Syntactic processing in English–Afrikaans bilinguals

by

Robyn Berghoff

Dissertation presented for the degree of Doctor of Philosophy (General Linguistics) in the Faculty of Arts and Social Sciences

at

Stellenbosch University

Supervisor: Prof Emanuel Bylund

March 2020

DECLARATION

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: March 2020

ABSTRACT

A fundamental question in the study of human language is why, compared to the acquisition of a first language (L1), second language (L2) acquisition should have such widely varying outcomes. Relatedly, there is a question regarding the upper limits on L2 acquisition, namely whether it is possible for learners who have not acquired a language from birth to perform identically to native speakers of that language.

Experimental psycholinguistic techniques offer insight into the moment-by-moment processes involved in language comprehension and production, and in recent years have increasingly been employed to investigate L2 and bilingual processing, both in their own right and in relation to L1 processing. In this dissertation, such techniques are employed to investigate L2 English syntactic processing among early L2 acquirers (L1 Afrikaans) who receive extensive naturalistic exposure to the L2 and have attained high proficiency therein. Second language populations with this combination of features, each of which has been shown to affect processing outcomes, are understudied in the literature, as are highly linguistically diverse settings such as South Africa. There is thus little information available regarding the consequences of this particular constellation of individual- and environmental-level characteristics on ultimate levels of L2 attainment.

The studies presented in the dissertation focus on three syntactic phenomena that have been shown to be processed in a non-nativelike fashion by L2 speakers. These are temporarily ambiguous or so-called garden-path sentences, pronouns, and long-distance wh-dependencies. The techniques of self-paced reading and eye-tracking-while-reading were utilized to obtain real-time processing data. These data were supplemented by measures of L2 proficiency and language background. First-language speakers of South African English were employed as a comparison group.

The findings show L1–L2 convergence for a subset of the L2 participants – those with a relatively earlier age of L2 acquisition – in the garden-path sentence processing experiment. In the pronoun resolution experiment, evidence of cross-linguistic influence at the verb level is observed, which subsequently affects processing at the sentence level. Finally, in the processing of long-distance wh-dependencies, the strategies employed by the L1 and L2 speakers differ, with awareness of an abstract syntactic cue being evident in the L1 but not the L2 speakers.

The results provide insight into the implications of the South African language acquisition and use contexts for L2 development. A more general consideration of these implications as they relate to other multilingual settings contributes to our knowledge of L2 attainment in linguistically heterogeneous environments.

OPSOMMING

'n Fundamentele vraag in die bestudering van menslike taal is waarom die verwerwing van 'n tweede taal (T2), in vergeleke met die verwerwing van 'n eerste taal (T1), sulke wyd uiteenlopende uitkomste het. Daar is ook 'n verwante vraag rakende die boonste perke van T2-verwerwing, naamlik of dit moontlik is vir leerders om identies met moedertaalsprekers te presteer as hulle nie die taal vanaf geboorte verwerf het nie.

Eksperimentele psigolinguistiese tegnieke bied insig in die oomblik-tot-oomblik prosesse wat by taalbegrip en -produksie betrokke is, en sulke tegnieke word toenemend gebruik om T2- en tweetalige verwerking, sowel in eie reg as in verband met T1-verwerking, te ondersoek. In hierdie proefskrif word sulke tegnieke gebruik om T2 Engelse sintaktiese verwerking onder vroeë T2-verwerwers (T1 Afrikaans) wat uitgebreide naturalistiese blootstelling aan die T2 het en wat 'n hoë vaardigheid in die T2 behaal het te ondersoek. In die bestaande literatuur is T2 populasies met hierdie kombinasie van eienskappe, wat individueel bewese invloed op verwerkingsuitkomste het, onderbestudeer. Verder is uiters linguisties-diverse omgewings soos Suid-Afrika ook onderbestudeer. Daar is dus min inligting rakende die gevolge van hierdie spesifieke konfigurasie van individuele- en omgewingseienskappe op uiteindelike vlakke van T2-bereiking.

Die studies wat in die proefskrif voorgelê word, fokus op drie sintaktiese verskynsels wat gedemonstreer is om deur T2 sprekers op 'n nie-moedertaalagtige wyse verwerk te word. Hierdie verskynsels sluit in: tydelik-dubbelsinnige of sogenaamde "garden-path"-sinne, voornaamwoorde en langafstand wh-afhanklikhede. Die tegnieke van eie-tempo lees en oognaspeuring-tydens-lees is gebruik om reële-tyd verwerkingsdata te bekom. Hierdie data is aangevul deur maatstawwe van T2-vaardigheid en taalagtergrond. Eerstetaalsprekers van Suid-Afrikaanse Engels is as kontrole groep gebruik.

In die eksperiment oor die verwerking van tydelik-dubbelsinnige sinne, is daar bevind dat van die T2-deelnemers – diegene met 'n betreklik vroeëre ouderdom van T2-aanleer – T1–T2 konvergensie vertoon. In die voornaamwoord-resolusie eksperiment is bewyse van kruis-linguistiese invloed op die werkwoordvlak, wat verdere verwerking op sinsvlak beïnvloed, waargeneem. Laastens verskil die strategieë wat deur die T1- en T2-sprekers aangewend word in die verwerking van langafstand wh-afhanklikhede, met T1-sprekers wat, anders as T2-sprekers, bewus is van 'n abstrakte sintaktiese leidraad.

Die resultate bied insig in die implikasies van die Suid-Afrikaanse taalverwerwing- en gebruikskontekste vir T2-ontwikkeling. 'n Algemene oorweging van hierdie implikasies, soos hulle verband hou met ander meertalige instellings, dra by tot ons kennis van T2-bereiking in linguisties heterogene omgewings.

ACKNOWLEDGEMENTS

My primary thanks go to Prof Emanuel Bylund, who provided invaluable guidance and support from the very beginning to the very end of this project. I have learned a great deal from you in the past three years. Thank you also to the members of the Department of General Linguistics at Stellenbosch University for their friendship, support and various forms of assistance.

I would like to thank Profs Claudia Felser and Harald Clahsen for welcoming me at the Potsdam Research Institute for Multilingualism (PRIM), and for their feedback on my work. Thanks also to the members of PRIM for their company, and especially to Dr Judith Schlenter for helping with the logistics of my stay. Further to the north, I would like to thank Prof Marianne Gullberg, both for the discussions we had and for connecting me with others in Lund with whom I could discuss this project.

Some of the studies reported on in this dissertation made use of experimental materials that were developed by other researchers. I would like to acknowledge the work of Dr Leah Roberts and Prof Claudia Felser in developing these materials and thank them for making these materials publically available.

Thank you to all the individuals who participated in the studies reported on here, and to those who assisted me with data collection, especially Jayde Mcloughlin and Carmel Carne.

This project was funded by the National Research Foundation (NRF) and the German Academic Exchange Service (DAAD). I gratefully acknowledge the support of both of these organizations, and I would also like to thank the DAAD for making possible my research stay in Potsdam.

Finally, thank you to my parents, for the years of care and support without which I would not be here today.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 General introduction

The focus of this dissertation is on syntactic processing in second language (L2) speakers. Syntactic processing involves the construction of a mental representation of how the elements in a sentence relate to one another. Previous studies, the majority of which have been conducted with post-childhood L2 learners living outside of a context in which the L2 is spoken, have identified persistent differences between first language (L1) and L2 processing for certain syntactic phenomena. In light of these findings, the aim of the three experiments presented here is to determine whether L1–L2 convergence is possible amongst early L2 learners (L1 Afrikaans–L2 English) who have received considerable naturalistic exposure to the L2 from a young age. The findings bear on a central question regarding the human capacity for language, namely whether it is possible to attain nativelike proficiency in a language that is not acquired from birth.

This first chapter introduces the topics of L2 acquisition and processing, identifies the factors that affect these, and presents a number of models that have been proposed to explain the observed differences between L1 and L2 processing. It then describes the context and participants of the present study and details the research questions the study sets out to answer.

1.2 Second language acquisition and L2 processing

First language acquisition has been said to be characterized by two factors: reliability, in that it reliably ends in mastery of the language; and convergence, in that the linguistic system the L1 acquirer develops is highly similar to those of other members of the same speech community (Bley-Vroman, 2009). These characteristics are not universally true of L2 acquisition, particularly when it begins after early childhood. Why this should be the case has been the subject of intense research interest for many decades.

A key question in this domain is whether it is in fact possible to obtain nativelike proficiency in a non-native language. Individuals who seem to have done so are sufficiently rare that they are the object of "much admiration and even astonishment" (Hyltenstam & Abrahamsson, 2003, p. 539).[1] Even these rare individuals, however, do not always resemble native speakers when their linguistic behavior is subjected to further scrutiny. In particular, sole reliance on offline data collection techniques such as speech production and grammaticality judgment tasks may yield false positives in this regard, as these techniques are often insufficiently sensitive to discern subtle differences between natives and non-natives. Therefore, to address the question of whether attaining nativelike L2 proficiency is possible, offline techniques are best complemented by examinations of language comprehension and production as it unfolds in real time. In recognition of this necessity, experimental psycholinguistic techniques have increasingly been used to study L2 processing in recent years.

[1] In this dissertation, use of the term "nativelike(ness)" is not intended to set up a hierarchy of superior and inferior performance. It is stressed that success in language learning can be defined in various ways, and if the key criterion is being able to negotiate the demands of everyday life in a non-native language, the world can be said to be filled with many highly successful multilinguals (see e.g., Ortega, 2019, for discussion of this point). Problems with the concept of the 'native speaker' (discussed extensively from a sociolinguistic perspective, see e.g., Rampton, 1990, Cook, 1999) are acknowledged. However, the text that follows aligns with the convention in the syntactic processing literature in referring to L1 speakers as "native speakers" and their performance as "nativelike".

Broadly speaking, two kinds of language processing can be distinguished: lexical processing and grammatical processing. The former, which involves accessing word-level knowledge stored in the lexicon, is typically considered to be less problematic for L2 learners, with differences between native and non-native speakers being relatively subtle (see e.g., Lemhöfer et al., 2008). Grammatical processing can be further divided into morphological processing, which encompasses the comprehension and production of morphologically complex words[2], and syntactic processing, which involves the construction of abstract, hierarchical representations of linguistic input. The focus of the present dissertation is on syntactic processing.

The domain of syntactic processing is itself broad, encompassing the wide range of syntactic phenomena found in natural language. As might be expected, the extent of L2 processing nativelikeness has been found to vary across different types of syntactic phenomena. The following section provides a brief overview of L2 syntactic processing research, with a focus on the points of convergence with and divergence from L1 processing that have been observed.

1.3 A brief overview of L2 syntactic processing research

One of the dividing lines across which L2 syntactic processing appears to become less nativelike is that between local and non-local syntactic dependencies. A local syntactic dependency is present when the formal features of one constituent impose requirements on those of a neighboring constituent. Such dependencies may occur

[2] See Clahsen, Felser, Neubauer, Sato and Silva (2010) for a review of research in this area.

within a phrase: for example, the presence of the auxiliary *had* in English requires the verb to take a particular ending (cf. *I had eaten the Medjool dates* vs. **I had eat the Medjool dates*). Alternatively, they may exist across a phrase boundary, as in the case of subject-verb agreement, where the person and number features of the subject determine those of the verb (cf. *She loves to eat Medjool dates* vs. **She love to eat Medjool dates*), and Bantu noun class agreement, where nominal modifiers, pronouns and verbs must all agree with the noun class of the head noun (see e.g., Katamba, 2003). In all cases, assuming incremental left-to-right processing (Marslen-Wilson 1973, 1975; see section 2.1 for further discussion) and given a set of constituents in a dependency relationship, the form of the constituent encountered first generates a prediction about the form of the upcoming element. When this prediction is not fulfilled, processing difficulty is incurred.

Local dependencies differ from non-local or discontinuous dependencies, where the elements in a dependency relationship are not linearly adjacent. Wh-fronting in English provides a prototypical example of such a non-local dependency. For example, *Which dates* in *Which dates does the girl like to eat __ ?* can only be interpreted when the reader or listener reaches the end of the sentence, as it is only at this point that *Which dates* can be identified as the object of *eat*. Again assuming incremental processing, the moved element *Which dates* must be held in working memory as the remainder of the sentence is processed, and finally reintegrated into the syntactic representation at the sentence's end, where it originated. A characteristic of human language is that such dependencies may in principle be unbounded, extending over many clauses, as in *Which dates did the girl think the hotel receptionist would advise the manager to purchase __ ?* Greater distance between a moved element and its site for reintegration has been shown to correlate with increased reintegration difficulty, which is attributed to the cost of processing the intervening linguistic material while holding the moved element in memory (Gibson, 1998, p. 11).

Wh-dependencies in which the wh-constituent is fronted may be described as “forward-looking”, in that the wh-element must be linked to a constituent that occurs at a later position in the sentence. Backwards-looking dependencies are also possible, and are found, for example, in sentences where an anaphoric element must be linked to a constituent that occurs earlier in the sentence. *Himself* in *John believes the rugby player admires himself* must be linked back to an earlier noun phrase in the sentence to be interpreted; the same is true for *him* in *John believes the rugby player admires him* (although *him* also allows for an antecedent previously identified in the discourse context).

In terms of comparisons between L1 and L2 syntactic processing, it seems that convergence is most likely with local dependencies. Advanced L2 learners, for example, have been found to show sensitivity to morphosyntactic agreement violations in their L2, regardless of differences in the agreement system of their L1 (Rossi, Gugler, Friederici, & Hahne, 2006; Rossi, Kroll, & Dussias, 2014; Sagarra & Herschensohn 2010, 2011). However, sensitivity to agreement violations seems to decrease as the distance between the constituents in a dependency relationship increases (Dowens, Vergara, Barber, & Carreiras, 2010; Keating, 2009), in accordance with the idea that syntactic predictions maintained in memory over longer temporal distances are more costly (Gibson, 1998, p. 8), and so may overburden an already taxed L2 processing system.

In sentences where a displaced element must be reintegrated into its canonical position in the sentence, L2 learners' processing behavior appears to be more nativelike given the availability of an overt marker of that canonical position. For example, in a sentence like *Which machine did the mechanic fix the motorbike with* __ *two weeks ago?*, the preposition *with* obligatorily takes an object, and so when the reader or listener encounters *with*, they are provided with a cue that *Which machine* should be reintegrated. Native and non-native convergence of the processing of such sentences has been observed, for example, by Williams, Möbius and Kim (2001) and Williams (2006). Further, non-native speakers have also been found to respect syntactic constraints that determine where a gap position can occur. So-called island constraints (Ross, 1967) prevent extraction from a complex noun phrase, such as *the author who wrote with exceptionally great skill*, where *the author* is modified by the relative clause beginning with *who*. In Felser, Cunnings, Batterham and Clahsen (2012), both native and non-native speakers were aware that in sentences of the form *Everyone liked the book that the author who wrote* __ *with exceptionally great skill saw* __ *while waiting for a contract*, the position after *wrote* is not a possible extraction site of *the book* (see Boxell and Felser, 2017, for similar results with another type of island constraint).

Despite this convergence, which shows L2 speakers' sensitivity to complex and nuanced aspects of the L2 grammar, differences between L1 and L2 behavior seem to persist in more subtle aspects of online long-distance discontinuous dependency processing. For native speakers, the presence of intervening clause boundaries appears to facilitate reintegration of a moved element, as reflected in shorter reading times in multi-clause compared to single-clause sentences of equivalent length in words (Gibson & Warren, 2004; Marinis, Roberts, Felser, & Clahsen, 2005; Pliatsikas & Marinis, 2013). This observation is in accordance with a theoretical proposal that the dislocated element moves through all clause boundary positions on the way to its landing site – so-called successive cyclic movement (Chomsky, 1986) – due to a syntactic principle that prohibits a moved element from crossing a clause boundary without first moving to that clause boundary (see Ross, 1967; Chomsky, 1977). In processing, successive cyclic movement entails the reactivation of the moved element at each clause boundary position, thus reducing the temporal distance over which it must be retrieved at the point of reintegration. In contrast to those of their native-speaker peers, the reading patterns of non-native speakers do not show a facilitative effect of the clause boundary, suggesting that they may not be equally sensitive – or able to react equally rapidly to – the presence of such boundaries, which are not marked by a subcategorizing element, during online dependency processing (Felser, 2015).

The extent of L2 nativelikeness has also been observed to vary in studies of syntactically constrained anaphora resolution.[3] For reflexive anaphors, ending in the suffix *–self* in English, syntactic constraints dictate that only the nearest c-commanding[4] NP can serve as antecedent. Thus, in *John believes the rugby player*

[3] I am setting aside here studies of the interpretation of ambiguous pronouns, where resolution is not syntactically constrained. For treatments of this topic, see e.g., Contemori, Asiri and Perea Irigoyen (2019), Cunnings, Foutiadou and Tsimpli (2017), Ellert (2013), Roberts, Gullberg and Indefrey (2008), Schimke, de la Fuente, Hemforth and Colonna (2018), and Sorace and Filiaci (2016).

[4] The term "c-command" denotes a structural relationship between constituents, where a constituent A c-commands a constituent B if A does not directly dominate B, and if every node that dominates A also dominates B (Chomsky, 1981). The local domain, defined loosely, denotes the nearest phrase headed by a subject.

admires himself, *himself* cannot refer to John. Online processing studies indicate that L1 speakers are sensitive to this constraint and do not consider other structurally illegal antecedents during processing. However, variation in L2 behavior has been observed. In two eye-tracking-while-reading experiments (see section 3.3.2 for further details on this method), L2 speakers were observed to consider an illegal antecedent in the early stages of processing, even when the same syntactic constraint applied in their L1 (Felser & Cunnings, 2012) and when they demonstrated knowledge of the constraint in an offline task (Felser & Cunnings, 2012; Felser, Sato, & Bertenshaw, 2009). In contrast, in a task using the visual world paradigm, L1 Korean–L2 English speakers, whose L1 binding constraints differ from those of the L2 (Kim, Montrul, & Yoon, 2010), patterned with native speakers, in that they seemed only to consider the structurally accessible antecedent (Kim, Montrul, & Yoon, 2015).

For non-reflexive anaphors – pronouns – a syntactic constraint rules out the nearest NP as a possible antecedent for the pronoun. Thus, in *John believes the rugby player admires him*, *the rugby player* cannot be the referent of *him*: *him* must either refer to John or to another male referent present in the discourse context. In contrast to reflexive interpretation, then, syntactic constraints here do not identify the correct antecedent, but merely serve to rule out an illegal antecedent.

Few studies to date have examined the online application of this constraint in non-native processing.[5] Patterson, Trompelt and Felser (2014) found that L1 German speakers of L2 English – whose L1 is again subject to the same syntactic constraint –

[5] There have however been offline studies that are not reviewed here; see e.g., Slabakova, White and Brambatti Guzzo (2017) and White (1998).

patterned equivalently to native speakers, in that they did not seem to consider the inaccessible NP as a possible antecedent for the pronoun. However, based on the findings for another experiment testing the interpretation of pronouns exempt from this constraint, these authors suggested that the L2 speakers' processing patterns were potentially also compatible with a strategy where the anaphor was simply linked to the most discourse-prominent antecedent. That is, in a sentence like *John believes the rugby player admires him*, *John*, by virtue of being in the subject position and being mentioned first in the sentence, is most prominent, and so therefore may be identified on discourse or pragmatic grounds as the most likely antecedent for the pronoun. While the non-native speakers' behavior was compatible with the use of such a strategy, that of the native speakers was not. A further indication that L2 pronoun resolution may not necessarily be nativelike is provided by Kim et al.'s (2015) study with L1 Korean speakers. Here, even high-proficiency L2 English speakers were shown to consider the structurally illegal NP as antecedent, both in their responses to an antecedent choice task and in their eye movements.

Taken together, these results indicate that there is still a lack of clarity regarding the extent to which L2 speakers can process reflexives and pronouns in a nativelike way. In this regard, Felser (2015) proposes that backwards-looking non-local dependencies may be especially challenging in L2 processing, as the reader or listener is not alerted beforehand to the presence of an element that will need to be linked to a constituent appearing earlier in the sentence. Thus, upon encountering such an element, they must access from memory the syntactic representation they have constructed thus far and recall the potential antecedents contained therein. In contrast, in forwards-looking dependencies, the identification of a moved element triggers a search for a potential gap later in the sentence, and reliance on previously built syntactic representations is diminished.

Felser (2015) relates this account of L2 speakers' difficulty with backwards-looking dependencies to another area in which L2 processing has been observed to be non-nativelike, namely the resolution of syntactic ambiguities. Syntactic ambiguity resolution is often tested in reading-time studies of so-called garden-path sentences, where an initially incorrect analysis of the input has to be revised. An example of such a sentence is *The old man the boat*. Upon the first read, *man* is likely to be understood as a noun, modified by the adjective *old*. When the end of the sentence is reached, however, it becomes clear that such an analysis results in ungrammaticality. To recover a grammatical interpretation of the sentence, *man* must be reanalyzed as a verb and *old* as a noun. Reanalysis of this nature also requires re-accessing and modifying portions of an already constructed syntactic representation. In alignment with the idea that this may be difficult for L2 speakers, comparisons between native and non-native syntactic ambiguity resolution have shown that non-native speakers are more susceptible to comprehension breakdown (Jacob & Felser, 2016; Juffs & Harrington, 1996; Roberts & Felser, 2011), and they also seem to experience difficulties with online reanalysis (Jessen & Felser, 2018; Roberts & Felser, 2011).

To summarize this sub-section: it seems relatively uncontroversial that L2 speakers can process local dependencies, such as morphosyntactic agreement across adjacent constituents, in a nativelike fashion. However, as the distance over which a dependency extends increases, processing difficulty seems to increase, potentially because the features of the first constituent must be maintained in memory until the subsequent constituent is encountered, at which point agreement between the two can be determined. In the case of forward-looking dependencies, L2 speakers have been observed to differ from native speakers in the use of an abstract syntactic cue – that of the clause boundary – when processing discontinuous dependencies that extend across multiple clauses. When determining the referent of a reflexive anaphor or a pronoun, conflicting results have been obtained regarding the nativelikeness of L2 processing. Non-nativelikeness in this domain has been linked to L2 speakers' difficulty in accessing and/or revising already constructed syntactic representations during online processing, which has been made apparent by studies of the processing of garden-path sentences. The present dissertation focuses on three of these phenomena that have yielded reliable L1–L2 processing differences: ambiguity resolution, pronoun resolution, and long-distance wh-dependency processing. As such, the theoretical review presented in the remainder of this chapter pays particular attention to these phenomena.

This sub-section has alluded to the somewhat obvious fact that there is variation not only across native and non-native syntactic processing, but within non-native syntactic processing as well. The following sub-section reviews the factors relevant to the present dissertation that seem to have the most influence on the nativelikeness of L2 processing.

1.4 Factors affecting L2 syntactic processing

A number of factors distinguish L2 speakers of a language from L1 speakers, all of which may affect the nature of L2 speakers' linguistic performance. This section discusses the age of acquisition (AoA) of the target language, L2 proficiency, working memory, cross-linguistic influence and the setting in which the L2 is acquired.

1.4.1 Age of acquisition

The acquisition of an L1, in most cases, begins at birth (and even before, see e.g., Oller, 2014). As noted at the beginning of this chapter, the end product of L1 acquisition is typically mastery of the language and a language system that is highly similar to that of adult speakers of the same linguistic community. Cases in which the outcome of L1 acquisition deviates seem to be characterized by a lack of sufficient linguistic input during early childhood. The more extreme examples of input-deprived childhoods are found with abused and so-called feral children, as well as deaf children who are only exposed to signed language relatively later in life. The case of Genie is a well-known example of the former: having received virtually no linguistic input from age 18 months to 13 years, her syntax and intonation remained underdeveloped despite extensive subsequent language exposure, although she was able to acquire vocabulary and general communicative skills (Curtiss, 1988). Deaf children born to hearing parents constitute a case of severely limited linguistic exposure under otherwise relatively normal social conditions. Again, even following many years of exposure to signed language after early childhood, the signing of such individuals lacks the syntactic and morphological complexity observed in that of individuals who received exposure to signing from birth (see e.g., Mayberry & Kluender, 2018; Mayberry, Lock, & Kazmi, 2002).

Another case of non-targetlike L1 acquisition is found with heritage language speakers, who acquire as L1 a language that is not the dominant language in their community. Observed differences between heritage language speakers and (monolingual) L1 speakers of a language have been attributed to sub-optimal input conditions during childhood, where the child learner may receive more and higher quality input in the societally dominant L2 than the L1. Again, these differences seem to be concentrated in the domain of morphosyntax, with some differences also occurring with syntax-pragmatics-interface phenomena (see Polinsky & Scontras, 2019, for discussion).

An explanation for the non-targetlike outcomes of L1 acquisition under the abovementioned conditions is provided by the Critical Period Hypothesis (CPH) (Lenneberg, 1967; Penfield & Roberts, 1959), which proposes the existence of a limited developmental period in which the language-learning system is most receptive and most plastic, thus allowing for language to be acquired as the result of mere exposure. Critical periods have been observed in the development of various other functions across species, some of which appear to bear some relation to language acquisition. For example, passerine songbirds such as the white-crowned sparrow appear to be unable to develop 'normal' song as produced by adult conspecifics if they do not receive exposure to this song within the first 50 or so days after birth (see Bolhuis, Okanoya, & Scharff, 2010; Mayberry & Kluender, 2018, for discussion).

As applied to human language development, two refinements of the CPH should be noted. Firstly, "critical" is often replaced with "sensitive", to indicate that some development of linguistic skills is possible outside of this window, although the learning involved is conscious and effortful and the individual's ultimate linguistic system is unlikely to be completely targetlike.[6] In the present dissertation, the term "critical period" is used as a generic term, also allowing for an understanding of this period as

[6] The term "optimal period" is also sometimes used to reflect that the beginning and end of such periods are not discrete, that they can to some extent be manipulated, and that some neural plasticity remains across the lifespan (Werker and Tees, 2005).

one of heightened sensitivity (see Birdsong, 2018). Secondly, as indicated in the above discussion of divergent L1 attainment, all aspects of language seem not to be subject to the same critical period. It appears that the window for nativelike development of phonology closes earliest, with that for morphosyntax remaining open somewhat longer (Long, 1990; Newport, Bavelier, & Neville, 2001).

The CPH has been extended to L2 acquisition as well, though not without controversy. There is widespread agreement that L2 learners who begin learning a language earlier in life outperform later learners in the long run (although this benefit is limited to immersion contexts, see Huang, 2015, and Munoz, 2008). Whether the relationship between ultimate attainment – that is, the final state of L2 proficiency, once learning has ceased – and AoA is linear or non-linear is however the subject of much debate. A linear decline in ultimate attainment with increasing AoA would be evidence against the existence of a critical period for language acquisition, suggesting rather a general decline in language acquisition ability across the lifespan. If a critical period exists, it should be reflected in a non-linear function explaining the relationship between AoA and ultimate attainment, with a discontinuity indicating the period's end. Three possibilities for the shape of this non-linear function have been proposed (Birdsong, 2006): one where the language acquisition ability is at its peak at birth and declines until the end of the critical period, after which it bottoms out (the stretched "L"); one where peak language acquisition ability is constant throughout the critical period and a linear decline across the lifespan begins at the end of the critical period (the stretched "7"); and one that combines both of these options, such that there is a critical period of maintained peak ability, followed by a linear decline for some years, which eventually bottoms out with the completion of brain maturation (the stretched "Z").

