistic information guides sentence processing, at what stage these sources of information are weighted by the processor, whether or to what extent they interact with each other during real-time comprehension.

Over the years, influential models of sentence processing have adopted various theoretical stances in an aim to answer these questions. Among these, serial or "syntax-first" approaches (Frazier & Fodor, 1978; Frazier & Rayner, 1982) posit that the parser constructs the representation of a given sentence by first deriving its syntactic derivation through structure-building operations and weights the non-syntactic cues such as plausibility, context and prosody only at the later stages of processing. In other words, these latter sources of information act somewhat as a filter or a check against the initially computed syntactic representation under these approaches. On the other hand, constraint-satisfaction approaches propose that the parser is not primarily limited to weighting any cue at a particular point; instead, it consults all possible sources of information simultaneously and follows an analysis that best fits the coalition of these available cues (e.g., Trueswell et al., 1993). As regards to the GP phenomena, the former class of accounts attribute the reason why the comprehenders initially consider a direct object reading to the parser's preference to build the simplest structure in line with Late Closure principle (Frazier & Fodor, 1978) which postulates that the parser incorporates an incoming string directly to the currently processed phrase instead of projecting a new one. The latter accounts, on the other hand, ascribe garden-pathing to

the interplay of multiple cues such as the transitivity of the verb, the plausibility of the initial parse, or the high distributional frequency of the subject-verb-object forms in English (MacDonald et al., 1994; Garnsey et al., 1997; McRae et al., 1998).

Despite their divergence in accounting for the psycholinguistic mechanisms underlying sentence comprehension with respect to the timing of various linguistic cues, these models can be said to converge in incorporating a covert assumption that sentence comprehension proceeds algorithmically in a way that the sentential representations built by the parser are usually accurate and detailed, and thus the final state of the interpretations is compositional in nature reflecting the genuine content of the input. In line with this assumption, comprehension breakdowns were usually neglected and tended to be dismissed as unsystematic, wild or “degenerate” cases as in MacDonald et al.’s (1994) terms for a considerably long time in the psycholinguistics literature. Against this assumption, however, dual-pathways or good-enough theories of sentence processing counter-argue that this may not be always the case (Christianson, 2016; Ferreira et al., 2002; Ferreira & Patson, 2007; Townsend & Bever, 2001). While these models acknowledge the existence of robust algorithmic routines that compute accurate and detailed representations of the input, broadly speaking, they also call into question the extent to which these routines are always within the parser’s reach and maintain that sentence processing can sometimes proceed with “rough-and-ready” representations that diverge with the actual content of the linguistic signal. With sentences “the dog was bitten by the man”, for instance, it was widely demonstrated that comprehenders often derive a reverse interpretation where the dog did the act of biting the man (e.g., Ferreira, 2003), suggesting that a simple heuristic operating on real-world knowledge can sometimes override the algorithmic parsing route that would otherwise yield the correct interpretation “the man bit the dog”.

Ironically, another key piece of empirical evidence that lend support to the idea of good-enough processing come from GP sentences, the very same phenomenon that has been often taken advantage of to choose between the aforementioned theories of sentence processing that incorporate the central assumption of compositionality. As mentioned in the previous chapter, a considerable bulk of work have so far showed that the initially adopted analysis is not fully abandoned to the extent that it can

determine the ultimate interpretation of the sentence in a stark contrast to what would be expected given the presence of a late-arriving disambiguating signal (Christianson et al., 2001; Chromý, 2022; Fujita & Cunnings, 2020, 2021; Huang & Ferreira, 2021; Karsenti & Meltzer-Asscher, 2022; Lau & Ferreira, 2005; Patson et al., 2009; Slattery et al., 2013; Qian et al., 2018; van Gompel et al., 2006; *inter alia*).

Although it is well-established that GP sentences are misinterpreted in accordance with the good-enough processing view, it is debated what exactly causes these lingering misinterpretations in the monolingual sentence processing literature. The arguments arising from these debates can be examined under two broad sides: (i) those attributing the lingering misinterpretations to failure to repair the underlying structural representation (Christianson et al., 2001; Chromý, 2022) and (ii) those attributing them to the interference of the initially constructed interpretation that is not discarded despite a successful syntactic repair (Slattery et al., 2013; Sturt, 2007). In the following section, these debates are briefly summarized by presenting a selection of studies that examined the post-repair stage of the GP comprehension in L1 speakers.

2.2. Comprehension of GP Sentences in L1 Speakers

Even though the reanalysis processes were widely studied in terms of several aspects (e.g., Altmann et al., 1992; Binder et al., 2001; Ferreira & Henderson, 1993; Pickering & Traxler, 1998; Spivey-Knowlton et al., 1993; Sturt et al., 1999), it was first Christianson et al. (2001) who systematically examined the ultimate interpretations derived from these sentences. In a series of off-line experiments, Christianson and his colleagues had the adult monolingual speakers of English read GP sentences and answer forced-choice questions tapping into the content of subordinate as well as matrix clauses of the stimuli (2):

(2) Did Mary dress the baby? (subordinate clause)

Did the baby spit up? (main clause)

With this design, Christianson et al. (2001) reasoned that the incorrect *Yes* responses should not be given to the subordinate-clause questions if the analysis of *the baby* as the object is successfully abandoned upon the disambiguating cue. In a similar vein,

such successful reanalysis would be further indexed by the correct *Yes* responses to the main clause questions, showing that *the baby* was not only relinquished as the object, but also fully revised as the subject of the matrix verb *spit up*. Against these predictions, it was found that the participants gave incorrect *Yes* responses to the subordinate clause questions after ambiguous sentences roughly 50% of the time, which indicates that the initial misinterpretation of Mary dressing the baby was not completely abandoned. In the main clause questions, on the other hand, the effects of ambiguity were much more attenuated, with the participants providing accurate *Yes* responses 90% of the time. This indicated that they entertained both interpretations “Mary dressed the baby” and “the baby spit up” contrary to what is permitted by the legal syntactic structure of the sentence. Taking these results as evidence that reanalysis may not be completed as previously assumed, Christianson et al. (2001) argued that their findings can be explained with a parsing failure with incomplete or partial reanalysis. Specifically, they maintained that the NP *the baby* may be successfully “stolen” from the subordinate clause, as evinced by the high accuracy rates to the matrix clause questions, but it may not be fully deleted from the object position of the subordinate clause, potentially in the form of a phonologically null copy (see Fodor & Inoue, 1998).

Although Christianson et al.’s (2001) findings have played a substantial role in motivating the accounts of good-enough language processing, their work has also been subject to a considerable amount of criticism due to their methodology. Central to a commonly-dubbed concern is that the explicit nature of the comprehension questions could have reactivated the initial misinterpretation even if the ambiguity is successfully repaired by the subjects. That is, the participants could have displayed a tendency to give *Yes* responses simply because the question itself reinstates the misinterpretation by containing the same surface strings “Mary”, “dress” and “the baby”. One influential study that has addressed this criticism was conducted by Kaschak and Glenberg (2004) who made use of a training design without using explicit comprehension probes. In the training phase of their critical experiment, the monolingual speakers of English had to learn either a novel *needs* construction or were supposed to be trained on its standard version (3).

(3) The meal needs (to be) cooked given that the dinner is in an hour.

The sentence in (3) is temporarily ambiguous in that the phrase *cooked* can be first understood as the modifier of an upcoming NP (e.g., *cooked vegetables*), but has to be revised as the subordinate predicate of a control construction found in some dialects of English. Its version with *to be*, on the other hand, is its unambiguous version that served as a baseline condition. Following the training session where the subjects were repeatedly exposed to these constructions, they read sentences that featured actual modifier constructions (4) in a subsequent self-paced reading experiment.

(4) This meal needs cooked vegetables to make it complete.

Although logical expectation dictates that repeated exposure to such structures should facilitate reanalysis and discourage the adoption of incorrect interpretation over time, Kaschak and Glenberg (2004) observed that the subjects who were exposed to the *needs* construction in the previous session read the post-participle NP (e.g., *vegetables*) faster than the other group exposed to the standard/unambiguous construction, indicating that the incorrect meaning was not deactivated entirely. Instead, it lingered to the extent that the participants kept expecting a modified NP in line with this interpretation. Arguing that these findings cannot be easily reconciled with constraint-satisfaction approaches according to which less probable interpretations are inhibited upon losing competition to the others generated in parallel (e.g., McRae & Matsuki, 2013), the authors claimed that their results instead point to the existence of a processing mechanism that operates on episodic memory traces of the initially selected interpretations.

In a similar vein, van Gompel et al. (2006) conducted three structural priming experiments to examine whether good-enough comprehension of GP sentences can be still observed without explicit comprehension probes. In the experiments, their participants had to provide verbal completions to a target fragment immediately after silently reading the experimental sentences:

(5) While the man was visiting(,) the children who were surprisingly pleasant and funny played outside.

The target fragment: When the doctor visi...

The rationale was that the subjects should utter more transitive continuations (e.g., *visiting the patient*) after reading ambiguous than after unambiguous condition if the initially adopted transitive interpretation is not fully deactivated and thus robust enough to prime a novel utterance with a transitive frame. If this misinterpretation is fully discarded, however, no such priming should be observed. The results of van Gompel et al. (2006) bore out the former prediction in accordance with Christianson et al.'s earlier findings (2001). Regarding what may cause the persistence of these misinterpretations, van Gompel et al. (2006) suggested that either incomplete syntactic reanalysis or memory traces of the initially adopted GP interpretation may be responsible for the transitive completions, between which they preferred to remain agnostic.

In a later study, Sturt (2007) showed that misinterpretations may persist even when the processing of the sentence is still ongoing and the ambiguity is relatively easier to resolve compared to the kind with preposed adjunct clauses (e.g., Christianson et al., 2001). Using eye-tracking technique, he had a group of L1 English individuals read sentences with complementizer ambiguity such as (6):

(6)

- a. The explorers found (that) the South Pole was actually right at their feet.
- b. The explorers found (that) the South Pole was actually impossible to reach.

In the GP condition where the complementizer *that* is absent, the NP *the South Pole* is initially understood as the object. This parse has to be, however, overridden by the auxiliary verb *was* since it requires a subject, and the South Pole should be revised as the subject of a new full CP. The reason why this ambiguity can be easier to resolve reduces to the scope of the thematic domain that needs to be revised. In (1), also known as NP/Z ambiguity, the reanalysis of the ambiguous NP involves a greater structural change since it should be removed from the domain where it receives the patient role of *dressee*. In complementizer or NP/S ambiguities (6), reanalysis of the NP does not induce a structural change as dramatic as the GP sentences with preposed adjunct clauses since *the South Pole* remains in the thematic domain of the verb *find* where it initially received the role of *findee* (see Sturt et al., 1999). To examine whether the initially adopted reading persists beyond the disambiguation, Sturt (2007) further

manipulated the semantic compatibility of the final region with that of a GP interpretation. The initial direct object reading is semantically congruent with the final segment *right at their feet* in (6a), while it is not with the final segment *impossible to reach* in (6b). In addition to the classic GP effects, Sturt (2007) observed that the final segment elicited longer go-past RTs in (6b) compared to (6a) when the sentences are ambiguous, suggesting that the initial misinterpretation was preserved after the disambiguation. Moreover, in the Ambiguous/Match condition (6a), the percentage of regressions was only 6% at the disambiguation region, indexing the relative ease with which reanalysis can be achieved with this type of GP sentences in the absence of a downstream semantic conflict. Based on these results, Sturt (2007) argued that lingering misinterpretations could be attributed to the difficulty with discarding the semantic content of the initial parse rather than a failure of syntactic repair, as he dubbed “semantic persistence”.

To further differentiate between the potential levels of linguistic representation that causes persistent misinterpretations, Slattery et al. (2013) conducted two eye-tracking while reading experiments. In the first one, the authors took advantage of Principle A that necessitates the reflexive pronouns to be coindexed with a c-commanding antecedent in the same local clause (Chomsky, 1981), and made use of gender mismatch paradigm by inserting a downstream reflexive pronoun to the experimental sentences. As part of this design, they manipulated the ambiguous NP’s match to the pronoun in terms of stereotypical gender (e.g., *David’s mother/father*):

(7) After the bank manager telephoned(,) David’s father/mother grew worried and gave himself approximately five days to reply.

Based on the line of research suggesting that real-time anaphora resolution respects hierarchical binding constraints (e.g., Sturt, 2003), Slattery et al. (2013) reasoned that the parser may fail to establish a binding domain for the reflexive pronoun if the ambiguous NP is not syntactically repaired (i.e. **[telephoned David’s father/mother_k] ... [himself_k]*), which would result in no or less reliable gender mismatch (GMM) effects in the ambiguous trials. On the other hand, GMM effects should be observed in both ambiguous and unambiguous trials if the repair of the ambiguity is successful and complete at the syntactic level of representation such that the region containing

the reflexive pronoun should elicit longer RTs when the temporarily ambiguous NP mismatches to the pronoun's gender than when it matches. This is because the temporarily ambiguous NP can sit in a c-commanding position relative to the reflexive pronoun in a fully repaired syntactic representation. Consistent with the second prediction, Slattery et al. (2013) observed GMM effects across the board, which they interpreted as evidence that the parser succeeds in deriving a correct representation that is in fact detailed enough to license a binding domain. They argued that the source of misinterpretations is thus unlikely to be the failure of syntactic repair contra Christianson et al. (2001), but instead may relate to the inability to dispose of the initially built syntactic parse as reanalysis presumably entails not only the building of the accurate structure but eliminating or inhibiting the incorrect one as well. In the next experiment, Slattery et al. (2013) investigated whether the effects of incorrect parse can be robust enough to interfere with the processing of the subsequent input, by using two-sentence long texts like (8):

(8) While Frank dried off(,) the truck/the grass that was dark green was peed on by a stray dog. Frank quickly finished drying himself off then yelled out the window at the dog.

In these texts, the follow-up sentence was congruent with the correct interpretation of the preceding sentence that was manipulated in terms of ambiguity, but incongruent with the incorrect GP interpretation. With the reflexive pronoun (e.g., *himself*), it was aimed to examine to what extent the erroneous parse is still retained and thus interferes with the processing of the subsequent input. A further layer of manipulation involved the plausibility of the ambiguous NP in order to examine how semantics could modulate reanalysis (see Pickering & Traxler, 1998). For instance, the NP *the truck* is a semantically plausible object of *dried off* in the above example, whereas *the grass* is not. In the former case, the readers can be more likely to be garden-pathed due to the semantic fit between the verb and the object, and thus have more difficulty in discarding the incorrect interpretation compared to the latter case where the implausibility of the NP can prevent the readers from adopting a direct object interpretation. In the experiment, Slattery et al. (2013) indeed observed robust effects of plausibility in several aspects. At the ambiguous NP region, first-pass RTs were found to be longer for implausible compared to the plausible NPs, indexing the parser's

immediate sensitivity to the semantic anomaly. These plausibility effects translated into the disambiguating region (i.e. *was peed*) in a reversed form such that the RTs were shorter when the previous NP was an implausible direct object. More importantly, the participants showed longer RTs at the reflexive region (e.g., *himself*) following the ambiguous than unambiguous sentences that contained a plausible NP, indicating that the incorrect interpretation “Frank dried off the truck” lingered and kept exerting an influence on the comprehension of an upcoming sentence.

Recently, Huang and Ferreira (2021) called the findings reported by Slattery et al. (2013) into question. Drawing attention to the fact that Slattery et al. (2013) did not ask comprehension questions in their first experiment, Huang and Ferreira (2021) raised the possibility that the observed GMM effects could stem from a portion of trials where the sentences are comprehended accurately since misinterpretations were observed to occur on almost half of the trials in previous research (Christianson et al., 2001). To address this alternative explanation, Huang and Ferreira (2021) conducted two experiments, one with eye-tracking and one with self-paced reading paradigm, in which the sentences were manipulated in a similar factorial design as in Slattery et al. (2013) and accompanied by end-of-trial comprehension questions. In both experiments, Huang and Ferreira (2021) found GMM effects regardless of the ambiguity, replicating the previous findings of Slattery et al. (2013). Moreover, in off-line accuracy, the ambiguity disadvantage was rather robust as in Christianson et al.’s (2001) study, with the participants displaying low accuracy rates of 46-48% in the GP condition(s). To further investigate the possibility that GMM effects reported in Slattery et al. (2013) are driven by the trials where the reanalysis ultimately succeeded, Huang and Ferreira (2021) conducted split-trial analyses where they examined GMM effects separately for the trials where the comprehension questions were responded correctly and incorrectly. Their analyses revealed that GMM effects are statistically preserved irrespective of the off-line response accuracy, hence ruling out the possibility that the results of Slattery et al. (2013) follow from a portion of stimuli being interpreted correctly. Based on these findings, Huang and Ferreira (2021) proposed two possible explanations regarding the source of the persistence of misinterpretations. In line with the Slattery et al.’s (2013) earlier conclusions, the first theoretical possibility could be grounded in a processing stage in which the attempts to inhibit or clean up the remnants of the erroneous interpretation may not succeed

despite the fact that the parser forms a fully reanalyzed syntactic representation. Maintaining that the GMM paradigm may not nevertheless rule out the possibility of the temporarily ambiguous NP being only partially reanalyzed, Huang and Ferreira (2021) also noted that the observed effects may stem from the co-existence of two locally legal structural representations, with the parser not deriving the globally correct one. More specifically, the syntactic reanalysis can be partially successful in moving the temporarily NP to the subject position and this may allow a binding domain to be established where *the baby* is in a c-commanding relation with the pronoun, as indexed by the GMM effects, which yields a local structure which is licit on its own right at the right clausal periphery. Meanwhile, another locally licit structure that is situated at the left periphery would be also a part of the overall representation of the sentence if the ambiguous NP is not fully erased from the object position (Christianson et al., 2001). Regardless of which scenario may hold, Huang and Ferreira (2021) concluded that no account that would predict full reanalysis can explain their results.

Apart from the syntactic diagnostics of the kind utilized by Slattery et al. (2013) and Huang and Ferreira (2021), some studies have attempted to elucidate the underlying cause of lingering misinterpretations by investigating the relationship between reanalysis effort and comprehension accuracy (Christianson & Luke, 2011; Wonnacott et al., 2016; Christianson et al., 2017; Qian et al., 2018). Based on the assumption that the distress observed in the disambiguating segment reflects efforts to conduct reanalysis, the rationale of such studies is that the amount of processing difficulty in this region should be able to predict comprehension accuracy positively if incomplete syntactic parsing is indeed responsible for good-enough interpretations. In other words, the greater time the parser spends to read the disambiguation region, the more likely it becomes to succeed in completing the syntactic repair and consequently arrive at the globally correct interpretation. One such study that examined this possibility is by Christianson and Luke (2011) whose primary interest was originally to investigate the influence of discourse context on the reanalysis of GP sentences. Across three self-paced reading experiments, they presented monolingual English speakers with target sentences with conjunction ambiguity that were preceded by a context sentence that is either GP-biased, Non-GP-biased or neutral. The texts were followed by forced-choice comprehension questions:

(9) Context sentence: There was a public outcry against the author (GP-biased) / the editor (Non-GP biased) / the publisher (Neutral) / of a racy new novel.

Target sentence: The publisher called up the editor (,) and the author refused to change the book's ending.

Question: How many people did the publisher call? (Answer: 1 or 2)

Having found little evidence for a selective effect of context on either RTs or off-line comprehension accuracy in the GP condition, Christianson and Luke (2011) also computed binomial models to probe whether RTs at disambiguation (e.g., *refused*) can predict the response accuracy for the comprehension questions, as part of their statistical analyses. These analyses revealed that the time the subjects spent does not modulate the likelihood to answer the comprehension questions accurately in any direction.

Likewise, Wonnacott et al. (2016) examined the nature of this relationship in an eye-tracking study in which the child native speakers of English who were aged 7-11 had to read GP sentences with preposed adjunct clauses. Even though the time spent reading the questions were found to predict comprehension accuracy for the GP sentences, it was observed that the go-past RTs at the disambiguating region was not associated with the off-line comprehension performance, which Wonnacott et al. (2016) interpreted as potential evidence against the idea that parsing failure is the source of lingering misinterpretations.