Birdsong and Vanhove (2016) (see also Vanhove, 2013) review offline studies that have set out to test for the existence of a critical period in L2 acquisition and tentatively conclude – sometimes based on reanalysis of the original data – that these studies provide no clear evidence of a non-linear relationship between AoA and ultimate attainment. However, recent studies in this domain with improved designs and/or a focus on online processing have indeed obtained results best explained by non-linear models. For example, Hartshorne, Tenenbaum and Pinker's (2018) study of grammaticality judgements obtained from close to 700,000 English speakers indicates that a period of peak learning ability ends in late adolescence, prior to which there is no difference in learning ability across AoAs. In their morphological priming study, Veríssimo, Heyer, Jacob and Clahsen (2018) found nativelike inflectional priming effects up to an AoA of 5–6 years of age, after which they observed a linear decline in nativelikeness (i.e., a stretched "7" geometry). Bosch, Veríssimo and Clahsen (2019), who examined L2 morpholexical representation and processing, found a non-linear pattern of a different shape, indicating a decline in nativelikeness beginning from birth (i.e., a stretched "L" geometry), again with a different offset age (viz. 11 years).

The abovementioned three studies all tap into some aspect of L2 morphosyntactic knowledge. How, then, to explain the divergent findings regarding the offset of the critical period in this domain? The extent to which Hartshorne et al.'s (2018) findings can be compared with those of Veríssimo et al. (2018) and Bosch et al. (2019) is limited, given that offline tasks provide insight into the outcome of processing rather than its progression. In such tasks, it is conceivable that even if some aspects of morphosyntactic knowledge are compromised in later L2 acquirers, they can rely on other intact linguistic knowledge and general reasoning abilities to arrive at nativelike outcomes (see e.g., Bley-Vroman, 1989; Green, Crinion, & Price, 2006). In the case of the two online processing studies, however, an explanation can be found in the notion that AoA effects may be highly selective even within a particular domain. This idea has been extensively developed within phonology, where distinct critical periods have been observed for phoneme perception, phonemic matching and audio-visual integration (see Werker & Hensch, 2015, for a review). Morphosyntactic processing too is believed to employ different mechanisms depending on the phenomenon at hand. In this regard, Bosch et al.'s (2019) study tested the processing of lexically conditioned inflection, examining German verbs that use an irregular stem form to express particular morphosyntactic features (for example, the verb stem *sterb*- 'die' becomes *starb*- in the past tense and *stürb*- in the subjunctive). Such irregular forms are stored in the lexicon rather than being computed by application of a regular inflectional rule and have been observed to take longer for native speakers to acquire. Bosch et al. (2019, p. 356) offer this point as support for the idea that there is "an extended period of sensitivity for the development and stabilization of complex lexical entries with their irregular subentries and corresponding morphosyntactic features". The decline in performance from an AoA of 0 to the end of the critical period, then, can be attributed to the reduced time available for later acquirers to establish these lexical entries before the period's offset.

Veríssimo et al.'s (2018) study, in contrast, tapped into the processing of a regular morphological operation, namely the affixation of the prefix *ge*- and suffix -*t* to a verb stem to form the German past participle (e.g., *prüfen* 'to check' – *geprüft* 'checked'). Referring to arguments that L2 grammars are either morphosyntactically impaired (Meisel, 1997) or fail to overtly realize inflectional morphology (Prévost & White, 2000), the authors propose that after early childhood, L2 speakers are no longer able to extract regular morphological rules from the input. Thus, in the absence of sufficient input before the age of 5–6, knowledge of such regular morphological operations may be compromised, leading to non-nativelike processing in this domain.

Investigations of the type reported above, which apply statistical techniques suited to identifying non-linear relationships between AoA and L2 performance and outcomes, have not yet been conducted to examine processing of the types of syntactic phenomena discussed in the previous sub-section. However, given the findings reviewed above for AoA effects on morphosyntactic processing in the single-word context, it is plausible that such effects would also be found in the sentence context. As such, AoA is a factor that the present study takes into account.

1.4.2 Second language proficiency

Second language proficiency appears to have a differential effect on semantic and syntactic processing. In neurophysiological studies, learners have been shown to be sensitive to semantic violations even in the very early stages of L2 learning (Osterhout, McLaughlin, Pitkänen, Frenck-Mestre, & Molinaro, 2006), and intermediate- and high-proficiency L2 learners have been found to show the same responses to such violations as native speakers (Ojima, Nakata, & Kakigi, 2005).

Sensitivity to syntactic violations has also been found to develop after only a few months of L2 exposure (Osterhout et al., 2006). However, amongst learners who have already successfully acquired the area of grammar under investigation, L2 proficiency still affects the extent of nativelikeness in the detection of morphosyntactic anomalies.

This has been shown, for example, by Rossi et al. (2006), who found that only high-proficiency L2 speakers showed the neural responses found in L1 speakers in response to phrase structure violations.

The effects of L2 proficiency on L2 parsing have also been investigated, albeit to a lesser extent. Hopp (2010), for example, had L2 speakers of two different proficiency levels (and from three different L1 backgrounds) process structurally ambiguous input that was disambiguated by means of case marking or verbal agreement. The L2 speakers were classified as either "advanced" or "near-native" on the basis of a C-test (see section 3.5.2 for further discussion of tests of this type), where the advanced group was made up of L2 speakers who scored below the median of 67%, and the near-native group of those who scored above this. The L2 speakers were additionally assessed on their performance on a picture description task, and speech samples taken from this task were also judged in terms of their nativelikeness by native speakers. The various assessments yielded two proficiency groups that were reliably different from each other.

Focusing specifically here on online performance, the advanced learners appeared not to use morphosyntactic information during incremental parsing: in two self-paced reading tasks testing different word order phenomena in German, they did not slow down upon encountering verbal agreement or case markers that ruled out the initial analysis of the sentence. The near-native group patterned with the L1 speakers in showing an immediate reaction to the morphosyntactic disambiguation in both tasks.

Effects of proficiency on performance also emerged in a speeded grammaticality judgment task, where the near-natives, unlike the advanced learners, exhibited increased response times when confronted with dispreferred word orders. Additionally, the grammaticality judgments of the near-natives were significantly more accurate than those of the advanced learners. However, on this task, L2 speakers whose L1 (Russian) did instantiate a rich case system outperformed speakers whose L1 (English or Dutch) did not, such that the L1 Russian speakers were the only L2 group to pattern with the native speakers. Hopp (2010) takes this finding to indicate that whilst greater proficiency leads to increased L2 processing nativelikeness, under a high processing burden, L2 speakers may still show non-targetlike performance on aspects of the L2 grammar that are not instantiated in their L1.

It should however be noted that the majority of the near-natives did not score within the native speaker range on the C-test, and the former group's mean C-test score was 10% lower than that of the natives. Further, they made an average of 1.37 mistakes per minute in their picture descriptions. Thus, while it is fair to refer to the more proficient L2 group as near-natives, they did not score at ceiling on the offline measures. Whether the proposed cross-linguistic influence (CLI) effect would still occur with L2 groups of even higher proficiency is unclear.

Hopp's (2010) findings indicate that highly proficient L2 speakers are able to deploy morphosyntactic information to guide the parsing of structurally ambiguous input in real time. For the other syntactic phenomena examined in this dissertation, possible effects of L2 proficiency have not been the object of independent examination, with studies in these areas generally aiming to control for proficiency differences rather than to investigate the potential influence of such differences. Furthermore, the comparability of L2 groups across studies is compromised by the variation in the kinds of proficiency measures that are employed. Nonetheless, the remainder of this section presents the findings for L2 speakers of various proficiency levels in these domains.

In the case of wh-dependencies, the L2 speakers that have been tested have been at least at an upper-intermediate level of proficiency. In Williams (2006), L2 learners with an average self-rated proficiency of 2.9 out of 4 patterned with native speakers in their reintegration of the moved element. However, a number of studies have found that highly proficient L2 speakers still do not show nativelike processing of long-distance wh-dependencies: L2 groups with average scores of 85% and 76%, respectively, on a standard proficiency test were found not to pattern with native speakers in Marinis et al. (2005) and Pliatsikas, Johnstone and Marinis (2017). Further, in their study using the same materials employed in the aforementioned two experiments, Pliatsikas and Marinis (2013) report that proficiency did not affect reading times or interact with any of the experimental manipulations. On the basis of these findings, it has been proposed that L2 proficiency does not affect parsing in this domain.

The role of L2 proficiency in L2 anaphora and pronoun resolution remains largely unclear. Felser et al. (2009) and Patterson et al. (2014) both tested L2 speakers with an average score of 78% on a standardized proficiency test, with non-nativelike performance being observed in the former study (on reflexive processing) and nativelike performance in the latter study (on pronoun processing). This suggests that the particular construction under study and/or other learner-specific factors may have more of an effect than L2 proficiency level in anaphora resolution. Additional tentative support for this hypothesis comes from Felser and Cunnings (2012), who contrasted the performance of their L2 speakers ranked as “upper advanced” to that of their less proficient L2 speakers and native speakers. They found only minor differences in the time course of processing between the two L2 groups, and the more advanced L2 speakers did not pattern with the native speakers.

Taken together, the findings reviewed above suggest that increased L2 proficiency leads to increased sensitivity to morphosyntactic information and also facilitates the use of such information during the online construction and revision of syntactic representations. However, although its influence has yet to be systematically investigated, L2 proficiency seems not to affect the processing of abstract syntactic structures (in long-distance wh-dependency processing) or anaphora/pronoun resolution.

These are two domains in which domain-general cognitive capacities have also been argued to play a role. The following sub-section reviews the relevant research in this area.

1.4.3 Working memory

Working memory, or the ability to temporarily store and manipulate information while performing higher-order cognitive tasks (Juffs & Harrington, 2011, p. 137), has been implicated in a range of cognitive processes, including reasoning, planning and problem-solving. A role for working memory has also been identified in language comprehension, where individuals with greater working memory capacity (WMC) are better at learning vocabulary (in the L1 and the L2) and at L1 reading and listening comprehension (see Daneman & Merikle, 1996, for discussion).

Working memory involves both the storage and processing of information (see Baddeley & Hitch, 1974, for discussion). Tests of WMC that tap into both of these functions have been found to be predictors of performance on sentence processing tasks, which require the individual to identify, store and integrate various sources of linguistic and contextual information while paying attention to upcoming parts of the sentence. One such test that is widely used in language processing research is the Reading Span Task (RST) (Daneman & Carpenter, 1980). The RST requires the individual to read a set of sentences and remember an element for each sentence, which may be a word of the sentence, a letter or a numeral. As a secondary processing task, they are also required to provide judgments of the semantic acceptability or grammaticality of the sentence. This secondary task obstructs the rehearsal of the elements to be recalled in memory, which prevents the RST from becoming a measure of short-term memory alone.

Overall, there seems to be a positive relationship between WMC and L2 processing: Linck, Osthus, Koeth and Bunting (2014), in their meta-analysis of 32 studies in this area, obtained an effect size of 0.24 that was confidently larger than zero (95% confidence interval: 0.18–0.30). However, Linck et al.'s (2014) study did not take into consideration the kinds of processing phenomena considered in this dissertation, where findings regarding the importance of WMC vary.

To begin with, effects of WMC on the processing of long-distance dependencies have been observed for both L1 adults and children, such that only participants with a relatively high RST score showed awareness of possible sites for reintegration that were not directly marked by a subcategorizing element (Roberts, Marinis, Felser, & Clahsen, 2007; see also Nakano, Felser, & Clahsen, 2002). However, WMC has not been found to affect L2 speakers' performance on such tasks (Felser & Roberts, 2007); a finding which has been attributed to qualitative differences between native and non-native processing for which increased WMC does not compensate (see section 1.5.1 for further discussion of this position).

The relationship between WMC and syntactic ambiguity resolution in the L1 has also been investigated. For example, when processing a sentence such as *The evidence examined by the lawyer convinced the jury*, where *examined* could be a main clause verb or the onset of a reduced relative clause, high-WMC readers have been found resolve the ambiguity more quickly. Working memory capacity is taken to be facilitative here because it allows for syntactic information to be rapidly integrated with pragmatic knowledge, which disfavors the implausible interpretation of *the evidence* being an agent of the verb *examine* (Juffs & Harrington, 2011, p. 148). High-WMC readers are also found to be more accurate in their responses to comprehension questions following temporarily ambiguous sentences such as *The soldiers warned during the midnight raid attacked after midnight*, where multiple syntactic parses are available (*warned* can be a main verb, an intransitive verb, or the onset of a reduced relative clause). Participants with higher WMC have also been found to be less likely to default to resolving ambiguities locally, which is thought to be a strategy that reduces processing costs (Swets, Desmet, Hambrick, & Ferreira, 2007; see also Felser, Marinis, & Clahsen, 2003, for findings with children).

In experiments where L2 speakers read sentences such as those in the previous paragraph, WMC effects have not reliably been obtained. Dussias and Piñar's (2010) high-WMC L2 speakers did pattern with their L1 group in their ability to employ plausibility information to guide the parsing of temporarily ambiguous structures. Some benefit of higher WMC was also observed by Havik, Roberts, van Hout, Schreuder and Haverkort (2009), who found L2 speakers with higher RST scores to perform similarly to low-WMC native speakers in the processing of subject-object ambiguities, but only when the additional semantic verification task that had to be performed directed their attention towards the experimental manipulation (see Williams, 2006, for a similar finding). However, Juffs (2004, 2005, 2006) found no effect of WMC on his L2 speakers' ability to recover from an initial misparse. Further, in the processing of globally ambiguous sentences where there is no single correct parse, Kim and Christianson's (2016) participants' WMC scores predicted sensitivity to ambiguity equivalently in their L1 and their L2. The authors thus suggest that the effect of WMC on processing strategies does not differ according to whether it is the L1 or L2 being processed.

With respect to anaphora resolution, it has been proposed that constraints on WMC may lead to a preference to keep dependencies short, such that the linearly nearest NP would be favored as an antecedent for a pronoun or reflexive. Some evidence for this proposal comes from Cunnings and Felser (2013), who found early signs of interference of the linearly closest antecedent during L1 reflexive processing in their low-WMC group. However, these individuals showed sensitivity to structural constraints on pronoun resolution at the same time. The authors suggest that although the low-WMC group was also aware of the applicable structural constraint, they had more difficulty suppressing the linearly closest (but structurally illegal) antecedent in memory.

Part of the variation in the results obtained in studies of the effect of WMC on L2 processing may be attributed to differing methods of administering RSTs. One potentially influential factor is that less proficient L2 learners may score more poorly on an RST administered in their L2, with higher scores on L2-administered RSTs thus potentially reflecting higher L2 proficiency rather than greater WMC. This concern is likely mitigated for more proficient L2 speakers, as robust correlations between WMC measured in the L1 and the L2 have been found (Osaka & Osaka, 1992). Related to this point is that as L2 proficiency increases, reliance on WMC during processing may decrease (e.g., Hummel, 2009; Sagarra, 2016; Serafini & Sanz, 2016), potentially due to increases in the integration of L2 knowledge and the automatization of processing.

Taken together, the findings from the studies reviewed above do not predict strong effects of WMC on L2 processing of the phenomena examined in the present study, particularly because the L2 participants included here were highly proficient L2 speakers. Nonetheless, potential WMC effects on the processing of long-distance wh-dependencies were examined for a subset of the L2 participants in Article 3 (see section 6.10). For these participants, an RST was administered in the L2. Working memory data were not collected for studies 1 and 2.

1.4.4 Cross-linguistic influence

A question that has received considerable attention in the literature is whether, for bilinguals, their knowledge of one language influences their knowledge of the other language. With respect to L2 processing, the question of interest, as formulated by Hopp (2017, p. 98), is "whether properties of the L1 are consulted in real-time production or comprehension, even if these effects may not persist to the final stage of word or sentence production or interpretation". In lexical processing, the existence of cross-linguistic influence (CLI) is well established. For example, even in monolingual contexts, bilinguals process cognate words faster than non-cognate words (Dijkstra, 2005). This cognate effect applies to both L1 and L2 processing, and it persists in a sentence context (e.g., van Assche, Duyck, Hartsuiker, & Diependaele, 2009), even when the target word in the language in question is highly predictable (e.g., van Assche, Drieghe, Duyck, Welvaert, & Hartsuiker, 2011, although see van Hell & de Groot, 2008).

In syntactic processing, most evidence of CLI comes from studies of production. Here, within a priming paradigm, bilinguals exposed to a particular sentence structure in one of their languages have been found to be more likely to subsequently reuse that same structure in their other language.[7] This however only applies when word order is shared between the two languages in question – exposure, for example, to a passive sentence where the *by*-phrase occurs before the verb (e.g., Afrikaans, *Die koerant was deur die seun gelees* 'The newspaper was by the boy read') does not result in the production of an ungrammatical passive in a language where the *by*-phrase must occur after the verb (e.g., English, *The newspaper was read by the boy*).

In comprehension studies, there is some evidence of CLI when it comes to the detection of morphosyntactic violations. Tolentino and Tokowicz (2011) review functional magnetic resonance imaging (fMRI) and event-related potential (ERP) studies and observe that when processing constructions dissimilar from the L1 or unique to the L2, L2 speakers' neural activation patterns differed from those of L1 speakers, whereas processing similarities were observed for constructions similar in

[7] For example, L1 Spanish–L2 English speakers, having just processed a Spanish passive sentence (e.g., *El camión es perseguido por el taxi* 'The truck is being chased by the taxi'), were subsequently more likely to describe a picture in English using a passive construction, compared to when they had processed an active sentence (e.g., *El taxi persigue el camión* 'The taxi chases the truck') or an intransitive sentence (e.g., *El taxi acelera* 'The taxi accelerates') (Bernolet & Hartsuiker, 2018).

the L1 and the L2. Self-paced reading studies have yielded the same findings: Avery and Marsden's (2019) meta-analysis of studies using this paradigm to test native and non-native sensitivity to morphosyntactic information finds that L2 speakers show greater sensitivity to morphosyntactic violations when the property being tested is similar in the L1 and the L2, although this increased sensitivity appears relatively slight.

In terms of parsing preferences, CLI, where observed, seems to be fleeting and limited to specific processing circumstances and/or learner populations. Some effects of L1 lexical information have been observed with respect to verb subcategorization frames (e.g., Dussias & Cramer Scaltz, 2008; Frenck-Mestre & Pynte, 1997), such that local processing difficulty – at the verb and directly following region – has been observed when these frames differ across the L1 and the L2.[8] In terms of the real-time recruitment of L1 parsing preferences during L2 processing, somewhat mixed evidence comes from studies of relative clause attachment preferences, which differ cross linguistically. For example, in a sentence such as *An armed robber shot the sister of the actor who was on the balcony*, English prefers to attach the relative clause beginning with *who* to the second noun phrase (*the actor*). Spanish, in contrast, prefers attachment to the first noun phrase (*the sister*) (see Cuetos, 1988, for discussion). Some processing studies where the preferences of the participants' L1 differ from those of the L2 suggest that L2 speakers with low L2 proficiency may transfer their L1 preferences during online L2 processing (Frenck-Mestre & Pynte, 1997), and speakers

[8] For example, Frenck-Mestre and Pynte (1997) found that, when reading a sentence such as *Every time the dog obeyed the pretty girl showed her approval*, French-dominant French–English bilinguals slowed down significantly more at the verb *obeyed* compared to English-dominant bilinguals. The authors attribute this to the difference in selectional restrictions between the English *obey* and French *obéir*, where the former is optionally transitive but the latter strictly intransitive.

who are L2-dominant have been observed to transfer L2 preferences to the L1 (Dussias & Sagarra, 2007). However, other studies have found no transfer of L1 attachment preferences to the L2 (e.g., Felser, Roberts, Marinis, & Gross, 2003; Papadopoulou & Clahsen, 2003). Further, L1–L2 morphosyntactic similarity, in terms of the presence or absence of a case and/or subject-verb agreement system, does not seem to modulate L2 speakers' ability to employ case and agreement cues during online disambiguation (Gerth, Otto, Felser, & Nam, 2017; see also Hopp, 2010).

Turning from ambiguity resolution to the processing of non-local dependencies: for sentences containing long-distance wh-movement, the reading patterns of L2 speakers with an L1 that does make use of wh-movement do not differ from those with an L1 that does not (i.e., so-called wh-in-situ languages such as Japanese) (Felser & Roberts, 2007; Marinis et al., 2005), suggesting no transfer of parsing strategies in this domain. Further, in the resolution of referential dependencies in the L2, speakers whose L1 has similar structural constraints to the L2 have not been found to show more nativelike behavior than speakers whose L1 has different constraints (re: reflexive processing, see Felser et al., 2009, for L1 Japanese–L2 English and Felser & Cunnings, 2012, for L1 German–L2 English). However, to date there has not been a study in this area that has compared multiple L2 groups with different L1s to native speakers, and so conclusions regarding CLI in this regard remain tentative.

Overall, then, while there is some evidence that differences in verb subcategorization preferences may fleetingly affect language comprehension, there is no strong evidence that the L1 grammar at a larger scale is recruited during real-time L2 sentence comprehension. The participants in the experiments reported on in the present dissertation all had Afrikaans as an L1. In the first and last experiments discussed –

on the processing of temporary subject-object ambiguities and long-distance wh-dependencies – the verbs used in the experimental stimuli had comparable selectional restrictions in English and Afrikaans, thereby controlling for potential effects of cross-linguistic differences in this regard. The second experiment reported on, on pronoun resolution, exploited cross-linguistic differences between English and Afrikaans in the interpretation of pronouns with particular verbs. The relevant differences between English and Afrikaans in this domain are explained in chapter 5.

1.4.5 Acquisition setting

Quantity and quality of language exposure play a crucial role in language acquisition. These factors differ across language acquisition environments. Typical classroom acquisition settings, as described by Munoz (2008), may be characterized by some or all of the following features: (i) instruction is limited to 2–4 sessions of approximately 50 minutes per week; (ii) exposure may be limited in terms of source (only the teacher may use the target language), quantity (the extent to which the target language is used in the classroom may vary) and quality (proficiency differs across teachers); (iii) students do not communicate amongst themselves in the target language; and (iv) the target language is not used outside of the classroom. In contrast, in immersion settings, the L2 learner has access to an abundance of unstructured L2 exposure from numerous different sources in a variety of settings, and interactions with native speakers offer the opportunity to negotiate for meaning in the L2, which some theories believe to be key for L2 acquisition (e.g., Krashen, 1992; Long, 1983).

There is however variation in the extent to which immersed L2 speakers are exposed to and make use of the L2. Flege and Liu (2001), for example, show that students benefited more than non-students from longer periods of residence in the target-language-speaking country. They attribute this finding to the increased amount of exposure to native L2 input students are likely to receive from their classroom peers and teachers and propose that non-students have comparatively fewer interactions with L2 speakers.[9]

In Flege and Liu's (2001) results, immersion had the strongest effect on phonetic perception. However, findings of other studies also show an effect of immersion on syntactic processing. In studies of relative clause attachment preferences, Frenck-Mestre (2002) found that L2 speakers with an immersion duration of five years patterned with native speakers in this regard, whereas L2 speakers who had been immersed for short periods (nine months) did not. A similar finding was obtained by Dussias (2003) for L2 speakers who had been immersed in an L2 environment for approximately eight years. Further, Pliatsikas and Marinis (2013) and Pliatsikas et al. (2017) found that L2 speakers who had been immersed in the L2 country made use of abstract syntactic structures in processing long-distance wh-dependencies, which distinguished them from classroom L2 learners and those who had spent less than a year in an L2-speaking country.

Findings such as the above relate to the argument that extensive L2 exposure is necessary for the development of implicit language knowledge, defined as "knowledge that the learner has no subjective awareness of, can access for spontaneous language

[9] A reasonable question about these findings is whether the student group may have benefited from explicit L2 instruction. The authors argue that this is unlikely, as the observed benefits applied not only to grammaticality judgments but also to the perception of word-final stops, which is unlikely to receive attention in a language classroom (Flege & Liu, 2001, p. 548).

use through automatic processing, and is unable to verbalise" (Ellis & Roever, 2018, p. 2). This assertion is supported by Caffrara, Molinaro, Davidson and Carreiras (2015), who in their review of ERP research found that L2 speakers who had spent more than five years in the L2-speaking country were more likely to show ERP patterns indicative of early, automatized responses to morphosyntactic violations (see Morgan-Short, Steinhauer, Sanz, & Ullman, 2012, for related findings in an artificial language-learning paradigm).

Implicit knowledge is understood to be the outcome of acquisition – an unconscious, relatively effortless procedure – which stands in contrast to deliberate, effortful learning and its outcome of explicit or conscious linguistic knowledge. Native speakers of a language, having begun the acquisition process at birth, have implicit knowledge of that language. It follows that because the extensive L2 exposure available in L2 immersion contexts promotes implicit learning, it can also lead to more nativelike L2 processing. The participants in the experiments reported on in this dissertation, although not inhabitants of a conventional L2 immersion environment (see section 1.6 for discussion), have all received extensive naturalistic exposure to the L2 from a young age. It remains to be seen whether extensive naturalistic exposure yields effects on L2 processing comparable to those produced by full L2 immersion.

1.4.6 Summary

This section has reviewed findings on the effects of AoA, L2 proficiency, WMC, CLI and acquisition setting on the extent of L2 processing nativelikeness. The following general remarks can be made regarding these factors' influence on the processing phenomena examined in this study. Firstly, in the processing of temporarily ambiguous sentences, differences between the disambiguation cues available in the L1 and the L2 seem not to affect parsing, and findings are mixed regarding the effect of WMC on the ability to recover from an initial misparse. Increased L2 proficiency does appear to facilitate the use of morphosyntactic disambiguation cues and may also facilitate online reanalysis. Secondly, in pronoun resolution, the little evidence that is available suggests that L2 non-nativelikeness cannot be attributed to a lack of proficiency or WMC, or to L1–L2 differences in structural constraints governing pronoun resolution. Further research is needed to examine the role of such factors in this domain. Thirdly, L2 processing of long-distance wh-dependencies appears to remain non-nativelike even at relatively high levels of L2 proficiency and does not appear to be affected by WMC or CLI. Nativelike L2 processing in this domain has not been observed outside of L2 immersion settings.

1.5 Models of L2 syntactic processing[10]

The previous two sections indicate that whilst native and non-native syntactic processing have been found to converge for certain morphosyntactic phenomena, differences seem to persist for others. Further, individual- and environmental-level factors may contribute to increased L2 processing nativelikeness, but it remains

[10] Two models that are not discussed here are Grüter, Rohde and Schafer's (2014, 2017) RAGE hypothesis, which proposes that non-nativelikeness in L2 processing stems from L2 speakers' reduced ability to generate expectations, and Hopp's (2018) Lexical Bottleneck Hypothesis, which puts forward that delayed lexical processing in L2 speakers has cascading effects on syntactic processing. These models are omitted from discussion because they have not yet been widely adopted in L2 processing research.

unclear whether all processing phenomena are equally susceptible to influence from these factors.

A number of models have been put forward to explain the observed differences between native and non-native processing and their susceptibility (or lack thereof) to the influence of individual and environmental variation. This section introduces approaches that place the focus on differences in (the deployment of) grammatical knowledge, in the efficacy of the retrieval of information from memory during processing, in the type of mental processes employed in processing, and in the computational burden imposed by processing.

1.5.1 Differences in (the deployment of) grammatical knowledge

To account for non-nativelike aspects of L2 syntactic processing that seem to persist even in the face of nativelike processing elsewhere, Clahsen and Felser (2006a, 2006b, 2006c, 2017) formulate the Shallow Structure Hypothesis (SSH). They draw on the proposal, previously made in relation to L1 processing (see e.g., Ferreira, Bailey, & Ferraro, 2002; Ferreira & Patson, 2007), that two routes are available during sentence processing: one involving the computation of a fully specified syntactic structure (termed “the grammatical route” in Clahsen & Felser, 2017, p. 6), and the other relying on lexical-semantic and surface information for the construction of a less detailed representation of the input (termed “the heuristic route”). In this account, neither L1 nor L2 speakers are proposed to be restricted to a particular processing route (Clahsen & Felser, 2017, p. 5). However, for L2 speakers, the SSH proposes that differences in underlying grammatical representations, in the weighting of grammatical constraints, and/or in the extent of reliance on grammatical information during processing increase the likelihood that the heuristic rather than the grammatical processing route will be employed. Addressing these possible differences more specifically, in the order of listing: firstly, L2 speakers may lack knowledge of a particular grammatical rule, or be in possession of incorrect knowledge regarding that rule. Secondly, they may prioritize non-grammatical constraints over grammatical constraints (see Goldrick, Putnam, & Schwarz, 2016, and Veríssimo, 2016, for a framework within which such an account can be implemented). Thirdly, they may have nativelike representations and prioritizations of grammatical information, but these information sources may not be relied upon to guide real-time processing.