Since these latter two studies may have had possible power issues due to insufficient number of items, Qian et al. (2018) undertook a larger-scale investigation of this hypothesized relationship by running three tasks, two with self-paced reading and one with ERP paradigm, where the number of experimental trials was set substantially high (N = 64). In the first reading experiment, Qian et al. (2018) found no reliable association between the RTs and response accuracy. In the second one, the authors observed a significant relationship between the two measures which was albeit in the reverse direction such that the RTs tended to be slower when the comprehension questions were answered incorrectly, indicating that more revision effort could

sometimes result in poorer comprehension. In the ERP experiment, larger P600 effects were observed at the disambiguation segment, in line with some previous studies that attributed P600 component to the processes related to grammatical repair (Osterhout & Holcomb, 1992; Osterhout et al., 1994; Hahne & Friederici, 2002). Similar to the first reading experiment, there was again no significant relation between the P600 amplitude and off-line accuracy. Taking these findings together, Qian et al. (2018) argued that their findings contest the incomplete reanalysis view due to the overall lack of an expected relationship between RT/P600 magnitude and off-line comprehension measures.

Against the relative unpopularity of the incomplete reanalysis view, Chromý (2022) has recently demonstrated that the reanalysis of some GP sentences may never occur, with comprehenders constructing structural representations that are completely wrong and incoherent. Across three self-paced reading experiments, he tested adult monolingual speakers of Czech on a type of GP sentence containing coordination ambiguity available in their native language:

(10)

a. Ambiguous Condition:

Kluci honili psa a kočku v podkroví
 Boy-NOM.M.PL chase-3PL.M.PST dog-ACC.M.SG and cat-ACC.F.SG in attic-LOC.N.SG
 znepokojovali šediví hlodavci.
 worry-3PL.M.PST grey-NOM.M.PL rodents-NOM.M.PL

“Boys chased a dog and grey rodents in the attic worried a cat.”

b. Unambiguous Condition:

Kluci honili psa a kočk-a v podkroví
 Boy-NOM.M.PL chase-3PL.M.PST dog-ACC.M.SG and cat-NOM.F.SG in attic-LOC.N.SG
 znepokojovala šedivé hlodavce.
 worry-3SG.F.PST grey-ACC.M.PL rodents-ACC.M.PL

“Boys chased a dog and a cat in the attic worried grey rodents.”

The syntactic ambiguity in (10/a) lies in the fact that the NP *kočku* “cat” in the second clause contains an accusative case marker and thus it temporarily qualifies as a suitable candidate for the conjoined object of the verb *chase* along with the NP *psa* “dog”. This incorrect parse can be adopted until the following verb (i.e. *znepokojovali* “worried”) is encountered, after which the parser should initiate reanalysis to repair *kočku* as an object of this second clause. On the other hand, such an ambiguity is absent in (10/b) due to the fact the NP *kočka* is marked with nominative case. Chromý (2022) predicted that this type of GP structure should be particularly difficult to process and consequently reanalyze since the second clause has a non-canonical and relatively infrequent OVS order in Czech, posing an additional source of surprisal to the parser. To tap into the ultimate interpretation derived by the participants, four types of comprehension questions were asked:

(11)

- a. Honili kluci kočku? (“Did the boys chase the cat?”)
- b. Znepokojovali hlodavci kočku? (“Did the rodents worry the cat?”)
- c. Honili kluci psa? (“Did the boys chase the dog?”)
- d. Znepokojovali hlodavci psa? (“Did the rodents worry the dog?”)

Among these questions, (11a) referred to the initial misinterpretation that would be adopted as a result of garden-pathing, while (11b) referred to the correct parse of second conjoined clause. (11c) targeted the correct parse of the first clause, whereas (11d) referred to an interpretation that is not licensed at any point of the sentence. Along with the classic GP effects at the disambiguating word, Chromý (2022) found that the participants were unable to abandon the initially adopted misinterpretation, as demonstrated by accuracy rates ranging from 42% - 62% to the question (11a) following the GP stimuli. More importantly, however, their off-line comprehension performance were also considerably poor in other questions including (11d) whose accuracy rates yielded statistically significant effects of ambiguity. In other words, the participants sometimes failed to reanalyze these sentences completely to the point that they entertained an interpretation that was never mentioned in the sentence although they did experience a processing difficulty in the disambiguating region. Chromý

(2022) argued that the range of reanalysis difficulty could substantially modulate the GP recovery such that the reanalysis of some GP structures could be easier, whereas some may leave the readers with representations that are completely distorted and wrong.

Taken together, even though the lingering misinterpretations of GP sentences is a well-attested exemplar phenomenon of good-enough comprehension, the debates are yet to be settled with regard to their exact cause. On the one side of these debates, the nature of syntactic representations is highlighted (Christianson et al., 2001; Chromý, 2022), whereas the sole persistence of the initial misinterpretations is argued to be the driving cause for the other side (Sturt, 2007; Slattery, 2013).

2.3. Parsing and Reanalysis in Second Language

2.3.1. Overview

An important bulk of work has been undertaken in an aim to elucidate the nature of L2 processing for almost 20 years. The core motivation of this line of research has been to understand the extent to which L2 processing is qualitatively similar to L1 processing and where the causes of observed differences may be rooted in, giving way to the formulation of several theoretical accounts over the years (see Cunnings, 2017; Hopp, 2022 for reviews).

The existing models of L2 processing can be collapsed under two broad classes: (i) those conceiving the observed L1-L2 differences as quantitative (McDonald, 2006; Hopp, 2010) and those as qualitative (Ullman, 2001; Clahsen & Felser, 2006a, 2018; Cunnings, 2017). The first class of approaches maintain that the quality of linguistic representations and the psycholinguistic mechanisms involved in computing these representations are not fundamentally different in L2 speakers. The observed differences are usually attributed to the less automatic and slower nature of L2 processing due to the capacity limitations which arise from the cognitive burden of processing an L2 that is acquired later in life (McDonald, 2006; Hopp, 2010). One theoretical reflex of this line of reasoning is that L2 processing could be qualitatively similar to native parsing especially if individual differences are taken into

consideration in some domains such as working memory capacity or speed of lexical access (Hopp, 2018) since superior cognitive abilities may arguably reduce the computational limitations induced by processing an L2. The second class of accounts, on the other hand, predict fundamental L1/L2 differences with regard to the mechanisms involved in parsing, further explaining the observed differences with how neurocognitive mechanisms operate (Ullman, 2001), the depth of syntactic representations (Shallow Structure Hypothesis: Clahsen & Felser, 2006a) and memory retrieval mechanisms subserving sentence comprehension (Cunnings, 2017). Before turning to the reanalysis processes in L2 speakers, a brief description of these L2 processing accounts would be in order.

2.3.2. Second Language Processing Accounts

2.3.2.1. Capacity Limitation Accounts

The seminal studies of McDonald (2006) and Hopp (2010) have long motivated the idea that the differences between the processing of native and non-native language can be reducible to the capacity limitations, with the latter one being more cognitively taxing¹.

In her study, McDonald (2006) had L1 and L2 speakers of English provide acceptability judgments for auditorily presented sentences whose grammaticality was manipulated with regard to some properties such as word order, question formation, past tense and articles. The measures of processing speed, decoding ability and working memory capacity were also collected from the participants to examine how cognitive factors could contribute to their judgment performances. McDonald's (2006) results revealed that L2 group were slower and their grammaticality judgments were poorer in comparison to L1 group. All the participants' decoding skills and working memory capacity were associated with their judgment performance, implicating the role of general cognitive resources in both populations. To establish a causal nature for these findings, McDonald (2006) administered the same task to two groups of L1 English participants. One of these groups completed the task under stress in the form of concurrent memory load, white noise or speeded response deadline, whereas the

¹ Similar arguments have been recently proposed for heritage speakers as well (see Polinsky & Scontras, 2020).

other L1 group performed the task in the absence of a particular cognitive stress. This experiment revealed that the L1ers who were under stress provided significantly less accurate judgments compared to those who were not, paralleling the performance of the L2 speakers in the previous experiment. Taking these results together, McDonald (2006) maintained that L2ers' weaker performance in grammatical processing can be accounted for through the general processing problems.

In a similar vein, Hopp (2010) conducted two speeded judgment experiments testing the knowledge of case marking in advanced to near-native L2 speakers of German whose L1 varied in terms of the realization of case, i.e. Russian, English and Dutch. Among these languages, Russian has a rich overt case system, whereas English and Dutch feature overt case marking only for personal pronouns. In the first experiment, the participants had to judge the grammaticality of the sentences whose well-formedness was manipulated with respect to the word order (subject-object: SO vs. object-subject: OS) and the case of the NPs (nominative vs. accusative). Hopp (2010) observed that the performance of near-native L2 speakers was characterized by the case of properties of their respective L1s. Although the L1 Russian subgroup performed similarly to the native group, the accuracy of the judgments provided by the L1 English and L1 Dutch groups was somewhat at chance level, suggesting the role of L1 transfer in the processing of case information. In the second experiment, Hopp (2010) had a group of L1 German speakers perform the speeded judgment task under varying presentation speeds, similar to the use of stressors in McDonald (2006). It was found that the L1 subjects who completed the task at the predesignated maximum speed (i.e. 71 ms) gave relatively poor grammaticality judgments, in a similar way to the L2 participants who demonstrated L1 transfer effects in the previous experiment with standard speed (i.e. 250 ms). Hopp (2010) took these results as evidence that native and non-native processing may be similar, with the latter being merely slower and computationally demanding.

2.3.2.2. Declarative/Procedural Model

In contrast to the accounts outlined above, others argue that the differences stem from more distinct psycholinguistic mechanisms rather than global processing deficits. Of these, perhaps the most radical position is adopted by Declarative/Procedural Model

that couches L1/L2 differences within differing reliance on declarative and procedural memory (Ullman, 2001, 2005). Within this model, the declarative memory is argued to subserve the associative component of language such as the storage of individual lexical entries or irregularly inflected forms, whereas the procedural memory is responsible for the application of grammatical rules or processing of other combinatorial phenomena such as derivational or regularly inflected forms. Ullman (2001) argues that L2 speakers may make use of declarative memory more potentially due to the changes in brain plasticity at or around puberty, a time window that roughly corresponds to the critical period of L2 acquisition, whereas both systems operate in a relatively proportionate way in L1 speakers who acquire their languages from birth.

The empirical evidence for Ullman's (2001) claims come from some earlier studies that examined the L2 speakers' brain responses to the linguistic stimuli through measures such as fMRI and/or ERP. In an fMRI study, for instance, Dehane et al. (1997) had a group of L1 French speakers of L2 English listen to stories first in their native language and then in second language, while their brain activity was monitored. Although most of the activation was observed in temporal lobe regions in both L1 and L2, it was substantially left-lateralized while the stories were listened to in L1. Notably, the direction of lateralization was rather dispersed and less consistent when L2 stories were listened, suggesting a lesser degree to which procedural memory system is utilized in L2 processing. Likewise, Hahne and Friederici (2001) run an ERP experiment where L1 and L2 speakers of German listened to sentences that were experimentally manipulated in terms of their semantic and syntactic well-formedness. While both participant groups were able to detect semantic anomalies, as indexed by a robust N400 amplitude, it was only the L1 group who evinced sensitivity to the syntactic anomalies in the form of LAN and P600 effects, implying the unstable nature of procedural mechanisms in L2 learners from Ullman's (2001) standpoint.

2.3.2.3. Shallow Structure Hypothesis

In the same "qualitative" side of the theoretical debates, Clahsen and Felser (2006a, 2018) argued that the differences may mostly lie in the type of linguistic information utilized during sentence processing. In what they dubbed as Shallow Structure Hypothesis (hereafter SSH), Clahsen and Felser (2006a) posited that L2 speakers can

integrate lexical-semantic, pragmatic or surface-level cues during real-time processing as successful as L1 speakers, but they may have difficulty in constructing detailed and hierarchical (morpho)syntactic representations. Although SSH was often misunderstood to have claimed that shallow processing is a unique property of the L2 processor (Birdsong & Gertken, 2013; Omaki & Schulz, 2011), Clahsen and Felser (2006b, 2018) emphasize that this account is actually built upon the existing dual-pathways theories. That is, the human parser may have a universal way of processing the linguistic input via different pathways, but “the likelihood of heuristic pathway dominating is greater in L2 compared to L1 processing” (Clahsen & Felser, 2018, p. 697), which ultimately restricts the availability of algorithmically-driven, full-parsing route to a particular extent as described in Figure 1:

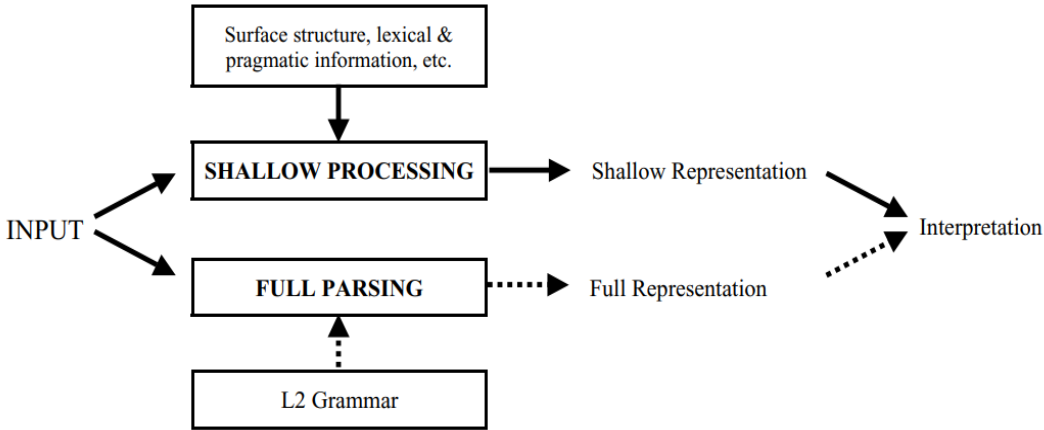


Figure 1. The architecture of L2 parser under SSH (Clahsen & Felser, 2006b, p. 118)

Importantly, SSH highlights that the observed cases of non-native-like processing cannot be solely ascribed to incomplete acquisition of grammar or L1 transfer. It could be maintained that the acquisition of a body of sufficient and rich grammatical knowledge is necessary to reach native-like parsing routines; however, the findings supporting SSH suggest that even advanced L2ers may process the grammatical information in a differing manner from L1ers on-line despite the fact that they demonstrate a native-like command of the relevant grammatical structures off-line (Clahsen et al., 2013; Felser et al., 2003; Felser & Cunnings, 2012; Boxell & Felser, 2017; Papadopoulou & Clahsen, 2003), with no compelling evidence of L1 interference.

Turning to the empirical evidence for SSH, consider the sentence (12) where the RC can be attached either to the NP *the secretary* (henceforth NP1) or *the professor* (henceforth NP2):

(12) The dean liked the secretary of/with the professor who was reading a letter.

In previous monolingual sentence processing studies, it was observed that young adult monolingual English speakers choose to attach the RC to the NP2 in line with the Recency Principle, although some languages such as Spanish and German were shown to have an NP1 preference against the prediction of this principle (Carreiras & Clifton, 1999; Cuetos & Mitchell, 1988). Importantly, it has also been demonstrated that the lexical-semantic property of the preposition can modulate the comprehenders' attachment preference such that the preposition *with* or its translation equivalents in other languages can strongly bias NP2 attachment even though the language in question favors NP1 attachment as a default parsing choice. Under construal theory (Frazier & Clifton, 1996), this is explained by the fact that attaching a modifier to an NP that is outside the thematic domain is too costly for the parser as the preposition *with* projects a thematic domain of its own. On the contrary, since *of* assigns syntactic case and cannot form a thematic domain, no such overriding in favor of NP2 is observed.

Using such sentences, Felser et al. (2003) investigated the global ambiguity resolution strategies of non-native speakers of English who had Greek and German as L1s. Among these languages, the former is argued to favor NP2 attachment (see also Papadopoulou & Clahsen, 2003), whereas the latter biases NP1 attachment. Despite replicating the aforementioned patterns in their L1 control group, Felser et al. (2003) observed that the L2 group showed a different tendency in that they displayed an NP2 preference only when the complex NP had the preposition *with*. However, there was no reliable attachment pattern when this container NP had the preposition *of*. In addition, the participants' parsing strategies were not modulated by the attachment preference of their respective L1s. Based on these findings, Felser et al. (2003) argued that L2 speakers are more guided by lexical-semantic cues than phrase structure-based parsing principles.

In a further study, Marinis et al. (2005) investigated the L2 processing of sentences that contain wh-gaps at the intermediate position as exemplified in (13), building on Gibson and Warren (2004):

(13)

- a. The nurse wh_i the doctor argued e_i that the rude patient had angered e_i is refusing to work late.
- b. The nurse wh_i the doctor's argument about the rude patient had angered e_i is refusing to work late.
- c. The nurse thought the doctor argued that the rude patient had angered the staff at the hospital.
- d. The nurse thought the doctor's argument about the rude patient had angered the staff at the hospital.

In (13a/b), the NP *the nurse* is base-generated as an object of the verb *angered* in the deep structure along with its wh-operator, and moves up to the specifier of the main clause. Since the Subjacency principle bans the movement across two clause boundaries in a single step (Chomsky, 1981), the NP in (13a) should first move to an intermediate landing site which corresponds to the specifier position of the lower complementizer phrase (e.g. *that*) to avoid ungrammaticality, whereas in (13b) there is no such structural requirement. From this formal logic, it is reasoned that the intermediate landing site can facilitate the resolution of such dependencies as the parser can reactivate the filler at this site before finally integrating it with its original thematic position *angered* __. In their experiment, Marinis et al. (2005) observed this expected parsing behavior in their L1 group, with elevated RTs at the intermediate gap position of (13a) relative to (13b), as an index of active-gap filling, whereas this was not observed in the L2 group who nevertheless comprehended the sentences accurately off-line. In line with Felser et al. (2003), there was no selective effect of L1 background in that the presence/lack of successive wh-movement in the participants' L1 did not modulate the extent of this sensitivity. Marinis et al. (2005) argued that syntactic information is under-utilized in non-native language processing.

In addition to global ambiguity and filler-gap dependencies, the predictions of SSH were later tested in other phenomena such as anaphora resolution as well. For example,

Felser and Cunnings (2012) examined to what extent the processing of reflexive pronouns is guided by the binding principles in native and non-native speakers of English. In an off-line antecedent selection task, the L2 group was observed to perform like the native speaker group in establishing a correct binding domain between the reflexive and its grammatically licit antecedent, indicating that they are aware of the requirements of Principle A. In the eye-tracking experiment, the gender congruence between the reflexive and the candidate antecedents were manipulated, following Sturt (2003):

(14) James/Helen has worked at the army hospital for years. He/She noticed that the soldier had wounded himself/herself while on duty in the Far East.

The NP *soldier* was the only licensed antecedent for the reflexive (e.g., *himself/himself*) in accordance with requirements of local binding constraints. The NP *James/Helen* was the discourse-salient antecedent despite being syntactically inaccessible to bind the pronoun. Against the off-line study, the L2ers had slower first-pass RTs at the segment of pronoun when it mismatched to the antecedent that was inaccessible, suggesting that they attempted to establish a binding relationship between these two at earlier stages of processing. It was only in the late reading measures (i.e. re-reading) that they showed evidence of establishing a licit local binding relationship in the form of GMM effects when the reflexive pronoun did not match to the syntactically accessible antecedent's gender. On the contrary, this was not the case for the L1 group who was observed to apply the local binding constraints immediately at first-pass RTs. This was interpreted as evidence that the application of binding conditions is delayed in non-native speakers to the point that they may initially resort to the discourse-driven coreference, in line with the predictions of SSH.

2.3.2.4. Interference Account

Recently, Cunnings (2017) outlined an alternative L2 processing account where he claimed that non-native speakers can create hierarchical and detailed syntactic representations in a similar fashion to the L1 speakers, but they could be more susceptible to interference of certain cues in the course of memory retrieval operations underlying on-line sentence comprehension.

This account is built upon the cue-based models of sentence processing whose implications were explored and tested across several phenomena mostly in adult L1 populations (e.g. Lewis et al., 2006; Lewis & Vasishth, 2005). Under such models, sentence comprehension is argued to proceed through continuous encoding, storage and retrieval of memory chunks containing linguistic information. To interpret the sentence “the key to the cabinets unsurprisingly was rusty from many years of disuse” (taken from Wagers et al., 2009, p. 211) successfully, for instance, the comprehenders have to first encode the NP *the keys*. When the verb *was* is encountered, they have to retrieve the memory representation of *the keys* since this NP matches an array of retrieval cues to compute an agreement relationship with the verb (e.g., [+singular] and [+nominative]). As all the encoded memory chunks are compared against each other in parallel for retrieval, one ramification of such architecture of sentence processing is that the comprehenders may sometimes be influenced by similarity-based interference since multiple chunks, rather than one, can provide a match for retrieval simultaneously. In the case of the agreement phenomenon mentioned above, for example, it was widely demonstrated that the readers fleetingly accept such sentences as grammatical when the distractor NP (e.g., *cabinets*) and the verb (e.g., *were*) are both encoded as plural, a phenomenon known as “agreement attraction” in the literature. Although debates are still ongoing with respect to the exact component of working memory that is responsible for these attraction effects, most studies place the blame on the fallibility of the retrieval operations, arguing that the backward memory search initiated by the verb *were* results in the occasional retrieval of the NP *the cabinets* due to the fact that both contain a common [+plural] cue (Lago et al., 2015, 2019; Wagers et al., 2009; Tanner et al., 2014; cf. Hammerly et al., 2019).