The SSH was initially based on the observations, discussed above, that L2 speakers do not seem to employ abstract syntactic structures in the processing of long-distance wh-dependencies (Felser & Roberts, 2007; Marinis et al., 2005) and also do not show structurally determined relative clause attachment preferences (Felser et al., 2003; Papadopoulou & Clahsen, 2003). A failure to compute fully specified syntactic representations would account for the former finding, in that it would predict that abstract syntactic elements would not be posited in the syntactic representation. In shallow processing, the filler is simply held in working memory until the subcategorizing element is encountered, at which point reintegration is triggered. As for the latter finding, it is proposed that in shallow processing, the internal structure of complex antecedents such as *the secretary of the professor* is not represented, which prohibits the application of structure-based disambiguation strategies – which, in English, favor the second NP as a relative clause attachment site – and leads to the absence of clear relative clause attachment preferences.

At the same time, the SSH accommodates observations that L2 speakers are sometimes more sensitive than L1 speakers to lexical and discourse cues in interpretation. For example, to resolve syntactic ambiguities online, L2 speakers have been shown to rely more than L1 speakers on plausibility information (Roberts & Felser, 2011) and referential context information (Pan & Felser, 2011; Pan, Schimke, & Felser, 2013). Such cues are taken to be prioritized over syntactic ones in shallow processing.

The proposal that L2 speakers are more prone to employ lexical and discourse-based information than grammatical information is directly related to claims regarding critical periods for language development. Clahsen and Felser (2017, p. 8) indeed identify AoA “as the crucial variable that distinguishes native and non-native grammatical processing”. In this regard, while critical period effects have been attested for morphosyntactic phenomena (see the discussion in section 1.4.1), the literature in this area has little to say about constrained periods for the development of (nativelike) lexical and pragmatic knowledge (although see Bylund, Abrahamsson, Hyltenstam, & Norrman, 2019, for recent evidence of AoA effects on lexical development).

It would however be inaccurate to claim that the SSH predicts more nativelike processing in earlier compared to later L2 learners. As discussed in section 1, the seemingly highly selective nature of critical periods within a particular domain (e.g., inflection vs derivation within morphology) and the non-linearity of age effects make it difficult to formulate predictions about the AoA at which L2 processing would become non-nativelike. Thus, although they believe AoA to have a central effect on the development and real-time deployment of L2 grammatical knowledge, Clahsen and Felser (2017) identify a need for additional research to be conducted in this area.

In addition to AoA, the SSH also affords a role for L2 proficiency and exposure in determining the likelihood that either the grammatical or heuristic processing route will be employed. Increases in both proficiency and exposure, in addition to developing grammatical knowledge, are expected to lead to the increased automatization or entrenchment of grammatical processing routines (Clahsen & Felser, 2017, p. 6), in turn making it easier to deploy such routines during online processing. Ultimately, it is proposed that the development of nativelike L2 processing is possible, given the acquisition of a nativelike grammar of the L2 (Clahsen & Felser, 2006a, p. 118).

In their most recent discussion of the SSH, Clahsen and Felser (2017) suggest that differences between native and non-native processing may be most persistent in cases where an already-constructed syntactic representation must be accessed and/or manipulated. This proposal is in alignment with another recent account of L1–L2 processing divergence, which is reviewed in the following section.

1.5.2 Differences in memory retrieval

Cunnings (2017a, 2017b) attributes the differences between L1 and L2 processing to divergent abilities to retrieve from memory information recently encountered during processing. Cunnings' account references theories that employ a bipartite model of the information processing system – divided into a narrow focus of attention and a long-term memory store – in which information that is not currently the focus of attention must be encoded, stored and retrieved when needed (e.g., McElree, 2006). Such models, applied to language processing, are interested in individual differences in these encoding, storage and retrieval processes, rather than in the amount of information that speakers can keep activated at any one point in time, which is the focus in traditional tripartite models of working memory (e.g., Baddeley & Hitch, 1974).

In theories that employ bipartite models of information processing, skilled memory retrieval is understood to be a key component of sentence processing (e.g., Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006). Two key factors influence this memory retrieval process. One is activation-based decay, where the activation level of an informational unit decreases over time once it has been encountered, with lower activation being associated with increased retrieval difficulty and latency. For example, in the case of long-distance dependency resolution, the moved element is stored in memory when it is first read or heard, and its level of activation in memory is assumed to decay over time as subsequent sections of the sentence are processed. This account aligns with the observation that reintegration difficulty is positively correlated with the distance between the surface position of a moved element and its ultimate reintegration site (Gibson, 1998).

The second factor that affects memory retrieval processes is similarity-based interference. The proposal in this regard is that sentence constituents that bear some relation to an earlier part of the sentence – for example, anaphoric elements – trigger a memory search, in which retrieval cues associated with the triggering element are compared with the features of the items encoded in memory. The item that best matches the triggering item's retrieval cues is then activated and retrieved. However, when multiple items that match these retrieval cues are encoded in memory, the incorrect item may be retrieved. For example, in a sentence like *The boy who greeted the man introduced himself*, *himself* triggers a search in memory for an item that matches the retrieval cue of [+male]. In the earlier part of the sentence, two such items have been encountered: *the boy* and *the man*. Interference between these two items may occur, which may lead to the illegal antecedent (*the man*) initially being considered as an antecedent for *himself*.

Cases such as the example above, in which more than one item in memory matches the cue provided by the anaphoric element, involve inhibitory interference, where competition between items in memory may lead to delayed processing. The reverse effect, termed facilitative interference, may also occur. For example, ungrammatical sentences in which no sentence constituent fully matches the relevant retrieval cues, but where a partial match can be made, may be processed faster than ungrammatical sentences in which no cue-matching of any sort is possible. To illustrate: in a sentence such as *John knew that Mary saw himself*, *himself* bears the retrieval cue of [+male], and a structural constraint dictates that *himself* must refer to the nearest syntactically legal antecedent. While there is a [+male] antecedent in the sentence that can be matched, an analysis in which *John* is the antecedent of *himself* violates this structural constraint. Nonetheless, the availability of a partial match may speed up the processing of this sentence compared to an alternative such as *Sue knew that Mary saw himself*, in which there is no possible antecedent matching the [+male] cue, and parsing should break down entirely.

Based on this account of sentence processing as skilled memory retrieval, Cunnings (2017a, 2017b) proposes that differences between L1 and L2 processing are not due to L2 speakers' failure to compute full syntactic parses (as posited by the SSH), but rather to their (i) increased susceptibility to interference during memory retrieval and (ii) greater reliance on discourse-based cues (as opposed to syntactic cues) during memory retrieval.

Cunnings provides support for the claim that L2 speakers can compute full syntactic parses with reference to studies indicating that that L1 and L2 speakers are equally sensitive to structural constraints dictating *when* a dependency may be formed (see e.g., the discussion of island constraints in section 1.3). The failure to make use of clause boundaries in order to reactivate a moved element in long-distance dependency processing he attributes to an underuse of the abstract syntactic cue identifying the clause boundary as a possible site for filler reactivation.

Cunnings' (2017a, 2017b) account is also applied to the (re)analysis processes involved in ambiguity resolution. In relation to the L2 processing of relative clause attachment ambiguities in sentences such as *The dean liked the secretary of/with the professor who was reading a letter*, he proposes that L2 speakers' attachment preferences are more strongly influenced by pragmatic or discourse information, thus leading them to vary across different sentences. With respect to the processing of temporarily ambiguous sentences, he attributes the comprehension breakdown observed in both L1 and L2 speakers to an effect of the interference between interpretations, stemming from the failure to erase a previously constructed syntactic representation from memory. This failure may lead to the persistence of incorrect interpretations even after a disambiguating cue has been encountered, such that after reading a sentence like *While Anna dressed the baby that was small and cute dropped the rattle*, speakers may still understand that Anna dressed the baby (see e.g., Christianson, Hollingworth, Halliwell, & Ferreira, 2001).

An obvious question, in response to this account, is why L2 speakers would be more susceptible to interference effects, and why differences in the weighting and/or utilization of memory retrieval cues might exist across natives and non-natives.

Regarding the former, it is suggested that the qualities of L2 speakers' representations in memory may be compromised in comparison to those of L1 speakers, and that differences in executive functioning – or the ability to resolve conflicting information during processing – may also reduce speakers' ability to manage interference. However, an account of why such differences in representation quality or executive functioning may arise is not provided. Regarding L1–L2 differences in cue weighting, Cunnings (2017a) puts forward that cues that are overtly marked on lexical items, such as morphosyntactic agreement features, are easier for L2 speakers to employ during parsing. However, L2 speakers are proposed to have greater difficulty with more abstract structural cues that are not overtly marked, such as those that identify legal antecedents in pronoun resolution. In this regard, Cunnings' account overlaps to some extent with the SSH, as syntactic information is expected to be underutilized in comparison to lexical and discourse information during L2 processing. This underutilization is however not attributed to critical period effects on L2 acquisition. Again, the source of the varying levels of difficulty presented by different kinds of cues is not explicitly discussed, although it is mentioned that the cues operative in one of a bilingual's languages may influence the weighting of cues in their other language (Cunnings, 2017b, p. 674), such that cues present in both languages may be easier for L2 speakers to master than cues present in only one language.

As a final point, Cunnings' (2017a, 2017b) account does not make explicit predictions about whether it is possible for L2 processing to become fully nativelike. There is some mention that early bilinguals would be expected to show more nativelike processing than late bilinguals (Cunnings, 2017a, p. 718), but that even highly proficient L2 speakers may implement retrieval cues differently from L1 speakers (Cunnings, 2017a, p. 716). Further research is needed to develop this proposal.

1.5.3 Differences in the mental pathways employed

One account of the differences between L1 and L2 processing references the notion that two distinct memory systems underlie the two kinds of linguistic knowledge, namely lexical knowledge and grammatical knowledge. The best-known proponent of this account is Michael Ullman, who has proposed the declarative/procedural model of language processing (see e.g., Ullman, 2001a, 2001b, 2016, and many others). According to this model, the lexicon contains knowledge of arbitrary sound–meaning pairings such as simplex words (e.g., *mango*), bound morphemes (e.g., the English past tense morpheme *–ed*) and idiomatic phrases whose meaning cannot be derived from their individual components (e.g., *kick the bucket*). Grammatical knowledge, in contrast, consists of rules and constraints that govern how words, phrases and sentences are constructed. Ullman proposes that underlying lexical knowledge is the declarative memory system, which has been implicated in the "learning, representation, and use of knowledge about facts" (Ullman, 2001b, p. 106). Grammatical knowledge, under particular circumstances discussed later in this sub-section, is said to be encoded in procedural memory (also referred to as the "skill" or "habit" system), which is implicated in the human learning of new motor and cognitive abilities, as well as in the maintenance of these abilities.

Support for this dual-system model comes, for example, from studies that show frequency effects for irregular forms but not regular forms (e.g., Ullman, 1999). That is, when the frequency of the stem is controlled for, irregular past tense forms that are frequent (e.g., *said*, from the stem *say*), are rated as more acceptable than irregular past tense forms that are infrequent (e.g *shrank*, from *shrink*). No such frequency effects are found for regular past tense forms, such that acceptability ratings of the infrequent *scoured* (from *scour*) are equivalent to those of the frequent *asked* (from *ask*). An explanation for this finding is that irregular forms are stored as is in the lexicon. If such forms are infrequent, they will be more difficult to retrieve from memory, leading the speaker to feel uncertain regarding their status as legitimate past tense forms and consequently to rate them as less acceptable. In contrast, for regular forms, only the stem (e.g., *scour* in *scoured*) is taken to be stored in the lexicon. Thus, while frequency effects may influence the ratings of stems as a result of memory retrieval effects, the addition of the –*ed* suffix to form the past tense should not result in any further acceptability differences.

Further support for the dual-system approach comes from examinations of processing in individuals with specific language impairment, a condition that is believed to compromise procedural learning abilities. Such individuals show frequency effects for both regular and irregular forms and also struggle to produce novel regular forms (e.g., *sprinked* from *sprink*) (e.g., Ullman & Gopnik, 1999; see also Conti-Ramsden, Ullman, & Lum, 2015). These findings suggest that rather than applying rules to procedurally combine discrete elements into larger structures, such individuals rely on declarative memory and learn compositional forms by rote.

In typically developing individuals, Ullman proposes that early childhood language acquisition relies upon the procedural system for encoding grammatical knowledge. As such, for L1 speakers and early bilinguals, grammatical knowledge will be proceduralized. However, it appears that declarative learning abilities increase during the course of childhood, and that this is accompanied by a decrease in procedural learning abilities (see Ullman, 2016, for discussion of the relevant neurobiological changes). Thus, it is proposed that later language learners rely more on the declarative memory system in language learning. Amongst later acquirers, then, grammar learning takes the form of memorizing structures (e.g., polymorphemic words such as *walked*) that would be compositionally constructed in earlier learners. In addition, representations of grammatical rules in memory may also differ across earlier acquirers and later acquirers, where amongst the latter group these rules are expected to take the form of explicit knowledge that is consciously applied (Ullman, 2001b, pp. 109–110).

Support for this understanding of age effects on the encoding of linguistic knowledge comes, for example, from studies of bilingual aphasics with damage to the brain areas implicated in procedural learning. Such individuals exhibit difficulty with grammatical computation in the L1, but not in the later-learned L2 (Fabbro, 2015). Conversely, damage to the brain areas underlying declarative learning has been found to result in poorer grammatical computation in the L2 than the L1. Neuroimaging studies also show distinct brain activation patterns for L2 compared to L1 grammatical processing, with L2 processing involving greater activation of the brain areas associated with declarative memory (see Ullman, 2001b, pp. 113–114 for discussion).

Importantly, however, the declarative/procedural model does not propose that the procedural system is completely unavailable in L2 grammar acquisition. Ullman (2001b, p. 110) notes a strong practice effect on procedural learning, such that learners who make frequent use of the L2 are expected to become more reliant on this system for performing grammatical computations. Empirical support for this assertion is again provided by neuroimaging studies, which show that more practiced L2 speakers have less activation of brain areas related to declarative memory during grammatical processing compared to their less practiced peers (see Ullman, 2001b, pp. 113–114 for discussion). A recent meta-analysis also indicates a strong correlation between procedural learning ability and grammar in more but not less proficient L2 learners (Hamrick, Lum, Jarrad, & Ullman, 2018).

Thus, within the declarative/procedural model, the key factors that drive L1–L2 processing similarity are understood to be AoA and L2 exposure. In principle, this model allows for the possibility that early L2 acquirers and/or frequent L2 users will show nativelike processing performance. Later L2 learners with lower levels of proficiency, however, are expected to experience particular difficulty with aspects of L2 grammar that are not amenable to memorization, such as complex syntax and non-local dependencies.

1.5.4 Differences in cognitive capacity

A final type of approach posits no fundamental or qualitative differences between L1 and L2 processing, but instead assumes that any differences observed are quantitative in nature. Such capacity-based approaches put forward that WMC and other cognitive functions are taxed more during L2 than L1 processing. Scores on measures of these capacities are hypothesized to correlate with individuals' performance on linguistic tasks, with reduced cognitive capacity causing particular problems in the processing of complex linguistic phenomena (McDonald, 2006, p. 383). General support for this proposal is provided by the often-observed difference between offline and online L2 performance, where L2 speakers typically perform better in offline tasks, which are less cognitively demanding due to the absence of time pressure and the reduction in memory demands. Further, in a targeted investigation of the relationship between cognitive capacity and L2 grammaticality judgment performance, McDonald (2006) found that WMC and L2 decoding ability were positively correlated with L2 judgment accuracy. This result aligns with the findings of Linck et al.'s (2014) meta-analysis, which indicates a small but reliable effect of WMC on L2 processing (see section 1.4.3 for discussion).

The capacity-based approach predicts that when capacity demands during L1 processing are increased, thus increasing the burden on WMC and other cognitive faculties, L1 speakers may show L2-like processing performance. Experimentally, capacity requirements can be increased by masking the input signal (e.g., by increasing noise), by adding time pressure to the processing task, or by requiring the speaker to perform an additional task during processing. McDonald (2006) shows that all these manipulations cause grammaticality judgment accuracy to decrease in native speakers, with processing under noisy input conditions in particular yielding L1 judgment patterns similar to those of L2 speakers. Hopp (2010) obtains similar findings in a speeded grammaticality judgment task testing the processing of inflection in German. In both of these studies, judgment accuracy was not equally affected across construction types: word order judgments remained unaffected, but judgements of morphosyntactic agreement phenomena (subject–verb agreement in McDonald, 2006; case and gender agreement in Hopp, 2010) were degraded. This finding is in accordance with the proposal that the processing of more complex phenomena is more difficult under conditions of cognitive strain.

Another resource-driven account of the variability observed in L1 processing, Ferreira and colleagues' 'good-enough' processing model (e.g., Ferreira, 2003; Ferreira et al., 2002; Ferreira & Patson, 2007), has also been extended to capture differences between L1 and L2 processing. The good-enough model, like the SSH discussed in section 1.5, is a dual-system model of language processing, which proposes the existence of two processing routes: one syntax-driven, and the other relying more on semantic, pragmatic and contextual information. In this model, parsing is said to proceed via both routes, yielding two representations of the input. The syntactically underspecified representation is proposed to be derived via 'fast and frugal' heuristics, which facilitate processing when there is more information available than can be consulted given constraints on time and cognitive resources.

During processing, these two representations can be integrated to yield a complete, fully specified parse of the input. However, this integration does not always occur. If, for example, syntactic information conflicts with semantic, pragmatic or contextual information, reconciliation between these information sources may not be attempted. In this regard, Lim and Christianson (2013) reference work showing that syntactic representations are 'fragile' (e.g., Sachs, 1967) and can be overridden by surface-level information, such that the ultimate parse of a sentence does not reflect its actual structure. An example of such a situation, where a heuristic interpretation wins out over a syntactically specified one, is provided in Ferreira (2003). Here, L1 English speakers were found to be more likely to misinterpret simple passive sentences when the syntactic structure was incompatible with world knowledge, leading to an outcome where *The cat is chased by the mouse* was interpreted as meaning *The cat chased the mouse*. To obtain the active interpretation, the noun–verb–noun structure of the sentence is mapped onto the canonical English agent–verb–patient representation to yield an interpretation that is more plausible. In contrast, in cases where the syntactic representation did not produce an implausible interpretation (e.g., *The mouse is chased by the cat*), comprehension accuracy rates were significantly higher, indicating that the syntactic structure of the sentence was incorporated into the final interpretation.

In the case of L2 processing, Lim and Christianson (2013) propose that syntactic representations may be especially vulnerable, particularly at lower levels of L2 proficiency. Second language speakers are then hypothesized to be especially susceptible to influence from semantic, pragmatic and contextual information during processing, especially if this information conflicts with that provided by the syntax. In addition, as the heuristic representation requires less processing effort to compute, L2 speakers – again, especially those of lower proficiency – may be expected to rely more heavily on such representations than L1 speakers. This proposal is in accordance with findings that L2 speakers are more sensitive to semantic, pragmatic and discourse information than L1 speakers (e.g., Lim & Christianson, 2013; Pan & Felser, 2011; Pan et al., 2013; Roberts & Felser, 2011).

The good-enough processing model bears obvious similarities to the SSH, as both predict L2 parsing to be guided by non-grammatical information to a greater extent than L1 parsing. However, there are essential differences between these accounts that should be emphasized here. Firstly, although the SSH acknowledges that both L1 and L2 speakers may make use of heuristic or shallow processing strategies, it predicts greater reliance on such strategies amongst L2 speakers. In contrast, the good-enough model assumes that all language processing – whether L1 or L2 – proceeds via both the heuristic and the syntactic route. Secondly, the SSH attributes L2 speakers' increased reliance on the heuristic route to non-nativelike aspects of the L2 grammar. The good-enough model, on the other hand, proposes that whether the heuristic and syntactic interpretations are integrated depends on various factors. Amongst L2 speakers, these factors include L2 proficiency, but task demands are also afforded a role, such that when integration is not necessary to achieve the current goal, or when there is insufficient time to perform integration, the heuristic representation may prevail.

Also in contrast to the SSH, the good-enough model does not reference critical period effects on the nature of L2 processing. This position is shared by general capacity-based accounts, such as those of McDonald (2006) and Hopp (2010). The factors that these accounts explicitly predict will affect L2 processing nativelikeness include L2 proficiency and general cognitive capacity. In this regard, both increased cognitive capacity and increased proficiency should reduce the computational demands of L2 processing, and increased proficiency should also facilitate the construction and maintenance of fully specified syntactic parses. The demands of the processing task are also expected to affect the nature of L2 processing. More demanding tasks may lead to computational breakdown, according to general capacity-based accounts. Alternatively, amongst speakers of sufficient L2 proficiency, the good-enough model proposes that tasks that direct the individual to pay attention to structural information may prompt the construction of fully specified syntactic parses. Finally, both accounts predict nativelike L2 processing to be possible.

1.5.5 Summary

This section has presented four different accounts of L2 processing which differ in the explanations they offer for non-convergence between L1 and L2 processing. On the one hand, the SSH and the declarative/procedural model can be grouped together, as both posit age-related effects on the representation and deployment of grammatical knowledge in L2 speakers. In another group, there are the capacity-based approaches and Cunnings' (2017a, 2017b) memory retrieval account, which posit that the construction and representation of syntactic parses is fundamentally similar across L1 and L2 speakers. These accounts attribute L1–L2 processing differences to cognitive limitations, interference between representations in memory, or differences in the weighting of cues that guide the retrieval of information from memory.

It is not the aim of the present dissertation to evaluate the merits of each of these accounts. Indeed, a great deal more empirical data will need to be collected in a variety of settings and amongst a variety of L2 populations in order to do so. The experiments presented here are one contribution towards this enterprise. The following section introduces the research context and the nature of the L2 population under study.

1.6 The context of the present study

1.6.1 Knowledge production in second language acquisition (SLA) and L2 processing research

The aim of the cognitive sciences, as applied to human cognition, is to understand the workings of the human mind. To obtain such an understanding, cognition must be studied across a wide range of human subpopulations, or at the very least within (a) subpopulation(s) that can be shown to be representative of the human population more broadly. In the last 15 years or so, scholarly failings on both of these fronts have received increasing attention. The most impactful denunciation was perhaps that of Henrich, Heine and Norenzayan (2010a, 2010b), who brought into focus what they termed the "WEIRD" (Western, Educated, Industrialized, Rich and Democratic) bias in the cognitive sciences. Henrich et al. (2010a, 2010b) showed that the overwhelming majority of findings in the behavioral sciences – amongst which they include psychology, cognitive science and economics – are based on experimentation on North American undergraduate students (often of psychology). Further, they showed that these participants, in their behavior across domains such as visual perception, numerical cognition, analytic reasoning and cooperation, are not only not representative of the wider human population, but are indeed outliers compared to other subpopulations that have been tested.

The fields of L2 acquisition and processing suffer from their own WEIRD bias. The author affiliations of publications in top-ranked journals in these fields show that the majority of research in SLA and psycholinguistics is being generated in the US and the UK (Bylund, in press), which may be considered archetypally WEIRD countries. Further, aside from the geographic and socio-political concentration, Ortega (2005, 2012, 2013, 2019; see also Bigelow & Tarone, 2004) identifies an overreliance on L2 learner populations who have been raised monolingually and subsequently "pursue elective bilingualism, whether by adding English as an international language or by adding a prestigious or a marketable foreign language other than English" (Ortega, 2012, p. 219).

There is an obvious connection between the dominance of WEIRD research locations in SLA and the over-investigation of certain types of L2 speakers. This is not to suggest that kinds of L2 learners and speakers other than those described in the Ortega (2012) quote in the above paragraph do not exist in WEIRD societies. However, for researchers in such societies, and particularly in the US and UK, the two most accessible L2 learner populations are native English speakers studying a foreign language at university and non-Anglophone (often international) students who have English as an L2. The convenience of sampling from university student populations may outweigh concerns regarding the epistemic limitations of doing so, particularly in the face of limited time and resources. While there have been attempts to expand the populations considered (see Andringa & Godfroid, 2019) – for example, the field of heritage language studies continues to grow (see Polinsky, 2018, for an introduction), and there has been some investigation of language learners of lower socio-economic status (e.g., Young & Astarita, 2013) and limited literacy (e.g., Tarone & Bigelow, 2012) – such studies still constitute a small minority of the available research.

Even looking beyond university student populations, however, there are differences in societies' linguistic makeup that limit the kinds of phenomena that can be studied in a particular location. One measure of such inter-societal differences is that of linguistic diversity, or the number of languages spoken within a particular society. Greenberg's (1956) linguistic diversity index represents the probability that, within a country, any two randomly selected individuals will have different L1s. Index values range from 0 to 1, with 0 indicating that all individuals have the same L1, and 1 that no two individuals have the same L1. At the extreme high end of the diversity spectrum is Papua New Guinea, with an index of 0.99. WEIRDer countries tend to have lower linguistic diversity indices; the US and the UK, for example, come in at 0.34 and 0.15, respectively, and all countries in Western and Northern Europe score below 0.7 (Eberhard, Simons, & Fennig, 2019).

Due to the overrepresentation of WEIRDer research sites, we know comparatively little about language acquisition and processing in settings that are highly linguistically diverse. With respect to language acquisition and use, linguistic diversity at the societal level has many potential consequences (see e.g., Canagarajah, 2007; Ortega, 2018, for discussion). First, in highly linguistically diverse settings, simultaneous childhood multilingualism is common, and an individual's 'native' language may be identified on the grounds of social identification rather than according to the language to which they were first exposed. Second, acquisition in general is likely to proceed at least in part through naturalistic exposure. Such acquisition may be context- and/or goal-driven, such that the individual masters the aspects of the language they need to perform a particular task. Third, the individual will be exposed to multiple languages, resulting in corresponding decreases in the extent of their exposure to any single language.

South Africa, the country in which the present study was conducted, is ranked 17th worldwide in terms of linguistic diversity, with an index of 0.87 (Eberhard, Simons, & Fennig, 2019). The characteristics mentioned in the previous paragraph are applicable to the country's linguistic setting, but the extent to which they apply differs across geographic areas and social groups (see e.g., Coetzee-Van Rooy, 2012, for discussion). Simultaneous multilingualism, for example, may be more common in low socio-economic-status areas, where population density increases cross-language contact (see e.g., Potgieter, 2014), and linguistic diversity in urban areas is likely to exceed that in rural areas (Posel & Zeller, 2016). Nonetheless, there are marked differences between the South African context and the WEIRD contexts in which most L2 processing research has been conducted to date.

1.6.2 Setting and participants

South Africa's linguistic diversity is reflected in the country's constitution, which recognizes 11 official languages – in alphabetical order, Afrikaans, English, isiNdebele, isiXhosa, isiZulu, Sepedi, Sesotho, Setswana, siSwati, Tshivenda, and Xitsonga – and requires the promotion of a number of additional minority languages, including the Khoi languages, South African Sign Language, German, Greek, Portuguese, Hindi, Gujarati and Tamil (see Mesthrie, 2009, for discussion).