In the context of L2 processing, Cunnings (2017) argues that potential divergence from native baseline can be attributed to how various kinds of retrieval cues are utilized, abstracting away from the quality of structure-building operations or grammatical knowledge underlying the L2 parser. To this end, he maintains that some previous findings taken to indicate shallow parsing can be re-interpreted in light of a new theoretical framework that capitalizes on retrieval mechanisms. For instance, the fact that the L2 speakers in Felser and Cunnings (2012) attempted to establish a discourse-driven coreference in violation of Principle A can be alternatively explained by the non-native speakers’ assigning a greater weight to a potential [+topic] cue for retrieval,

whereas clausal proximity could be a more utilized cue in the native speakers for anaphora resolution. This reasoning is based on the assumption of cue-based parsing models that the interpretation of referential dependencies takes places through the search of the previously encoded potential antecedents in memory and subsequent retrieval of the one that provides the best match for the reflexive pronoun with respect to certain retrieval cues (e.g., [+/- masculine], [+/- singular], [+/- 3rd person]).

Likewise, in an ambiguity resolution study, Pan et al. (2015) had previously reported that referential context strongly affected the L2 speakers' strategies of RC attachment on-line such that they displayed NP1 preference in NP1-favoring and NP2 preference in NP2-favoring contexts, while the L1 control group was not reliably affected by the context manipulation. Although the authors originally argued that such findings could support the idea that L2 parsing could be more likely to be guided by the extra-sentential information in line with SSH, Cunnings (2017) noted that this particular finding can be also re-captured with a potential L2 susceptibility to retrieval interference. To be more precise, as cue-based parsing models would expect the relative pronoun *who* to trigger the search and retrieval of previously encoded NPs to attach, it could be that L2 speakers may have been more prone to retrieving the one that was favored by the discourse context relative to the L1ers, eliminating the need to argue for a difference based on structure building operations contra what is argued by Pan et al. (2015).

Another important set of findings that Cunnings (2017) cited as evidence of L2ers' being disadvantaged for retrieval are those suggesting that they experience greater difficulty in abandoning the incorrect interpretation of GP sentences (Jacob & Felser, 2016; Pozzan & Trueswell, 2016), as will be reviewed in the following text. Based on Slattery et al.'s (2013) earlier work, Cunnings (2017) claimed that the source of such difficulty may not relate to the shallow parsing in that L2 speakers may actually construct the globally correct syntactic representation of such sentences as successful as L1 speakers, but they could be rather more susceptible to misretrieve the initial parse that lingers in memory compared to the L1 speakers².

² Since an extensive summary of the findings claimed to reflect an L2 vulnerability to retrieval interference would be beyond the scope of this chapter, interested readers are referred to Cunnings' (2017) keynote article for further information.

2.3.3. Reanalysis Processes in L2 Speakers

As in the L1 processing literature, GP sentences have long served as a useful phenomenon to shed light on the parsing routines of L2 speakers. It has been shown that L2 speakers experience reanalysis difficulty similar to the L1 speakers when they encounter disambiguating information (e.g., Juffs & Harrington, 1996; Frenck-Mestre & Pynthe, 1997; Juffs, 2004; Rah & Adone, 2010). In addition to these studies whose sole focus was on the L2 parser's on-line sensitivity to the disambiguating information, another important line of research has taken advantage of GP phenomenon to explore whether L2 speakers can utilize various linguistic cues incrementally like L1 speakers, such as plausibility (Roberts & Felser, 2011; Hopp, 2015), verb bias (Frenck-Mestre & Pynthe, 1997; Dussias & Cramer-Scaltz, 2008; Şafak & Hopp, 2022), subcategorization (Brothers et al., 2021), referential context (Pozzan & Trueswell, 2016), case (Jackson, 2008; Gerth et al., 2017) as well as aspect information (Roberts & Lizska, 2021). Among the studies that focused on plausibility, for example, Roberts and Felser (2011) investigated how the plausibility of the ambiguous NP modulates reanalysis. Following the original design of Pickering and Traxler (1998), they tested native and advanced L2 speakers of English on the following “strong” and “weak” GP sentences (15) in a self-paced reading experiment, by manipulating the semantic fit between the ambiguous NP and the preceding verb:

- (15) a. While the band played the song (the beer) pleased all the customers.
b. The inspector warned the boss (the crimes) would destroy many lives.

For both the strong and weak GP sentences, the authors observed that the effects of (im)plausibility were more immediate and pronounced for the L2 group such that they showed greater sensitivity to the implausible NPs relative to the L1ers, as evinced by the RTs in this segment. While the learners showed the reversal of these plausibility effects at the disambiguating region for the weak GPs (15b), these effects were absent in the strong GP sentences. Moreover, the off-line comprehension data revealed that the L2ers displayed significantly lower accuracy to the plausible strong GPs (15a), whereas L1ers performed similarly in both conditions (see Felser et al., 2012 for similar plausibility effects in island phenomena).

Similarly, Dussias and Cramer-Scaltz (2008) examined the incremental use of verb frequency information in native and Spanish L2 speakers of English using GP phenomena. Across two self-paced reading experiments, the participants read sentences where the verbs with direct object (i.e. *confirm*) and sentential complement bias (i.e. *admit*) took either a direct object or sentential complement. With this crossed design, the post-verbal NPs either matched or mismatched to the sub-categorization bias of the preceding verb. Dussias and Cramer-Scaltz (2008) report that both groups were slower when they read the disambiguation region in the mismatch conditions, indicating that they anticipated an NP following the direct-object bias verb and a full clause following the verbs with sentential complement bias. The authors took these patterns as evidence that L2 speakers can use verb bias incrementally as reliably as native speakers.

As the ultimate focus of such studies was mostly to investigate the L2 speakers' ability to integrate a range of cues incrementally during real-time processing, they were not specifically designed to examine the ultimate interpretation that native and non-native speakers arrived, with no unambiguous baseline condition. Still, a few studies examined the success with which L2ers can conduct reanalysis and arrive at the globally correct interpretation by using various measures. For instance, Pozzan and Trueswell (2016) tested imperative *put* sentences that contain syntactic ambiguity between a modifier interpretation and goal interpretation (see. Trueswell et al., 1999):

(16) Put the frog (that is) on the napkin onto the box.

Their subjects who were L1 and intermediate L2 speakers of English listened to these sentences, while they viewed a visual scene that featured referents of the corresponding sentences, and then acted out these commands on the screen through mouse. The eye-movement patterns revealed that both groups were lured into the goal interpretation; however, the latter group performed significantly less accurate actions overall, particularly in the ambiguous condition, which suggests an increased L2 difficulty with reanalysis. Based on the similar findings in children (Trueswell et al., 1999; Choi & Trueswell, 2010), Pozzan and Trueswell (2016) speculated that the difference between their groups could be accounted for through the overloading of inhibition resources while processing an L2.

Adopting an experimental design similar to Sturt's (2007), Jacob and Felser (2016) explored the reanalysis behavior of L2 speakers, testing the sentences (17) below in an eye-tracking experiment:

(17) While the gentleman was eating(,) the hamburgers were still being reheated in the microwave.

As part of this design, the ambiguous version of the sentence is disambiguated in two steps. In the main clause, the segment with the first auxiliary verb (e.g., *were still*) constituted syntactic disambiguation, whereas the region *being reheated* served as a form of semantic disambiguation due to the fact that something cannot be eaten and reheated at the same time. Jacob and Felser (2016) reasoned that the readers should experience processing difficulty at the latter region if the incorrect interpretation is not fully abandoned. If the reanalysis is successfully abandoned upon disambiguation, on the other hand, no such difficulty should be observed. The authors also introduced end-of-trial comprehension questions to further probe the final interpretation. Consistent with the prior research, the authors observed that the first pass RTs were longer at the syntactic disambiguation for the ambiguous relative to the unambiguous condition. Nevertheless, these GP effects were smaller for the L2 group in later reading measures (e.g., regression-path duration and total RTs), which was taken as evidence that L2 speakers could be more prone to leave the revision uninitiated. While there was also evidence of semantic persistence on-line at the semantic disambiguation region (Sturt, 2007), the L2 group was found to be overall less accurate in their responses. Highlighting the fact that the GP meaning is highly plausible across their experimental items, Jacob and Felser (2016) concluded that L2 comprehension could employ the good-enough route more frequently.

Fujita and Cunnings (2020, 2021) have recently carried out a series of studies targeting the reanalysis processes in L2 speakers. In one such study, Fujita and Cunnings (2020) tested temporarily ambiguous sentences with filled (18) and non-filled-gaps (19):

(18)

a. The child noticed the brush which the maid was cleaning the floor with very carefully.

b. The child noticed the brush with which the maid was cleaning the floor very carefully.

(19)

a. The girl was in the school bus which Alan was driving very slowly near earlier today.

b. The girl was in the school bus near which Alan was driving very slowly earlier today.

In (18-19/a), the filler NPs *the brush* or *the school bus* can be initially integrated at a potential gap position following the relativized verbs (e.g., *clean/drive*); however, the preposition (e.g., *with/near*) is later understood to be the true ultimate gap. In (18a), a following NP immediately causes garden-pathing through a filled gap effect (Stowe, 1986), whereas the initial misanalysis is canceled only when the preposition is encountered further downstream in (19a), hence containing non-filled-gaps. The items (18/19b) served as unambiguous conditions since the preposition is fronted. In the first eye tracking experiment, the authors further manipulated the consistency of the correct meaning of the target sentence with that of a follow-up sentence.³

(20) Some chores needed to be done. The child noticed (with) the brush which the maid was cleaning the floor (with) very carefully. It seemed that the maid was cleaning the floor / the brush while thinking about dinner.

In the third sentence, the object NP served as a probe to tap into the persistence of misanalysis from the preceding sentence. In the inconsistent continuations, this NP (e.g., *the brush*) was incongruent with the correct meaning of the preceding sentence, whereas it was congruent with the initial misanalysis. In consistent continuation, the NP (e.g., *the floor*) was congruent with the globally correct parse; however, it was incongruent with the initial misanalysis. In addition to having experienced filled-gap effects in the second sentence, both participant groups read the inconsistent NP quicker following the ambiguous sentences, whereas this pattern was reversed for the

³ For the sake of brevity, the prepositions are put in parentheses to mark ambiguity and consistency manipulations.

consistent NPs such that they elicited longer total RTs after the ambiguous sentences, indicative of lingering misinterpretations. In the second eye-tracking experiment, Fujita and Cunnings (2020) made use of the same design for the non-filled-gaps:

(21) The girl was in the school bus (near) which Alan was driving very slowly (near) earlier today. Alan was driving (near) the school bus very patiently on the road. It was extremely crowded.

For these sentences, the presence or absence of the preposition served as a probe of lingering misinterpretations in the second sentence. The L2 group showed an overall effect of consistency only in the regression path measure, whereas the native control group did in other measures as well at the NP following the preposition (e.g., *the school*). As regards the source of this pattern that was not present in the first experiment, Fujita and Cunnings (2020) argued that the L2 speakers could be more likely to fail to complete the reanalysis for this particular structure probably because the preposition may not be a cue that is as informative as an overt NP that immediately disambiguates the structure in the filled-gap configuration.

In another study with classic GP sentences containing preposed adjunct clauses, Fujita and Cunnings (2021) run experiments to elucidate the potential source of increased reanalysis difficulty in L2 speakers. Based on Slattery et al.'s (2013) earlier work, they identified two possibilities. The first one relates to the incomplete nature of syntactic reanalysis in L2 speakers, which they took to be broadly in accordance with the predictions of SSH (Clahsen & Felser, 2006a; 2018) as well as the assumption that lingering misinterpretations in L1ers stem from memory traces rather than the parsing failure. The other possibility, on other hand, could pertain to the L2ers' elevated difficulty in inhibiting the traces of the initial parse despite having computed a globally correct structure, as predicted by Cunnings' (2017) interference account. To tease apart between these possibilities, Fujita and Cunnings (2021) used the structural diagnostic first employed in Slattery et al. (2013)'s first experiment. Replicating Slattery et al. (2013), Fujita and Cunnings (2021) observed GMM effects irrespective of the ambiguity and group factor, which they took to suggest that both groups complete syntactic reanalysis successfully. The authors also conducted two off-line experiments, one with comprehension questions following the full presentation of the sentences and

the other with sentence-picture matching. In the first one, the participants answered forced-choice questions that tap into the subordinate (e.g., *What happened?*) and matrix clause (e.g., *Who laughed very happily?*). The subordinate clause questions had two answer choices in declarative form (e.g., *the mother dressed herself* vs. *the mother dressed the baby*), whereas the one for matrix clause had the options of referent NPs in the sentence (e.g., *the mother* vs. *the baby*):

(22) After the mother dressed(,) the baby in the living room laughed very happily.

Subordinate clause: What happened?

Matrix clause: Who laughed very happily?

As for the picture-matching task, the subjects had to select the picture that they thought best fit the content of sentence they read in full-form (22). For both subordinate and matrix clause, the picture pairs either depicted the correct or incorrect action in the corresponding clauses. In both experiments, although there was evidence of lingering misinterpretations in the subordinate clause probes, Fujita and Cunnings (2021) also found effects of ambiguity in the matrix clause probes, indicating that the subjects sometimes failed to reanalyze the GP completely —contrary to the eye-tracking experiment with reflexives. Moreover, these effects were marginally found to be qualified by the group factor such that L2ers were more likely to respond inaccurately to the questions following GP trials. One issue with Fujita and Cunnings' (2021) materials is however that they are somewhat confounded with an additional layer of ambiguity within the GP interpretation itself. Note, for instance, that the above sentence (22) also contains another ambiguity where either the (initially understood) act of the baby being dressed can occur in the living room or the baby herself can be the one located in the living room in line with the disambiguated parse of the sentence. If the participants in Fujita and Cunnings (2021) mostly parsed the prepositional phrase as the modifier of dress in line with the former interpretation, then the surprisal induced by disambiguation could have been aggravated since not only the noun should be repaired as the subject, but also the prepositional phrase as the modifier of this noun. In this respect, it may well be the case that the authors could have found ambiguity effects in both subordinate and main clause probes simply because being forced to

resolve two ambiguities at the same time overtaxed the subjects' ability to repair the ambiguity, potentially causing them to stop parsing the sentences sometimes.

Cunnings and Fujita (2021) further examined the potential role of the individual differences in L2 proficiency in recovering from GP effects and arriving at the globally correct interpretation. To this end, they had their participants read sentences with coordination ambiguity and answer comprehension questions (e.g., *Did Ken wash the cat?*):

(23) Ken washed the dog (and/while) the cat in the garden played with a ball.

The L2 group also had to complete a placement test as an independent measure of proficiency. There was no difference between the groups in terms of the either size of GP effects or off-line accuracy. As regards the proficiency, Cunnings and Fujita (2021) found that the individual differences in proficiency scores did not affect either on-line or off-line performance of their L2 group in terms of GP recovery.

CHAPTER 3

THE PRESENT STUDY

3.1. Aim of the Study

In the previous chapters, the literature on processing and comprehension of GP sentences by native and non-native speakers was summarized. While it is well-established that adult L1 speakers do not completely abandon the initial GP misinterpretation in line with the good-enough processing models, less is known with regard to how L2 speakers parse and comprehend these sentences, with the existing evidence coming from only a handful of studies (e.g., Jacob & Felser, 2016; Fujita & Cunnings, 2021). As mentioned before, the existing studies were mostly limited to the utilization of GP phenomena as a diagnostic to test whether a variety of cues are used incrementally on-line by L2 speakers. To address this gap, the present study examines the processing of GP sentences by L1 and L2 speakers of English in a web-based self-paced reading experiment. In the experiment, off-line comprehension accuracy is tested using comprehension questions that target both the initial misinterpretation and the success of downstream syntactic repair in a similar fashion to Christianson et al. (2001). These questions respectively target subordinate and main clauses of the experimental sentences as exemplified below:

While Anna dressed(,) the baby who was cute and small fell off the bed.

Subordinate clause question: Did Anna dress the baby?

Matrix clause question: Did the baby fall off the bed?

In addition, how individual differences in proficiency may modulate GP recovery among L2 speakers will be explored.

3.2. Research Questions and Predictions

To reiterate from Chapter 1, the research questions of the current study are as follows:

1. Do L2 speakers of English differ from L1 speakers of English in terms of GP recovery?
2. Does the level of proficiency modulate the degree to which GP recovery is successful within L2 speakers?

For the first research question, in line with the predictions of SSH (Clahsen & Felser, 2006a, 2018), it is predicted that L2 speakers of English will experience more failure of GP recovery relative to the L1 control group. Since SSH posits that the shallow or heuristic route of processing is more likely to dominate in L2 speakers compared to their L1 counterparts, it could be thus predicted that L2ers will be more strongly committed to the semantically plausible GP interpretation such that they should commit greater amount of comprehension errors in the ambiguous condition⁴. On the other hand, if there is no difference between L1 and L2 speakers in terms of relying on good-enough route of processing, we expect comparable between-groups patterns in our on-line and off-line indices of reanalysis.

For the second research question, it is expected that the increases in proficiency should be able to predict the success of GP recovery from lower to higher proficiency levels, in accordance with previous studies suggesting that parsing routines develop with increasing L2 proficiency (e.g., Cheng et al., 2021; Dallas et al., 2013; Jackson & van Hell, 2011). Regarding the performance at the higher levels, finding evidence that the highly proficient L2 speakers commit similar amount of errors with the L1 speakers in the ambiguous condition would lend support to the idea that native and non-native language processing can be similar especially with sufficient amount of L2 proficiency.

⁴ If such increased failure is only observed in the subordinate but not matrix clause questions, this could be also expected by Cunnings' (2017) interference account in addition to SSH. However, due to lack of solid basis for why L2ers should be more vulnerable to retrieval interference in general, we prefer to remain agnostic towards such idea for present purposes.

3.3. Method

3.3.1. Participants

The study included two groups of participants, namely 37 L2 speakers of English (21 males, 15 females, one other, mean age = 21.08, range = 18-32) and 28 native speakers of English (13 females, 13 males, one other, one preferred not to disclose, mean age = 29.07, range = 18-50). In the L1 group, two additional participants who reported to be bilingual speakers of Spanish and German respectively were excluded as their status of nativeness could not be ascertained. The participants were recruited over the Internet, with the L2 participants mostly from METU student groups in social media and the L1 participants from the subforum r/SampleSize in Reddit as well as the expat native speakers of English who reside in various cities of Turkey. To increase the variability among L2 participants in terms of proficiency, intermediate and upper-intermediate level students from METU English Preparatory School were also included.

The L2 participants were all native speakers of Turkish and started to learn English predominantly in school environment at a mean age of 8.86 (range = 4-16; SD = 2.71). To further establish the level of proficiency, the L2 group was asked to rate their each of the four skills in English; namely reading, speaking, writing and listening, on a nine-point scale through a demographic form administered prior to the experiment. For the current study, the application of self-ratings was chosen for practical reasons. Since participants may sometimes tend to attrite by not completing a separate test on Internet-based studies, the current study did not make use of a standardized test of proficiency in order to prevent such cases of attrition. Despite the fact that they are subjective measures in nature, their scores were frequently observed to correlate with those of standardized placement measures (Marian et al., 2007; Sabourin et al., 2014), so it was reasoned that they can be used to determine the proficiency level of the L2 group.

For each participant in the L2 group, general proficiency scores were computed by calculating the average of these skill ratings. In other words, a composite measure of proficiency was established by taking the average of listening, speaking, writing and

reading ratings out of nine for every participant. The descriptive information including participants' mean skill self-ratings and age of acquisition of English can be seen in Table 1:

Table 1. Background of the L2 participants

Characteristic	Mean	SD	Min	Max
AoA ¹	8.86	2.71	4.00	16.00
Reading	7.97	1.07	5.00	9.00
Writing	7.16	1.36	4.00	9.00
Speaking	6.95	1.63	3.00	9.00
Listening	7.73	1.30	5.00	9.00
Overall ²	7.45	1.16	4.50	9.00

1. Age of Acquisition of English
2. Mean of the four skills

All participants reported themselves to have normal or corrected-to-normal vision, and except for nine of them, the majority was right-handed. To ensure that the collected data would reflect a behavior typical of the tested populations, extra care was taken not to recruit individuals who have a formal background of linguistics (e.g. students of a language-related programme) due to the popularity of the investigated phenomenon in the field⁵. As an incentive as well as a form of compensation, the participants in the L2 group were included in a raffle to grant randomly selected three of them with discount vouchers. This thesis was approved by the Human Subjects Ethics Committee of METU (Appendix A).

3.3.2. Materials and Design

The critical items consisted of 20 pairs of sentences that were manipulated with respect to the ambiguity and the type of the comprehension question asked, yielding a 2 x 2 factorial design, as repeated from the previous example:

While Anna dressed(,) the baby who was cute and small fell off the bed.

⁵ This decision was further motivated by the fact that two senior students of Foreign Language Education programme at METU were able to name the manipulation in terminological words when asked in an earlier pilot study — despite a filler to total item ratio of greater than 2:3. These two individuals did not take part in the current experiment.

Subordinate clause question: Did Anna dress the baby?

Matrix clause question: Did the baby fall off the bed?

As part of the ambiguity, the presence of comma between the subordinate clause verb (e.g. *dress*) or the ambiguous NP (e.g. *the baby*) was manipulated. In the ambiguous condition, the NP is initially understood as the object of the previous verb, but it has to be revised as the subject of the upcoming verb *fell* due to the lack of a comma. In the unambiguous condition, this ambiguity was not present since the sentences were disambiguated by a comma. As for the comprehension questions, they either targeted the content of the subordinate clause or main clause, with the former tapping into the persistence of the direct object misinterpretation and the latter into whether the ambiguous NP was correctly repaired as the subject of the late-coming verb. The correct answer was *No* for the subordinate clause questions, while it was *Yes* for the matrix clause questions.

All the stimuli including filler items were adapted from Huang and Ferreira (2021) and the third experiment in Christianson et al.'s (2001) study. The reason why these materials are chosen is that the subordinate clauses are always predicated by the so-called reflexive absolute transitive (RAT) and reciprocal verbs such as *dress* or *meet* rather than optionally transitive verbs (OPT) like *hunt*. When used intransitively, these verbs are most commonly interpreted in accordance with a reflexive or reciprocal argument frame. For instance, "Anna dressed" is understood as "Anna dressed herself", or "the grandparents met" mean that the action of meeting was undertaken mutually by the grandparents. In the case of GP phenomena, RAT verbs result in the initial misanalysis and globally correct interpretation having two distinct semantic interpretations with the former being "Anna dressed the baby" and the latter "Anna; dressed \emptyset_i " in a non-logophoric configuration of reflexivity. With OPT verbs, the observed cases of reanalysis failure could be potentially confounded by pragmatic inference. In the sentence "While the man hunted the deer ran into the woods", for example, the participants may tend to opt for *Yes* responses simply because a fully revised interpretation requires positing an unspecified direct object (e.g., hunted *something*). In an aim to refrain from postulating such an additional non-existent referent, they can consequently endorse the GP interpretation based on pragmatic

reasoning as there is no huntable entity in the context other than the NP *the deer*⁶. By exclusively using the items with RAT verbs, we aimed to minimize the likelihood that the participants would give such inference-based responses (see Christianson et al., 2001 for an exhaustive discussion of RAT vs. OPT verbs).

In all critical items, the temporarily ambiguous NP was modified by a relative clause (e.g., who was cute and small) as a means of lengthening the ambiguity since previous research suggest that reanalysis failure is more likely when subjects commit to the ambiguity longer (e.g., Ferreira & Henderson, 1991). In both studies from which the materials were adapted, the relativizer was originally *that* for most of the items. In the present study, this pronoun was replaced with *who* for the ambiguous nouns that refer to a human entity (e.g., the baby, the boy) in order to make the sentences sound more natural. Furthermore, to allow for spillover effects to surface in on-line reading, an adverbial or prepositional phrase (e.g., quite happily) was added to the end of some sentences taken from Christianson et al. (2001) who did not take this issue into consideration probably because their experiments were mostly off-line.

Following such adaptations, the critical items were distributed over four lists in a Latin Square design such that there were five items per each possible ambiguity+question combination in each list. In this way, each version of the items was seen only once. All the lists were pseudorandomized manually to ensure that the items from the same condition would not appear consecutively. 40 filler items were added for a total of 60 trials per list and they appeared at least once between the experimental sentences. Containing a variety of syntactic structures in general, a portion of the filler items resembled to the experimental sentences in syntactic complexity by containing subordinate clauses with unambiguous direct object NPs (e.g., When the customer ordered a dessert the coffee that the waiter was preparing boiled rapidly). Half of the filler sentences were accompanied by a comprehension question, whereas the other were not. In the filler trials where comprehension questions were asked (N = 20), the correct answer was *Yes* for half of the items and *No* for the other half in order to balance the yes/no bias along with the critical items. These comprehension questions targeted

⁶ Note also that OPT verbs were often observed to incur numerically high rates of inaccurate responses on unambiguous trials as well in previous research (e.g. Christianson et al., 2001; Qian et al., 2018), which is another reason why we avoided using them in our stimuli.

the various parts of the filler sentences, mimicking the manipulation of question type in the critical items.

3.3.3. Procedure

The sentences were presented in a self-paced reading experiment programmed in PciBex (Zehr & Schwarz, 2018), using a non-cumulative moving window paradigm. As part of this paradigm, the subjects read the sentences on word-by-word basis where one word was visible at a time with the previous one being masked with each button press.

The participants completed the study on their computers remotely. They first had to read a consent form (Appendix B and C) and tick a following box in order to give their consent. In the consent form, they were specifically asked to minimize potential sources of distraction around themselves and keep their attention on the screen in the course of the experiment. After the informed consent, a demographic form was administered in order to collect biodata such as age, gender, and self-rated English proficiency for the L2 group (Appendix D and E).

Following these, a set of detailed written instructions were shown. The participants were informed that they were supposed to read sentences on a word-by-word basis using the Space button to uncover each word and sometimes answer comprehension questions with the keys 1 (for Yes) and 2 (for No) in the number pad. They were asked to read the sentences at their own natural reading speed as much as possible. In order to be familiarized with the requirements of the experiment, they were later routed to a practice session that contained five trials, three of which featured comprehension questions. The items presented in this part bore no resemblance to the critical items in terms of syntactic structure. When the practice session ended, the participants were instructed to press Enter to initiate the main session of the experiment or Backspace if they would like to take the practice session again. The sentences were displayed in black 14pt Times New Roman font on a light gray background. All of the experiment took approximately 15 minutes to complete. Even though the duration of the experiment was relatively short, an optional self-timed break was still offered after the first half of the trials in order to alleviate the impact of fatigue and/or boredom effects

resulting from word-by-word reading and minimize the resultant probability of attrition.

To be able to provide the participants with necessary directives in their own dominant language, especially for the convenience of the L2ers with lower proficiency, two instances of the experiment were programmed. One of these provided the written instructions, consent and demographic form in Turkish, whereas the other did in English. The former link was intended for the L2 group and the latter for the native control group. Except for this distinction, the procedures were identical by all means across the two experiment instances.

3.4. Data Analysis

For the analysis of RTs, three regions of interest were defined. The disambiguating verb (e.g., *fell*) was determined as critical region, the following word (e.g., *off*) as the first spillover region, and the one after (e.g., *the*) as the second spillover region. In the analyses, these are respectively labeled as Region 1, Region 2 and Region 3. Data trimming was conducted in two steps. First, the participants who displayed lower than 75% accuracy in the filler items were excluded from the dataset since the lower end of this threshold usually suggest that the participants did not attend to the task sufficiently (e.g., Huang & Ferreira, 2021; Cunnings & Fujita, 2021). This resulted in the exclusion of two subjects in the L2 group. The second step involved outlier data points in RTs. To approximate normal distribution, RTs were first log-transformed (Vasisth & Nicenboim, 2016). Following Keating and Jegerski's (2015) suggestions, a combined approach of outlier treatment was employed based first on absolute and then on variable cut-off points. For the absolute cut-off, (raw) RTs that are below 100 ms and above 3000 ms for the native and 4000 ms for non-native speakers were removed in each region of interest as such data points tend to constitute either unconscious reading behavior in the form of rapid button presses or loss of attention respectively. Such distinction on group basis was made since L2 speakers are known to be typically slower (Keating & Jegerski, 2015). As for the treatment based on variable cut-off points, log RTs that are below and above 3.5 standard deviations (SDs) from the mean were removed per each region for each level of the ambiguity factor (separately by the groups). All of the trimming affected nearly 3% of the data.

The continuous log RT data were analyzed with linear mixed effects models (Baayen et al., 2008; see also Cunnings, 2012) using *lmerTest* package in R environment (Kuznetsova et al., 2017; R Core Team, 2021). These models included the sum-coded (-1/1) fixed effects of group (L1 vs. L2) and ambiguity (ambiguous vs. unambiguous) as well as their interaction for each segment. For the question accuracy data, logit mixed effects models of binomial family were fit with the sum-coded fixed effects of ambiguity (ambiguous vs. unambiguous), group (L1 vs. L2), question type (subordinate vs. main clause) and their interaction(s). In all analyses, the models initially comprised of a maximal random effects structure that contained all the fixed effects and their interactions, random intercepts for participants and items as well as random slopes for the fixed effects⁷ (Barr et al., 2013). If the maximal model did not converge, random slopes were gradually removed starting from the one that explained the least amount of variance until convergence was achieved. Follow-up pairwise comparisons were conducted using the package *emmeans* (Lenth, 2022) for the interactions that were significant.

A main analysis was first conducted as described above to test for between-group differences. To explore the potential role of individual differences in L2 proficiency, additional models for L2 speakers were fit with centered global proficiency scores, as had been calculated by averaging over four skill ratings for each L2 participant.

⁷ Random slopes for group were not included in these initial models since such construct, namely an individual being an L1 or L2 speaker, cannot vary over subjects or items by its nature (see Winter, 2020: 243). The same applied to the continuous scores of proficiency in the second set of analyses.

CHAPTER 4

RESULTS

4.1. Between-Groups Analyses

4.1.1. RT Analyses

The participants' mean RTs are presented in Figure 2 per each group. For ease of readability, the back-transformed RTs are plotted in the figure although all the inferential analyses were conducted on log-transformed RTs. Table 2 further reports the summary of these analyses for each region.

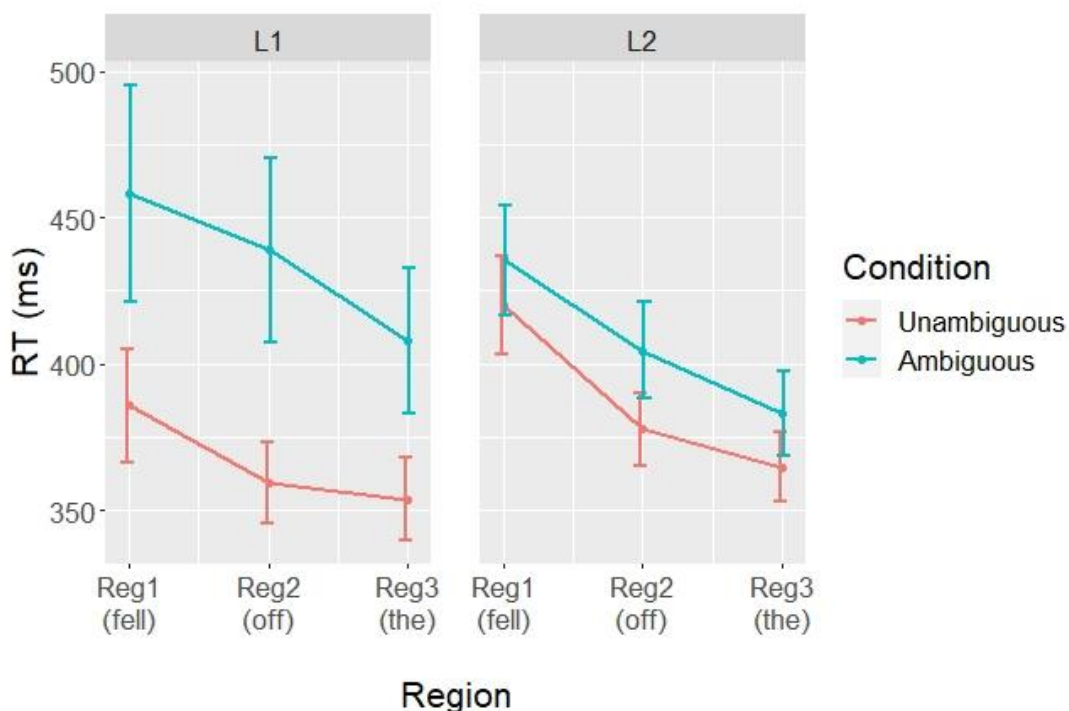


Figure 2. Mean reading times (ms) across groups and conditions

At the critical segment, namely Region 1, there was a main effect of ambiguity, with the temporarily ambiguous sentences eliciting significantly longer RTs than the unambiguous sentences (estimate = 0.036, SE = 0.011, $t = 3.114$, $p < 0.01$). Neither

the main effect of group (estimate = 0.032, SE = 0.036, $t = 0.891$, $p = 0.38$) nor its interaction with ambiguity (estimate = -0.016, SE = 0.009, $t = -1.736$, $p = 0.088$) was significant in this segment. At the first spillover region, namely Region 2, the main effect of ambiguity was still significant (estimate = 0.048, SE = 0.011, $t = 4.183$, $p < 0.001$) and it was further qualified by group in the form of a statistically significant interaction (estimate = -0.018, SE = 0.008, $t = -2.277$, $p < 0.05$). Pairwise comparisons at the each level of group yielded ambiguity effects for both the L1 (estimate = -0.132, SE = 0.030, $t = -4.468$, $p < 0.001$) and the L2 group (estimate = -0.060, SE = 0.028, $t = -2.166$, $p = 0.04$), although the size of GP effects seemed to be smaller for the latter group (L1 effect = 79.54 ms, L2 effect = 26.90 ms). The main effect of group was not significant (estimate = 0.006, SE = 0.030, $t = 0.213$, $p = 0.83$). At Region 3, this interaction between ambiguity and group was still present (estimate = -0.017, SE = 0.008, $t = -2.043$, $p = 0.05$) along with a main effect of ambiguity (estimate = 0.036, SE = 0.010, $t = 3.555$, $p < 0.01$). This time, follow-up pairwise comparisons revealed that the effect of ambiguity was not significant for the L2 group (estimate = -0.038, SE = 0.026, $t = -1.478$, $p = 0.15$), whereas for the L1 group it was (estimate = -0.107, SE = 0.028, $t = -3.795$, $p < 0.001$). The main effect of group was not significant as in the previous regions (estimate = 0.002, SE = 0.027, $t = 0.077$, $p = 0.94$).

Table 2. Summary of the RT analyses across the groups

	Beta	95% CI¹	p-value
Region 1			
Ambiguity	0.04	0.01, 0.06	< 0.01
Group	0.03	-0.04, 0.10	0.376
Ambiguity * Group	-0.02	-0.03, 0.00	0.088
Region 2			
Ambiguity	0.05	0.02, 0.07	<0.001
Group	0.01	-0.05, 0.07	0.832
Ambiguity * Group	-0.02	-0.03, 0.00	0.026
Region 3			
Ambiguity	0.04	0.02, 0.06	< 0.01
Group	0.00	-0.05, 0.06	0.939
Ambiguity * Group	-0.02	-0.03, 0.00	0.045

1. Confidence Interval

4.1.2. Accuracy Analyses

The participants' mean accuracy rates are presented in Figure 3.

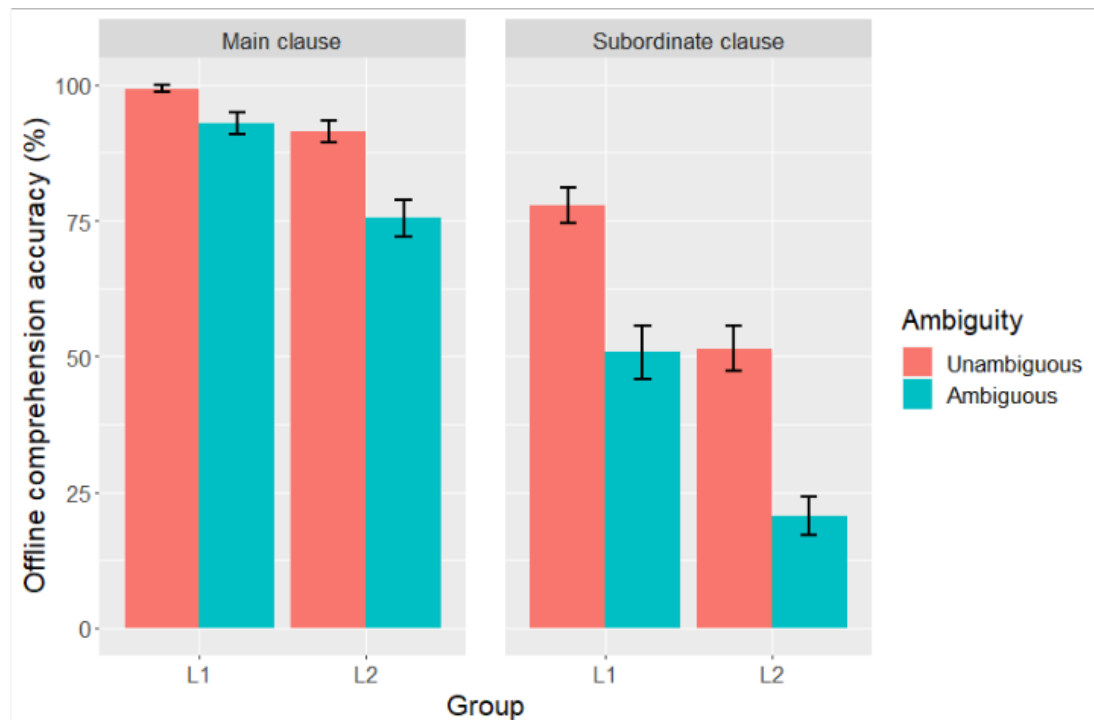


Figure 3. Mean accuracy rates across question type and group

The results from the best fit model yielded a statistically significant main effect of ambiguity, with lower accuracy in the ambiguous sentences compared to the unambiguous ones overall (estimate = -0.974, SE = 0.161, $z = -6.063$, $p < .001$). There was also a main effect of question type, indicating that the main clause questions elicited overall significantly higher rates of accuracy compared to the subordinate clause questions (estimate = -1.692, SE = 0.203, $z = -8.343$, $p < .001$). The main effect of group was also significant, with the L2 group having committed significantly higher amount of errors across the board compared to the native speakers, as can be seen in Figure 3 (estimate = -1.008, SE = 0.183, $z = -5.521$, $p < .001$). Among these factors, however, none of the interactions reached statistical significance, including ambiguity and question type (estimate = 0.031, SE = 0.153, $z = 0.202$, $p = 0.84$), ambiguity and group (estimate = 0.061, SE = 0.15806, $z = 0.388$, $p = 0.70$) along with the three-way interaction between ambiguity, group and question type (estimate = -0.157, SE = 0.152, $z = -1.034$, $p = 0.30$). An extensive summary of these accuracy analyses can further inspected in Table 3.

Table 3. Summary of accuracy analyses between groups

Characteristic	log(OR) ¹	95% CI ²	p-value
Ambiguity	-0.97	-1.3, -0.66	<0.001
Question Type	-1.7	-2.1, -1.3	<0.001
Group	-1.0	-1.4, -0.65	<0.001
Ambiguity * Question Type	0.03	-0.27, 0.33	0.840
Ambiguity * Group	0.06	-0.25, 0.37	0.698
Question Type * Group	0.10	-0.24, 0.44	0.570
Ambiguity * Question Type * Group	-0.16	-0.45, 0.14	0.301

1. Odds Ratio

2. Confidence Interval

4.2. Individual Differences in L2 Proficiency

4.2.1. RT Analyses

To explore whether or how the individual differences in the level of L2 proficiency can modulate the success of GP recovery within the L2 speakers, the analyses conducted above were repeated for the L2 group's subsetted data only in which the composite general proficiency scores were treated as a (centered) continuous predictor.