First language speakers of the country's official languages are not evenly distributed in terms of geography. For example, isiZulu, which has the largest number of L1 speakers (approximately 23% of the population), is the L1 of 0.4% of individuals in the Western Cape and 78% of individuals in Kwa-Zulu Natal. In accordance with this uneven distribution, the government allows provinces to select, from these official languages, at least two in which to conduct their business. This selection should be made in line with the demographics of the province in question.

Despite the country's linguistic diversity and the constitutional promotion of multilingualism, English, the L1 of only around 9% of the population, remains "the language of power and access – economically, politically, socially" (Probyn, 2001, p. 250). The language's elevated status is reflected in the education system, where although mother tongue education is provided for until the end of the Foundation

Phrase (Grade 3, or typically age 9)[11], English is the language of education for the majority of learners from Grade 4 onwards. The most recent available figures indicate that more than 80% of learners write their secondary school exit examinations in English (Department of Basic Education, 2010; see also Plüddemann, 2015). The socio-economic importance of having at least some proficiency in English is also indicated by the fact that the rise in the proportion of individuals who speak English as an L2 only plateaus at around age 30, suggesting that those who do not learn English in a classroom setting do so once they enter the labor market (Posel & Zeller, 2016).

The present study was conducted in the Western Cape Province of South Africa, where the languages with the largest proportions of L1 speakers are Afrikaans (approximately 50%), isiXhosa (approximately 25%) and English (approximately 20%) (Census, 2011) and where 71% of households report using more than one language at home (Hill & Bekker, 2014). The participants in the experiments belonged to one of two groups: L1 Afrikaans–L2 English or L1 English. The L1 English participants in the experiments reported on in chapters 4 and 6 were L2 speakers of Afrikaans. For the experiment reported on in chapter 5, no data on the L1 English participants' L2 knowledge was collected, but as discussed in that chapter, it is likely that they had some knowledge of Afrikaans, given that it is often a compulsory subject in Western Cape schools. The L2 English group were all early acquirers of English (AoA ≤ 10 years) and were highly proficient in the language. Given the importance of English in education and South

[11] There are caveats here as well; for example, not all learners live in an area that has a school offering education in their L1. Further, school governing bodies are allowed to determine the language of learning and teaching, and they may opt for English regardless of the learner demographics (Posel & Zeller, 2016).

African society at large, they had received extensive exposure to the language from a young age, both naturalistically and through the schooling system. At the time of testing, they were also enrolled at a university where the majority of academic activity is conducted in English.

This population of L2 speakers differs from those conventionally tested in L2 processing studies in a number of ways. To begin with, most L2 processing research to date has focused on later L2 acquirers. Many of the studies that have considered early acquirers have focused on heritage speakers, whose L1 input is typically compromised in terms of quality and quantity. The early acquirers in this study, in contrast, are not L1 speakers of a minority language: their L1 acquisition proceeded in a 'normal' fashion and the L1 has been maintained throughout their lives. This L1 maintenance also sets them apart from immersed L2 speakers, for whom it is presupposed that the L2 is "exclusively or mostly used" (Pliatsikas & Chondrogianni, 2015, p. 2). Secondly, the present L2 group differs from conventionally studied L2 populations in the nature of the L2 input they receive. As discussed above, English is a dominant societal language, but less than 10% of South Africa's population are L1 English speakers. Many communicative interactions therefore rely on lingua franca English, a variety that, like other linguae francae, is characterized by variation rather than stability and may be influenced by transfer and simplification processes (Canagarajah, 2007; Meierkord, 2004). Significant exposure to lingua franca English, with its sometimes idiosyncratic lexico-grammatical features, may have consequences for the nature of the L2 English system that is developed (Bylund, in press).

The present study also departs from the conventional paradigm in L2 processing studies in terms of the L1 group that is tested. Typically, the native speakers that serve as the control group in L1–L2 comparisons are monolinguals. The practice of comparing monolingual L1 speakers to bilingual L2 speakers has been called into question, given the observed consequences of bilingualism for language processing (see Birdsong & Gertken, 2013; Ortega, 2013, for discussion; but also Hyltenstam, Bylund, Abrahamsson, & Park, 2009; Bylund, Abrahamsson, & Hyltenstam, 2012; Bylund, Hyltenstam, & Abrahamsson, 2013 for an alternative perspective). These include, but are not limited to, potential CLI (Jarvis & Pavlenko, 2010) and the need to suppress the competing language system during processing (e.g., van Assche et al., 2009; van Hell & Dijkstra, 2002; van Heuven, Dijkstra, & Grainger, 1998), as well as delayed lexical access and elevated reading times (e.g., Gollan, Montoya, Fennema-Notestine, & Morris, 2005; Gollan et al., 2011; Whitford & Titone, 2012, 2015). Given these differences, across-the-board convergence between bilingual L2 speakers and monolingual L1 speakers may be considered unlikely. In the present study, concerns regarding unfair comparisons are mitigated to some extent[12] by making use of a bilingual L1 control group and keeping the language pair constant across the L1 and L2 groups.

Additionally, the nature of the L1 group in the present study may make convergence between their processing and that of previously tested monolingual British or American English speakers improbable. Alongside the abovementioned consequences of bilingualism for language processing, the present group’s frequent contact with lingua

[12] These concerns are mitigated rather than eliminated entirely because to isolate the effect of L2 status on processing, the L1 group would need to be composed of simultaneous bilinguals (see e.g., Bylund, Hyltenstam, & Abrahamsson, 2013, for discussion). In addition, L2 proficiency differed across the two groups, and this factor is likely to affect the extent to which CLI occurs.

franca English may have affected their L1 system, such that their grammatical representations may differ from those of L1 speakers raised in more linguistically homogenous environments (Bylund, in press).

Taken together, the nature of the L1 and L2 groups' language acquisition and use trajectories distinguish them from the participants relied upon in most WEIRD studies of syntactic processing. Nonetheless, it must be acknowledged that the current participants remain WEIRD in terms of socio-economic background and their status as university students. Future studies should aim to extend research of this nature to other segments of the South African population.

1.7 Research questions

It has been the aim of this chapter to provide a broad overview of the current state of our knowledge regarding L2 syntactic processing. Section 1.3 introduced phenomena that are persistently processed in a non-nativelike fashion by L2 speakers, and section 1.4 identified factors that may influence the degree of L2 processing nativelikeness. It is proposed that the linguistic setting and populations described in section 1.6 present an opportunity to test whether given an early age of L2 acquisition, extensive naturalistic L2 exposure and high L2 proficiency, L1–L2 processing of the phenomena identified in section 1.3 can converge.

As such, the research questions of the present study are:

1. To what extent are there differences between L1 and L2 speakers' abilities to reanalyze temporarily ambiguous input during real-time processing?

2. To what extent is bilingual processing of English pronouns guided by English structural constraints on pronoun reference assignment?
3. Do L1 and L2 speakers rely equally on abstract syntactic structures to process sentences involving long-distance wh-dependencies?

1.8 Outline of the dissertation

This dissertation takes the form of a dissertation by articles. Chapter 2 presents the theoretical framework of the dissertation, in which a general introduction to syntactic parsing is provided. Chapter 3 explains the methods employed for the dissertation's three experiments and describes the analyses that were conducted. Chapters 4–6 are the research articles written as part of the dissertation. Chapter 4 presents a self-paced reading experiment conducted to answer research question 1. An article based on this chapter is under revision for a journal with a broad focus on experimental psycholinguistics. Chapter 5 presents the eye-tracking-while-reading experiment that was undertaken to address research question 2. Additional data on this topic are being collected, and the resulting paper will be submitted to a journal with a primary interest in language comprehension and production and the cognitive processes involved therein. Chapter 6 presents a self-paced reading experiment that was conducted to answer research question 3. A modified version of this chapter has been submitted to a journal that primarily publishes theoretically grounded empirical studies of L2 acquisition and processing. The chapter also contains a supplementary section that examines the potential effect of WMC on this experiment's findings. Finally, chapter 7 concludes the dissertation with a discussion of the findings, which are considered in relation to one another and to the models of L2 processing presented earlier in this chapter. The limitations of the study are also addressed, and suggestions for further research are provided.

2. THEORETICAL FRAMEWORK

Given a string of lexical items, syntactic processing, or parsing, is the process of constructing a hierarchical representation of how these items relate to one another, or, more informally, of figuring out “who did what to whom” (Traxler, 2014, p. 605). It is syntactic processing that makes accessible the meaning difference between *The mouse ate the cheese* and *The cheese ate the mouse*, and that makes apparent the availability of two interpretations of *The hunter killed the poacher with the rifle* (Van Gompel & Pickering, 2007, p. 289). This chapter provides an introduction to the nature and mechanisms of syntactic processing, with a particular focus on the mechanisms relevant to the syntactic phenomena investigated in this dissertation.

2.1 Prediction and incrementality

Hale (2011) identifies two characteristics of the human sentence processing mechanism that are well established. One is its predictive nature: rather than waiting passively for a signal to appear and then reacting to it, comprehenders anticipate upcoming words. Evidence for this claim comes from eye-tracking-while-reading studies, where words that are more predictable based on the context are skipped more often than those that are less predictable (Rayner, 1998) (see chapter 3 for discussion), and from studies using the visual world eye-tracking paradigm, where participants make anticipatory eye movements towards depictions of entities that are likely to be mentioned next in the sentence (Altmann & Kamide, 1999).

A second characteristic of sentence processing is that it is incremental. That is, during parsing, the processor does not wait until the end of the sentence to put together a representation thereof. Rather, it attempts to incorporate each word it encounters into the developing syntactic representation. Evidence for the incrementality of processing first came from Marslen-Wilson’s (1973, 1975) series of speech-shadowing experiments. In a speech-shadowing task, the subject is required to repeat spoken input as they hear it, with as little delay as possible. The time that elapses between the presentation of a word and the subject’s production is measured. Prior to Marslen-Wilson’s work, psycholinguistic theory put forward that speech perception was partially reliant on syntactic information, and that such information was only available for units at least as large as a phrase (as an individual word, considered in isolation, has no syntactic structure). This proposal predicts larger response latencies to be found in speech-shadowing tasks than in tasks requiring the repetition of isolated words, as in the former case the subject must wait until an entire phrase has been processed before they have sufficient information to guide their production (Marslen-Wilson, 1973, p. 522). However, Marslen-Wilson’s experiments established that it is possible to shadow speech at very short response latencies, similar to those observed in isolated word repetition tasks. Indeed, a subset of the participants could produce clearly intelligible material following a delay of around 250 ms, which indicated that they were using only the information contained in the first syllable of an incoming word (with a duration of approximately 200 ms) to guide processing. An analysis of the shadowers’ errors showed that the vast majority of erroneous productions were syntactically and semantically appropriate, and further that shorter latencies were not associated with more grammatical errors. Production of grammatically suitable elements can only be explained if the syntactic and semantic information of the foregoing words is relied upon to integrate subsequent words into the developing representation. The key contribution of Marslen-Wilson's studies, then, is their indication that parsing happens for each word, immediately upon encountering that word.

Since Marslen-Wilson (1973, 1975), a large body of research on the processing of temporarily ambiguous sentences has also shown that syntactic representations of input are constructed well before the end of a sentence is reached. During parsing of a temporarily ambiguous sentence, more than one syntactic analysis is available. For example, in (1), two analyses are possible when the word *examined* is encountered: either *examined* is a main clause verb, in which case the processor expects an object to follow, or it marks the beginning of a reduced relative clause, in which case *examined* is part of a phrase that modifies *the defendant*.

(1) The defendant examined by the lawyer turned out to be unreliable.

Numerous reading time studies (e.g., Clifton Jr et al., 2003; Ferreira & Clifton Jr, 1986; Rayner, Carlson, & Frazier, 1983; Trueswell, Tanenhaus, & Garnsey, 1994) have shown that readers slow down upon encountering *by the lawyer*. This slowdown suggests that *examined* is first analyzed as a main clause verb, and processing difficulty is subsequently experienced when a prepositional phrase rather than an object follows. The difficulty that arises when an initial analysis turns out to be incorrect and must be discarded is termed a garden-path effect.

Current understandings of syntactic processing assume one of two means by which the incremental construction of syntactic representations proceeds. The first, termed the serial account, proposes that only one representation is constructed, which necessitates revision and repair processes if this representation later turns out to be incorrect. According to the second, the ranked parallel account, multiple representations are constructed simultaneously and ranked according to the extent to which they are supported by the information available during parsing. Re-analysis, in this view, involves the re-ranking of representations.

Accounts of syntactic processing differ further according to their assumptions regarding the kinds of linguistic information that are consulted from the outset of parsing. In this regard, modular and interactive accounts of parsing can be distinguished. The following section introduces these accounts.

2.2 Inputs to syntactic processing

Modular understandings of the mind propose the existence of stores of specific knowledge (related, for example, to visual perception or language) that are separate from a central repository of general knowledge. Fodor (1983) originally proposed that language existed as an independent module, but applied to syntactic processing, the point of interest regarding modularity has been whether the various components of language (syntax, semantics, prosody, and so forth) are independent modules themselves.

Modular accounts of processing put forward that parsing first proceeds solely according to syntactic information, ignoring other considerations such as the semantic plausibility of an analysis. Because plausibility and other related information is understood to be taken into consideration at a later stage of processing, modular processing accounts are also “two-stage” accounts, with the first stage drawing on structural information and other information only being considered during the subsequent stage.

The predominant two-stage model in the literature is Frazier's (1979) Garden-Path Model. The Garden-Path Model is serial in nature, and so it assumes that a single analysis is constructed during parsing. The model incorporates a number of structural principles, aimed at reducing processing effort, according to which this single analysis is shaped.

The principle of Minimal Attachment, to begin with, prioritizes a minimal syntactic structure, such that the analysis that requires the fewest nodes in a phrase structure tree will be preferred. In example (1) above, the main clause verb analysis of *examined* has fewer nodes than the reduced relative clause analysis, and so is predicted to be favored.[13] A second principle, that of Late Closure, applies when two analyses have equal numbers of nodes. Late Closure dictates that incoming material is attached to the constituent currently being processed, such that in a sentence such as *The steak with the sauce that was tasty*, the relative clause *that was tasty* attaches to the lower NP, *the sauce*, rather than the higher one, *the steak*.[14] A third principle is the Active Filler Strategy (e.g., Frazier & Clifton Jr, 1989; Frazier & d'Arcais, 1989), which puts

---

[13] According to the assumptions in the Garden-Path Model – which, it should be noted, do not necessarily align with those of more recent syntactic theory – additional nodes are required in the reduced relative clause analysis to construct a complex NP. In the main clause verb analysis, only two nodes are required upon reaching *The defendant examined:* an NP node and a VP node (in addition to the parent node of the sentence) (Van Gompel & Pickering, 2007, p. 290).

[14] Late Closure runs counter to the Spanish preference, noted in chapter 1, to attach modifiers to the more distant noun phrase. The so-called "Tuning Hypothesis" (Mitchell, Cuetos, Corley, & Brysbaert, 1995) proposes that individuals' parsing preferences are shaped by the frequencies of particular structures in their own languages. Because this hypothesis allows for an effect of non-structural information (i.e., frequency) in determining initial parsing preferences, it is not compatible with a strict two-stage account of parsing.

forward that when a displaced element is encountered, the parser tries to integrate it into the developing syntactic representation at the earliest opportunity. Accordingly, when processing a sentence such as *Who did the housekeeper from Germany urge the guests to consider?*, where the moved element *who* is encountered at the beginning of the sentence, integration is predicted to be attempted following *urged*, which is able to take a direct object. This is despite the fact that the position following *consider* turns out to be the site at which the moved element originated.

The Garden-Path Model's contention that non-structural information is ignored during construction of the initial parse has been challenged by a considerable amount of empirical research. It has been found, firstly, that lexical frequency affects ambiguity resolution preferences. Trueswell, Tanenhaus and Kello (1993) compared processing difficulty across pairs of temporarily ambiguous sentences such as those in (2).

(2a) The student forgot the solution was in the book.

(2b) The student hoped the solution was in the book.

Both *forget* and *hope* can take either a direct object complement or a sentential complement. Minimal Attachment predicts that during processing, the direct object analysis should be preferred, as it requires the postulation of fewer nodes in the representation. It follows from this prediction that equal processing difficulty – manifested in elevated reading times – should be incurred in (2a) and (2b) upon encountering *was*, which disambiguates both sentences towards the sentential complement analysis. This prediction was however not borne out in Trueswell et al.'s (1993) study: compared to an unambiguous control sentence, (2a) was more difficult to process, but (2b) was not. The explanation the authors offer for this finding is that *hope* is used more frequently with a sentential than a direct object complement. Immediate access to this frequency information should favor a sentential complement analysis of *the solution*, thus preventing downstream processing difficulty. In contrast, *forget* is used more frequently with a direct object complement, which should result in initial favoring of the direct object analysis of *the solution* and bring about later processing difficulty when this analysis must be revised.

Trueswell et al.'s (1993) findings suggest that lexical frequency information is drawn upon from the very early stages of parsing. This idea has received subsequent support from a number of other studies (e.g., Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Snedeker & Trueswell, 2004; Traxler, 2005; Wilson & Garnsey, 2009); although some investigations have also found that verb bias does not serve to eliminate effects of structural ambiguity (e.g., Kennison, 2001; Pickering, Traxler, & Crocker, 2000). Nonetheless, the finding that verb bias can affect initial parsing preferences is not compatible with the Garden-Path Model of processing, or with other modular accounts. It is however accommodated by so-called interactive accounts of sentence processing, which assume that all information sources are drawn upon throughout parsing. Interactive models, also referred to as constraint-based (or constraint-based lexicalist) models (e.g., Garnsey et al., 1997; Gibson & Pearlmutter, 2000; MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell et al., 1993), assume that all possible representations of a sentence are activated simultaneously, with the level of activation of any one analysis depending on the amount of support it receives from the information available during parsing.

Interactive accounts afford a role in parsing for a number of non-syntactic information sources in addition to frequency information. One such source is prosody. Evidence that prosody can affect initial parsing decisions comes from examples such as (3), where if an intonational phrase boundary – involving the lengthening of the verb and appropriate pitch modifications – is inserted after *leaves*, the direct object analysis of *the house* is blocked (Kjelgaard & Speer, 1999; see also Steinhauer, Alter, & Friederici, 1999; as well as Cutler, Dahan, & van Donselaar, 1997; Frazier, Carlson, & Clifton, 2006, and Warren, 2016, for reviews).[15]

(3) When Roger leaves the house is dark.

Discourse context is another source of information that has been shown to guide parsing decisions. One influential proposal relating to the influence of discourse information on syntactic processing is referential theory (Altmann & Steedman, 1988; Crain & Steedman, 1985). Referential theory incorporates the so-called principle of parsimony, which puts forward that the parser adopts the analysis that requires the fewest number of presuppositions to be postulated. In sentences such as (1), Altmann and Steedman (1988) point out that the main clause and reduced relative clause analyses differ not only in the number of nodes in their respective phrase structure trees, but also in their presuppositional complexity. That is, the main clause analysis presupposes the existence of only one defendant, the agent of *examined*. In contrast,

[15] Even during silent reading, there is some evidence (e.g., Kentner, 2019; Kentner and Vasishth, 2016; see Breen, 2014, for a review) that so-called "implicit prosody" can guide parsing. Such findings are of course also incompatible with a strict two-stage model of parsing.

the reduced relative clause reading presupposes the existence of at least two defendants – one that was examined by the lawyer, and at least one other. According to the principle of parsimony, when such a sentence is presented in isolation, the main clause verb analysis is preferred not because it involves fewer structural nodes, but because it does not require the unsupported postulation of multiple defendants in the discourse context.

Crain and Steedman (1985) tested the principle of parsimony by presenting garden-path sentences within contexts that would support the presuppositions of either the main clause or the reduced relative clause analysis. Consider the examples in (4):

(4a) Main-clause-verb-supporting context

A psychologist was counselling a man and a woman. He was worried about one of them but not the other.

(4b) Reduced-relative-supporting context

A psychologist was counselling two women. He was worried about one of them but not the other.

(4c) Main-clause-verb target

The psychologist told the woman he was having trouble with her husband.

(4d) Reduced-relative target

The psychologist told the woman he was having trouble with to visit him again.

The context in (4a) supports the main clause verb reading that is ultimately correct in (4c), and the context in (4b) supports the reduced relative clause reading that is ultimately correct in (4d). Thus, the pairs (4a)–(4c) and (4b)–(4d) should result in mitigated garden-path effects in the target sentences. If, however, contexts are paired with target sentences whose presuppositions they do not support, exacerbated garden-path effects are predicted.

This hypothesis has received support from the findings of a number of studies (e.g., Altmann & Steedman, 1988; Altmann, Garnham, & Dennis, 1992; Crain & Steedman, 1985; Van Berkum, Brown, & Hagoort, 1999). However, it has also been found that when a strong bias towards one analysis exists – for example, on the basis of verb subcategorization preferences – this bias is not overridden by contextual information (e.g., Britt, 1994; Britt, Perfetti, Garrod, & Rayner, 1992; Spivey & Tanenhaus, 1998). These findings can be accommodated within an account in which discourse-pragmatic information is a cue that can affect parsing, but can also be overridden by other, stronger cues.

The final information source to be considered in this section is plausibility. Effects of plausibility on ambiguity resolution have been examined, for example, by manipulating the animacy of the noun phrase preceding the temporarily ambiguous verb in sentences like (1), repeated here as (5a). In this regard, the main clause verb analysis is more plausible in (5a) than (5b), because inanimate *evidence* cannot be the agent of *examined* (Ferreira & Clifton Jr, 1986).

(5a) The defendant examined by the lawyer turned out to be unreliable.

(5b) The evidence examined by the lawyer turned out to be unreliable.

If plausibility immediately effects parsing, (5b) should be no more difficult to process than an unambiguous control sentence, and the opposite should be true for (5a). Most empirical research, however, has shown that processing difficulty in sentences such as (5b) is not eliminated (e.g., Ferreira & Clifton Jr, 1986; Rayner et al., 1983), although it may be mitigated (Clifton Jr et al., 2003).

Filler reintegration also seems to be attempted regardless of the plausibility of the filler as an argument of the relevant subcategorizing verb. Support for this position comes from Traxler and Pickering's (1996) comparison of reading times for sentences like (6a) and (6b).

(6a) We like the book that the author wrote unceasingly and with great dedication about while waiting for a contract.

(6b) We like the city that the author wrote unceasingly and with great dedication about while waiting for a contract.

Each of these sentences contains a moved element: *the book* in (6a) and *the city* in (6b). In both sentences, the filler's original position is after *about*. In Traxler and Pickering's (1996) experiment, participants' reading times were elevated at *wrote* in both (6a) and (6b) compared to equivalent sentences in which structural constraints ruled out *wrote* as a possible gap location. In accordance with a principle aimed at keeping filler–gap distances short, this indicates that participants attempted reintegration at *wrote* in (6a) and (6b). However, reading times were longer in the latter than the former. The elevated reading time in (6b) indicates that the implausibility of *city* as an object of *wrote* resulted in processing difficulty.

On the whole, then, plausibility information seems to play a relatively weak role in guiding initial syntactic analysis. However, as shown by Traxler and Pickering's (1996) results, having adopted a particular analysis, the parser runs into difficulty if the resulting structure yields an implausible interpretation. The same finding was obtained by Pickering and Traxler (1998), who showed that disruption is incurred immediately upon encountering an NP that is implausible as the direct object of a preceding potentially transitive verb, such that *the magazine* is easier to process in (7a) than in (7b).

(7a) As the woman edited the magazine about fishing amused all the reporters.

(7b) As the woman sailed the magazine about fishing amused all the reporters.

Together, the findings reviewed in this section indicate that it is not only syntactic information that guides the initial parsing process, although structural cues may sometimes be weighted more strongly than other sources of information as initial parsing decisions are made. In this dissertation, effects of plausibility information are expected to be observed in the self-paced reading study presented in chapter 4, which manipulated a temporarily ambiguous NP's plausibility as a direct object of a preceding verb. In addition, effects of non-structural information on the early stages of syntactic processing may occur in the eye-tracking-while-reading experiment presented in chapter 5. The expected effects here relate to interference driven by a gender match between a pronoun and a nearby NP, which may lead to this NP being incorrectly considered as a possible antecedent for the pronoun.

2.3 Processing linguistic dependencies

Syntactic structures differ in the ease with which they are processed. This observation constitutes a central *explicandum* for a number of approaches to syntactic processing. Amongst these are information processing approaches and accounts that relate processing difficulty to the memory operations that subserve sentence processing.

Information processing approaches – which include surprisal (Hale, 2001; Levy, 2008) and entropy reduction (Hale, 2006, 2016), as well as Bayesian accounts of sentence processing (e.g., Narayanan & Jurafsky, 1998) – generally do not commit to particular mechanisms by means of which parsing proceeds. Rather, they predict the difficulty of processing a particular structure by taking into consideration properties of the input such as frequency, with less frequent structures being assumed to pose more processing difficulty than more frequent ones.[16]

The focus of this section is on one specific source of processing difficulty, namely the resolution of linguistic dependencies, to which information processing approaches have not been explicitly applied. Memory-based approaches, in contrast, have paid particular attention to this aspect of processing. Thus, the ensuing discussion takes as its subject two such approaches, namely Gibson's (1998) Dependency Locality Theory

[16] The reader is referred to Hale (2016) for discussion of surprisal and entropy reduction, and to Traxler, Hoversten and Brothers (2018) for discussion of the Bayesian approach. Traxler et al. (2018) (also Traxler, 2014) additionally discuss the recent noisy channel approach to sentence processing, which is also Bayesian in nature (see Gibson, Bergen, and Piantadosi, 2013).

(DLT) and Lewis and Vasishth's (2005) Adaptive Control of Thought–Rational (ACT-R) model.

A central consideration of Gibson's DLT is the locality of a linguistic dependency, where less local dependencies – or those that extend across a larger distance – are taken to cause increased processing difficulty. By means of illustration, consider the well-known processing effort asymmetry across subject and object relative clauses. In this regard, (8a), a subject relative clause, is easier to process than (8b), an object relative clause.[17]

(8a) The lawyer that criticized the doctor has an office on Strand Street.

(8b) The lawyer that the doctor criticized has an office on Strand Street.

The DLT accounts for this asymmetry based on the relative distance between the moved element and its site of reintegration in subject and object relative clauses. Specifically, a key factor in determining processing difficulty is the locality of a dependency, which is defined according to the number of discourse referents that intervene between the two elements in a dependency relationship. In the DLT, reduced locality is tied to two sources of complexity in a sentence: storage costs and integration costs. Storage costs are incurred when, given a dependency between two elements in a sentence, the first element must be held in working memory until it can be integrated

[17] This generalization seems to hold for many languages, including English, Dutch, French, German, Japanese and Korean. One study has reported the opposite pattern for Mandarin (Hsiao & Gibson, 2003), although this finding has subsequently been called into question (Vasishth, Chen, Li, & Guo, 2013).

with the second element. Integration costs arise when the dependency between these two elements is resolved. Storage costs increase in parallel with the duration for which a prediction regarding the reintegration of an element must be maintained, and integration costs increase with the distance across which an element must be retrieved from memory. To minimize both storage and integration costs, the parser aims to keep dependencies local.

In the object relative clause in (8b), *the lawyer* must be integrated following *criticized*. As processing proceeds, *the lawyer* is held in working memory, and *the doctor* – another discourse referent – is also encountered before the integration site is reached. The dependency in (8b) is therefore less local than that in (8a), where *the lawyer* is integrated before *criticized*, and no intervening discourse referents are encountered.