It was found that the main effect of proficiency was not significant in any of the regions of interest (Region 1 estimate = -0.030, SE = 0.038, $t = -0.789$, $p = 0.44$; Region 2 estimate = -0.016; SE = 0.032, $t = -0.502$, $p = 0.62$; Region 3 estimate = -0.011, SE = 0.030; $t = -0.373$; $p = 0.71$). The main effect of ambiguity displayed a somewhat different pattern across the regions compared to the previous analyses that tested for the RT differences at the level of group. Specifically, there was no statistically significant effect of ambiguity at the critical region (estimate = 0.020, SE = 0.013, $t = 1.551$, $p = 0.14$). However, they were significant only at the spillover regions (Region 2 estimate = 0.029; SE = 0.012, $t = 2.513$, $p = 0.02$; Region 3 estimate = 0.020, SE = 0.009, $t = 2.162$, $p = 0.04$). For none of the regions of interest, the interaction between proficiency and ambiguity was significant (Region 1 estimate = 0.008, SE = 0.010, $t = 0.768$, $p = 0.45$; Region 2 estimate = -0.001, SE = 0.009, $t = -0.097$, $p = 0.92$; Region 3 estimate = -0.003; SE = 0.008; $t = -0.420$, $p = 0.67$). The analyses can be inspected further in detail in Table 4.

Table 4. Summary of the proficiency analyses for RTs

	Beta	95% CI¹	p-value
Region 1			
Proficiency	-0.03	-0.11, 0.05	0.435
Ambiguity	0.02	-0.01, 0.05	0.137
Proficiency * Ambiguity	0.01	-0.01, 0.03	0.448
Region 2			
Proficiency	-0.02	-0.08, 0.05	0.619
Ambiguity	0.03	0.00, 0.05	0.022
Proficiency * Ambiguity	0.00	-0.02, 0.02	0.923
Region 3			
Proficiency	-0.01	-0.07, 0.05	0.712
Ambiguity	0.02	0.00, 0.04	0.044
Proficiency * Ambiguity	0.00	-0.02, 0.01	0.674

1. Confidence Interval

4.2.2. Accuracy Analyses

The logit model fitted to test for the impact of proficiency is summarized in Table 5.

Table 5. Summary of the proficiency analyses for accuracy rates

Characteristic	log(OR)¹	95% CI²	p-value
Ambiguity	-0.84	-1.1, -0.59	<0.001
Proficiency	0.06	-0.24, 0.37	0.697
Question Type	-1.5	-1.9, -1.2	<0.001
Ambiguity * Proficiency	0.19	-0.05, 0.43	0.119
Ambiguity * Question Type	-0.09	-0.31, 0.14	0.439
Proficiency * Question Type	0.10	-0.20, 0.40	0.519
Ambiguity * Proficiency * Question Type	-0.04	-0.26, 0.18	0.739

1. Odds Ratio

2. Confidence Interval

The main effects of ambiguity (estimate = -0.844, SE = 0.130, $z = -6.485$, $p < 0.001$) and the question type (estimate = -1.512, SE = 0.176, $z = -8.605$, $p < 0.001$) as well as the non-interaction of ambiguity and question type (estimate = 0.089, SE = 0.115, $z = -0.775$, $p = 0.44$) were retained from the previous model. There was no significant main effect of proficiency (estimate = 0.061, SE = 0.156, $z = 0.390$, $p = 0.70$) and it did not further interact with ambiguity (estimate = 0.191, SE = 0.122, $z = 1.558$, $p = 0.12$), question type (estimate = 0.099, SE = 0.153, $z = 0.645$, $p = 0.52$) or

participated in a three way interaction with ambiguity and question type (estimate = -0.038, SE = 0.114, $z = -0.333$, $p = 0.74$).

CHAPTER 5

DISCUSSION

The current study investigated the nature of GP comprehension by L1 and L2 speakers of English by combining on-line and off-line indices of reanalysis in the same experimental design. To this end, a self-paced experiment was conducted in which the sentences were manipulated with respect to the ambiguity and the type of the comprehension question in a similar manner to Chromý's (2022) study for Czech monolinguals.

Several key findings were obtained in this study. Firstly, both groups showed GP effects at or following the disambiguation region, suggesting that they adopted a direct object interpretation up to this point, although these effects were more elusive and weaker for the L2 group as can be also seen in the discrepant results of the two analyses. Specifically, while the GP effects start to emerge at the critical region for both groups according to the first analyses that tested for between-groups differences, they manifest only at the spill-over regions for the L2 group per the models that examined the role of individual differences in proficiency. Secondly, the performance in main clause questions was generally better and the initial misinterpretations lingered to a certain extent, which is reflected in the low accuracy rates to the subordinate clause questions following ambiguous trials. However, the ambiguity effects were statistically present in the main clause probes alike. Thirdly, the groups did not significantly differ off-line in GP recovery despite the fact that the accuracy rates of the latter group were overall more depressed. Finally, individual differences in the level of L2 proficiency did not predict either the amount of processing difficulty that ensues at the disambiguation region or off-line comprehension performance among the L2 speakers. In the following, these findings are discussed against the previous research on both good-enough comprehension and the mechanisms of second language processing.

5.1. The Nature of Lingering Misinterpretations in GP Sentences

In the experiment, we replicated the well-attested phenomenon of lingering misinterpretations resulting from GP sentences (e.g., Christianson et al., 2001). Both participant groups displayed a failure to abandon the direct object interpretation induced by garden-pathing, with the native group performing a chance level of comprehension accuracy in the subordinate clause questions following the GP condition. However, similar to Chromý's (2022) experiments in Czech, ambiguity effects were also observed in the main clause probes, the numerical trend of which was greater for the L2 speakers as illustrated in Figure 3 although no interaction with group reached statistical significance. This suggests that both groups at least sometimes failed to repair the ambiguity entirely and could not achieve a correct syntactic representation at all in which the ambiguous NP was reassigned the correct grammatical role. Such pattern may in part contrast with both the incomplete/partial reanalysis view (Christianson et al., 2001; Huang & Ferreira, 2021) and the view that lingering misinterpretations are (exclusively) caused by the interference from the meaning associated with earlier incorrect parse (Slattery et al., 2013; Sturt, 2007) as both these accounts predict that the parser should at least succeed in moving the NP to the subject position. The current findings suggest that a complete parsing failure can be responsible for the lingering misinterpretations on some trials, resonating with the earlier claims of Stevenson (1998) who had proposed that an unbound null subject may occupy the subject position of the matrix clause which would otherwise host the reanalyzed ambiguous NP (e.g., *[CP[CP While Mary dressed the baby][CP Ø fell off the bed]]) at the post-repair stage.

One caveat to the current findings is however that there is a possible incompatibility between the numerical accuracy rates and the results of inferential analyses. Note that the accuracy rates of 75% (for L2ers) - 92% (for L1ers) for main clause questions following ambiguous trials are actually parallel with those of Christianson et al.'s (2001) study where the nature of the effects emerged as an interaction between ambiguity and question type (in L1 speakers) such that ambiguity effects were smaller in magnitude for the main clause probes. In contrast, analogous numerical patterns are expressed as a mere main effect of ambiguity in the present study. Given wide confidence intervals in the accuracy analyses, we acknowledge that this discrepancy

could stem from insufficient sampling in the current study. Nevertheless, the observed effects still beg the question of why the parser completely fails to repair the GP on a subset of trials, whereas no such failure occurs in others potentially with only a lingering semantic effect of the initial misinterpretation. Although the present study was not systematically designed to elucidate this issue, its further investigation may be potentially a fruitful avenue of research towards a more graded account of lingering misinterpretations.

5.2. Reanalysis and Mechanisms of L2 Sentence Processing

Our findings are reminiscent of those reported by Jacob and Felser (2016) who found smaller GP effects in total reading times and regression-path duration as well as generally low accuracy rates for their L2 group in an eye tracking experiment. Note, however, that the current study differed in design from theirs since they only introduced subordinate clause questions to their participants, whereas there were matrix clause questions as well in our study. As off-line ambiguity effects were not qualified by the group, we failed to replicate Fujita and Cunnings (2021) who found larger ambiguity effects for the L2 speakers across their off-line experiments, the statistical significance of which was nevertheless “marginal” per their report.

Turning to theoretical implications of our findings, the smaller and less consistent GP effects observed in the L2 group may suggest at face value that non-native speakers are less likely to initiate syntactic reanalysis in line with the predictions of SSH (Clahsen & Felser, 2006a; 2018) along with the claims of Jacob and Felser (2016); however, the fact that such interactive pattern did not reliably persist further into the comprehension accuracy may be challenging to reconcile with the predictions of this account, with L2ers simply experiencing more frequent comprehension breakdowns across the board. That is, the parsing difficulty of the L2 speakers was not particularly localized to an environment where a syntactic repair is required and a semantically plausible parse should be abandoned, but also present in an otherwise scenario where they similarly performed worse than the native control group, which in turn makes it difficult to argue for a selective ambiguity disadvantage off-line for L2 speakers. In this respect, we maintain that the current findings may not fully support the predictions of SSH due to the lack of a clear relevant pattern and suggest that L2 processing does

not rely on the heuristic pathway any more than L1 processing. Note that even though Jacob and Felser (2016) themselves previously concluded that L2 speakers may assign a greater weight to the semantic fit between the verb (e.g., dress) and noun (e.g., the baby) as an interpretation cue compared to the L1 speakers and consequently be more likely to utilize the good-enough/semantic route of parsing, such argument is somewhat on shaky grounds since their own findings suffered from a similar issue to ours – (simple) main effect of group in the off-line comprehension data.

Another account that is relevant to the current findings is Cunnings' (2017) interference account. Cunnings (2017) claimed that the observed differences in parsing can be explained by the L2ers' increased susceptibility to retrieval interference rather than the quality of the syntactic representations built. Within the comprehension of GP sentences, this corresponded to a prediction that both populations should construct correct and sufficiently detailed syntactic representation of such sentences, but L2 speakers may be more vulnerable to the interference from the earlier interpretation that is not deleted from memory (see. Slattery et al., 2013). Our findings do not confirm the broad predictions of this account either since there was no increased ambiguity effect for the L2 group specifically in the subordinate clause probes – which would under this account tap into the memory interference from the incorrect parse that co-exists and competes with the correct and fully repaired one. In addition, it should be also noted that Cunnings' (2017) account has some conceptual issues such as lack of conclusive evidence from other phenomena suggesting an L2 susceptibility to retrieval interference or the unclarity of why L2ers should be more vulnerable to it, as also noted by Jacob et al. (2017) in their commentary to Cunnings (2017).

Elsewhere, it has been argued that the source of non-native-like processing can be ascribed not to the ability to utilize a particular cue or structure building routines but to a more rapid depletion of cognitive resources during processing an L2 (McDonald, 2006; Hopp, 2010) such that that the limited availability of these resources may often prevent L2 speakers from integrating multiple cues in real-time processing (Sorace, 2011; Şafak & Hopp, 2022). Unlike the two accounts discussed above, capacity limitation approaches may have a greater potential to capture the findings reported in this study in certain aspects. As noted above, since L2 participants' difficulty with parsing was not only observed in the sentences in the ambiguous condition, it can be

maintained that this may implicate a more global processing deficit that goes beyond a dependence on semantic heuristics (Clahsen & Felser, 2006a) or cue interference (Cunnings, 2017). In other words, the exhaustion of cognitive resources imposed by processing an L2 may have impacted the performance of our participants severely to the degree that they failed to understand even unambiguous sentences some of the time. In fact, this reasoning is supported by the fact that the L2 group's performance in the unambiguous probes numerically parallels that of the L1 group in the ambiguous probes, as revealed by closer inspection of each panel in Figure 3, which may suggest that the same processing effort devoted to parse the ambiguous sentences in L1 amounts to the parsing of the unambiguous sentences in L2. As far as the L2 group of the current study are concerned, this burden on computational resources may have been aggravated in part by the task demands as the participants were not allowed to re-read the words in non-cumulative display and they had to further reconstruct the representation of considerably long and complex sentences modified by a relative clause at the question-answering phase.

Note that such explanation that capitalizes on resource limitations can likewise capture the fleeting and weaker GP effects experienced by the L2 speakers during on-line reading due to the fact that they may not allocate sufficient resources to notice the syntactic error signal in a native-like fashion, with a form of sensitivity that is weaker and delayed. Especially, the second set of analyses that revealed significant GP effects starting only from the first spillover region for the L2 group may attest to this view, whereas the effects started to arise immediately at the critical region for the L1 speakers.

In spite of being theoretically more elegant, albeit post-hoc, in accounting for the majority of our findings, this explanation is not without its limitations either. As revealed by the latter analyses, individual differences in proficiency were not associated with the on-line and off-line performance including overall RTs and comprehension accuracy among the L2 participants – in contrast to Cunnings and Fujita (2020) who found that increasing proficiency went hand in hand with the generally shorter RTs and overall enhanced comprehension accuracy using coordination ambiguity. From the point of the resource limitation approaches, this may be somewhat unexpected since such theories usually anticipate that the increases in L2

proficiency and experience should at least predict the (overall) processing speed, partly as a theoretical reflex of the oft-observed correlation between L2 proficiency and working memory span (e.g., Hopp, 2015; Jackson & Bobb, 2009; Miyake & Friedman, 1998) and the related rationale that increasing proficiency should relieve the burden on computational resources (Service et al., 2002; see also Abutalebi, 2008). Although the lack of a pattern to this effect arguably poses a challenge to these theories, we acknowledge that relatively low sample size of the L2 group could have caused the current study to have fallen short of being a well-controlled investigation of individual differences, ultimately yielding null findings. Therefore, we do not entirely rule out the possibility that we would have found effects of proficiency had we included a larger pool of L2 participants.

5.3. Conclusion

In the present study, the processes of reanalysis were explored in L1 and L2 speakers of English. Generally, both groups displayed GP effects upon having encountered the syntactic error signal, i.e. disambiguating information; however, the nature of these on-line effects was delayed and more fragile in the L2 group. For the off-line comprehension, the participants could not generally abandon the initial misinterpretation induced by garden-pathing, resulting in the replication of well-documented effects of lingering misinterpretations that long underpinned the good-enough comprehension literature along with other phenomena (e.g., Ferreira et al., 2002). Regarding the performance of L2 participants, they were more likely to experience comprehension breakdowns across the board compared to the L1 control group, and their level of proficiency did not modulate their on-line and off-line comprehension of experimental sentences in any way. Following the discussion of the findings with respect to the prominent models of L2 processing, it is tentatively concluded that L2 parsing may not be fundamentally from L1 parsing, with the observed differences being traceable to the capacity limitations. In the context of the phenomenon under investigation, it is maintained that such global processing limitations may give rise to the weaker sensitivity to the disambiguating cues along with more frequent comprehension failures than L1 speakers although this explanation is admittedly post-hoc and may need further testing.

Taken together, the current study addressed the gap in the literature as to comprehension of GP sentences in L2 speakers by exploiting on-line and off-line measures in the same design. As noted in the Chapter 2, most of the existing studies exclusively utilized GP ambiguities to diagnose whether L2 speakers can integrate some types of cues incrementally like L1 speakers, with the resulting interpretation mostly being left unexplored except for a few studies (e.g., Pozzan & Trueswell, 2016; Fujita & Cunnings, 2021). Against this background, the lingering misinterpretations and on-line processing of GP sentences were systematically examined in a self-paced reading experiment combined with off-line comprehension questions. While the phenomenon of lingering misinterpretations was generally replicated (e.g., Christianson et al., 2001; Slattery et al., 2013), it was observed that the L2 speakers were generally more prone to the comprehension errors and less sensitive to syntactic error signal (i.e. Jacob & Felser, 2016; Pozzan & Trueswell, 2016) —irrespective of their proficiency level.

5.4. Limitations and Suggestions for Further Research

One important limitation of the present investigation is that the number of participants is relatively smaller compared to the similar studies in the literature, especially in the L2 group (Fujita & Cunnings, 2021; Şafak & Hopp, 2022). At a general level, this may have brought about the large confidence intervals in the accuracy analyses which are commonly thought to indicate sampling error, while also resulting in a limited statistical power to reliably capture the effects of L2 proficiency at a specific level. This issue could highlight the need to replicate current findings through a larger sample size.

Another potential limitation of this study was that the L2 participants' level of proficiency was determined using self-rating. Although the use of self-rating measures of proficiency is fairly common in both L2 acquisition and processing research (e.g., Favier et al., 2019; Luk & Bialystok, 2013; Tan & Foltz, 2020) and their scores often correlate with those of objective proficiency measures (Marian et al., 2007; Sabourin et al., 2014), the possibility that they could have at least contributed to the lack of proficiency effects in the observed findings as a subjective measure is not completely

disregarded. Therefore, further studies should incorporate formal measures as well to investigate the impact of L2 proficiency in a more comprehensive way.

Due to the time restrictions, it was possible to test only one type of GP sentences as well as the individual differences in one factor, namely proficiency. To arrive at more generalizable findings, other types of GP sentences such as those with reduced relative clauses (e.g., The horse raced past the barn fell: Bever, 1970) or coordination ambiguity can be tested in future studies. Likewise, the exploration of individual differences in other domains such as working memory span (Christianson et al., 2006), executive control (Novick et al., 2005; Vuong & Martin, 2014) and reading skills (Wonnacott et al., 2016) can be useful to broaden our understanding about the degree to which native and non-native processing are governed by similar individual differences and thus can be qualitatively similar.

REFERENCES

- Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychologica*, 128(3), 466–478. <https://doi.org/10.1016/j.actpsy.2008.03.014>
- Altmann, G. T. M., Garnham, A., & Dennis, Y. (1992). Avoiding the garden path: Eye movements in context. *Journal of Memory and Language*, 31(5), 685–712. [https://doi.org/10.1016/0749-596X\(92\)90035-V](https://doi.org/10.1016/0749-596X(92)90035-V)
- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247–264. [https://doi.org/10.1016/S0010-0277\(99\)00059-1](https://doi.org/10.1016/S0010-0277(99)00059-1)
- Altmann, G. T. M., & Mirković, J. (2009). Incrementality and Prediction in Human Sentence Processing. *Cognitive Science*, 33(4), 583–609. <https://doi.org/10.1111/j.1551-6709.2009.01022.x>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–362). Wiley.
- Binder, K. S., Duffy, S. A., & Rayner, K. (2001). The Effects of Thematic Fit and Discourse Context on Syntactic Ambiguity Resolution. *Journal of Memory and Language*, 44(2), 297–324. <https://doi.org/10.1006/jmla.2000.2754>
- Birdsong, D., & Gertken, L. M. (2013). In faint praise of folly: A critical review of native/non-native speaker comparisons, with examples from native and bilingual processing of French complex syntax. *Language, Interaction and Acquisition*, 4(2), 107–133. <https://doi.org/10.1075/lia.4.2.01bir>
- Boxell, O., & Felser, C. (2017). Sensitivity to parasitic gaps inside subject islands in native and non-native sentence processing. *Bilingualism: Language and Cognition*, 20(3), 494–511. <https://doi.org/10.1017/S1366728915000942>

- Brothers, T., Hoversten, L. J., & Traxler, M. J. (2021). Bilinguals on the garden-path: Individual differences in syntactic ambiguity resolution. *Bilingualism: Language and Cognition*, 24(4), 612–627. <https://doi.org/10.1017/S1366728920000711>
- Carreiras, M., & Clifton, C. (1999). Another word on parsing relative clauses: Eyetracking evidence from Spanish and English. *Memory & Cognition*, 27(5), 826–833. <https://doi.org/10.3758/BF03198535>
- Cheng, Y., Rothman, J., & Cunnings, I. (2021). Parsing preferences and individual differences in nonnative sentence processing: Evidence from eye movements. *Applied Psycholinguistics*, 42(1), 129–151. <https://doi.org/10.1017/S014271642000065X>
- Choi, Y., & Trueswell, J. C. (2010). Children’s (in)ability to recover from garden paths in a verb-final language: Evidence for developing control in sentence processing. *Journal of Experimental Child Psychology*, 106(1), 41–61. <https://doi.org/10.1016/j.jecp.2010.01.003>
- Chomsky, N. (1981). *Lectures on government and binding: The Pisa lectures*. Foris Publications.
- Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good-enough, underspecified, or shallow language processing. *Quarterly Journal of Experimental Psychology*, 69(5), 817–828. <https://doi.org/10.1080/17470218.2015.1134603>
- Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic Roles Assigned along the Garden Path Linger. *Cognitive Psychology*, 42(4), 368–407. <https://doi.org/10.1006/cogp.2001.0752>
- Christianson, K., & Luke, S. G. (2011). Context Strengthens Initial Misinterpretations of Text. *Scientific Studies of Reading*, 15(2), 136–166. <https://doi.org/10.1080/10888431003636787>
- Christianson, K., Luke, S. G., Hussey, E. K., & Wochna, K. L. (2017). Why reread? Evidence from garden-path and local coherence structures. *Quarterly Journal of Experimental Psychology*, 70(7), 1380–1405. <https://doi.org/10.1080/17470218.2016.1186200>

- Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006). Younger and Older Adults' "Good-Enough" Interpretations of Garden-Path Sentences. *Discourse Processes*, 42(2), 205–238. https://doi.org/10.1207/s15326950dp4202_6
- Chromý, J. (2022). When readers fail to form a coherent representation of garden-path sentences. *Quarterly Journal of Experimental Psychology*, 75(1), 169–190. <https://doi.org/10.1177/17470218211037152>
- Clahsen, H., Balkhair, L., Schutter, J.-S., & Cunnings, I. (2013). The time course of morphological processing in a second language. *Second Language Research*, 29(1), 7–31. <https://doi.org/10.1177/0267658312464970>
- Clahsen, H., & Felser, C. (2006a). Grammatical processing in language learners. *Applied Psycholinguistics*, 27(1), 3–42. <https://doi.org/10.1017/S0142716406060024>
- Clahsen, H., & Felser, C. (2006b). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, 27(1), 107–126. <https://doi.org/10.1017/S0142716406060206>
- Clahsen, H., & Felser, C. (2018). Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition*, 40(3), 693–706. <https://doi.org/10.1017/S0272263117000250>
- Cuetos, F. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition*, 30(1), 73–105. [https://doi.org/10.1016/0010-0277\(88\)90004-2](https://doi.org/10.1016/0010-0277(88)90004-2)
- Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition*, 30(1), 73–105. [https://doi.org/10.1016/0010-0277\(88\)90004-2](https://doi.org/10.1016/0010-0277(88)90004-2)
- Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, 28(3), 369–382. <https://doi.org/10.1177/0267658312443651>
- Cunnings, I. (2017). Parsing and Working Memory in Bilingual Sentence Processing. *Bilingualism: Language and Cognition*, 20(4), 659–678. <https://doi.org/10.1017/S1366728916000675>

- Cunnings, I., & Fujita, H. (2021). Quantifying individual differences in native and nonnative sentence processing. *Applied Psycholinguistics*, 42(3), 579–599. <https://doi.org/10.1017/S0142716420000648>
- Dallas, A., DeDe, G., & Nicol, J. (2013). An Event-Related Potential (ERP) Investigation of Filler-Gap Processing in Native and Second Language Speakers: ERP Study of L2 Filler-Gap Processing. *Language Learning*, 63(4), 766–799. <https://doi.org/10.1111/lang.12026>
- Dehaene, S., Dupoux, E., Mehler, J., Cohen, L., Paulesu, E., Perani, D., van de Moortele, P.-F., Lehericy, S., & Le Bihan, D. (1997). Anatomical variability in the cortical representation of first and second language. *Neuroreport*, 8(17), 3809–3815.
- Dussias, P. E., & Cramer Scaltz, T. R. (2008). Spanish–English L2 speakers’ use of subcategorization bias information in the resolution of temporary ambiguity during second language reading. *Acta Psychologica*, 128(3), 501–513. <https://doi.org/10.1016/j.actpsy.2007.09.004>
- Favier, S., Wright, A., Meyer, A., & Huettig, F. (2019). Proficiency modulates between- but not within-language structural priming. *Journal of Cultural Cognitive Science*, 3(S1), 105–124. <https://doi.org/10.1007/s41809-019-00029-1>
- Felser, C., & Cunnings, I. (2012). Processing reflexives in a second language: The timing of structural and discourse-level constraints. *Applied Psycholinguistics*, 33(3), 571–603. <https://doi.org/10.1017/S0142716411000488>
- Felser, C., Cunnings, I., Batterham, C., & Clahsen, H. (2012). The timing of island effects in nonnative sentence processing. *Studies in Second Language Acquisition*, 34(1), 67–98. <https://doi.org/10.1017/S0272263111000507>
- Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, 24(3), 453–489. <https://doi.org/10.1017/S0142716403000237>
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, 47(2), 164–203. [https://doi.org/10.1016/S0010-0285\(03\)00005-7](https://doi.org/10.1016/S0010-0285(03)00005-7)

- Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-Enough Representations in Language Comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. <https://doi.org/10.1111/1467-8721.00158>
- Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*(3), 348–368. [https://doi.org/10.1016/0749-596X\(86\)90006-9](https://doi.org/10.1016/0749-596X(86)90006-9)
- Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, *30*(6), 725–745. [https://doi.org/10.1016/0749-596X\(91\)90034-H](https://doi.org/10.1016/0749-596X(91)90034-H)
- Ferreira, F., & Henderson, J. M. (1993). Reading processes during syntactic analysis and reanalysis. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *47*(2), 247–275. <https://doi.org/10.1037/h0078819>
- Ferreira, F., & Patson, N. D. (2007). The ‘Good Enough’ Approach to Language Comprehension: The ‘Good Enough’ Approach. *Language and Linguistics Compass*, *1*(1–2), 71–83. <https://doi.org/10.1111/j.1749-818X.2007.00007.x>
- Fodor, J. D., & Inoue, A. (1998). Attach Anyway. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in Sentence Processing* (Vol. 21, pp. 101–141). Springer. https://doi.org/10.1007/978-94-015-9070-9_4
- Frazier, L., & Clifton, C. (1996). *Construal*. MIT Press.
- Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291–325. [https://doi.org/10.1016/0010-0277\(78\)90002-1](https://doi.org/10.1016/0010-0277(78)90002-1)
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178–210. [https://doi.org/10.1016/0010-0285\(82\)90008-1](https://doi.org/10.1016/0010-0285(82)90008-1)
- Frenck-Mestre, C., & Pynte, J. (1997). Syntactic Ambiguity Resolution While Reading in Second and Native Languages. *The Quarterly Journal of Experimental Psychology A*, *50*(1), 119–148. <https://doi.org/10.1080/027249897392251>

- Fujita, H., & Cunnings, I. (2020). Reanalysis and lingering misinterpretation of linguistic dependencies in native and non-native sentence comprehension. *Journal of Memory and Language*, *115*, 104154. <https://doi.org/10.1016/j.jml.2020.104154>
- Fujita, H., & Cunnings, I. (2021). Reanalysis processes in non-native sentence comprehension. *Bilingualism: Language and Cognition*, *24*(4), 628–641. <https://doi.org/10.1017/S1366728921000195>
- Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The Contributions of Verb Bias and Plausibility to the Comprehension of Temporarily Ambiguous Sentences. *Journal of Memory and Language*, *37*(1), 58–93. <https://doi.org/10.1006/jmla.1997.2512>
- Gerth, S., Otto, C., Felser, C., & Nam, Y. (2017). Strength of garden-path effects in native and non-native speakers' processing of object–subject ambiguities. *International Journal of Bilingualism*, *21*(2), 125–144. <https://doi.org/10.1177/1367006915604401>
- Gibson, E., & Warren, T. (2004). Reading-Time Evidence for Intermediate Linguistic Structure in Long-Distance Dependencies. *Syntax*, *7*(1), 55–78. <https://doi.org/10.1111/j.1368-0005.2004.00065.x>
- Hahne, A. (2001). What's Different in Second-Language Processing? Evidence from Event-Related Brain Potentials. *Journal of Psycholinguistic Research*, *30*(3), 251–266. <https://doi.org/10.1023/A:1010490917575>
- Hahne, A., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition*, *4*(2), 123–141. <https://doi.org/10.1017/S1366728901000232>
- Hahne, A., & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research*, *13*(3), 339–356. [https://doi.org/10.1016/S0926-6410\(01\)00127-6](https://doi.org/10.1016/S0926-6410(01)00127-6)
- Hammerly, C., Staub, A., & Dillon, B. (2019). The grammaticality asymmetry in agreement attraction reflects response bias: Experimental and modeling evidence. *Cognitive Psychology*, *110*, 70–104. <https://doi.org/10.1016/j.cogpsych.2019.01.001>

- Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, 22(3), 369–397. <https://doi.org/10.1191/0267658306sr272oa>
- Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, 120(4), 901–931. <https://doi.org/10.1016/j.lingua.2009.06.004>
- Hopp, H. (2015). Individual differences in the second language processing of object–subject ambiguities. *Applied Psycholinguistics*, 36(2), 129–173. <https://doi.org/10.1017/S0142716413000180>
- Hopp, H. (2018). The bilingual mental lexicon in L2 sentence processing. *Second Language*, 17, 5–27. https://doi.org/10.11431/secondlanguage.17.0_5
- Hopp, H. (2022). Second Language Sentence Processing. *Annual Review of Linguistics*, 8(1), 235–256. <https://doi.org/10.1146/annurev-linguistics-030821-054113>
- Huang, Y., & Ferreira, F. (2021). What causes lingering misinterpretations of garden-path sentences: Incorrect syntactic representations or fallible memory processes? *Journal of Memory and Language*, 121, 104288. <https://doi.org/10.1016/j.jml.2021.104288>
- Jackson, C. (2008). Proficiency Level and the Interaction of Lexical and Morphosyntactic Information During L2 Sentence Processing. *Language Learning*, 58(4), 875–909. <https://doi.org/10.1111/j.1467-9922.2008.00481.x>
- Jackson, C. N., & Bobb, S. C. (2009). The processing and comprehension of *wh* - questions among second language speakers of German. *Applied Psycholinguistics*, 30(4), 603–636. <https://doi.org/10.1017/S014271640999004X>
- Jackson, C. N., & van Hell, J. G. (2011). The effects of L2 proficiency level on the processing of *wh*-questions among Dutch second language speakers of English. *IRAL - International Review of Applied Linguistics in Language Teaching*, 49(3). <https://doi.org/10.1515/iral.2011.012>
- Jacob, G., & Felser, C. (2016). Reanalysis and semantic persistence in native and non-native garden-path recovery. *Quarterly Journal of Experimental Psychology*, 69(5), 907–925. <https://doi.org/10.1080/17470218.2014.984231>

- Jacob, G., Lago, S., & Patterson, C. (2017). L2 processing and memory retrieval: Some empirical and conceptual challenges. *Bilingualism: Language and Cognition*, 20(4), 691–693. <https://doi.org/10.1017/S1366728916000948>
- Juffs, A. (2004). Representation, Processing and Working Memory in a Second Language. *Transactions of the Philological Society*, 102(2), 199–225. <https://doi.org/10.1111/j.0079-1636.2004.00135.x>
- Juffs, A., & Harrington, M. (1996). Garden Path Sentences and Error Data in Second Language Sentence Processing. *Language Learning*, 46(2), 283–323. <https://doi.org/10.1111/j.1467-1770.1996.tb01237.x>
- Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49(1), 133–156. [https://doi.org/10.1016/S0749-596X\(03\)00023-8](https://doi.org/10.1016/S0749-596X(03)00023-8)
- Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly Journal of Experimental Psychology*, 69(5), 1013–1040. <https://doi.org/10.1080/17470218.2015.1053951>
- Karsenti, L., & Meltzer-Asscher, A. (2022). Prediction of successful reanalysis based on eye-blink rate and reading times in sentences with local ambiguity. *Language and Cognition*, 1–23. <https://doi.org/10.1017/langcog.2022.17>
- Kaschak, M. P., & Glenberg, A. M. (2004). This construction needs learned. *Journal of Experimental Psychology: General*, 133(3), 450–467. <https://doi.org/10.1037/0096-3445.133.3.450>
- Keating, G. D., & Jegerski, J. (2015). Experimental designs in sentence processing research: A Methodological Review and User's Guide. *Studies in Second Language Acquisition*, 37(1), 1–32. <https://doi.org/10.1017/S0272263114000187>
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146, 23–49. <https://doi.org/10.1016/j.brainres.2006.12.063>

- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13). <https://doi.org/10.18637/jss.v082.i13>
- Lago, S., Gračanin-Yukse, M., Şafak, D. F., Demir, O., Kırkıcı, B., & Felser, C. (2019). Straight from the horse's mouth: Agreement attraction effects with Turkish possessors. *Linguistic Approaches to Bilingualism*, 9(3), 398–426. <https://doi.org/10.1075/lab.17019.lag>
- Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement attraction in Spanish comprehension. *Journal of Memory and Language*, 82, 133–149. <https://doi.org/10.1016/j.jml.2015.02.002>
- Lau, E. F., & Ferreira, F. (2005). Lingering effects of disfluent material on comprehension of garden path sentences. *Language and Cognitive Processes*, 20(5), 633–666. <https://doi.org/10.1080/01690960444000142>
- Lenth, R. V. (2022). *Emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.8.3. <https://CRAN.R-project.org/package=emmeans>
- Lewis, R. L., & Vasishth, S. (2005). An Activation-Based Model of Sentence Processing as Skilled Memory Retrieval. *Cognitive Science*, 29(3), 375–419. https://doi.org/10.1207/s15516709cog0000_25
- Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, 10(10), 447–454. <https://doi.org/10.1016/j.tics.2006.08.007>
- Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, 25(5), 605–621. <https://doi.org/10.1080/20445911.2013.795574>
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4), 676–703. <https://doi.org/10.1037/0033-295X.101.4.676>
- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and*

Hearing Research, 50(4), 940–967. [https://doi.org/10.1044/1092-4388\(2007/067\)](https://doi.org/10.1044/1092-4388(2007/067))

- Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, 27(01). <https://doi.org/10.1017/S0272263105050035>
- McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, 55(3), 381–401. <https://doi.org/10.1016/j.jml.2006.06.006>
- McRae, K., & Matsuki, K. (2013). Constraint-based models of sentence processing. In R. P. G. van Gompel (Ed.), *Sentence Processing* (pp. 51–77). Psychology Press.
- McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the Influence of Thematic Fit (and Other Constraints) in On-line Sentence Comprehension. *Journal of Memory and Language*, 38(3), 283–312. <https://doi.org/10.1006/jmla.1997.2543>
- Miyake, A., & Friedman, N. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. Healy & L. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339–364). Erlbaum.
- Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, 5(3), 263–281. <https://doi.org/10.3758/CABN.5.3.263>
- Omaki, A., Davidson White, I., Goro, T., Lidz, J., & Phillips, C. (2014). No Fear of Commitment: Children's Incremental Interpretation in English and Japanese Wh-Questions. *Language Learning and Development*, 10(3), 206–233. <https://doi.org/10.1080/15475441.2013.844048>
- Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island constraints in second-language sentence processing. *Studies in Second Language Acquisition*, 33(4), 563–588. <https://doi.org/10.1017/S0272263111000313>

- Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31(6), 785–806. [https://doi.org/10.1016/0749-596X\(92\)90039-Z](https://doi.org/10.1016/0749-596X(92)90039-Z)
- Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 786–803. <https://doi.org/10.1037/0278-7393.20.4.786>
- Özge, D. (2020). Tümce İşleme Süreçleri. *Dil Dergisi*, 171(2), 46-69. <https://doi.org/10.33690/dilder.693690>
- Pan, H.-Y., Schimke, S., & Felser, C. (2015). Referential context effects in non-native relative clause ambiguity resolution. *International Journal of Bilingualism*, 19(3), 298–313. <https://doi.org/10.1177/1367006913515769>
- Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: A Study of Relative Clause Attachment in Greek. *Studies in Second Language Acquisition*, 25(4), 501–528. <https://doi.org/10.1017/S0272263103000214>
- Patson, N. D., Darowski, E. S., Moon, N., & Ferreira, F. (2009). Lingering misinterpretations in garden-path sentences: Evidence from a paraphrasing task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(1), 280–285. <https://doi.org/10.1037/a0014276>
- Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(4), 940–961. <https://doi.org/10.1037/0278-7393.24.4.940>
- 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(3), 447–475. <https://doi.org/10.1006/jmla.2000.2708>
- Polinsky, M., & Scontras, G. (2020). Understanding heritage languages. *Bilingualism: Language and Cognition*, 23(1), 4–20. <https://doi.org/10.1017/S1366728919000245>

- Pozzan, L., & Trueswell, J. C. (2016). Second language processing and revision of garden-path sentences: A visual word study. *Bilingualism: Language and Cognition*, *19*(3), 636–643. <https://doi.org/10.1017/S1366728915000838>
- Qian, Z., Garnsey, S., & Christianson, K. (2018). A comparison of online and offline measures of good-enough processing in garden-path sentences. *Language, Cognition and Neuroscience*, *33*(2), 227–254. <https://doi.org/10.1080/23273798.2017.1379606>
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rah, A., & Adone, D. (2010). Processing of the reduced relative clause versus main verb ambiguity in L2 learners at different proficiency levels. *Studies in Second Language Acquisition*, *32*(1), 79–109. <https://doi.org/10.1017/S027226310999026X>
- Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics*, *32*(2), 299–331. <https://doi.org/10.1017/S0142716410000421>
- Roberts, L., & Liszka, S. A. (2021). Grammatical aspect and L2 learners' online processing of temporarily ambiguous sentences in English: A self-paced reading study with German, Dutch and French L2 learners. *Second Language Research*, *37*(4), 619–647. <https://doi.org/10.1177/0267658319895551>
- Sabourin, L., Brien, C., & Burkholder, M. (2014). The effect of age of L2 acquisition on the organization of the bilingual lexicon: Evidence from masked priming. *Bilingualism: Language and Cognition*, *17*(3), 542–555. <https://doi.org/10.1017/S1366728913000643>
- Şafak, D. F., & Hopp, H. (2022). Verb bias and semantic persistence effects in L2 ambiguity resolution. *Second Language Research*, *38*(4), 705–736. <https://doi.org/10.1177/0267658321997904>
- Service, E., Simola, M., Metsänheimo, O., & Maury, S. (2002). Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology*, *14*(3), 383–408. <https://doi.org/10.1080/09541440143000140>
- Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic

- representations. *Journal of Memory and Language*, 69(2), 104–120. <https://doi.org/10.1016/j.jml.2013.04.001>
- Sorace, A. (2011). Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to Bilingualism*, 1(1), 1–33. <https://doi.org/10.1075/lab.1.1.01sor>
- Spivey-Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, 47(2), 276–309. <https://doi.org/10.1037/h0078826>
- Stevenson, S. (1998). Parsing as Incremental Restructuring. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in Sentence Processing* (Vol. 21, pp. 327–363). Springer. https://doi.org/10.1007/978-94-015-9070-9_10
- Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, 1(3), 227–245. <https://doi.org/10.1080/01690968608407062>
- Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48(3), 542–562. [https://doi.org/10.1016/S0749-596X\(02\)00536-3](https://doi.org/10.1016/S0749-596X(02)00536-3)
- Sturt, P. (2007). Semantic re-interpretation and garden path recovery. *Cognition*, 105(2), 477–488. <https://doi.org/10.1016/j.cognition.2006.10.009>
- Sturt, P., Pickering, M. J., & Crocker, M. W. (1999). Structural Change and Reanalysis Difficulty in Language Comprehension. *Journal of Memory and Language*, 40(1), 136–150. <https://doi.org/10.1006/jmla.1998.2606>
- Tan, M., & Foltz, A. (2020). Task Sensitivity in L2 English Speakers’ Syntactic Processing: Evidence for Good-Enough Processing in Self-Paced Reading. *Frontiers in Psychology*, 11, 575847. <https://doi.org/10.3389/fpsyg.2020.575847>
- Tanner, D., Nicol, J., & Brehm, L. (2014). The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language*, 76, 195–215. <https://doi.org/10.1016/j.jml.2014.07.003>

- Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension: The integration of rules and habits*. MIT Press.
- Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, 73(2), 89–134. [https://doi.org/10.1016/S0010-0277\(99\)00032-3](https://doi.org/10.1016/S0010-0277(99)00032-3)
- Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(3), 528–553. <https://doi.org/10.1037/0278-7393.19.3.528>
- Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, 4(2), 105–122. <https://doi.org/10.1017/S1366728901000220>
- Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In C. Sanz (Ed.), *Mind and Context in Adult Second Language Acquisition: Methods, Theory, and Practice* (pp. 141–178). Georgetown University Press.
- van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Jacob, G. (2006). The activation of inappropriate analyses in garden-path sentences: Evidence from structural priming. *Journal of Memory and Language*, 55(3), 335–362. <https://doi.org/10.1016/j.jml.2006.06.004>
- Vasishth, S., & Nicenboim, B. (2016). Statistical Methods for Linguistic Research: Foundational Ideas - Part I: Statistical Methods for Linguistics - Part I. *Language and Linguistics Compass*, 10(8), 349–369. <https://doi.org/10.1111/lnc3.12201>
- Vuong, L. C., & Martin, R. C. (2014). Domain-specific executive control and the revision of misinterpretations in sentence comprehension. *Language, Cognition and Neuroscience*, 29(3), 312–325. <https://doi.org/10.1080/01690965.2013.836231>
- Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61(2), 206–237. <https://doi.org/10.1016/j.jml.2009.04.002>

Winter, B. (2020). *Statistics for linguists: An introduction using R*. Routledge, Taylor & Francis Group.