Gibson's (1998) account is also able to accommodate a number of other parsing preferences that have been observed. For example, the Active Filler Strategy, discussed in section 2.2, can be understood as aimed at keeping dependencies as local as possible. The DLT also accounts for the Late Closure principle, where newly encountered material is attached to the most recently processed, or most local, constituent.[18]

In the literature, some attention has been devoted to the mechanism that ties an increase in the number of discourse referents encountered to an increase in processing difficulty. The specific question posed is whether processing difficulty

[18] Again in relation to the cross-linguistic variation observed in attachment preferences, MacDonald and Hsiao (2018) raise the possibility that innate predispositions to keep dependencies short may be overridden by language-specific preferences.

increases simply because memory capacity is being overburdened, or whether the form of the discourse referents that intervene between the two elements in a dependency matters. This question was investigated by Warren and Gibson (2002), who found that, all things being equal, sentence complexity was lower when intervening discourse referents were pronouns (e.g., *I, you*) but higher when they were indefinite (e.g., *a reporter*) or definite noun phrases (e.g., *the reporter*). Warren and Gibson (2002) explain this finding on the basis of accessibility, where pronouns are said to refer to highly accessible discourse referents and indefinite and definite noun phrases to less accessible referents. According to this account, sentence complexity increases when a dependency extends over discourse referents that, by virtue of being less accessible, require more effort to process.

Another perspective on the importance of the form of the intervening discourse referents is offered by Gordon, Hendrick and Johnson (2001, 2004; also Gordon, Hendrick, Johnson, & Lee, 2006). These authors propose that it is not the accessibility of the intervening discourse referents that matters, but their similarity to the referents already being held in working memory. In this regard, when an element needs to be retrieved from memory, interference results when two similar elements are available. Gordon et al.'s account can explain why (9a) and (9b) are easier to process than sentences containing either two definite noun phrases or two proper names.

(9a) It was the barber that Bill saw in the parking lot.

(9b) It was John that the lawyer saw in the parking lot.

Similarity-based interference also plays a crucial role in other theories of parsing that foreground the operation of memory retrieval as the point at which processing difficulty occurs. Lewis and Vasishth's (2005) ACT-R model is one such theory. The ACT-R model, in simple terms, proposes that as parsing progresses, elements of a sentence are encoded in declarative memory. Then, as parsing and time progress, an element's level of activation declines. When a syntactic dependency must be resolved, the relevant elements must be retrieved from memory. Those elements encountered more recently have higher activation levels, and so can be more easily retrieved. This fact can explain the difference in processing difficulty across subject and object relative clauses. However, a further consideration is that the presence of similar items in memory makes retrieving the correct element more difficult. Specifically, interference occurs when more than one item in memory presents a (partial) match to retrieval cues encoded on the item that triggers retrieval.

Lewis and Vasishth's (2005) account forms the foundation of Cunnings' (2017a, 2017b) memory retrieval account discussed in section 1.5.2. The ACT-R model predicts that processing will be affected by the presence of more than one noun phrase in a sentence that is a potential antecedent for an anaphoric element. Partial cue matches between an anaphor and a noun phrase may result in inhibitory interference, when an illegitimate antecedent must be suppressed in order to successfully retrieve the legitimate antecedent. Alternatively, facilitatory interference may occur in the absence of an antecedent that matches all the retrieval cues, when a partially matching illegitimate antecedent may be retrieved instead of no reference assignment occurring at all.

The ACT-R model constitutes the theoretical grounding for the experiment described in chapter 5, and Gibson’s (1998) DLT is particularly relevant to the experiment presented in chapter 6. The following section describes the theory related to the process of reanalysis, which is a central point of investigation in the experiment reported on in chapter 4.

2.4 Analysis and reanalysis

As indicated in section 2, serial and ranked parallel accounts of sentence processing differ in their understandings of what happens during reanalysis. In serial accounts, reanalysis requires abandoning or revising the incorrect structure and constructing a new, correct structure. In ranked parallel accounts, reanalysis involves promoting another structure over the one that was initially favored. This is strictly speaking not reanalysis, then, but modifying the activation levels of structures that have already been computed (van Dyke, 2003, pp. 289–290). However, in the present section, this process is considered together with understandings of reanalysis captured by the term’s literal meaning.

Van Dyke and Lewis (2003, p. 290) identify a set of functions that they deem necessary to a reanalysis process, regardless of the parsing theory in question. These functions are (i) identifying the location of the incorrect attachment in the structure; (ii) breaking this attachment; (iii) revising the subcategorization frame of the constituent involved in the erroneous attachment; and (iv) making the correct attachment. The authors illustrate these steps with reference to the example in (10).

(10) When the boy strikes the dog kicks.

Assume that *the dog* is initially analyzed as the direct object of *strikes*, and reanalysis is triggered when the parser encounters *kicks*. Step (i) involves identifying the incorrect attachment in the structure, which is the analysis of *the dog* as the complement of the verb *strikes*. Step (ii) requires *the dog* to be excised from the verb phrase. In step (iii), the subcategorization frame of *strikes* must be changed, such that it becomes intransitive rather than transitive. Finally, in step (iv), *the dog kicks* must be attached to the parent node of the sentence in order to yield a two-clause analysis. In a ranked-parallel processing account, these steps should be understood not as being applied to a single analysis, but rather as guiding the evaluation of a set of available analyses, such that the analysis that matches the outcomes of these steps is promoted.

As in the case of analysis itself, reanalysis varies in difficulty across different sentences. Empirical examinations of reanalysis have attempted to identify the source(s) of reanalysis difficulty. One finding yielded by such examinations is that the greater the extent of the structural modifications that need to be made to the initial analysis, the greater the processing difficulty involved. In line with this observation, it has been proposed that the parser treats reanalysis as a last resort. According to this 'Revision as Last Resort' principle (Frazier & Clifton, 1998; Frazier & Fodor, 1980), if reanalysis cannot be avoided, the parser prefers to keep modifications to the initial analysis as minor as possible.

The effect of the extent of the required reanalysis on processing difficulty can be illustrated with reference to the sentences in (11). Sturt, Pickering and Crocker (1999) (see also Holmes, Kennedy, & Murray, 1987; Holmes, Stowe, & Cupples, 1989; Pickering & Traxler, 1998) show that reanalysis in sentences such as (11a) is easier than in sentences such as (11b).

(11a) The Australian woman saw the famous doctor had been drinking quite a lot.

(11b) Before the woman visited the famous doctor had been drinking quite a lot.

The difference between these two reanalysis processes is that in (11a), the direct object analysis of *the famous doctor* is rejected in favor of a complement clause analysis, whereas in (11b), *the famous doctor* must be reanalyzed as the subject of a separate main clause. When revision occurs, the complement clause in (11a) remains the direct object of *saw*, whereas in (11b), *the famous doctor* must be moved out of the thematic domain of *visited* and thematic roles must be assigned anew (Pritchett 1988, 1992). The additional operations involved in (11b) are taken to impose an additional reanalysis cost.

Reanalysis difficulty has also been found to be positively correlated with the length of the ambiguous region in a sentence. A potential explanation for this finding is that the longer the incorrect analysis is maintained, the greater the extent of the parser's commitment to this analysis, and the more difficult it is to revise. As reviewed earlier in this chapter, commitment to an incorrect analysis is also strengthened when this analysis receives support from non-structural sources of linguistic information, such as verb bias (Trueswell et al., 1993) and plausibility (Pickering & Traxler, 1998). As such, the length manipulation makes (12a) easier to reanalyze than (12b), and the plausibility manipulation makes (13a) easier to reanalyze than (13b).

(12a) As the woman edited the magazine amused all the reporters.

(12b) As the woman edited the magazine about fishing amused all the reporters.

(13a) As the woman sailed the magazine amused all the reporters.

(13b) As the woman edited the magazine amused all the reporters.

An alternative perspective on the differences in reanalysis difficulty in (12) and (13) is offered by van Dyke and Lewis (2003). These authors, working within a framework akin to Lewis and Vasishth's (2005) ACT-R model, propose that the already constructed representation of a sentence decays in memory as parsing proceeds. When the disambiguating constituent is encountered, a cue-driven search in memory is initiated in order to locate the correct site for attachment. Reanalysis may prove more difficult when retrieval must occur over a longer temporal distance (as in 12b compared to 12a). Alternatively, interference may occur when there is a strong semantic association between the verb whose subcategorization frame must be revised and the following NP (as in 13b compared to 13a), which heightens the difficulty of retrieving the intransitive analysis of the verb.

The experiment reported on in chapter 4 of this dissertation examines the processing of two kinds of temporarily ambiguous sentences: one where the temporarily ambiguous NP must be reanalyzed as the subject of a complement clause (as in 11a) and one where it must be reanalyzed as the subject of a main clause (as in 11b). Reanalysis difficulty is expected to be greater in the latter than the former. In addition, this experiment manipulated the plausibility of the temporarily ambiguous NP as a direct object of the preceding verb, as in (13). Reanalysis is expected to be more difficult in the 'plausible' than the 'implausible' condition.

2.5 The relationship between the grammar and the parser

There is some debate in the literature regarding the relationship between the grammar and the parser. The debate centers on whether these constitute two separate systems, or different aspects of the same system. A system here is understood as "a collection of cognitive mechanisms with a distinct purpose, operating over representations of a distinct kind" (Lewis & Phillips, 2015, p. 28).

The two-system perspective divides language into competence (grammar) and performance (comprehension and production), and posits that the performance system(s) draw on competence in order to function. From the one-system perspective, the grammar constitutes an abstract description of the representations that the processing system constructs during comprehension and production (Lewis & Phillips, 2015, p. 30).

It is beyond the scope of this dissertation to argue in favor of one of these perspectives. The relevant consideration is that both accounts allow for processing to occasionally deviate from the rules laid down by the grammar. Thus, both accounts can explain why the results of offline tests (such as grammaticality judgement tasks), which are taken to reflect what is licensed by the grammar, may diverge from those of online tests, which are understood to provide insight into the processing mechanisms at play (Lewis & Phillips, 2015, p. 30).

A two-system perspective allows for the processing system to employ parsing principles, such as the ambiguity resolution strategies discussed above in relation to the Garden-Path Model, that exist separately from the grammar and may produce ungrammatical outcomes. In a one-system view, online responses may deviate from offline ones not because processing mechanisms exist separately from the grammar, but because capacity-based limitations may cause processing errors. That is, because real-time processing occurs under time and resource constraints and recruits domain-general resources such as working memory and cognitive control, the parser may sometimes produce unintended outputs.

With reference to L2 processing, the above accounts can also explain why L2 speakers sometimes show nativelike knowledge of a particular grammatical phenomenon offline, but exhibit non-nativelike online processing patterns. It is generally assumed that L1 and L2 processing utilize the same cognitive architecture and parsing mechanisms during syntactic processing (Clahsen & Felser, 2017, p. 1). Differences between L1 and L2 processing, then, are not attributed to different parsing mechanisms employed to process L2 input. Rather, these differences may lie in the way or order in which grammatical knowledge is deployed during processing. Alternatively, they may be caused by external factors, such as resource limitations, that produce discrepancies between grammatical knowledge and processing outcomes, and that may affect L2 processing to a greater extent than L1 processing (see section 1.4.3 for discussion).

2.6 Summary

The current dissertation adopts the uncontroversial understanding of parsing as predictive and incremental, and as being open to influence from non-structural information sources from the outset. The theoretical aspects relevant to the three experiments presented in chapters 4 to 6 are as follows. First, in the processing of temporary subject–object ambiguities (chapter 4), where the plausibility of a temporarily ambiguous NP is manipulated, plausibility information is expected to influence both incremental interpretation and reanalysis, and the extent of reanalysis difficulty is expected to differ across the different construction types examined. Secondly, in reference resolution (chapter 5), a similarity-based interference account (Lewis & Vasishth, 2005) is adopted. The presence of multiple NPs with overlapping features may affect pronoun reference assignment, such that the incorrect referent may sometimes be retrieved. Finally, in the processing of long-distance wh-dependencies (chapter 6), storage and integration costs, as described in Gibson's (1998) DLT, are expected. Throughout, the same parsing mechanisms are taken to be operative in L1 and L2 processing, with deviations between these two groups' processing patterns being attributed to differences in grammatical knowledge and/or the real-time deployment thereof.

3. METHODOLOGY

3.1 Participants

As noted in the previous section, all participants in the experiments for this dissertation were students at Stellenbosch University. They ranged in age from 18–23 years. Individuals of two linguistic profiles were recruited. One group consisted of L1 Afrikaans–L2 English speakers. These participants were first exposed to English during early childhood (AoA ≤ 10 years) and were highly proficient in the language. The second group were L1 English–L2 Afrikaans speakers, who also typically had a relatively early age of first L2 exposure. As it was peripheral to the focus of the dissertation, objective data on these participants' L2 proficiency was not collected, but self-reports indicated that they had fairly high proficiency in Afrikaans. Background data on the participant groups for each experiment are provided in the relevant chapters.

3.2 Research methods

The experiments reported on in this dissertation made use of online data collection methods, which, as they relate to sentence interpretation, involve capturing information as the participant incrementally processes a sentence. Such methods contrast with offline approaches, where a participant reads or hears a sentence in its entirety before they provide a response (such as a grammaticality judgment), and where this response at the end of the sentence is the only data point that is captured. The collection of incremental data as parsing progresses is one of the central benefits of online techniques, as it allows the researcher to examine processing patterns at the region(s) targeted by the experimental manipulation. Online techniques also measure participants' immediate responses without allowing delays for reflection on the stimulus or revision of an initial analysis. Thus, such techniques are considered to be less susceptible to influence from metalinguistic or explicitly learned knowledge and to provide a window into implicit language processing (Keating, Jegerski, & VanPatten, 2016, p. 3).[19] Researchers may also combine online and offline techniques in an experiment in order to contrast participants' offline interpretations with their online processing patterns. An example of such an approach commonly employed in L2 research is to have participants read sentences containing grammatical violations in an online task, and then to compare their processing behavior at the critical regions of these sentences to their ratings of similar sentences in an offline grammaticality judgement task. A possible finding of such an approach would be for participants to display sensitivity to grammatical violations offline but not online, suggesting that they may have explicitly mastered certain grammatical rules of the language, but are not yet able to deploy these rules in real time.

Various methods are available to examine real-time sentence processing. Two of these were employed in this dissertation: self-paced reading (chapters 4 and 6) and eye-tracking-while-reading (chapter 5). Brief descriptions of these methods and their underlying assumptions are provided in what follows.

[19] Although Marsden, Thompson and Plonsky (2018, p. 866) point out that this assumption may be less valid in tasks such as self-paced reading, where the written modality and the participant's ability to control the rate of presentation themselves might lead to greater reliance on explicit knowledge.

3.2.1 Self-paced reading

Self-paced reading is a computer-based research technique in which participants read a sentence in a word-by-word or phrase-by-phrase fashion. It is termed “self-paced” because the participant controls the rate of presentation of each word/phrase by pressing a button. Each button press is recorded, and the time that elapses between button presses (reaction time [RT]) indicates how long the participant spent reading a particular word/phrase. Underlying the interpretation of these RT data is the assumption that the time taken to read a word/phrase is a direct indication of the time taken to process it. Thus, relatively longer RTs are taken to indicate relatively greater processing difficulty, and relatively shorter RTs to mark facilitation. For segments where greater processing difficulty is experienced, so-called spillover effects may also be observed, such that increased RTs are found for both the difficult segment and the immediately following segment. Two reasons for the delay on the spillover segment have been proposed. Firstly, this segment may make available information needed to fully process the previous segment. Secondly, readers may program motor movements such as button presses in advance, such that they press to reveal the following segment before they have fully processed the current one (Just, Carpenter, & Woolley, 1982, p. 232).

An important component of self-paced reading experiments is how the stimuli are presented on screen. A key distinction here is between cumulative and non-cumulative presentations. In the former, each segment of the stimulus that a participant reveals remains on screen as they proceed to subsequent segments, and so at the end of a trial the entire sentence is visible. In the non-cumulative mode, by contrast, the preceding segment is masked each time the participant moves on, and so there is never more than one segment visible on screen. Experimenters may also opt for either a centered or a linear presentation mode, where in the former each segment appears in the center of the screen – thus overwriting the previous segment – and in the latter segments extend in a line across the screen, as in normal text. Cumulative displays have been found to produce a particular reading strategy in participants, such that they press to reveal multiple segments on screen and only then read what is displayed (Just et al., 1982). This strategy renders the RT data uninterpretable, and so the non-cumulative approach is generally preferred. In addition, linear displays are typically favored over centered ones, as the former resemble naturalistic reading conditions more closely. For the self-paced reading experiments reported on in this dissertation, the non-cumulative linear paradigm was employed. This technique is also referred to as the “moving window” paradigm (Just et al., 1982), because each button press causes the ‘window’ of visible text to move across the screen.

Self-paced reading was first used in the 1970s to investigate various facets of L1 processing (Marsden et al., 2018). Its use has however subsequently been extended to investigations of L2 processing, beginning with Juffs and Harrington (1995). In this domain, its relative ease of use and cost-effectiveness have made it one of the most widely used online data collection techniques.

Marsden et al.’s (2018) review of the use of self-paced reading in L2 research indicates that many of these studies aim to address the question of whether and to what extent L1 and L2 processing differ. This question is also the focus of the self-paced reading experiments presented in this dissertation. The linguistic phenomena of interest in these experiments are syntactic ambiguity resolution (chapter 4) and long-distance wh-dependencies (chapter 6). Further details on the use of self-paced reading to investigate processing in these domains are provided in the relevant chapters.

3.2.2 Eye-tracking-while-reading

Eye movements have been studied as a means of gaining insight into cognitive functioning since the late nineteenth century. In the early days of research in this area, methods for capturing eye-movement data were intrusive, and the data yielded were coarse-grained. At present, video-based eye-tracking systems can capture eye movements with essentially no disruption to the subject and can sample at rates of up to 2,000 Hertz, providing extremely fine-grained information on the location of the gaze on a moment-by-moment basis. In addition to the unobtrusive nature of present-day eye-tracking systems, a further benefit of this method is that it can collect data on cognitive processes without the subject having to perform an additional task, such as the button press that is necessary to obtain an RT measure in self-paced reading. Thus, eye-tracking is considered one of the most ecologically valid means of collecting data on language processing.

Eye-tracking methods can be divided into two categories: so-called visual world approaches, which capture the movement of the eyes as they pass over a visual scene while the subject is processing auditory input, and eye-tracking-while-reading, where the subject's eye movements are recorded as they read a text. Both methods have been extensively used to study language processing in L1 and L2 populations.[20] As it is the eye-tracking-while-reading method that is used in the present dissertation, the remainder of this section focuses on this method.

In eye-tracking-while-reading, the stimulus as a whole – which may be a phrase, a sentence or a longer text – is visible on screen throughout the duration of the trial. As the participant reads, two types of eye-movement data are captured: fixations and saccades. Fixations are "relatively stable state[s] of eye movement" (Lai et al., 2013, p. 92), and saccades are rapid movements between fixations. However, as Rayner (1998, p. 373) points out, even during a fixation, the eyes are never completely still, as tremors and other small movements produce some instability. For this reason, particularly in reading research, researchers typically pool short fixations (below 80 ms is standard practice) with a longer neighboring fixation, if the distance between the two is within a defined range.

The interpretation of eye-movement data rests on the so-called eye-mind assumption (Just & Carpenter, 1984), which posits that the location of a fixation indicates where the individual's attention is directed at that moment. With reference to reading, this is taken to mean that when the eye is fixated on a word, the reader is processing that word. Gaze duration is therefore interpreted as an indication of how long processing

[20] Roberts and Siyanova-Chanturia (2013) and Conklin and Pellicer-Sánchez (2016) provide useful overviews of the use of eye-tracking in linguistic research, with specific reference to studies focused on SLA and L2 processing. Huettig, Rommers and Meyer (2011) focus specifically on the use of the visual world paradigm to study language processing.

takes, with, as in the self-paced reading technique, longer gaze durations indexing greater processing difficulty.

However, not all processing during reading is reflected directly in fixations. The term “perceptual span” denotes the amount of information a reader can extract from any one fixation (Roberts & Siyanova-Chanturia, 2013, p. 216). For English specifically, the perceptual span has been proposed to extend from 3-4 characters to the left of a fixation to 14-15 characters to the right of the fixation (McConkie & Rayner, 1975), with this range decreasing as text difficulty increases. The considerable rightwards extension of the perceptual span means that readers are able to extract some information about upcoming words during reading without moving their eyes. In the case of short and/or frequent words, this information may be sufficient to allow the reader to skip the upcoming word altogether. Function words in particular are skipped more often than content words (Carpenter & Just, 1983).

In the case of words or regions that are more difficult to process, two further effects are observed. The first is the spillover effect – also discussed above in relation to self-paced reading – where inflated reading times are observed on the region following one that is more difficult to process. The second are backwards movements of the eyes to earlier parts of the text. According to Rayner (Rayner, 1998), these backwards movements, termed “regressions”, account for 10–15% of saccades. They are also associated with processing difficulty. Depending on their landing site – an earlier position in the word, the sentence or the text – regressions can provide insight into processing difficulties at the word level, the sentence level, or the level of a text as a whole, respectively.

A key advantage of eye-tracking, alongside its ecological validity, is that, unlike self-paced reading, it provides information on both the early and later stages of processing. Early measures, which capture the first movements of a subject's gaze over a region, have been argued to reflect the processes involved in the initial comprehension of a text, such as lexical access and the early stages of information integration (Roberts & Siyanova-Chanturia, 2013, p. 217). Later measures, which include regression path time, re-reading time and total reading time, are believed to be sensitive to later processes related to the comprehension of a text, including reanalysis and recovery from processing difficulties (Roberts & Siyanova-Chanturia, 2013, p. 217). It is standard practice to analyze both early and late measures in eye-tracking studies, in order to provide an overall indication of the progression of processing. The analysis, furthermore, is typically focused on specific regions or areas of interest that are defined according to the study's research question.

The eye-movement measures employed in the present study, which are commonly employed elsewhere, are first fixation duration and first-pass reading time (early measures), as well as selective and non-selective regression path duration, re-reading time, and total reading time (late measures). First fixation duration captures the length of a subject's first fixation in an interest area and is taken to be the earliest measure able to reflect an effect of an experimental manipulation. First-pass reading time provides the sum of the durations of all fixations in an interest area during the subject's first reading of that area, that is, before their eyes move to another interest area. First-pass reading time should be considered in conjunction with first fixation duration in regions larger than single words, as here the subject is likely to make multiple fixations.

As for the later measures, non-selective regression path duration is the sum of the durations of all fixations from the first fixation in an interest area up to and excluding the first fixation to the right of this region. Thus, regression path duration indicates the time a subject spent reading an interest area, as well as the time they spent reading text to the left of the interest area after that interest area had been entered. Selective regression path duration only considers regressions within the interest area in question, thus excluding regressions to earlier parts of the text. Re-reading time is calculated by subtracting first-pass reading time from regression path duration. Finally, total reading time includes the sum of the duration of all fixations made within an area of interest. Effects observed in this measure but not earlier ones may be attributed to later stages of processing (Roberts & Siyanova-Chanturia, 2013).

In chapter 5 of this dissertation, eye-tracking-while-reading was used to examine the pronoun processing in bilingual Afrikaans–English speakers. As pronouns are short and likely visible within the rightwards extension of the perceptual span, the area of interest here was defined so as to include the last three letters of the word preceding the pronoun. The sentence regions following the pronoun were also examined to provide an indication of downstream processing difficulty. Further details on the method of this experiment are provided in the relevant chapter.

3.3 Experimental software

3.3.1 PsychoPy

The two self-paced reading experiments were administered using the experimental software PsychoPy (version 1.85.2; Peirce et al., 2019). PsychoPy has a number of advantages: it is free, open-source, and runs on multiple platforms. PsychoPy experiments can be built in two ways. The software provides a graphical user interface, the 'Builder', which makes available pre-designed experiment components from which users can select. Alternatively, the user can code their experiment from scratch, using Python syntax. This allows for extensive customization, making it possible for researchers to design experiments tailored to their exact needs.

For the two self-paced reading experiments reported on in this dissertation, the Builder interface was used to construct the main scaffolding of the experiments. The presentation of the stimuli was however coded manually by inserting sections of Python code into the Builder-generated script.

3.3.2 SR Research Experiment Builder

The eye-tracking experiment reported on in chapter 5 of this dissertation was built using the Experiment Builder software (version 2.1.14; SR Research, 2017) made available by SR Research, the manufacturer of the eye-tracker employed for this experiment. SR Research provides a wide range of templates for experimental designs commonly employed in eye-tracking. A template for eye-tracking-while-reading ("TextLine") was modified to create the experiment reported on here.

3.4 Materials

The materials for the two self-paced reading experiments and the eye-tracking experiment are described at length in the relevant chapters. A description of the background data collection instruments that were used in all three experiments is provided below.

3.4.1 Language background questionnaires

Two of the three experiments conducted for the dissertation made use of the Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian, Blumenfeld, & Kaushanskaya, 2007) to collect data on participants' language backgrounds. This questionnaire is widely used in psycholinguistic research and is available for free download.[21] The questionnaire was administered in a Microsoft Word form, and the data were subsequently exported to Microsoft Excel for analysis.

Background data for the experiment reported on in chapter 5 were collected using the Language History Questionnaire (LHQ) 3.0 (Li, Zhang, Tsai, & Puls, 2014). The LHQ makes use of a web-based interface for data collection, and as such the data are captured in a downloadable and analyzable format. The questionnaire was developed to serve as a standardized tool for the collection of language background data, and the authors based its contents on the questions most commonly asked across a sample of 41 published linguistic studies. As such, it was possible to use the LHQ to collect the same data captured by the LEAP-Q, while eliminating one of the data pre-processing steps required by use of the LEAP-Q.

3.4.2 C-test

English proficiency was assessed by means of a C-test. In a C-test, words are deleted from a text and replaced by gaps, and participants are provided with (some of) the first letters for each gap. They are then required to fill in the remainder of the missing letters.

---

[21] From the following URL, at the time of writing: https://bilingualism.northwestern.edu/leapq/.

C-tests are useful measures of global language proficiency, as they assess morphosyntactic, lexical, and discourse competence (McNamara, 2000). In particular, for the experiments reported on here, the C-test consisted of three texts. Each of these included 20 incomplete words, with the first 50% of the letters of each missing word provided. Two of the texts were developed by Keijzer (2007) in order to assess L1 attrition in L1 English speakers, and so are relatively difficult. They were therefore deemed appropriate for assessing the L2 speakers in the population under study, who are highly proficient in their L2. The third text was a modified version of a Wikipedia article about a South African university (see Appendices). The L1 English speakers also completed the test to provide a baseline indicator of native-speaker performance and insight into the test's validity.

For scoring of the C-test, any form that was both morphosyntactically and semantically appropriate was accepted. Spelling mistakes were ignored, as participants' spelling abilities are considered not to be indicative of their language proficiency.

3.5 Analysis

The analyses conducted for each of the dissertation's three articles adhere to what are currently considered best practices for dealing with psycholinguistic data. Firstly, to analyze proportion data, for example when comparing comprehension accuracy across participant groups, logit mixed-effects models were employed, to account for the fact that the nature of such data may cause t-tests or analyses of variance (ANOVAs) to yield spurious results (see Jaeger, 2008). Secondly, all RT data were analyzed by means of linear mixed-effects models (see Baayen, Davidson, & Bates, 2008). Due to their right-skewed nature, the RT data were log-transformed to render them suitable for analysis using linear models (see Baayen & Milin, 2010; Vasishth & Nicenboim, 2016).

Chapter 6 also makes use of a Bayesian approach to evaluate the extent to which the data collected support a particular hypothesis. This is an approach that is gaining traction in applied linguistic and psycholinguistic research, as indicated by several recent publications calling for an extension of its use in these fields (e.g., Norouzian, Miranda, & Plonsky, 2018; Norouzian, Miranda, & Plonsky, 2019; Ross & Mackey, 2015). The key affordances of the Bayesian approach compared to the conventional null hypothesis significance testing (NHST) framework are outlined below.

Null hypothesis significance testing has long been the dominant approach to quantitative data in linguistic research (Norris, 2015, p. 97). Within this framework, when a researcher obtains an effect from a study with a particular sample of participants, they calculate the likelihood of obtaining an effect of this size or larger if in reality no effect exists in the true population. The assumption against which the obtained effect is evaluated – that there is no effect in the actual population – is referred to as the null hypothesis. In commonly accepted practice, if the researcher finds that an effect of the size they obtained would occur less than 5% of the time while the effect in the actual population is zero, they reject the null hypothesis (that there is no effect). If an effect of the obtained size would occur more than 5% of the time while the actual effect is zero, the researcher fails to reject the null hypothesis. Importantly, the logic of the NHST framework dictates that nothing more than this can be done. That is, it is not possible for the researcher to conclude that there is no effect; they can only state that they failed to detect an effect. This failure may have occurred for a variety of reasons

– for example, a lack of statistical power may have rendered an existing effect undetectable – and so absence of evidence is not accepted as evidence of absence.

The Bayesian hypothesis testing approach, in contrast, allows any alternative value to serve as the value against which the obtained effect is evaluated. In this approach, the probability of obtaining the observed effect if the null hypothesis were true is calculated, as is the probability of obtaining that same effect if any of the specified alternative hypotheses were true. These two sets of probabilities are then compared, by dividing the probability of the observed effect under the alternative hypotheses by the probability of the observed effect under the null hypothesis. The result of this procedure yields a Bayes Factor, which is a numerical indicator of the extent to which the observed data provide evidence for the alternative rather than the null hypothesis. Bayes Factors are typically interpreted on a gradient, as shown in Table 3-1.

Table 3-1: Bayes Factor Classificatory Scale (Jeffreys, 1998)

Bayes Factor (Alternative/Null)	**Strength of Evidence**
> 100	Decisive evidence for alternative
10 – 30	Very strong evidence for alternative
3 – 10	Substantial evidence for alternative
1 – 3	Anecdotal evidence for alternative
1	Hypothesis-insensitive alternative
1/3 – 1	Anecdotal evidence for null
1/10 – 1/3	Substantial evidence for null
1/30 – 1/10	Strong evidence for null
1/100 – 1/30	Very strong evidence for null
< 1/100	Decisive evidence for null

Bayes Factors, then, make it possible not only to reject but also to accept the null hypothesis. For example, a Bayes Factor of 0.01 indicates that the null hypothesis is 100 times more likely to be true given the observed data. Chapter 6 employs a Bayesian approach with precisely the aim of quantifying support for the alternative over the null hypothesis. Further details on the Bayesian analysis are provided in the chapter.

Finally, all analyses presented in this dissertation were conducted in the R environment for statistical computing (R Core Team, 2018). Each article specifies the version of R and of the R packages that were employed for the relevant analyses. The raw data and the R scripts used for the analyses are available on request from the author.

3.6 Ethical considerations

Ethical clearance for all the experiments reported on in this dissertation was obtained from the Research Ethics Committee: Human Research (Humanities) of Stellenbosch University (see Appendices). In addition, because all participants were students at Stellenbosch University, institutional permission from the university was also obtained (see Appendices). Participants partook in the experiments on a voluntary basis, and were informed that they could withdraw their participation at any time without any adverse consequences. At the time of testing, the procedure of the experiment was explained to them by the experiment administrator, and they were given the opportunity to ask questions. Subsequently, all participants signed an informed consent form indicating their willingness to participate. They were also provided with the principal investigator's contact details to allow them to ask questions about the experiment and/or their data at a later stage. Each experiment incorporated multiple self-timed breaks (as reported in the relevant chapters), and participants were allowed to take additional breaks if they wished to. The experiments ranged in duration from approximately 50 minutes (Experiments 1 and 3) to 80 minutes (Experiment 2).

All data were anonymized for analysis and storage.

4. THE PROCESSING OF OBJECT–SUBJECT AMBIGUITIES IN EARLY L2 ACQUIRERS

4.1 Study abstract

Studies of the second-language (L2) processing of object-subject ambiguities (OSA), where the parser must revise an initially incorrect analysis of the input, have yielded two findings that the present paper addresses. Firstly, L2 speakers are guided more strongly by plausibility information than L1 speakers in OSA processing. Secondly, L2 speakers generally do not perform reanalysis in real time, and their comparably poor performance on subsequent comprehension questions suggests ultimate failure to derive a grammatical parse. Because previous studies have targeted later L2 acquirers living outside of an L2 context, the present study tested whether the above two findings hold for childhood L2 acquirers with extensive naturalistic L2 exposure, hypothesising that these factors would engender more nativelike processing. A self-paced reading task involving two kinds of OSA construction was conducted. Sensitivity to plausibility information differed across the L1 and L2 groups per construction type. Within the L2 group, there is evidence that the early childhood acquirers initiated online reanalysis. Comprehension question accuracy did not differ across participant groups, suggesting that all participants succeeded in reanalysis to the same extent. The role of age of onset in L2 processing is discussed, as is the findings' relevance to current models of L2 processing.

4.2 Introduction

Acquiring a language entails developing the ability to process linguistic input in real-time. As syntactic representations of incoming input are constructed under conditions of uncertainty, without access to the entire sentence being available, language processing also occasionally demands on-the-fly reanalysis of incorrect interpretations. This is challenging, and sometimes fails even during first language (L1) processing (Christianson et al., 2001; Patson, Darowski, Moon, & Ferreira, 2009). Examinations of mis- and reanalysis in second language (L2) processing therefore have much to tell us about the potential limits on ultimate L2 attainment.

To date, a number of such examinations have been conducted with late L2 acquirers, and on the whole have observed marked differences between native and L2 processing of temporarily ambiguous sentences (Jacob & Felser, 2016; Juffs, 1998a; Juffs, 1998b; Juffs & Harrington, 1996; Roberts & Felser, 2011; for an overview, see Papadopoulou, 2005). Attainment of nativelikeness in an L2, however, has been demonstrated to be more likely amongst childhood L2 acquirers (for an overview, see Long, 2013), with success being most likely when first exposure is received in the early childhood years (Long, 1990). In light of these considerations, the present study set out to determine whether childhood L2 acquirers process temporarily ambiguous input in a nativelike fashion, and furthermore whether a distinction can be observed between the processing patterns of early and late childhood acquirers.

4.3 L1 and L2 processing of object–subject ambiguities

During language comprehension, the individual must not only access the meaning of the lexical items they encounter, but also compute a representation of the syntactic structure of the input via syntactic parsing. It is well established that syntactic representations are constructed incrementally, that is, upon encountering each subsequent word, the parser attempts to integrate it into the current representation (Marslen-Wilson, 1973, 1975). A phenomenon that has been used to investigate the incremental nature of parsing and its interaction with other contributors to linguistic interpretation is the garden-path sentence. Consider (1): the comma following *drank* clearly indicates that the noun phrase (NP) *the beer* marks the beginning of a separate clause, and so parsing presents minimal difficulty.

(1) While the men drank, the beer pleased all the customers.

(2a) While the men drank the beer pleased all the customers.

(2b) While the men sang the beer pleased all the customers.

In (2a), readers generally prefer an analysis in which *the beer* is the direct object of *drank*. Different sentence comprehension models explain this preference in different ways. The garden-path model holds that incoming material is preferentially attached inside the current clause (Frazier & Fodor, 1978). Constraint-based accounts attribute this preference to the fact that *drink* is more frequently transitive than intransitive and new clauses are typically signalled by disambiguating punctuation (MacDonald et al., 1994; Spivey-Knowlton, Trueswell, & Tanenhaus, 1993). In online studies, the preference for the direct-object analysis is reflected by longer reading times (RTs) at *pleased*, which is the earliest point in the sentence that serves to reject a direct-object analysis of the preceding NP (Frazier & Rayner, 1982; Warner & Glass, 1987). What happens in the parsing of (2a) is that the parser is led down the metaphorical garden path, initially favoring an analysis of the input that later turns out to be incorrect.

That parsing interacts with other linguistic aspects is indicated by the fact that the strength of the garden-path effect in (2) is modulated by how likely the temporarily ambiguous NP is as a direct object of *drank*. As such, the parser commits less strongly to an analysis in which *the beer* is the direct object of *sang* (cf. 2b). Evidence in support of this claim is provided in an eye-tracking study by Pickering and Traxler (1998). Here, participants were found to have greater difficulty processing NPs that were implausible as direct objects of the preceding verb, as indicated by a larger number of regressions in the NP region in the implausible condition. The reverse effect was evident at the disambiguating region, where more processing difficulty was observed in the plausible condition. The latter finding indicates that the parser commits more strongly to a plausible than an implausible analysis, leading to greater processing difficulty when an initially plausible analysis must be abandoned.

Regarding the mechanisms involved in recovery from misanalysis, different models of sentence comprehension again make different proposals. One account holds that re-ranking of a selection of possible parses occurs, such that the initial parse is demoted and a grammatical alternative is promoted (Spivey & Tanenhaus, 1998); and another that the initial parse is repaired (Frazier & Clifton, 1998). In (2a), repair would entail detaching the NP from the verb phrase (VP) headed by *drank*, reanalysing it as the subject of the main clause, and making the necessary case and thematic role modifications. In instances where the parser was strongly committed to the initial analysis, reanalysis may be sufficiently taxing that the reader is unable to abandon the incorrect parse completely (Christianson et al., 2001; Patson et al., 2009).

Although investigations of garden-path processing provide useful information on how the parser deals with misleading constructions, their utility also extends beyond this. By examining the (re)analysis processes and ultimate outcomes of garden-path comprehension, insight can be gained into the mechanisms employed in 'normal' sentence comprehension as well. For this reason, studies of garden-path processing have proven valuable in comparisons between L1 and L2 sentence processing. Because they shed light on how parsing proceeds in each group, such studies contribute to addressing the central question of whether and under what circumstances L1 and L2 processing can converge. In this regard, three central findings have emerged.

The first is that, like L1 processing, L2 processing is also incremental. When processing temporarily ambiguous sentences, L2 speakers have been found to show sensitivity to the error signal provided by the disambiguating region (Jackson, 2008; Jacob & Felser, 2016; Jegerski, 2012; Juffs, 1998a; Juffs, 1998b; Juffs & Harrington, 1996).

The second finding is that L2 parsing appears to be more sensitive to semantic and probabilistic information than L1 parsing. This is suggested by Roberts and Felser (2011), who examined the processing of two types of garden-path construction in L1 and L2 English (L1 Greek) speakers. The first type of construction had the form exemplified in (2). The second type, in which the temporarily ambiguous NP marks the beginning of a complement clause, is illustrated in (3).

(3a) The boy wrote the report would start an important debate.

(3b) The boy thought the report would start an important debate.

Roberts and Felser's (2011) L2 group showed significantly shorter RTs in the NP region when the NP was a plausible rather than an implausible object of the directly preceding verb in both construction types, whereas this effect only surfaced weakly in the complement-clause constructions for the native speakers. The heightened sensitivity to plausibility information observed in the non-native group aligns with two accounts of L2 processing – Clahsen and Felser's (2006a, 2006b, 2006c, 2017) Shallow Structure Hypothesis (SSH) and Cunnings' (2017a, 2017b) cue-based parsing model – which propose that non-native speakers are generally more reliant on surface-level probabilistic information and underuse morphosyntactic cues in comparison with native speakers.

The third finding is that L2 speakers are typically less successful than L1 speakers in arriving at a correct analysis of temporarily ambiguous input. In terms of initiating reanalysis at the disambiguating region of a temporarily ambiguous construction, self-paced reading and eye-tracking studies have found that L2 speakers either do not attempt reanalysis or abandon attempts at reanalysis earlier than L1 speakers (Jacob & Felser, 2016; Roberts & Felser, 2011). In Roberts and Felser (2011) specifically, the L2 group only showed evidence of initiating reanalysis in the complement-clause constructions. This finding aligns with those of previous studies that have observed minimal processing difficulty in garden-path sentences of this type (Holmes et al., 1987; Sturt et al., 1999). The reduced reanalysis cost in constructions like (3) has been attributed to the fact that the temporarily ambiguous NP remains within the thematic domain of the preceding verb in the grammatical parse, and so the structural revisions required to derive this parse are relatively minor (Pritchett, 1992). In terms of the ultimate outcome of reanalysis processes – as tested by comprehension questions specifically probing the interpretation of the ambiguous region – L2 speakers have been found to perform more poorly than L1 speakers. This applies both to their responses to questions following ambiguous versus unambiguous sentences and to those following sentences where the direct-object analysis had been plausible rather than implausible. Overall, their comparably poor performance suggests that L2 speakers are more susceptible to parsing breakdown in garden-path processing.

Whilst the incremental nature of L2 parsing holds for L2 speakers in general, the other two findings reviewed above appear to vary in accordance with individual differences within L2 speakers. For one, in a study by Hopp (2015), those L2 speakers who scored higher on a measure of syntactic integration ability used both morphosyntactic and semantic information in their incremental analysis of the input, thus showing the more syntactically driven processing typically assumed to be characteristic of native speakers. Additionally, in Roberts and Felser (2011), the faster L2 readers did seem able to initiate reanalysis online.

These results notwithstanding, variation *across* different L2 populations is comparatively underexamined within the syntactic processing literature (although see Gerth et al., 2017) for a cross-linguistic study of garden-path processing). Studies to date have primarily tested late L2 acquirers who have received limited naturalistic exposure to the L2, and little attention has been paid to the processing of early L2 acquirers who have received prolonged, extensive L2 exposure. It is well established that early L2 acquirers are more likely to attain nativelikeness in L2 morphosyntax

(Hartshorne et al., 2018; Johnson & Newport, 1989). With respect to online processing specifically, Clahsen and Felser (2017, p. 9) cite results (Clahsen & Veríssimo, 2016; Veríssimo et al., 2018) suggesting that later L2 acquisition results in less nativelike grammatical processing. Regarding contact with the L2, these authors state that "increased exposure may well lead to the increased automatization or entrenchment of grammatical processing routines". Recent findings indeed suggest a correlation between exposure and processing nativelikeness: specifically, naturalistic L2 exposure seems to play a decisive role in engendering more syntactically rather than semantically driven L2 processing (Pliatsikas et al., 2017; Pliatsikas & Marinis, 2013). In light of these findings, it is plausible that early L2 acquirers who have received considerable naturalistic L2 exposure would show more nativelike garden-path processing.

4.4 The present study

The present paper reports on a self-paced reading study on garden-path processing in English. The participants were L1 Afrikaans–L2 English speakers and a control group of L1 English–L2 Afrikaans speakers. In contrast to Roberts and Felser's (2011) L2 speakers, who had first been exposed to English at age eight in a classroom setting and had received an average of 2.3 years of naturalistic L2 exposure, our L2 group's L2 acquisition, on average, began before age five. They have received naturalistic exposure to English for most of their lives, and currently their exposure is approximately equally divided between their L1 and L2. The study's primary aim is to determine whether garden-path processing in the two groups is similar. A secondary aim is to investigate whether the L2 participants show more sensitivity to semantic information than the L1 participants do. The question here is whether prolonged naturalistic L2 exposure commencing at an early age serves to reduce sensitivity to semantic cues during L2 processing.

The experiment reported on is a modified version of Roberts and Felser (2011), which is in turn based on Pickering and Traxler (1998). It focuses on the influence of plausibility information on the (re)analysis of constructions involving object–subject ambiguities (OSAs), where an NP that is initially taken to be the object of a preceding verb in fact turns out to be the subject of a separate clause. In this regard, as indicated above, Pickering and Traxler (1998) found that: (i) readers have greater difficulty processing a syntactically ambiguous region when the analysis they adopt is implausible rather than plausible, and (ii) readers commit less strongly to an implausible analysis, which entails reduced reanalysis difficulty at the disambiguating region of the sentence.

As in Pickering and Traxler (1998) and Roberts and Felser (2011), the present study considers two kinds of OSA constructions. The first has the form exemplified in (4a), where a preposed adjunct clause is followed by a temporarily ambiguous NP that is the subject of the main clause. In the second type, exemplified in (4b), the temporarily ambiguous NP marks the beginning of a complement clause that is the argument of the verb. The plausibility of the direct-object analysis of the temporarily ambiguous NP was manipulated across both construction types, such that the sections within parentheses in (4a) and (4b) represent the items in the Implausible condition.

(4a) While the child climbed the ladder against the wall (the housewife in the kitchen) fell to the ground (adjunct-clause item).

(4b) The student wrote the report about the budget (the issue with the budget) would start an important debate (complement-clause item).

To enable a comparison between our results and those of previous studies, we employed the procedure used in Roberts and Felser (2011), as well as modified versions of their stimuli. The modification entailed extending the ambiguous region of the experimental items by adding a three-word phrase between the temporarily ambiguous noun and the disambiguating verb. Longer ambiguous regions have been found to result in stronger garden-path effects and thus greater processing difficulty (Christianson et al., 2001; Ferreira & Henderson, 1991; Tabor & Hutchins, 2004), and so Roberts and Felser (2011) shortened their ambiguous region to avoid placing undue demands on their L2 group. However, we assumed our highly proficient L2 participants would not have difficulty comprehending the garden-path sentences, and so we retained the modifying phrase used in Pickering and Traxler's (1998) stimuli. As such, we also extended the window during which a plausibility effect could be observed in the native-speaker group, allowing for the possibility that greater automaticity in L1 processing may delay responses to experimental manipulations (Kaan, Ballatyne, & Wijnen, 2015).

Our predictions were as follows: we expected more nativelike L2 processing patterns, due to our L2 group's early L2 acquisition, prolonged and extensive L2 exposure, and high L2 proficiency. Specifically, we predicted (i) that the L2 group would not show heightened sensitivity to plausibility information compared to the L1 group; and (ii) that the L2 speakers would show a reverse plausibility effect at the disambiguating region of both the adjunct-clause and complement-clause items, such that processing would be easier in the Implausible than the Plausible condition. The latter prediction was expected to result in similarly high accuracy scores on the comprehension questions in the L1 and L2 groups across both construction types.

4.5 Method

Participants

Thirty-three L1 Afrikaans–L2 English speakers (mean age: 20.6 years, SD: 1.98 years) and 34 L1 English–L2 Afrikaans speakers (mean age: 20 years, SD: 0.98 years) participated in the study. The participants were students at a university in South Africa's Western Cape province and all had normal or corrected-to-normal vision. They received course credit for their participation in the experiment.

The LEAP-Q (Marian et al., 2007) was administered to assess participants' language backgrounds, and a C-test of consisting of three texts, each of which contained 20 gaps, was used to obtain a measure of English proficiency. Two of the texts in the C-test were developed by Keijzer (2007) in order to assess L1 attrition in L1 English speakers, and so are relatively difficult. They were therefore deemed appropriate for assessing the L2 speakers in the population under study. The third text was a modified version of a Wikipedia article about a South African university. The two groups scored comparably (L1: 75%, L2: 71%; $p < .001$, $d = 0.29$), but not at ceiling level.

The background data of the L2 participants are summarised in Table 4-1. It should be noted at this point that the L1 group were also functionally bilingual, as they were raised in the bi-/multilingual South African context and at the time were enrolled at a university that uses both English and Afrikaans in learning and teaching. Data on the L1 group's L2 exposure and proficiency is provided in Table 4-8 in the appendix.

Table 4-1: L2 participants' background data (standard deviation)

Age of L2 acquisition	4.45 (2.4)
English exposure (%)	48 (15.1)
Afrikaans exposure (%)	46.76 (14.7)
L2 speaking self-rating*	8.15 (1.42)
L2 reading self-rating	8.6 (1.23)
L2 spoken comprehension self-rating	8.75 (1.35)
C-test score (%)	71.97 (13.7)

Note: *Self-ratings are on a scale from 1–10.

As in Roberts and Felser (2011), the adjunct-clause items and complement-clause items were included in the same experiment, with the one type of construction serving as a filler for the other type. However, because structural differences prevent direct comparison of the processing of these two construction types, we present the results separately, beginning with those for the adjunct-clause items.

4.6 Adjunct-clause items

4.6.1 Materials

The adjunct-clause garden-path items consisted of 20 sentence pairs containing preposed adjunct clauses. They were constructed from 15 optionally transitive verbs, each used no more than three times, which had comparable selectional restrictions in English and Afrikaans to avoid potential effects of lexical transfer. The critical nouns were all matched for syllable number and frequency, based on data from the Celex database (Centre for Lexical Information, 1993). Each noun appeared once in a plausible direct object condition and once in an implausible direct object condition. As in Roberts and Felser (2011), to verify that the plausibility manipulation worked as intended, eight English–Afrikaans bilinguals drawn from the same population as the participants in the main experiment, but who did not participate in the main experiment, conducted an offline rating task in which they rated each verb–noun pairing in terms of its plausibility. The participants were presented with sentences such as *The men drank the beer* and *The men sang the beer* and asked to rate them in terms of plausibility on a scale from 1 (*very plausible*) to 7 (*very implausible*). The plausible sentences received a mean rating of 1.37 and the implausible sentences a mean rating of 6.3. According to a t-test, this is a significant difference ($p < .001$), indicating that the plausibility manipulation functioned as intended.

Each of the experimental items was followed by a comprehension question, where the correct answer was “yes” for one half of the questions and “no” for the other half. In so far as possible, the comprehension questions targeted a correct interpretation of the ambiguous NP, such that if the sentence was *While the child climbed the ladder against the wall fell to the ground*, the comprehension question would be *Did the child fall to the ground?* A correct response of “no” to such a question indicates that the participant has recovered an analysis where *the child* is not understood as the subject of the main clause. The experimental sentences were pseudorandomized and combined with the 20 complement-clause items and 44 filler sentences of different structural types. Twelve of the fillers also contained preposed adjunct clauses where the direct object interpretation turned out to be correct. These were included to prevent participants from becoming aware of the experimental manipulation. One third of the fillers were followed by comprehension questions, again with the “yes” and “no” answers counterbalanced. Finally, eight practice items were included at the beginning of the task.

4.6.2 Procedure

The self-paced reading experiment made use of the non-cumulative moving window procedure (Just et al., 1982) and was administered using PsychoPy (version 1.85.2; Peirce et al., 2019) on a 15-inch laptop (resolution: 1366 x 768). All sentences were presented in a word-by-word fashion in black type (font: Consolas) on a pale blue background. The sentences all fit onto one line, and the end of each sentence was marked by a full stop.

Participants used the space bar to initiate the display of each successive word. The comprehension questions were answered using the ‘z’ and ‘m’ keys on the laptop. The experiment administrator provided verbal instructions to each participant, which were accompanied by written instructions on the screen. Participants were given the opportunity to ask questions before and after completing the practice session. Three self-timed breaks were included in the experiment, so that participants could pause after each block of 22 sentences. The self-paced reading task took approximately 30 minutes, after which participants filled in the LEAP-Q and completed the C-test. An entire testing session lasted approximately 50 minutes.

4.6.3 Analysis

All analyses were conducted in the R environment for statistical computing (version 3.5.1; R Core Team, 2018). Question response accuracy was examined using generalized linear mixed effects models, fit using the lme4 package (version 1.1.17;

Bates, Mächler, Bolker, & Walker, 2015). *P*-values for the regression coefficients were obtained using the lmerTest package (version 2.0.36; Kuznetsova, Brockhoff, & Christensen, 2017). All models included Plausibility (Implausible or Plausible, contrast coded as -1 and 1) and Group (L1 Afrikaans or L1 English, contrast coded as -1 and 1) as fixed effects. As a model with random intercepts for both participants and items would not converge, Akaike Information Criterion (AIC) comparisons were used to determine the optimal random effects structure, which included random intercepts for item.[22]

In accordance with standard practice in sentence processing studies, we only analyzed RT data from trials where the comprehension question was answered correctly. We opted for minimal data trimming, in line with the recommendations in Baayen and Milin (2010). Thus, at the outset of the RT analysis, we only removed extreme values of more than 4,000 ms. This affected 7.5% of the L1 English data and 8.6% of the L1 Afrikaans data.

In order to meet the normality requirement for linear regression, we ran analyses on the log-transformed RTs. We analyzed the data for each segment from the first word of the temporarily ambiguous NP (segment 6) onwards (eight single-word segments in total for each item). To investigate possible differences across conditions and groups in the processing of the comprehension questions, we also examined the question response times. Effects on RT were analyzed with linear mixed effects models using the same package configurations reported above. In all models, Plausibility

[22] AIC value for model with random intercepts for participants = 1091.74; AIC value for model with random intercepts for items = 1011.69.

(Implausible or Plausible; contrast coded as -1 and 1) and L1 Group (L1 Afrikaans or L1 English, contrast coded as -1 and 1) were fixed effects. For all models except that for the segment 12 RTs, the maximal random effects structure that would converge included random intercepts for participants and items, and by-participants random slopes for Plausibility. The segment 12 model included only random intercepts. Model outputs are provided in Table 4-2.

At the segments where there was an L1 Group x Plausibility interaction, we ran separate analyses for each L1 Group. These models included Plausibility (Implausible or Plausible, contrast coded as -1 and 1) as a fixed effect and random intercepts for participants and items. The inclusion of random slopes is specified in footnotes to the relevant text.

4.6.4 Results

Accuracy: The accuracy of both groups' responses to the experimental items was high, with the L1 English group answering 88.5% of the questions correctly, and the L1 Afrikaans group 88.8%. A generalized linear mixed effects model indicated that accuracy was significantly lower for the Plausible vs the Implausible items (Plausibility: $\beta = -0.04$, SE = 0.02, $p = .03$). There was no main effect of group ($\beta = -0.01$, SE = 0.08, $p = .9$), nor a Group x Plausibility interaction ($\beta = 0.09$, SE = 0.08, $p = .2$).

RTs: An example of an adjunct clause item divided into segments is provided below to serve as a reference for the presentation of the RT analyses.

While / the / men / drank / the / beer / imported / from / Europe / pleased / all / the /

1 2 3 4 5 6 7 8 9 10 11 12

customers.

13

Table 4-2: Model outputs, adjunct-clause items

	SEGMENT 6 *beer*		**SEGMENT 7** *imported*		**SEGMENT 8** *from*		**SEGMENT 9** *Europe*		**SEGMENT 10** *pleased*		**SEGMENT 11** *all*		**SEGMENT 12** *the*		**SEGMENT 13** *customers*	
Parameter	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***
Plausibility	-0.03 (0.05)	0.548	-0.15 (0.05)	0.002	-0.13 (0.04)	0.001	-0.01 (0.05)	0.804	0.13 (0.07)	0.070	0.14 (0.06)	0.009	0.06 (0.04)	0.146	0.14 (0.06)	0.011
Group	-0.19 (0.11)	0.075	-0.12 (0.09)	0.163	-0.08 (0.07)	0.271	-0.15 (0.09)	0.097	-0.07 (0.10)	0.482	-0.05 (0.06)	0.407	-0.02 (0.06)	0.708	-0.11 (0.08)	0.181
Plausibility x Group	0.05 (0.06)	0.377	0.05 (0.05)	0.344	0.09 (0.05)	0.046	0.10 (0.06)	0.081	0.11 (0.08)	0.175	0.14 (0.04)	0.002	0.06 (0.04)	0.094	0.03 (0.06)	0.594

The reading speeds of the L1 English and L1 Afrikaans groups did not differ for the most part: the L1 English group were slightly faster than the L1 Afrikaans group only at segment 6, the first word of the temporarily ambiguous NP. At segment 7, which contained the first word of the phrase modifying the noun, there was a significant main effect of Plausibility, where participants read faster when the NP was a plausible direct object of the verb in segment 4. The effect of Plausibility carried over to segment 8. At this segment, there was also an interaction between L1 Group and Plausibility. Separate analyses indicated that whereas the L1 Afrikaans group read significantly faster in the Plausible condition at this segment ($\beta = -0.17$, SE = 0.04, $p < .001$), the Plausibility effect was only a weak trend for the L1 English group ($\beta = -0.09$, SE = 0.05, $p = .08$).[23]

Segment 10 included the main clause verb, which served to indicate that the direct-object analysis of the preceding NP was incorrect. At this segment, a marginal effect of Plausibility appeared, where participants slowed down more upon encountering the disambiguating verb when the direct-object analysis of the preceding NP had been plausible. This effect was also present at the following segment. At segment 11, in addition to the main effect of Plausibility, there was also a significant L1 Group x Plausibility interaction. Separate analyses[24] revealed that the RTs of the L1 English group were significantly higher in the Plausible condition ($\beta = 0.2$, SE = 0.06, $p = .001$), whereas there was no effect of Plausibility on the RTs of the L1 Afrikaans group (β =

[23] The L1 Afrikaans model included by-participants random slopes for Plausibility. The L1 English model would not converge with any random slopes.