Wonnacott, E., Joseph, H. S. S. L., Adelman, J. S., & Nation, K. (2016). Is children's reading "good enough"? Links between online processing and comprehension as children read syntactically ambiguous sentences. *Quarterly Journal of Experimental Psychology*, 69(5), 855–879.
<https://doi.org/10.1080/17470218.2015.1011176>

Zehr, J., & Schwarz, F. (2022). *PennController for Internet Based Experiments (IBEX)*. <https://doi.org/10.17605/OSF.IO/MD832>

APPENDICES

A. APPROVAL OF THE METU HUMAN SUBJECTS ETHICS COMMITTEE

UYGULAMALI ETİK ARAŞTIRMA MERKEZİ
APPLIED ETHICS RESEARCH CENTER



ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

DUMLUPINAR BULVARI 06800
ÇANKAYA ANKARA/TURKEY
T: +90 312 210 22 91
F: +90 312 210 79 59
ueam@metu.edu.tr
www.ueam.metu.edu.tr

Sayı: 28620816 /

14 NİSAN 2022

Konu : Değerlendirme Sonucu

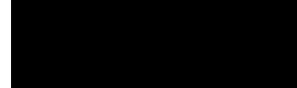
Gönderen: ODTÜ İnsan Araştırmaları Etik Kurulu (İAEK)

İlgi : İnsan Araştırmaları Etik Kurulu Başvurusu

Sayın Duygu ÖZGE

Danışmanlığımı yürüttüğünüz Süleyman Yaman'ın "İkinci dil İngilizce konuşucularında geçici olarak sözdizimsel belirsizlik içeren tümcelerin işlemlenmesi ve yorumlanması" başlıklı araştırmanız İnsan Araştırmaları Etik Kurulu tarafından uygun görülmüş ve **234-ODTÜİAEK-2022** protokol numarası ile onaylanmıştır.

Saygılarımızla bilgilerinize sunarız.


Prof. Dr. Mine MISIRLISOY
İAEK Başkanı

B. CONSENT FORM FOR THE L1 GROUP

This study is carried out by the graduate student Süleyman Yaman at Middle East Technical University in Turkey. This form was prepared to inform you about the study and its procedure.

What is the purpose of the study?

The current study investigates how first and second language speakers of English comprehend some sentences, using an experimental approach.

How can you help us?

If you agree to participate in the study, you will be asked to complete a self-paced reading task that approximately takes 15 minutes to complete. In the task, you are required to read sentences in chunks on your computer screen and sometimes answer comprehension questions for these sentences. **The task can be taken only through computer.** We kindly ask you to minimize the distractions around yourself and keep your attention on the screen throughout your participation.

How will we use your responses?

No identificatory information are requested. All your responses will be kept confidential and evaluated only by the researcher. The information to be obtained from the participants will be evaluated collectively and used in scientific publications.

What you need to know about your participation:

Your participation should be completely on voluntary basis. The task does not contain disturbing questions in general, but you are free to withdraw and exit the task in case you feel uncomfortable for any reason.

If you would like to learn more about this research:

Thank you for your participation in advance. If you have any questions about this study, you can contact the researcher (e-mail: [REDACTED]) or Assoc. Prof. Duygu Özge (e-mail: [REDACTED]).

I have read the information above and consent to take part in this study.

C. CONSENT FORM FOR THE L2 GROUP

Bu araştırma, ODTÜ yüksek lisans öğrencisi Süleyman Yaman tarafından yürütülmektedir. Bu form sizi araştırma koşulları hakkında bilgilendirmek için hazırlanmıştır.

Çalışmanın Amacı Nedir?

Bu araştırma, birinci ve ikinci dil İngilizce konuşucusu bireylerin bazı cümleleri nasıl işlemlediklerini deneysel bir yöntemle incelemeyi amaçlamaktadır.

Bize Nasıl Yardımcı Olmanızı İsteyeceğiz?

Araştırmaya katılmaya gönüllü olmanız durumunda, yaklaşık 15 dakika sürecek bir kendi hızında okuma testini tamamlamanız istenecektir. Bu test kapsamında, bilgisayar ekranında parçalar halinde sunulan cümleleri okumanız ve zaman zaman sorulacak anlam sorularını cevaplamanız beklenmektedir. **Çalışmaya sadece bilgisayar üzerinden katılabiliyorsunuz.** Katılımınız süresince çevrenizde dikkat dağıtıcı görüntü/ses olmamasına özen göstermenizi ve dikkatinizin ekranda olmasını rica ediyoruz.

Sizden Topladığımız Bilgileri Nasıl Kullanacağız?

Araştırmaya katılımınız tamamen gönüllülük temelinde olmalıdır. Çalışmada sizden kimlik belirleyici hiçbir bilgi istenmemektedir. Cevaplarınız tamamıyla gizli tutulacak, sadece araştırmacı tarafından değerlendirilecektir. Katılımcılardan elde edilecek bilgiler toplu halde değerlendirilecek ve bilimsel yayınlarda kullanılacaktır. Sağladığımız veriler gönüllü katılım formlarında toplanan kimlik bilgileri ile eşleştirilmeyecektir.

Katılımınızla ilgili bilmeniz gerekenler:

Çalışma, genel olarak kişisel rahatsızlık verici sorular içermemektedir. Ancak, katılım sırasında sorulardan ya da herhangi başka bir nedenden ötürü kendinizi rahatsız hissederseniz cevaplama işini yarıda bırakıp çıkmakta serbestsiniz. Böyle bir durumda çalışmayı uygulayan kişiye, çalışmadan çıkmak istediğinizi söylemek yeterli olacaktır. Araştırma sonlandıktan sonra yapılacak çekilişle 3 katılımcıya D&R mağazalarında kullanılabilir 100 TL değerinde hediye çeki vereceğiz. Çekiliş hakkı kazanmak için çalışmayı tamamlamanız gerekmektedir.

Araştırmayla ilgili daha fazla bilgi almak isterseniz:

Bu çalışmaya katıldığınız için şimdiden teşekkür ederiz. Çalışma hakkında daha fazla bilgi almak için ODTÜ yüksek lisans öğrencisi Süleyman Yaman (e-posta: [REDACTED]) ya da öğretim üyesi Doç. Dr. Duygu Özge (e-posta: [REDACTED]) ile iletişim kurabilirsiniz.

Yukarıdaki bilgileri okudum ve bu çalışmaya tamamen gönüllü olarak katılıyorum.

D. BACKGROUND QUESTIONNAIRE FOR THE L1 GROUP

Participant Background Information

Please complete the short form below. The information requested in this section are for demographic purposes only, and will remain confidential and anonymous. The fields marked with an asterisk (*) are required.

Age*:

.....

Gender:

Male Female Other Prefer not to say

Is English your native language?*

Yes No

Do you have other native languages? If yes, please indicate.

.....

What is your dominant hand?*

Left Right

Do you use glasses or contact lenses?*

Yes No

E. BACKGROUND QUESTIONNAIRE FOR THE L2 GROUP

Katılımcı Bilgi Formu

Lütfen aşağıdaki formu doldurunuz. Bu bölümde istenen bilgiler demografik amaçlıdır ve gizli tutulacaktır. Yıldız(*) ile işaretlenen alanların doldurulması zorunludur.

Yaşınız*:

.....

Cinsiyetiniz:

Erkek Kadın Diğer Belirtmek İstemiyorum

Anadil(ler)iniz nedir?*

.....

Kaç yaşında İngilizce öğrenmeye başladınız?*

.....

Hangi ortamda (ev, okul vb.)?*

.....

Baskın olarak hangi elinizi kullanıyorsunuz?*

Sol Sağ

Gözlük veya kontakt lens ile düzeltilen bir görme bozukluğunuz var mı?*

Evet Hayır

E-mail adresiniz (Çekilişe katılabilmemiz ve kazanmanız durumunda sizinle iletişime geçebilmemiz için gereklidir.)

.....

İngilizce Özyeterlilik Değerlendirmesi*

Nasıl değerlendirirsiniz (1=çok zayıf, 9=çok iyi)

İngilizce okuma becerinizi: 1 2 3 4 5 6 7 8 9

İngilizce yazma becerinizi: 1 2 3 4 5 6 7 8 9

İngilizce konuşma becerinizi: 1 2 3 4 5 6 7 8 9

İngilizce dinleme becerinizi: 1 2 3 4 5 6 7 8 9

F. TURKISH SUMMARY / TÜRKÇE ÖZET

İKİNCİ DİL İNGİLİZCE KONUŞUCULARINDA GEÇİCİ OLARAK SÖZDİZİMSEL BELİRSİZLİK İÇEREN TÜMCELERİN İŞLEMLENMESİ VE YORUMLANMASI

Giriş

İnsanlar, dili anlayarak ve üreterek iletişim kurma yeteneğine sahiptir. Bu becerinin altında dilsel girdinin gerçek zamanlı olarak yapısal temsillerle dönüştürüldüğü çözümleme (İng., *parsing*) süreçleri yatmaktadır (Altmann ve Kamide, 1999; Omaki vd, 2014; güncel bir özet için bkz., Özge, 2020). Çözümlemenin önemli bir özelliği, bireylerin tümce sonunu beklemelerine gerek kalmadan her dilsel birimi otomatik olarak işlemleyip bu birimleri anlık olarak mevcut temsile dahil edebiliyor olmalarıdır. Bu durum her ne kadar çoğu zaman avantaj sağlasa da, *While Mary dressed the baby fell off the bed* gibi geçici sözdizimsel muğlaklık içeren tümcelerde olduğu gibi işleme güçlüğüne sebep olup çözümleme süreçlerini anlık olarak sekteye uğratabilmektedir. Bu gibi muğlak tümceler ilk okunduğunda *the baby* isim öbeği *dress* eyleminin nesnesi olarak anlaşılmaktadır. Ancak, bireyler hemen sonra gelen *fell off* fiil öbeği ile karşılaştıklarında bu birimin mevcut temsile nasıl dahil edilebileceğini ilk başta anlayamadıkları için tümcede dilbilgisel bir bozukluk olduğunu düşünüp işleme güçlüğü (İng., *garden-path effect*) yaşarlar. Bu belirsizliğin giderilmesi için, *the baby* isim öbeğinin *fell off* eyleminin öznesi; *dress* eyleminin ise dönüşlü-geçişsiz fiil olarak revize edilmesi gerekmektedir.

Psikodilbilim alanyazınında, bu tarz tümceler çözümleyicinin dilsel girdiden nasıl gerçek zamanlı yapısal analizler ürettiğine ilişkin çeşitli tümle işleme kuramlarını motive etmede oldukça önemli bir rol oynamıştır (Frazier ve Rayner, 1982; Ferreira ve Clifton, 1986; MacDonald vd., 1994). Bu kuramlar, genel düzeyde tümce işleme mekanizmalarının farklı dilsel ipuçları nasıl kullandığını açıklamada; özelde ise

bireylerin neden ilk olarak yanlış bir şekilde nesne okuması (*dressed the baby*) takip ettiği konusunda farklılıklar gösterse de, çözümleyicinin her zaman dilsel girdinin doğru ve detaylı bir temsilini oluşturduğu fikri üzerinde üstü kapalı bir uzlaşma göstermiştir. Bu doğrultuda ise muğlaklığa işaret eden hata sinyaliyle (örn., *fell off*) karşılaşıldıktan sonra son doğru anlama başarılı bir şekilde ulaşıldığı varsayılmıştır. Ancak, özellikle Christianson ve arkadaşlarının (2001) çalışmalarından beri bu varsayım ciddi bir şekilde sorgulanmıştır. Anadil İngilizce konuşucusu yetişkinlerle yaptıkları deneylerde, Christianson ve arkadaşları (2001) muğlaklık barındıran tümcelerden sonra yöneltilen anlam sorularına (örn., *Did Mary dress the baby?; Mary bebeği giydirdi mi?*) katılımcıların sadece şans düzeyi %50'ye yakın bir oranda doğru cevaplar verdiğini gözlemlemiştir. Buna karşılık muğlak isim öbeğinin kendisini takip eden fiil öbeğinin öznesi olarak revize edilip edilmediğini sözdizimsel bir düzeyde test eden sorularda da (örn., *Did the baby fall off the bed?; Bebek yataktan düştü mü?*) %90'a yakın doğruluk oranları gözlemlemiştir. Christianson ve arkadaşlarının (2001) çıkarımlarına göre bu bulgular önceden varsayıldığı gibi sözdizimsel muğlaklığın tamamen revize edilmediğini, aksine ilk benimsenen *Mary bebeği giydirdi* şeklindeki dolaysız nesne okumasının önemli bir ölçüde korunduğuna işaret etmektedir. Bu bakımdan sözdizimsel revize sürecinin ancak kısmen yapıldığı anlaşılmaktadır. Her ne kadar bu önemli bulgular ilk başta açık bir şekilde yöneltilen anlam sorularının yanlış anlamı tekrar aktifleştirmiş olabileceği şeklinde eleştirel bir alternatif açıklamayla karşılanmış olsa da (bkz., Tabor vd., 2004); benzer bulgular yapısal hazırlama (van Gompel vd., 2006), metin okuma sırasında göz hareketlerini izleme (Slattery vd., 2013; Fujita ve Cunnings, 2020, 2021; Sturt, 2007) ve tekrar yazma (Karsenti ve Meltzer-Asscher, 2022; Patson vd., 2009) gibi daha örtülü yöntemlerle de gözlemlenmiştir.

Başka psikodilbilimsel olguların yanı sıra, muğlaklığın anlam düzeyinde tamamen revize edilmediğini gösteren bu bulgular yüzeysel (İng., *good-enough*) ya da iki-yollu tümce işleme kuramlarını motive etmede kritik bir rol oynamıştır (Ferreira vd., 2002; Kuperberg, 2007; Karimi ve Ferreira, 2016; Christianson, 2016). Bu kuramlara göre, dil işleyici birisi detaylı ve hiyerarşik yapısal temsillerin oluşturulduğu tam çözümleme, diğeri ise gerçek dünya bilgisi gibi sözdizimsel olmayan ipuçların anlamlandırma sürecini daha ağırlıklı belirlediği sıg/yüzeysel çözümleme yolu olmak üzere birbirinden bağımsız ve paralel çalışan iki yoldan yararlanmaktadır. Bu şekilde

ikili bir mekanizma üzerine kurulu dil işleme mimarisinde, yüzeysel çözümleme yapan yolun bazı durumlarda tam çözümleme yoluna göre daha ağır basıp Christianson vd. (2001) çalışmasında olduğu gibi yüzeysel anlamlandırmaya sebep olduğu düşünülmektedir.

Sıg işleme yolunun ne ölçüde işlev gösterdiği, farklı popülasyonlar arasında da değişiklik gösterebilmektedir. Clahsen ve Felser (2006a, 2018) tarafından ortaya atılan Sıg Yapı Hipotezi yukarıda bahsedilen teorileri temel alarak yetişkin ikinci dil (D2) konuşucularının gerçek zamanlı işleme esnasında sözdizimsel olarak sıg temsiller oluşturup daha çok anlamsal ve edimsel ipuçlarından yararlandığını ileri sürmektedir. Clahsen ve Felser'a (2006b, 2018) göre yüzeysel çözümleme yolunun ağır basma olasılığı D2 konuşucularında daha yüksek olabilmektedir. Mevcut çalışmada, bu hipotez ele alınarak D1 ve D2 konuşucularının revizyon süreçlerinde birbirinden farklılık gösterip göstermediği incelenmektedir.

Çalışmanın Önemi

Mevcut çalışmanın, hem D2 işleme hem de yüzeysel anlamlandırma hakkındaki alanyazına birkaç şekilde katkı sağlaması beklenmektedir. Birincisi, D2 alanyazınındaki birçok çalışma geçici muğlaklığa sahip tümceleri D2 konuşucularının fiil yanlılığı (Dussias ve Cramer-Scaltz, 2008; Şafak ve Hopp, 2022), durum imleme (Jackson, 2008) ve anlamsal yatkınlık (Roberts ve Felser, 2011) gibi ipuçlarını D1 konuşucuları gibi artımlı işleyip işleyemediğini incelemek için sadece bir tanı aracı olarak kullanmıştır. Bu tarz çalışmalar her ne kadar D1 ve D2 işlemlenin ne ölçüde benzer olduğu konusundaki tartışmaları bilgilendirecek önemli bulgular ortaya koymuş olsa da, söz konusu tümcelerin son anlamlandırma süreçlerini sistematik olarak inceleyen çalışma sayısı D2 alanyazınında nispeten azdır (Jacob ve Felser, 2016; Pozzan ve Trueswell, 2016; Fujita ve Cunnings, 2021). İkincisi, gerek D1 gerek D2 işleme alanyazınında yapılmış çalışmalar ağırlıklı olarak tümcenin sadece ilk benimsenen anlamını test eden anlam soruları kullanıp, tümcenin diğer kısımlarını doğrudan test etmemiştir. Her ne kadar bu eğilime istisna oluşturan bazı çalışmalar bulunsa da (örn., Christianson vd., 2001; Chromý, 2022), böyle çalışmaların sayısı muğlaklığın anlamlandırma süreçlerinin çok boyutlu bir şekilde anlaşılabilmesi için halen görece sınırlıdır. Bu bağlamda, mevcut çalışmada hem ilk anlamı hem de muğlak

isim öbeğinin sözdizimsel düzeyde başarılı bir şekilde revize edilip edilmediğini test eden iki tip anlam sorusu kullanılmaktadır. Son olarak, mevcut çalışma aynı zamanda D2 yeterlik seviyesinin revizyon süreçlerindeki etkisini keşfetmeyi hedeflemektedir. Yeterlik seviyesinin etkisinin incelenmesinin gerek geliştirmekte olan D2 çözümleyicisinin özellikleri ve nasıl bir gelişimsel yörünge izlediği, gerek D2 işlemedeki bireyler arası farklılıklar hakkında önemli kuramsal çıkarımlar sunması beklenmektedir.

Araştırma Soruları ve Tahminler

Mevcut çalışma aşağıdaki iki soruya cevap aramaktadır:

1. D2 İngilizce konuşucuları, geçici muğlaklığı çözümleme konusunda D1 konuşucularından farklılık gösteriyor mu?
2. İngilizce yeterlik düzeyindeki bireysel farklılıklar, D2 konuşucularının muğlaklık çözümleme becerileri etkiliyor mu?

Bu sorulara yanıt aramak için tümceleri anlam sorularının takip ettiği internet tabanlı bir kendi hızında okuma deneyi tasarlandı. Bu deney kapsamında, katılımcıların her sözcükteki okuma süresi kaydedilmiş ve her deneysel tümceden sonra anlam soruları sorulmuştur. İlk araştırması sorusu için, Sığ Yapı Hipotezi'nin öngörülerini kapsamında D2 konuşucularının D1'lere göre muğlak koşulda yöneltilen çevrimdışı anlam sorularına daha sık yanlış yanıtlar vermesi beklenmektedir. Böyle bir durumun yanıtlar üzerine yapılan istatistiksel analizlerde grup ve muğlaklık faktörleri arasında anlamlı bir etkileşim olarak kendini göstermesi beklenmektedir. Bu hipotezin öngörülerinin aksine, iki popülasyonun işleme rutinleri arasında sığ yola bağlı kalma konusunda önemli bir farklılık yoksa, D1 ve D2 konuşucularının benzer oranlarda yanlış yanıtlar vermesi beklenmektedir. İkinci araştırma sorusu için, çözümleme rutinlerinin D2 yeterliği ile geliştiğini gösteren önceki çalışmalar (örn., Cheng vd., 2021; Dallas vd., 2013) ile uyumlu olarak İngilizce yeterliğinin muğlaklık çözümleme becerilerini de etkileyeceği beklenmektedir. Özellikle üst yeterlik düzeylerinde, D1 grubuyla benzer örüntülerin gözlemlenmesinin, D1 ve D2 tümce işleme mekanizmaları arasında

önemli bir farklılık olmadığı görüşüne belli bir düzeyde destek vermesi beklenmektedir.