[24] No random slopes were included in these models, as neither model would converge with them.

0.07, SE = 0.06, *p* = .2). No significant effects were observed at segment 12. At segment 13, the final segment, participants were slower in the Plausible than the Implausible condition. This likely reflects the heightened cost of reanalysis in the Plausible condition, where participants took longer to settle on a final analysis of these items.

Response times across all comprehension questions were also examined.[25] Both groups were significantly slower in the Plausible compared to the Implausible condition (β = 0.2, SE = 0.07, *p* = .005). There was no effect of L1 Group on response times (β = -0.04, SE = 0.07, *p* = .6), nor was there an L1 Group x Plausibility interaction (β = 0.005, SE = 0.06, *p* = .9).

Summary – Adjunct-clause items: Both groups showed a response to the Plausibility manipulation, where the temporarily ambiguous NP was easier to process if it was plausible as a direct object of the preceding verb. However, the effect of Plausibility lasted longer in the L1 Afrikaans group, persisting into segment 8. At the disambiguating region of the sentence, there was some evidence that both groups initiated reanalysis at segment 10, but the effect of Plausibility on RTs only carried over to the following segment in the L1 English group. The groups did not differ in their overall performance on the comprehension questions, either in terms of the time taken to answer or accuracy.

[25] The model included fixed effects of L1 Group and Plausibility, random intercepts for participants and items, and by-participants and by-items random slopes for Plausibility.

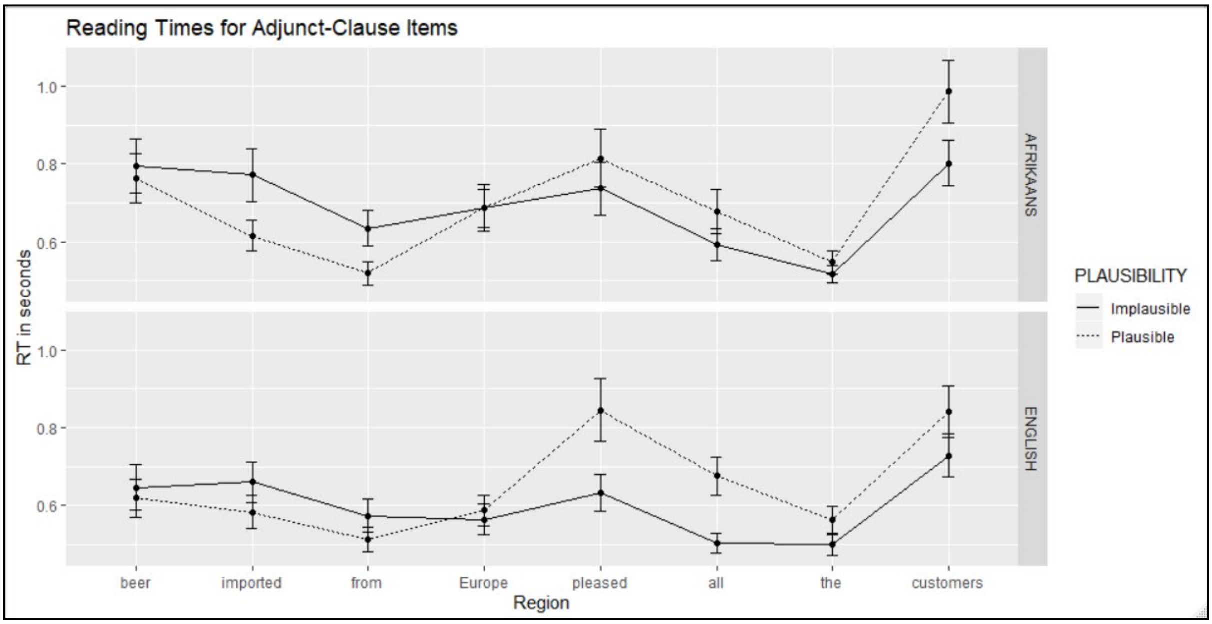

Figure 4-1: Reading times for adjunct-clause items. Error bars represent standard error.

4.7 Complement-clause items

4.7.1 Materials

The complement-clause items consisted of 20 sentence pairs, each containing an optionally transitive verb followed by a finite complement clause. Eleven verbs were selected, none of which occurred more than three in the stimulus list. The critical nouns were again matched for syllable length and frequency, and each noun occurred once in a plausible direct object condition and once in an implausible direct object condition. In the same offline rating task discussed in relation to the adjunct-clause garden-path items, eight English–Afrikaans bilinguals who did not participate in the main experiment rated the plausibility of each verb–noun pair, with a rating of 1 indicating that the noun was a very plausible object of the verb, and a rating of 7 indicating that the pairing was implausible. The plausible items received a mean rating of 1.9 and the implausible items a mean rating of 4.9. These are significantly different ($p < .001$), again indicating that the plausibility manipulation functioned as intended.

The 20 experimental items, all of which were followed by a comprehension question with "yes" and "no" answers counterbalanced, were pseudorandomized and combined with the 20 adjunct-clause garden-path items and the 44 fillers. The fillers included 12 items where the direct-object analysis was actually correct in order to disguise the experimental manipulation.

4.7.2 Analysis

As with the adjunct-clause items, a generalized linear mixed effects model was fit to investigate possible differences between the groups and conditions. The maximal random effects structure that would converge included random intercepts for participants and items.

Reading times were analyzed for each segment from the first word of the temporarily ambiguous NP (segment 5) onwards (nine single-word segments in total for each item). In addition, we analyzed the groups' RTs to the comprehension questions following the experimental items. Only RTs from trials where the comprehension question was answered accurately were analyzed, and RTs over 4,000 ms were removed. This resulted in the removal of 7.5% of the data for the L1 Afrikaans group and 5.9% of the data for the L1 English group.

As in the previous analysis, separate linear mixed effects models were computed for each segment, with log-transformed RT as the dependent variable and L1 Group (Afrikaans or English, contrast coded as -1 and 1) and Plausibility (Implausible or

Plausible, contrast coded as -1 and 1) as fixed effects. For all but two models, the maximal random effects structure that would converge included random intercepts for participants and items, as well as by-participants random slopes for Plausibility. The segment 5 and 7 models converged with both sets of random intercepts, as well as by-items random slopes for Plausibility. Model outputs are provided in Table 4-3.

At segments where there was an L1 Group x Plausibility interaction, we computed separate linear mixed effects models for each group, with Plausibility as a fixed effect and random intercepts for participants and items. The inclusion of random slopes is specified in footnotes to the relevant text.

Table 4-3: Model outputs, complement-clause items

	SEGMENT 5 *Report*		SEGMENT 6 *about*		SEGMENT 7 *the*		SEGMENT 8 *budget*		SEGMENT 9 *would*		SEGMENT 10 *start*		SEGMENT 11 *an*		SEGMENT 12 *important*		SEGMENT 13 *debate*	
Parameter	**β (SE)**	***P***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***
Plausibility	-0.03 (0.04)	0.339	-0.20 (0.04)	< .001	-0.09 (0.04)	0.015	-0.08 (0.04)	0.035	-0.00 (0.05)	0.980	-0.10 (0.05)	0.058	0.01 (0.04)	0.813	-0.03 (0.03)	0.421	-0.05 (0.05)	0.303
Group	-0.15 (0.10)	0.120	-0.13 (0.09)	0.168	-0.12 (0.07)	0.120	-0.13 (0.09)	0.148	-0.13 (0.09)	0.156	-0.06 (0.08)	0.471	-0.10 (0.07)	0.118	-0.04 (0.06)	0.540	-0.05 (0.08)	0.537
Plausibility x Group	-0.07 (0.04)	0.100	-0.06 (0.05)	0.184	-0.06 (0.04)	0.089	-0.07 (0.04)	0.092	-0.05 (0.06)	0.355	0.02 (0.06)	0.781	-0.03 (0.04)	0.488	-0.01 (0.03)	0.826	0.05 (0.05)	0.412

4.7.3 Results

Accuracy: The accuracy of both groups' responses to the comprehension questions was again high, with the L1 English group answering 88% of the questions correctly, and the L1 Afrikaans group answering 91.8% of the questions correctly. A generalized linear mixed effects model indicated no significant difference in performance across the groups (β = -0.2, SE = 0.14, *p* = .1) or the conditions (β = -0.16, SE = 0.2, *p* = .4), nor any L1 Group x Plausibility interaction (β = -0.004, SE = 0.12, *p* = .9).

RTs: An example of a complement-clause item divided into segments is provided below to serve as a reference for the presentation of the RT analyses.

The	student	wrote	the	report	about	the	budget	would	start	an	important
1	2	3	4	5	6	7	8	9	10	11	12

debate.
13

Overall reading time did not differ between the L1 groups, as indicated by the absence of a main effect of L1 Group at all segments. At segment 6, the first word of the phrase modifying the temporarily ambiguous noun, a significant main effect of Plausibility was observed. Here, participants read faster when the noun was a plausible direct object of the verb in segment 3. This effect persisted at segment 7, where there was also a weak L1 Group x Plausibility trend. Separate analyses of the groups[26] indicated that

[26] The L1 Afrikaans model included by-participants random slopes for Plausibility. The L1 English model did not converge with any random slopes.

the L1 English group were significantly faster in the Plausible condition at this segment (β = -0.12, SE = 0.04, *p* = .004), whereas the RTs of the L1 Afrikaans group did not differ across the Plausible and Implausible conditions (β = -0.06, SE = 0.04, *p* = .1). At segment 8, participants still read faster in the Plausible than the Implausible condition, and the weak Plausibility x L1 Group interaction remained. The separate group analyses[27] again showed shorter RTs in the L1 English group in the Plausible condition (β = -0.12, SE = 0.04, *p* = .01), but no effect of Plausibility in the L1 Afrikaans group (β = -0.05, SE = 0.04, *p* = .3). There were no significant effects at segment 9, but at segment 10, which contained the disambiguating verb, a marginal main effect of Plausibility was present. Participants were slower in the Plausible than the Implausible conditions here, indicating the increased difficulty of reanalysis when the direct-object analysis of the temporarily ambiguous NP had been plausible. There are no further significant effects at segments 10–13.

In their responses to the comprehension questions, there were no significant differences in RTs across groups (β = 0.03, SE = 0.08, *p* = .7) or across conditions (β = 0.12, SE = 0.08, *p* = .1), nor was there an L1 Group x Plausibility interaction (β = -0.07, SE = 0.06, *p* = .2).

Summary – Complement-clause items: As with the adjunct-clause items, both groups showed an effect of Plausibility, where the temporarily ambiguous NP was read more quickly when it was plausible as a direct object of the preceding verb. In contrast to the adjunct-clause results, however, the Plausibility effect was more persistent in

[27] The L1 Afrikaans model included by-participants random slopes for Plausibility. The L1 English model included by-participants and by-items random slopes for Plausibility.

the L1 English speakers in this condition. At the point in the sentence signalling the need for reanalysis, both groups slowed down slightly more in the Plausible compared to the Implausible condition. This effect was marginal and did not persist to the following segments, suggesting that reanalysis of these items did not pose much difficulty for the participants. The groups' comprehension question accuracy and response times again did not differ: they were equally successful in recovering the correct interpretations of the experimental items, and the duration of this recovery process did not differ according to L1 status.

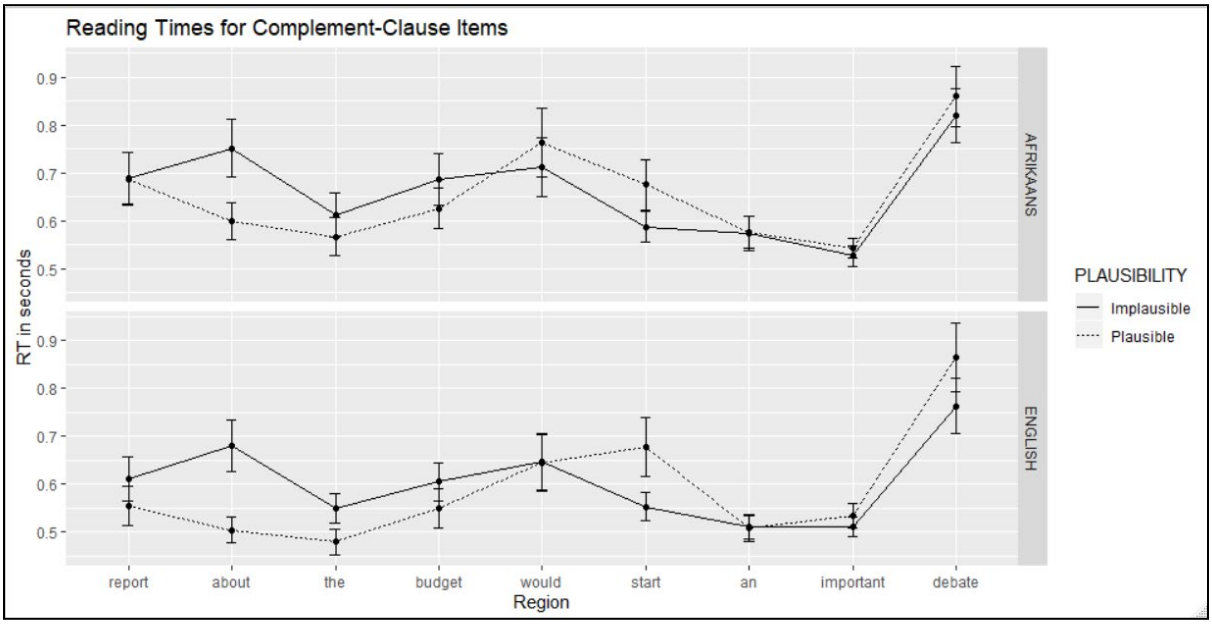

Figure 4-2: Reading times for complement-clause items. Error bars represent standard error.

4.8 Analysis of age effects

As noted earlier in the paper, L2 speakers with earlier ages of L2 acquisition (AoA) are typically assumed to have a greater likelihood of attaining L2 nativelikeness. Specifically in terms of L2 grammatical processing, nativelike performance has been found in L2 speakers with an AoA of 5–6 years or lower (Granena & Long, 2013; Veríssimo et al., 2018; Weber-Fox & Neville, 1996), with nativelikeness subsequently decreasing as AoA increases. As a follow up to the main analyses presented above, we investigated whether there were AoA influences on processing within our L2 group, whose AoAs ranged from 1–10 years. We computed linear mixed effects models for each segment of interest in both construction types. In all models, the dependent variable was log-transformed RT. Fixed effects were Plausibility (Implausible or Plausible, contrast coded as -1 and 1), AoA (centred around the mean) and the interaction of these two variables. Random intercepts were included for participants and items. For the adjunct-clause models, the maximal random effects structure that would converge included by-participants random slopes for Plausibility for all but three segments. At segment 10, by-participants and by-items random slopes for Plausibility were included. At segments 7 and 11, no random slopes were included. For the complement-clause items, the segment 5–7 models included only random intercepts. By-participants random slopes for Plausibility were included in the remaining models. Model results are presented in Table 4-4 and Table 4-5.

Table 4-4: Age effects on L2 processing of adjunct-clause items

	SEGMENT 6 *beer*		SEGMENT 7 *imported*		SEGMENT 8 *from*		SEGMENT 9 *Europe*		SEGMENT 10 *pleased*		SEGMENT 11 *all*		SEGMENT 12 *the*		SEGMENT 13 *customers*	
Parameter	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***
Plausibility	-0.03 (0.03)	0.270	-0.08 (0.02)	0.000	-0.09 (0.02)	0.000	-0.03 (0.03)	0.318	0.03 (0.04)	0.387	0.03 (0.03)	0.258	0.01 (0.02)	0.575	0.06 (0.04)	0.078
AoA	0.13 (0.08)	0.109	0.06 (0.06)	0.369	0.06 (0.05)	0.263	0.08 (0.06)	0.188	0.14 (0.07)	0.038	0.06 (0.05)	0.194	0.04 (0.04)	0.358	0.06 (0.07)	0.391
Plausibility x AoA	-0.01 (0.02)	0.687	0.00 (0.02)	0.954	0.01 (0.02)	0.607	0.00 (0.02)	0.863	0.00 (0.02)	0.860	-0.02 (0.02)	0.184	-0.02 (0.01)	0.084	0.00 (0.02)	0.984

Table 4-5: Age effects on L2 processing of complement-clause items

	SEGMENT 5 *report*		SEGMENT 6 *about*		SEGMENT 7 *the*		SEGMENT 8 *budget*		SEGMENT 9 *would*		SEGMENT 10 *start*		SEGMENT 11 *an*		SEGMENT 12 *important*		SEGMENT 13 *debate*	
Parameter	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***
Plausibility	0.00 (0.02)	0.894	-0.08 (0.02)	0.000	-0.03 (0.02)	0.126	-0.02 (0.02)	0.292	0.01 (0.03)	0.680	0.04 (0.03)	0.134	0.00 (0.02)	0.873	0.02 (0.01)	0.258	0.01 (0.03)	0.712
AoA	0.11 (0.07)	0.117	0.06 (0.07)	0.416	0.09 (0.05)	0.096	0.07 (0.07)	0.266	0.09 (0.07)	0.178	0.06 (0.06)	0.276	0.06 (0.05)	0.265	0.04 (0.04)	0.315	0.07 (0.06)	0.203
Plausibility x AoA	0.01 (0.02)	0.396	-0.02 (0.01)	0.257	0.02 (0.01)	0.155	0.02 (0.02)	0.264	0.03 (0.02)	0.187	0.01 (0.02)	0.793	0.01 (0.02)	0.455	0.01 (0.01)	0.324	0.02 (0.02)	0.219

For the complement-clause items, there were no significant main effects of AoA, nor any interactions between AoA and Plausibility, at any of the segments. However, for the adjunct-clause items, there was a significant main effect of AoA at segment 10, the disambiguating verb, and a Plausibility x AoA trend at segment 12. To further examine this AoA effect, a median split was performed on the adjunct-clause data for the L2 speakers, yielding one group with an AoA below 5 (n = 16; AoA range: 1–4 years), and one group with an AoA equal to or greater than 5 (n = 17; AoA range: 5–10 years). We then ran separate linear mixed effects models per group for each segment of the adjunct-clause items, with log-transformed RT as the dependent variable, Plausibility as a fixed effect (Implausible or Plausible, contrast coded as -1 and 1), and random intercepts for Participant and Item. The models for the earlier acquirers included by-participants random slopes for Plausibility at all but three segments – 7, 9 and 12 – where segments 7 and 12 only had the random intercepts included in the other models, and segment 9 only had by-participants random intercepts. The models for the later acquirers had by-participants random slopes for Plausibility at segments 5, 10, 12 and 13, by-items random slopes for Plausibility at segments 6 and 7, both by-items and by-participants random slopes for Plausibility at segment 9, and only random intercepts at segments 8 and 11. Model results are presented in Table 4-6 and Table 4-7.

Table 4-6: Later L2 acquirers, adjunct-clause items

	SEGMENT 6 *beer*		**SEGMENT 7** *imported*		**SEGMENT 8** *from*		**SEGMENT 9** *Europe*		**SEGMENT 10** *pleased*		**SEGMENT 11** *all*		**SEGMENT 12** *the*		**SEGMENT 13** *customers*	
Parameter	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***
Plaus.	-0.04 (0.03)	0.132	-0.09 (0.03)	0.001	-0.07 (0.02)	0.001	-0.03 (0.04)	0.503	0.00 (0.05)	0.941	0.01 (0.04)	0.806	-0.02 (0.03)	0.531	0.05 (0.04)	0.174

Table 4-7: Earlier L2 acquirers, adjunct-clause items

	SEGMENT 6 *beer*		**SEGMENT 7** *imported*		**SEGMENT 8** *from*		**SEGMENT 9** *Europe*		**SEGMENT 10** *pleased*		**SEGMENT 11** *all*		**SEGMENT 12** *the*		**SEGMENT 13** *customers*	
Parameter	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***	**β (SE)**	***p***
Plaus.	-0.02 (0.04)	0.583	-0.08 (0.03)	0.005	-0.10 (0.04)	0.004	-0.02 (0.02)	0.447	0.07 (0.04)	0.042	0.06 (0.03)	0.060	0.04 (0.02)	0.049	0.07 (0.05)	0.135

4.9 Discussion

Previous investigations of L2 OSA processing have found L2 speakers (i) to be more sensitive to plausibility information than L1 speakers and (ii) to have greater difficulty than L1 speakers in recovering a correct parse of the input. In the present study, the results regarding sensitivity to plausibility information were mixed, with L2 speakers showing greater sensitivity than L1 speakers in the adjunct-clause condition, and the reverse being true in the complement-clause condition. Regarding garden-path recovery, we observed a clear slowdown at the disambiguating region of the Plausible adjunct-clause items in the L1 speakers. Among the L2 speakers, we found an effect of AoA, where the group of early childhood acquirers (AoA < 5) showed a nativelike slowdown effect. In the late childhood acquirers (AoA ≥ 5), no such effect was present. Furthermore, in contrast to previous findings, the L1 and L2 groups did not differ in the accuracy of their responses to the comprehension questions, and so the non-native speakers were evidently not more susceptible to parsing breakdown.

The interaction between L1 Group, Plausibility and Construction Type observed in our data is not present in Roberts and Felser (2011), where only the L2 speakers showed a strong response to the Plausibility manipulation, and where this response was present across both types of construction. In our data, there was an initial strong response to the plausibility of the temporarily ambiguous NP in the L1 and the L2 group. However, although both groups were sensitive to the Plausibility manipulation regardless of the construction type, they differed in each showing a stronger reaction to plausibility information in a particular construction type.

Findings from the monolingual processing literature offer some insight into why the L1 speakers' sensitivity to the plausibility manipulation may have differed across the two construction types. A close examination of the adjunct- and complement-clause items reveals that they differed in terms of the subcategorization information provided by the initial verb. Subcategorization information specifies whether a verb can take only an NP complement, only a sentential complement, either type of complement, or no complement at all (Chomsky, 1965). Among verbs of the penultimate sort, information on verb bias – the tendency for a verb to take a particular kind of complement – can also serve to generate expectations about upcoming input. For example, while *hit* can only take an NP complement, and *proved* can only take a sentential complement, *wrote* can take either type: both *John wrote the manuscript* and *John wrote the manuscript had been destroyed in the fire* are possible (Garnsey et al., 1997, p. 59). Information on the frequency with which *wrote* occurs in direct-object versus sentential complement constructions in a particular speech community can be used to classify it in terms of subcategorization bias.

Verb bias has been found to interact with plausibility information (Garnsey et al., 1997), such that plausibility only has a clear effect under certain verb bias conditions. For example, in Garnsey et al. (1997), readers only showed sensitivity to the plausibility of a temporarily ambiguous NP as a direct object of the preceding verb in the absence of clear verb bias information. A possible explanation for this finding is that when verb bias information is available, readers adopt the corresponding structural analysis of the temporarily ambiguous section, and so rely less on the plausibility of the NP to guide parsing. However, in the absence of verb bias information, neither a direct-object nor a sentential-complement analysis is favored by the available cues, and the reader may rely more heavily on plausibility information in deciding which parse to adopt.

In the present study, clear differences in subcategorization biases are evident across the adjunct-clause and complement-clause conditions. All but two of the verbs in the former condition (*called* and *painted*) can only take an NP-complement (see Table 4-9 in the appendix); arguably, the two verbs that constitute the exceptions also have a heavy direct-object bias. These verbs therefore provide strong cues in favor of a direct-object analysis of the following NP. In contrast, in the complement-clause items, none of the verbs are restricted to taking a particular kind of complement. While some of the verbs may have been biased in this regard[28], what is important is that the cue provided by the subcategorization information is considerably weaker in this condition.

The strong initial reaction to plausibility observed in the L1 group aligns with previous findings that readers attempt to integrate a temporarily ambiguous NP as the direct object of a preceding verb regardless of the verb's bias (Traxler, 2002; Traxler, Pickering, & Clifton Jr, 1998; Van Gompel, Pickering, & Traxler, 2001). However, Traxler (2005) also provides evidence that verb bias information is accessed rapidly as parsing proceeds, which would account for the rapid change in the strength of the plausibility effect in our data. Following Garnsey et al. (1997), the presence of clear verb bias information in the adjunct-clause condition may have led the L1 group to rely

[28] Based on Garnsey et al.'s (1997) ratings, four of the complement-clause verbs would be classified as having a sentential-complement bias (*confessed*, *believed*, *proved* and *suggested*), and three as having a direct-object bias (*warned*, *wrote* and *confirmed*). Again, though, the point is that, in contrast to the adjunct-clause verbs, none of the complement-clause verbs that were used was restricted to taking a particular kind of complement.

less on plausibility information here, whereas the absence of verb bias information in the complement-clause condition elicited the opposite response.

Although some studies have found L2 speakers to be able to both learn verb bias information and employ it during online parsing (Dussias & Cramer Scaltz, 2008; Jegerski, 2012; although cf. Frenck-Mestre & Pynte, 1997, Lee, Lu, & Garnsey, 2013, and Qian, Lee, Lu, & Garnsey, 2018, for alternative findings), previous experiments with L2 speakers have not explicitly manipulated plausibility in conjunction with verb bias. Thus, while it is evident that L2 speakers *can* use verb bias information during online processing, it is not clear whether they do so preferentially, given the simultaneous availability of semantic cues. The pattern of results observed in the present study could be explained if the L2 speakers did employ verb subcategorization information to guide parsing, but experienced greater processing disruption if the temporarily ambiguous NP was implausible given their expectations regarding the upcoming input. For example, supposing that the L2 group also accessed the subcategorization information provided by the verb in the adjunct-clause condition, they would have expected the following NP to be the direct object of the verb. Subsequently, assuming a heightened sensitivity to plausibility information in this group – as is suggested by Roberts and Felser's (2011) results – they may have reacted strongly when the NP was implausible in this role. In the complement-clause condition, the absence of unequivocal subcategorization information may have prevented strong expectations regarding the upcoming input from being generated. In this case, there would be a reduced clash between the reader's expectations and the plausibility of the input in this condition, leading to a mitigated response to implausible NPs.

The speculative nature of this explanation is acknowledged, but it suggests potentially fruitful avenues for further investigation of the relative weightings of syntactic versus semantic cues in L1 and L2 processing, in line with recent cue-based approaches to L2 processing such as that of Cunnings (2017a, 2017b). In particular, the fact that both the early and late L2 acquirers showed the same differential sensitivity to plausibility information across the two construction types may suggest that AoA is not the sole determinant of non-nativelikeness in L2 speakers' cue-weighting.

We turn now to the second main finding of this study, namely the difference in behavior at the disambiguating region across the L1 and L2 groups. One of the study's predictions was that given our L2 participants' early age of L2 acquisition, extensive naturalistic L2 exposure, and high L2 proficiency, they would exhibit nativelike processing patterns. This prediction was borne out at the disambiguating region of the complement-clause items, where both the natives and non-natives showed only a marginal slowdown in the Plausible condition. The mild nature of the processing disruption observed at this point is in line with the findings of previous studies with monolingual speakers, and, as discussed above, has been attributed to the fact that the structural revisions required to arrive at a grammatical parse of garden-path constructions of this sort occur within the thematic domain of the predicate. In spite of this, L2 speakers with a later AoA and limited naturalistic L2 exposure have been found to slow down significantly at the disambiguating region of complement-clause OSA constructions (Roberts & Felser, 2011), which reflects difficulty in revising their initial parse. Our results, however, indicate that nativelike processing of these constructions is possible in L2 speakers of the profile examined here.