Katılımcılar

Çalışma, D1 ve D2 İngilizce konuşucuları olmak üzere iki gruptan oluşmuştur. Bütün katılımcılar çalışmaya internet üzerinden katılmıştır. D1 grubu 28 kişiden (yaş ortalaması : 29.07; aralık : 18-50), D2 grubu ise anadili Türkçe olan 37 kişiden (yaş ortalaması: 21.08; aralık : 18-32) oluşmaktadır. İlk grup ağırlıklı olarak sosyal medya platformu Reddit'in r/SampleSize alt forumu üyelerinden ve Türkiye'de yaşayıp anadili İngilizce olan bireylerden; ikinci grup ise ağırlıklı olarak ODTÜ öğrencisi ve mezunlarından oluşmuştur. D2 grubu içerisinde yeterlik düzeyi bakımından çeşitliliği arttırabilmek için 2022-2023 Sonbahar Dönemi itibariyle ODTÜ Temel İngilizce Bölümü'nde orta ve orta-üst kurlarda öğrenim görmekte olan öğrenciler de katılmıştır.

D2 grubu, dokuzluk bir Likert ölçeği üzerinde İngilizce dinleme, yazma, konuşma ve okuma becerilerini puanlamış; bu puanların ortalaması katılımcıların genel İngilizce skoru olarak belirlenmiştir (ortalama: 7.45; aralık : 4.50 - 9.00). Toplanan verinin olabildiği kadar doğal ve tipik bir davranış yansıtabilmesi için, katılımcıların İngilizce Öğretmenliği bölümü öğrencileri gibi dilbilim eğitimi almış bireyler arasından seçilmemesine dikkat edilmiştir. Katılımı teşvik edebilmek için D2 grubundaki üç katılımcıya çekiliş ile 100₺ değerinde hediye çeki verilmiştir.

Materyaller ve Desen

Deneysel uyaranlar 20 çift tümce setinden oluşmaktadır. Bu tümceler, Christianson vd. (2001) ve Huang ve Ferreira'nın (2021) çalışmalarından adapte edilerek, muğlaklık ve anlam sorularının tümce içinde hedeflediği kısım olmak üzere 2 x 2 faktöriyel desen kullanılarak aşağıdaki gibi manipüle edilmiştir:

Muğlak: While Anna dressed the baby who was cute and small fell off the bed.

Açık: While Anna dressed, the baby who was cute and small fell off the bed.

Yan tümce sorusu: Did Anna dress the baby?

Ana tümce sorusu: Did the baby fall off the bed?

Muğlak koşuldaki tümcelerde, *the baby* ad öbeği ilk başta *dress* eyleminin nesnesi olarak anlaşılırken tümcenin devamında karşılaşılan *fell off* eylemi bu anlamı geçersiz kılmaktadır. Muğlaklığın dilbilgisel olarak doğru çözümlenmesi ise *the baby* öbeğinin *fell off* eyleminin öznesi olarak, bir önceki *dress* eyleminin ise dönüşlü-geçişsiz fiil olarak revize edilmesiyle mümkün olmaktadır. Açık koşuldaki tümcelerde bulunan virgül ise bu geçici muğlaklığın ortaya çıkmasını engellediği için kontrol koşulu işlevi görmüştür. Daha önceki çalışmalarda (Christianson vd, 2001; Ferreira ve Henderson, 1991) bireylerin ilk çözümlenen yanlış anlama ne kadar uzun süre bağlı kalırlarsa revize süreçlerinde de o kadar başarısız olduğu gözlemlendiği için bütün deneysel tümcelerde muğlak olan isim öbeği birer ilgi tümcecisi ile nitelenmiştir (örn., *who was cute and small*). Belirsizliğin giderildiği bölgeden hemen sonra oluşabilecek çevrimiçi taşma etkilerinin gözlemlenebilmesi için bazı orijinal tümceler sonuna bir edat ya da zarf öbeği eklenerek adapte edilmiştir.

Yan tümce soruları, muğlaklığın sebep olduğu yanlış anlamın başarılı bir şekilde terk edip edilmediğini, ana tümce soruları ise muğlak isim öbeğinin (örn., *the baby*) sözdizimsel olarak revize edip edilmediğini test etmektedir. Yan tümce sorularının doğru cevabı Hayır, ana tümce sorularının ise Evet idi. Gerekli adaptasyonlardan sonra tümceler Latin karesi deseni takip edilerek dört listeye dağıtılmıştır. Deneydeki manipülasyonun katılımcılar tarafından fark edilmemesi için, her listeye 40 adet dikkat dağıtıcı tümce eklenmiştir. Bu tümcelerin bir kısmı yapı olarak test edilmekte olan olguya sözdizimsel karmaşıklık bakımından benzerlik göstermiştir. Bu tümcelerin yarısından sonra anlam soruları yöneltilmiştir. Evet/hayır yanlılığının deney içerisinde dengelenebilmesi için soru bulunan dikkat dağıtıcı tümcelerin yarısı için doğru yanıt Evet, diğer yarısı için Hayır idi. Soru tipi manipülasyonuna benzer bir şekilde dikkat dağıtıcı sorular da sık sık tümcelerin farklı kısımlarının doğru anlaşılıp anlaşılmadığını test etmiştir.

Yöntem

Tümceler, PciBex (Zehr ve Schwarz, 2018) platformu kullanılarak internet tabanlı bir kendi hızında okuma deneyinde sunuldu. Deneye kendi bilgisayarlarıyla uzaktan

katılan katılımcılar, ilk önce bir gönüllü katılım formunu okuyup onayladılar. Deney internet tabanlı olduğu için katılımcılardan etraflarında dikkat dağıtabilecek uyarıları olabildiği kadar azaltmaları istendi. Bu kısmı ise kısa bir demografik form takip etmiş, sonrasında detaylı olarak deneyde ne yapılması gerektiğine dair yazılı yönergeler sunulmuştur. Beş tümceden oluşan bir alıştırmaya bölümünden sonra da deneyin ana kısmı başlamıştır. Her sunulan tümce ilk önce sözcüklerin yerinde alt çizgiler olacak şekilde maskelenmiş, katılımcılar Boşluk tuşuna basarak her sözcüğü sırasıyla açarken bir önceki sözcük tekrar eski haline dönmüştür. Bu esnada her sözcükte katılımcıların okuma süreleri milisaniye cinsinden deney programı tarafından kaydedilmiştir. Tümceler en sonuna kadar okunduktan sonra anlam soruları gösterilmiş; Evet cevapları için katılımcılardan klavyede 1 tuşuna, Hayır cevapları için ise 2 tuşuna basmaları istenmiştir. Tümceler açık gri bir arkaplanda 14 boyutlu siyah Times New Roman fontunda sunulmuştur. Deney yaklaşık 15 dakika sürmüş, bilişsel yorgunluğun etkisini azaltmak için deney ortasında isteğe bağlı ara verilmiştir.

Veri Analizi

Toplanan veriler R programında okuma süreleri için lineer karma modeller; çevrimdışı soruların doğrulukları için ise genelleştirilmiş lineer karma modeller kullanılarak analiz edilmiştir (Baayen vd., 2008). Dikkat dağıtıcı öğelerde %75'in altındaki doğruluk oranları genellikle deneyde istenenlerin yeterince dikkatli bir şekilde yapılmadığına işaret ettiği için (bkz., Huang ve Ferreira, 2021; Cunnings ve Fujita, 2021) bu oranın altındaki iki D2 katılımcısı veri setinden çıkarılmıştır.

Okuma sürelerinin analizi için üç bölge belirlenmiş; bunlar sırasıyla sözdizimsel belirsizliğin ilk giderildiği kritik bölge (örn., *fell*) ve bu bölgede ortaya çıkacak işleme güçlüğüne taşması beklenen sonraki iki sözcüktür (örn., *off* ve *the*). Bu bölgelerde ilk önce 100 ms'den düşük ve D1 grubu için 3000, D2 grubu için 4000 ms'den yüksek okuma süreleri kaldırılmıştır. Bunun ardından okuma sürelerine log dönüşümü uygulanarak bütün bölgelerde her muğlaklık faktörü içinde 3.5 standard sapmanın üstünde ve altında kalan log okuma süreleri kaldırılmıştır. Bütün bu işlemler yaklaşık %3'lük bir veri kaybına sebep olmuştur.

İlk olarak birinci araştırma sorusunda odaklanılan gruplar arası farklılıkları test etmek için grubun (D1 ve D2) bağımsız değişken olduğu analizler yapılmıştır. İkinci araştırma sorusu için ise sadece D2 grubunun verisi temel alınarak bileşik genel İngilizce skorları bağımsız değişken alınarak aynı analizler tekrarlanmıştır.

Bulgular

Log dönüşümlü okuma süreleri ve anlam sorularına verilen yanıtların doğruluğu üzerine yapılan analizler sonucu birkaç bulgu öne çıkmaktadır. Birincisi, her iki grup da belirsizliğin giderildiği bölgede çevrimiçi işleme güçlüğü yaşamıştır. Bu işleme güçlüğü, muğlak koşuldaki tümcelerın açık koşuldakilere göre istatistiksel olarak daha uzun süreli okuma sürelerine sebep olması şeklinde ortaya çıkmaktadır. Ancak D2 grubunun yaşadığı işleme güçlüğü genel olarak daha zayıf olup analizler arasında değişkenlik göstermiştir. Daha açık ifade etmek gerekirse, grup farklılıklarına yönelik yapılan analizlerde hem D1 hem D2 grubu kritik bölgeden (örn., *fell*) itibaren anlamlı bir işleme güçlüğü yaşarken, bireysel farklılıklara yönelik yapılan analizlere göre D2 grubu sadece bir sonraki (örn., *off*) sözcükten itibaren anlamlı bir işleme güçlüğü yaşamaya başlamaktadır. İkincisi, muğlak koşuldaki yan tümce sorularına verilen yanıtların düşük doğruluk oranlarının da gösterdiği gibi ilk takip edilen anlam tamamen terk edilmemiş olsa da, ana tümce sorularında da muğlaklığın istatistiksel olarak anlamlı bir etkisi gözlemlenmiştir. Üçüncüsü, D2 grubu D1 grubuna göre çevrimdışı anlam sorularında genel olarak daha çok hata yapmıştır. Muğlaklığın çözümü konusunda ise iki grup birbirlerinden istatistiksel olarak anlamlı bir farklılık göstermemiştir. Son olarak İngilizce yeterliğindeki bireysel farklılıkların, belirsizliğin giderildiği bölgede yaşanan çevrimiçi işleme güçlüğüne ya da anlam sorularına verilen yanıtlara anlamlı bir etkisi gözlemlenmemiştir.

Tartışma

Çalışmada, tümce işleme alanyazınında iyi bilinen yanlış yorumlama (İng., *lingering misinterpretations*) olgusu replike edilmiştir. D1 grubu, muğlak koşulda bulunan tümceleri takip eden yan tümce sorularına %50, ana tümce sorularına ise %92 oranlarında doğru cevaplar vererek Christianson ve arkadaşlarının (2001) çalışmasındaki katılımcılarla benzer davranışı göstermiştir. Bununla birlikte

muğlaklığın istatistiksel etkisi soru tipinden anlamlı olarak etkilenmediği için bu durum katılımcıların muğlaklığı revize ederken sözdizimsel düzeyde bazen tamamen başarısız olduğuna işaret etmektedir. Her ne kadar analizlerde güven aralıklarının genişliği bu istatistiksel etkileşimsizliğin katılımcı sayısının yetersiz olabileceğinden kaynaklandığı izlenimini verse de, neden küçük bir deneme kümesinde bu çözümleme başarısızlığı görülürken diğerlerinde görülmediği gelecek çalışmalarda incelenebilecek önemli bir araştırma alanı olabilir.

D2 grubunda gözlemlenen sonuçlar Jacob ve Felser (2016) tarafından raporlanan örüntüye benzerlik göstermektedir. Bu grup daha zayıf ve değişkenlik gösteren çevrimiçi işleme güçlüğü yaşarken aynı zamanda anlam sorularına genel olarak daha çok yanlış cevaplar vermiştir. D2 grubunun çevrimdışı veride zaman zaman açık koşuldaki tümcelerde de D1 grubuna göre güçlük yaşadığı anlaşılan bu örüntü Sığ Yapı Hipotezi'ni (Clahsen ve Felser, 2006) tamamıyla desteklemiyor olabilir. Bu tablo, sözdizimsel revizyon ve anlamsal olarak akla yatkın bir çözümlemenin terk edilme ihtiyacının olmadığı durumlarda da D2 konuşucularının bazen güçlük yaşadığına işaret etmektedir. Yine D2 grubu yan tümce sorularında D1 grubuna göre muğlak koşulda daha fazla anlam hataları yapmadığı için Sığ Yapı Hipotezi'ne alternatif olarak Cunnings'in (2017) ortaya attığı bozucu etki (İng., *interference*) kuramı da destek görmemektedir. Nitekim bu kuramda D2 konuşucularında görülebilecek olası bir yanlış yorumlama dezavantajının sığ çözümleme rutinlerinden ziyade D2 konuşucularının bellek geri çağırma operasyonlarında ilk çözümlemenin bozucu etkisine karşı daha yatkın olmalarından kaynaklandığı öne sürülmektedir (bkz., Slattery vd., 2013; Lewis ve Vasishth, 2005). Bu şekilde bir örüntünün olmayışının yanı sıra, alanyazında başka olgulardan Cunnings'in (2017) iddialarını destekleyecek önemli ölçüde ampirik bir kanıt olmadığı ve D2 konuşucularının niçin geri çağırma operasyonlarında böyle bir güçlük yaşayabileceğine ilişkin açık bir neden olmaması da göz önünde bulundurulmalıdır.

Bu iki modelin aksine, bilişsel kapasite sınırlılığına dikkat çeken modellerin (McDonald, 2006; Hopp, 2010) elde edilen bulguları daha kapsamlı açıklayabilme potansiyeli vardır. Bu modeller, D1 ve D2 konuşucuları arasında belirgin olarak farklı sözdizimsel çözümleme mekanizmalarının bulunmadığını, sonradan edinilmiş bir D2'yi işleme sadece daha yavaş olduğunu ve işler bellekteki mevcut kaynakları

daha hızlı tükettiğini ileri sürmektedir. Bu durumun sözdizim ile sınırlı kalmayıp bütün dilsel alanları benzer şekilde etkilemesi beklendiği için mevcut çalışmada D2'lerin muğlaklıktan bağımsız gözlemlenen genel çevrimdışı işleme güçlüğü bu şekilde bir genel işleme sorunuyla açıklanabilir. Ayrıca, deneyde katılımcılar ilgi tümceciği barındıran nispeten uzun ve karmaşık cümleleri akıllarında tutarak anlam sorularını cevaplamak zorunda olduğu için bu durum bellek kaynakları üzerinde var olan yükü daha da arttırmış olabilir. Benzer bir şekilde, söz konusu kuramlar zayıf ve değişken çevrimiçi işleme güçlüğü de D2'lerin sözdizimsel hata sinyali D1'ler gibi anında fark edebilmek için yeterli işler bellek kaynaklarına da sahip olamayabileceği üzerinden açıklayabilir.

Ancak, D2 yeterliğindeki bireysel farklılıkların tümceler işlenmesine hiçbir etkisininin olmaması bu kuramlar tarafından açıklanamayabilir. Nitekim, söz konusu kuramlar artan D2 yeterliği ile birlikte bilişsel kaynaklar üzerindeki yükün de hafifleyeceğini ve dolayısıyla D2 işlememenin daha hızlı olmasını bekler (bkz., Abutalebi, 2008). Elde edilen veride, özellikle tepki sürelerinde, böyle genel bir örüntünün olmaması kapasite kuramları için sorun oluştursa da, D2 grubundaki katılımcı sayısı yetersizliğinin söz konusu tablonun ortaya çıkmasına sebep olmuş olabileceği de göz önünde bulundurulmalıdır.

Mevcut çalışmanın birtakım önemli kısıtları da bulunmaktadır. Yukarıda da bahsedildiği gibi bu kısıtlardan birisi örneklem büyüklüğünün alanyazındaki benzer çalışmalara göre nispeten sınırlı olmasıdır. Bu durum genel düzeyde çevrimdışı doğruluk analizlerindeki güven aralıklarının geniş olmasına sebep olurken daha spesifik düzeyde de D2 yeterliğindeki bireysel farklılıkların ölçümü için yetersiz bir istatistiksel güce sebep olmuş olabilir. Bu durum, çalışmanın daha büyük örneklerle replike edilmesi gerekliliğini ortaya koyabilir. Bununla birlikte çalışmada D2 katılımcılarının yeterlik düzeyi standardize edilmiş testler olmaksızın sadece öz-yeterlik ölçekleri kullanılarak ölçülmüştür. Söz konusu ölçeklerin kullanımı her ne kadar alanyazında son derece yaygın olup (Favier vd., 2019; Luk ve Bialystok, 2013; Tan ve Foltz, 2020) mevcut çalışmada pratik sebeplerden ötürü seçilmiş olsa da, yapıları gereği öznel bir ölçü olmaları bireysel farklılıkların anlamlı bir şekilde yakalanamamasına neden olmuş olabilir. Bundan ötürü gelecek çalışmaların, standardize edilmiş testler kullanmayı da göz önünde bulundurması gerekebilir. Son

olarak, D1 alanyazınında baskılama (Novick vd., 2005) ve işler bellek uzamı (Christanson vd., 2006) gibi bilişsel faktörlerdeki bireysel farklılıkların da sözdizimsel muğlaklığın çözümlenmesinde önemli olduğu bilindiği için, bunların D2 yetişkinlerde de incelenmesi önemli bir araştırma alanı olabilir.

Sonuç

Mevcut çalışma, D1 konuşucularında görülen yanlış yorumlama olgusunu replike etmiştir. D1 grubundaki bulgular, Christianson ve arkadaşlarının (2001) bulgularıyla büyük ölçüde örtüşmektedir. Ayrıca, D2 konuşucusu bireylerin İngilizce yeterlik seviyesinden bağımsız olarak D1 kontrol grubuna göre daha zayıf ve değişken işleme gücüyle yaşayıp çevrimdışı işlemede de daha çok hata yaptığı saptanmıştır. Mevcut bulguların alanyazında önde gelen D2 işleme modelleri çerçevesinde tartışılmasının ardından, kapasite sınırlılığı kuramlarının elde edilen sonuçları en kapsamlı bir şekilde açıklayabileceği sonucuna varılmıştır. Ancak, getirilen açıklama belirli bir ölçüde post hoc olduğu için gelecek çalışmalarda daha sık test edilmeye ve kapsamlı araştırmaya ihtiyaç duyabilir.

G. THESIS PERMISSION FORM / TEZ İZİN FORMU

(Please fill out this form on computer. Double click on the boxes to fill them)

ENSTİTÜ / INSTITUTE

- Fen Bilimleri Enstitüsü / Graduate School of Natural and Applied Sciences
- Sosyal Bilimler Enstitüsü / Graduate School of Social Sciences
- Uygulamalı Matematik Enstitüsü / Graduate School of Applied Mathematics
- Enformatik Enstitüsü / Graduate School of Informatics
- Deniz Bilimleri Enstitüsü / Graduate School of Marine Sciences

YAZARIN / AUTHOR

Soyadı / Surname : Yaman
Adı / Name : Süleyman
Bölümü / Department : İngiliz Dili Öğretimi / English Language Teaching

TEZİN ADI / TITLE OF THE THESIS (İngilizce / English): Processing and Interpretation of Garden-Path Sentences in L2 Speakers of English

TEZİN TÜRÜ / DEGREE: Yüksek Lisans / Master Doktora / PhD

1. **Tezin tamamı dünya çapında erişime açılacaktır.** / Release the entire work immediately for access worldwide.
2. **Tez iki yıl süreyle erişime kapalı olacaktır.** / Secure the entire work for patent and/or proprietary purposes for a period of **two years**. *
3. **Tez altı ay süreyle erişime kapalı olacaktır.** / Secure the entire work for period of **six months**. *

* Enstitü Yönetim Kurulu kararının basılı kopyası tezle birlikte kütüphaneye teslim edilecektir. / A copy of the decision of the Institute Administrative Committee will be delivered to the library together with the printed thesis.

Yazarın imzası / Signature

Tarih / Date

(Kütüphaneye teslim ettiğiniz tarih. Elle doldurulacaktır.)
(Library submission date. Please fill out by hand.)

Tezin son sayfasıdır. / This is the last page of the thesis/dissertation.