The study's prediction of L1–L2 convergence in reanalysis behavior was not entirely borne out for the adjunct-clause items, where more substantial structural revisions are required to derive a grammatical parse. Specifically, the L1 group exhibited a significantly stronger slowdown at the disambiguating region in the Plausible condition compared to the L2 group. Closer scrutiny of the L2 group revealed an effect of AoA at this region: whereas the early childhood acquirers (AoA < 5) exhibited longer RTs here in the Plausible condition, the late childhood acquirers (AoA ≥ 5) did not. This suggests that only the early childhood acquirers initiated reanalysis of these constructions online.

The present data do not allow for strong claims to be made regarding the precise age at which the ability to reanalyze L2 input in real-time begins to decline. However, the difference between the processing patterns of the early and late childhood L2 acquirers in our study is in accordance with previous observations of a distinction between L2 acquirers with AoAs on either side of approximately six years of age, which has been cited as the offset of the period in which ultimate attainment of nativelike L2 morphosyntax is most likely (Flege, Yeni-Komshian, & Liu, 1999; Granena & Long, 2013; Hyltenstam, 1992; Johnson & Newport, 1989; Long, 1990). Studies of online processing specifically have observed a decline in nativelikeness beginning at an AoA of ages four to five (Veríssimo et al., 2018; Weber-Fox & Neville, 1996). Regardless of the exact age at which processing nativelikeness begins to decline, it is nonetheless striking that age effects are observed in the present sample of L2 speakers, who on the whole have had extensive exposure to the L2 from a young age.

An additional novel finding of the present study is that the non-native group as a whole performed within the native-speaker range on the offline measure of comprehension, contra some previous findings (e.g., Jacob & Felser, 2016; Roberts & Felser, 2011). In this regard, the L1 and L2 groups were equally accurate in their responses to the comprehension questions following the experimental items, and there was also no difference in accuracy across the early and late childhood L2 acquirers. The pattern observed across the board was lower accuracy in the Plausible compared to the Implausible condition. Evidently, then, all participants were equally able (or unable, as the case may have been) to recover a correct parse of the adjunct-clause items. In terms of RTs to the comprehension questions, both the L1 and L2 group were significantly slower in responding to the questions following the Plausible items. However, this delay was not extended amongst the late childhood L2 acquirers. It therefore appears that, although the late childhood acquirers did not seem to initiate reanalysis of the Plausible items online, they were still able to arrive quickly at a grammatical parse of these constructions.

This finding suggests that even for late childhood L2 acquirers, there is some benefit to extensive naturalistic L2 exposure and high L2 proficiency in facilitating the revision of structural analyses. Whilst online reanalysis may still be beyond reach in particularly difficult garden-path sentences, it seems that L2 speakers of this profile are nonetheless able to perform rapid offline reanalysis, ultimately leaving them on equal footing with native speakers and early childhood L2 acquirers in terms of the final outcome of comprehension.

4.10 Conclusion

The results of the present study have several implications. Firstly, they suggest an early cut-off point in terms of AoA for nativelike L2 processing of structurally complex garden-path constructions. Secondly, although L2 speakers first exposed to the L2 beyond this point may not be equipped to perform reanalysis in real-time, they may still be able to conduct rapid offline reanalysis. This assertion is supported by the equivalent performance of our L1 and L2 groups on the comprehension questions following the experimental items, despite the lack of evidence of online reanalysis amongst the later L2 acquirers. Thirdly, contra certain models of L2 processing, L2 speakers are not always more sensitive than L1 speakers to semantic information. To explain the interaction between construction type, L1 versus L2 status and plausibility observed in our study, we have suggested an interplay between two different cue types – syntactic and semantic – that may be weighted differently across L1 and L2 speakers, sometimes leading to surprising results. Further research should aim to shed light on this interplay in both native- and non-native language processing.

4.11 Appendix

Table 4-8: L1 participants' background data (standard deviation)

Age of L2 acquisition	5.7 (3)
English exposure (%)	74 (14.6)
Afrikaans exposure (%)	21.89 (15.45)
L2 speaking self-rating*	6.15 (1.52)
L2 reading self-rating	6.58 (1.59)
L2 spoken comprehension self-rating	7.4 (1.39)
C-test score (%)	75.3 (9.2)

Note: *Self-ratings are on a scale from 1–10.

Table 4-9: Verb biases for complement- vs. adjunct-clause constructions

Construction	Verb
Complement clause	confessed
	warned
	cautioned
	wrote
	resolved
	believed
	read
	advised
	proved
	confirmed
	suggested
Adjunct clause	sailed
	climbed
	painted
	drank
	played
	baked
	parked
	called
	walked
	polished
	rode
	ate
	flew
	sang

5. CROSS-LANGUAGE ACTIVATION CAUSES FACILITATORY INTERFERENCE IN BILINGUAL PRONOUN RESOLUTION

5.1 Study abstract

This paper investigates whether cross-language activation of verb reflexivity information affects pronoun resolution among bilinguals. In an eye-movement monitoring experiment, we examined English pronoun resolution in a group of Afrikaans–English bilinguals (n = 36). Cross-linguistic differences were exploited by pairing pronouns with two sets of English verbs: one set whose Afrikaans translation equivalents require an accompanying pronoun to be resolved in the English fashion, and another set whose translation equivalents can take as a complement a simplex reflexive that is formally indistinguishable from a pronoun, but takes a local rather than a non-local antecedent. Reading times indicate early facilitatory interference of a structurally inaccessible feature-matching antecedent, but only in sentences where the verb's Afrikaans translation equivalent would license this antecedent. The results suggest that cross-language activation at the word level may influence structurally constrained reference processing.

5.2 Introduction

Referential expressions such as reflexives (e.g., *himself*, *herself*) and pronouns (e.g., *him*, *her*) are widespread in language. Successful interpretation of an utterance containing a referential expression requires a comprehender to determine the antecedent of this expression. One input to the reference resolution process are structural constraints, such as the binding principles (Chomsky, 1981), which determine which noun phrases (NPs) in a sentence are possible antecedents. Binding Principle A, for example, dictates that a reflexive must take an antecedent within its local domain, and that this antecedent must c-command the reflexive, whereas Binding Principle B prohibits a pronoun from referring to a c-commanding noun phrase within its local domain. Another input to reference resolution is verb argument structure, where, for example, the implicit causality biases of verbs may favor either a subject or an object NP as a pronoun antecedent (Garvey & Caramazza, 1974; Garvey, Caramazza, & Yates, 1974).

Within the study of language processing, the timing of these different information sources during real-time comprehension is of central interest. Further, because a general theory of parsing should be able to account for how reference resolution unfolds across both mono- and bilinguals, it is essential to study the use of structural and other cues during antecedent search in speakers of more than one language. In this regard, there is evidence that L2 speakers do not immediately respect the L2 structural constraints on reference resolution during online processing (Felser & Cunnings, 2012; Felser et al., 2009; Kim et al., 2015). Divergence from monolingual reference assignment behavior has also been observed in early bilinguals (Kwon, Cunnings, & Lesmana, 2013; Serratrice, Sorace, & Paoli, 2004) and in L1 attriters

(Tsimpli, Sorace, Heycock, & Filiaci, 2004), suggesting that this non-nativelikeness may relate to the relationship between language systems and not simply to a later age of acquisition.

To date, there has been relatively little investigation into the contributors to bilinguals' divergent reference assignment behavior. One recent finding in this regard is that verb biases in the L1 can influence reference assignment preferences in sentences containing translation equivalents in the L2 (Kim & Grüter, 2019). This finding has been attributed to so-called "lemmatic transfer" (Jarvis, 2009), which is assumed to result from the sharing of syntactic and semantic representations across languages at the lemma level in the lexicon. Effects such as these provide valuable insight into word-level influences on bilingual syntactic processing. However, a number of aspects regarding such cross-linguistic word-level influences on reference resolution remain unclear. Firstly, the existing literature has only examined implicit causality biases, and so it is uncertain whether other types of verb information are also candidates for lemmatic transfer. Secondly, it is unclear how such word-level influences interact with structural constraints on reference resolution, and whether these two information sources act sequentially or in parallel to guide antecedent search during real-time processing. This latter point is of particular relevance in light of recent models of non-native language comprehension (e.g., Clahsen & Felser's 2006a, 2006b, 2006c, 2017 Shallow Structure Hypothesis and Cunnings' 2017a, 2017b cue-based memory retrieval account), which propose that non-native speakers rely more on semantic and discourse-level information than structural information during real-time processing.

The focus of the present study is on an aspect of verb argument structure that has not received previous attention in the literature, namely reflexivity. Unlike implicit causality biases, which seem to be largely consistent across languages (Hartshorne, Sudo, & Uruwashi, 2013), there is “daunting diversity” (Dimitriadis, Everaert, & Reuland, 2017, p. 1) in the cross-linguistic expression of reflexivity, making this domain a fertile testing ground for potential effects of lemmatic transfer. This paper investigates whether cross-language activation of reflexivity information in one of a bilingual’s languages can influence real-time, structurally constrained reference resolution in their other language in such a way that the structural constraints on reference resolution are violated. The findings are relevant to our understanding of the potential effects of cross-language activation at the verb level on processing at the sentence level, and to our understanding of the relative prioritization of structural constraints in bilingual reference resolution.

5.3 Structural constraints on reference resolution

The proposal that the binding principles act immediately during reference resolution to rule out structurally illegal antecedents from consideration was first made by Nicol and Swinney (1989) based on findings from a series of cross-modal priming experiments. These authors used sentences such as those in (1) to determine whether structurally illegal antecedents are activated during the processing of a referential expression.

(1a) The boxer told the skier that the doctor for the team would blame himself for the recent injury.

(1b) The boxer told the skier that the doctor for the team would blame him for the recent injury.

Semantic associates of the candidate NPs were presented at the offset point of the referential expression, and participants had to perform a lexical decision task. For (1a), reduced decision latencies were found for semantic associates of *doctor*, but not for associates of *skier* and *boxer*. The reverse pattern was obtained for (1b), where facilitation was observed for associates of the first two NPs, but not for those of *doctor*. Based on these results, Nicol and Swinney (1989) conclude that when the referential expression was being processed, only the structurally legal antecedent(s) had been reactivated, and diffusion of this activation to semantic neighbours produced the facilitative effect observed in the lexical decision task. Assuming that the presentation of the lexical decision task immediately following the referential expression tapped into early referential processing, the authors propose that the binding constraints apply immediately to restrict antecedent search, such that illegal antecedents are not considered at any point during reference resolution. This is the so-called Binding as Initial Filter (BAIF) hypothesis.

Aside from the binding principles, another established influence on reference resolution are (mis)matches between the morphosyntactic features of a referential expression and those of potential antecedent NPs (see e.g., Carreiras, Garnham, Oakhill, & Cain, 1996; Osterhout & Mobley, 1995). It has been found that when an NP is considered as an antecedent for a referential expression, feature mismatches between these two items cause processing difficulty. Assuming, as per the BAIF, that the binding principles apply immediately to exclude structurally illegal antecedents from consideration, this mismatch effect is predicted to occur only when the candidate antecedent is structurally accessible. A number of studies have fulfilled this prediction (e.g., Chow, Lewis, & Phillips, 2014; Clifton, Frazier, & Deevy, 1999; Clifton, Kennison,

& Albrecht, 1997; Dillon, Mishler, Sloggett, & Phillips, 2013; Xiang, Dillon, & Phillips, 2009).

Subsequent investigations using a feature-mismatch paradigm have however called into question the inviolability of the binding constraints. In particular, the use of eye-tracking, which provides more finely grained data on the time course of referential processing, has made possible more precise estimates of when the binding constraints come into play. Sturt's (2003) eye-tracking results, for example, demonstrated that although Binding Principle A acted to constrain antecedent search during early reading measures, the presence of an inaccessible feature-matching NP caused difficulty during later reading measures. The relevant contrast is between the sentences in (2) (with the ellipses on the left and right replacing the lead-in and wrap-up sentences, respectively):

(2a) … {She/He} remembered that the surgeon had pricked himself with a used syringe needle. …

(2b) … {She/He} remembered that the surgeon had pricked herself with a used syringe needle. …

Early reading time measures in (2a), where the gender of reflexive matched the stereotypical gender of the structurally licensed antecedent (*the surgeon*), were shorter than for (2b), where there was a mismatch in stereotypical gender between these two items. In later measures, reading times on the reflexive in sentences like (2b) were elevated when the reflexive matched the structurally inaccessible antecedent (*She*) compared to when it did not (*He*).

Additionally, in a follow-up offline experiment, Sturt's (2003) participants were found to be especially likely to settle on an ungrammatical interpretation of a reflexive given (i) a mismatch in features between the reflexive and the accessible antecedent and (ii) the availability of a feature-matching inaccessible antecedent. In light of the late effects of the inaccessible antecedent observed in the eye-tracking data, this offline finding was interpreted as the outcome of a recovery strategy initiated during processing wrap-up. Sturt's results contributed to a modification of the BAIF hypothesis, such that the binding constraints were to be understood as 'defeasible' filters, operational during early-stage processing, but no longer constraining antecedent search during later processing stages.

A final perspective on the role of the binding constraints posits that they might rather act as one cue among many that operate from the outset of the antecedent search process (e.g., Badecker & Straub, 2002). Cue-based retrieval theories of memory (e.g., Lewis & Vasishth, 2005; Lewis et al., 2006) propose that from the outset of reference resolution, the parser uses both structural cues (such as the binding constraints) and non-structural cues (such as gender) to evaluate potential antecedent NPs by comparing these NPs' features against those of the referential expression. Support for this model comes, for example, from Patil, Vasishth, and Lewis (2016), where participants showed sensitivity to a gender-matching but structurally inaccessible antecedent during the early stages of processing. Similar patterns are observed in Choy and Thompson (2010), Clackson, Felser and Clahsen (2011), Clackson and Heyer (2014), Cunnings and Felser (2013), Kennison (2003) and Thompson and Choy (2009).

The inconstancy of the results reviewed above has made the question of whether early-stage antecedent search is governed by "strictly syntactic" (Patil et al., 2016) principles a controversial one. Amidst this controversy, interference effects such as those described in the previous two paragraphs, where participants show sensitivity to the features of structurally illegal candidate antecedents, have emerged as a touchstone. Two kinds of interference effects have been distinguished. Inhibitory interference, as defined by Jäger, Engelmann, and Vasishth (2017), is an effect of the availability of multiple feature-matching candidate NPs, which results in a processing slowdown. Facilitatory interference (also termed "intrusion" or "attraction"), on the other hand, accounts for the accelerated processing of referential expressions across sentences that should be equally acceptable or unacceptable (Parker & Phillips, 2017, p. 273). The contrast between these two types of interference can be illustrated with reference to the sentences in (3) (taken from Chow et al., 2014):

(3a) John thought that Bill liked him a lot.

(3b) John thought that Mary liked him a lot.

(3c) Jane thought that Bill liked him a lot.

(3d) Jane thought that Mary liked him a lot.

According to Binding Principle B, only the subject NP in (3) is a legal antecedent for the pronoun *him*. The absence of a feature-matching structurally legitimate antecedent in (3c) and (3d), then, is predicted to lead to processing difficulty.[29] Further, the features of the embedded subject should have no effect on processing times, making (3a) and (3b) equally acceptable and (3c) and (3d) equally unacceptable. An inhibitory interference effect would be reflected in increased reading times in (3a) compared to (3b), where in the former sentence the gender match between the pronoun and the embedded subject may delay identification of the correct antecedent. A facilitatory interference effect, resulting from the gender match between the embedded subject and the pronoun in (3c), would manifest as reduced reading times in this sentence compared to (3d).

It has been argued that inhibitory interference does not constitute evidence against a strictly syntactic antecedent search procedure. Dillon et al. (2013), for example, attribute such effects to encoding interference. Their account relies on an understanding of memory encoding in which already encoded items that possess common features compete for the representations of these features in memory. Consequently, an already encoded item can be degraded if its features are lost to the representation of another item (Bancroft, Jones, Ensor, Hockley, & Servos, 2016; Nairne, 1990). According to this account, inhibitory interference may then result from the features of the structurally accessible subject NP being overwritten by those of the structurally inaccessible NP and does not necessarily indicate that the inaccessible

[29] As Chow et al. (2014) remark, the pronoun can take an antecedent outside of the sentence, but when sentences such as those in (3c) and (3d) are presented out of context, they are usually first considered to be ungrammatical.

antecedent has been retrieved.[30] Facilitatory interference, in contrast, has been taken as clear evidence that structurally illegitimate antecedents are accessed early during referential processing (Dillon, 2011). It has therefore been argued (e.g., Chow et al., 2014; Dillon, 2011; Dillon et al., 2013) that only the occurrence of such interference effects constitutes evidence in favor of structurally unconstrained antecedent retrieval.

Whilst facilitatory interference has occasionally been observed in reflexive resolution (Cunnings & Felser, 2013; King, Andrews, & Wagers, 2012; Parker & Phillips, 2017; Patil et al., 2016), Chow et al. (2014, p. 3) note that "a facilitative interference effect has never been observed for pronouns". The relative paucity of observed facilitatory interference effects constitutes a weak point in the argument for structurally unconstrained antecedent retrieval. Further, the robustness of facilitatory interference effects observed in the resolution of other dependencies (e.g., subject–verb agreement, Dillon et al., 2013; Patson & Husband, 2016; Tanner, Nicol, & Brehm, 2014; Wagers, Lau, & Phillips, 2009) challenges parsing models which posit a uniform mechanism by means of which all linguistic dependencies are resolved (Chow et al., 2014; Jäger et al., 2017). The question of whether facilitatory interference does in fact occur thus has direct bearing on current understandings of the human parsing mechanism.

Of central interest to this paper is the relationship between structural constraints on pronoun resolution and another established input to reference resolution, namely verb

[30] See however the discussion in Patil et al. (2016) regarding the validity of this point; Jäger, Benz, Roeser, Dillon, and Vasishth (2015) also argue against invoking encoding interference in explaining inhibitory interference effects.

argument structure. It has been established that a verb's argument structure acts to restrict the domain of reference during parsing (Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003). One example in this regard is the phenomenon of implicit causality. For example, in (4), where the two sentences differ only in the verbs that are used, the pronoun interpretation preferences differ. In (4a), it is Henry who is judged to be the antecedent of *he*, whereas in (4b), it is John.

(4a) John feared Henry$_i$ because he$_i$...

(4b) John$_i$ frightened Henry because he$_i$...

It is generally accepted that the differing resolution preferences in (4) are the consequence of differing verb argument structures. The idea here is that, given a particular discourse relationship across clauses (e.g., *Explanation* or *Result*, see Kehler, 2002), the verb's argument structure guides antecedent search. Psychological verbs, which show the most robust implicit causality effects (Hartshorne & Snedeker, 2013), can be classified in terms of structure as either experiencer–stimulus or stimulus–experiencer verbs (Brown & Fish, 1983). *Fear* is an experiencer–stimulus verb, where NP1 is the experiencer and NP2 is the stimulus. *Frighten,* in contrast, is a stimulus–experiencer verb, where NP1 is the stimulus and NP2 the experiencer. As the connective *because* signals that the relationship between the two clauses is one of *Explanation*, the antecedent preference is for the NP that is the causer of the action being described, that is, the stimulus, which corresponds to NP2 in (4a) and NP1 in (4b).

A number of studies have shown that speakers make rapid use of information of this nature to guide antecedent search (e.g., Koornneef & Van Berkum, 2006; McDonald & MacWhinney, 1995). In one theoretical account of this process (Kehler, Kertz, Rohde, & Elman, 2008), verb argument structure shapes expectations about how a sentence will continue, and these expectations, which are updated as parsing progresses, shape further processing.

5.4 Reference resolution in L2 speakers

Among other topics, examinations of how reference resolution unfolds in L2 and other bilingual populations have considered (i) whether such individuals are equally sensitive to structural constraints in this domain, and (ii) whether they are capable of deploying verb-level information such as implicit causality to guide antecedent search.

Although a number of studies indicate that L2 speakers pattern with native speakers in the offline application of the binding principles (Bertenshaw, 2009; Cook, 1990; Drummer & Felser, 2018; Felser & Cunnings, 2012; Felser et al., 2009; Matsumura, 1994; Patterson & Felser, 2019; Patterson et al., 2014; Thomas, 1991; White, Hirakawa, & Kawasaki, 1996; White, 1998), the majority of online results suggest that non-native reference resolution is subject to interference from structurally inaccessible antecedents. For example, in reflexive processing, L2 speakers have been found to consider structurally inaccessible antecedents in early reading measures. Felser et al. (2009) observed inhibitory interference amongst their L2 speakers, and Felser and Cunnings' (2012) participants were affected by the presence of an inaccessible antecedent that mismatched the reflexive in gender.[31] Further, in a visual world experiment, Kim et al.'s (2015) lower proficiency L2 speakers were delayed in their application of Binding Principle A, initially considering an inaccessible NP as a possible antecedent.

Although less evidence is available regarding L2 pronoun resolution, findings in this domain also suggest that the binding principles are not immediately applied. Patterson et al. (2014) found similar processing patterns in their L1 and L2 groups, in that both groups appeared to respect Binding Principle B from the earliest stages of processing. However, a follow-up experiment testing the interpretation of short-distance pronouns (e.g., *him* in *Nick saw David put the cat beside him*), which are generally considered exempt from Binding Principle B, showed that the L2 speakers maintained a preference for linking the pronoun to the non-local antecedent. The L1 speakers, in contrast, demonstrated awareness that the local referent could also serve as an antecedent in these conditions. The authors propose that the L2 speakers' behavior is consistent with a strategy in which the pronoun is simply linked to the most discourse-prominent antecedent, which in this case was the non-local referent mentioned first in the sentence.[32] Consistent with the idea that L1 and L2 pronoun resolution may differ

---

[31] Felser et al.'s study (2009) was not set up to test for facilitatory interference effects. There appears to be a numerical trend towards a facilitatory interference effect in Felser and Cunnings (2012), but the authors do not discuss this result.

[32] A subject/first-mention preference in ambiguous pronoun resolution has also been discussed in the monolingual literature, see Crawley, Stevenson, and Kleinman (1990), Gernsbacher, Hargreaves, and Beeman (1989), and Järvikivi, Van Gompel, Hyönä, and Bertram (2005).

are the findings of Kim et al. (2015), where both higher and lower proficiency L2 speakers were found to consider a structurally illegitimate referent as a possible pronoun antecedent during online processing.

Finally, in the processing of cataphora, where so-called Binding Principle C precludes co-reference between a pronoun and an NP that it c-commands[33], interference from structurally illegitimate antecedents has also been observed. In both Drummer and Felser (2018) and Patterson and Felser (2019), reading times were longer in the presence of a gender mismatch between a pronoun and a structurally inaccessible antecedent. Although this finding held for both native and non-native speakers, it is another indication that application of the binding principles may be delayed during non-native processing.

Because the inventories of referential expressions and the constraints on their use vary cross-linguistically, it has been questioned (see e.g., the discussion in Patterson et al. 2014, p. 15) whether L1–L2 differences in online reference resolution might be due to transfer of the L1 binding constraints. However, for structurally constrained reference resolution, there is limited evidence for this position. For example, although Felser et al.'s (2009) L2 group was made up of L1 Japanese speakers, whose native language does allow long-distance reflexive binding, they showed similar behavior to Felser and Cunnings' (2012) L2 group, whose L1 (German) has identical binding constraints to English.

[33] As, for example, in *He got depressed when the lord married the duchess,* where *he* cannot refer to *the lord*.

There is however some support in the literature for the notion that strategies for ambiguous reference resolution in one of a bilingual's languages may influence this process in their other language. This support comes particularly from studies of reference resolution in null-subject languages, where bilingual speakers of a null-subject language and a non-null-subject language diverge from their monolingual peers in their interpretations of overt pronouns. For example, in (5a) (from Carminati, 2002), monolingual Italian speakers interpret *Marta* as the antecedent of the null pronoun, whereas in (5b) the preferred antecedent is *Piera*. In contrast, simultaneous Italian–English bilinguals show a preference for *Marta* as the antecedent of *lei* in (5b) (Serratrice, 2007).

(5a) *Marta scriveva frequentemente a Piera quando Ø era negli Stati Uniti.*

Marta wrote frequently to Piera when was in.the States United

"Marta wrote frequently to Piera when Ø was in the United States."

(5b) *Marta$_i$ scriveva frequentemente a Piera$_j$ quando lei$_{*i/j}$ era negli Stati Uniti.*

Marta wrote frequently to Piera when she was in.the States United

"Marta wrote frequently to Piera when she was in the United States."

The reverse finding, where L1 speakers of a null-subject language interpret pronouns in a non-null-subject L2 as marking a topic shift (e.g., the results for L1 Turkish–L2 Dutch in Roberts et al., 2008), has also been observed.

It thus seems that macro-level strategies for the resolution of ambiguous referential expressions may be subject to cross-linguistic influence. On the word level, recent evidence also indicates that properties of verbs in the L1 can influence reference resolution in L2 sentences containing translation equivalents. The relevant findings are from Kim and Grüter (2019), who examined ambiguous pronoun resolution in L1 Korean–L2 English speakers. Because certain Korean verbs incorporate an explicit causality marker, these authors hypothesized that in L2 English, the translation equivalents of these verbs would yield stronger implicit causality biases than the translation equivalents of verbs without explicit causality markers. Kim and Grüter's (2019) findings supported this hypothesis, leading to the authors' conclusion that word-level transfer of semantico-syntactic properties can influence processing at the sentence and discourse levels. It is this conclusion that underpins the present study.

5.5 Reflexivity and reference resolution in English and Afrikaans

The languages under study in this paper are English and Afrikaans, where the latter is a Germanic language spoken primarily in Southern Africa. This section provides an introduction to reflexivity and reference resolution in these languages.

Anaphoric expressions are standardly divided into pronouns and anaphors, where the former can refer independently to an entity in the world, and the latter can only refer by means of co-indexation with an antecedent in the same sentence. Reinhart and Reuland (1993) distinguish two kinds of anaphors: complex anaphors such as *himself* and *herself*, which they term SELF-anaphors, and simplex anaphors (such as the Dutch *zich* or Norwegian *seg*), which they term SEs (for "simplex expressions"). In some languages, SEs are formally indistinguishable from pronouns. This is the case in Afrikaans.

The experiments reported on in this paper take as a starting point the fact that Afrikaans has both SELF-anaphors and SEs, whilst English has only SELF-anaphors. As a result, in a number of Afrikaans verbal constructions[34], items that are formally indistinguishable from pronouns can serve as reflexives, taking a local rather than a non-local antecedent. The taxonomy of SEs and SELF-anaphors in Afrikaans is presented in Table 5-1, along with their English equivalents.

Table 5-1: Anaphors in Afrikaans

Person	**Number**	**Gender**	**SE**	**SELF-anaphor**	**English equivalent**
1	Singular	–	*my*	*myself*	*myself*
1	Plural	–	*ons*	*onsself*	*ourselves*
2	Singular	–	*jou*	*jouself*	*yourself*
2	Plural	–	*julle/jul*	*julleself/julself*	*yourselves*
3	Singular	Masculine/Neuter	*hom*	*homself*	*himself*
3	Singular	Feminine	*haar*	*haarself*	*herself*
3	Plural	Masculine/Feminine/Neuter	*hulle/hul*	*hulleself/hulself*	*themselves*

Verbs in Afrikaans that can take an SE as their complement include (i) so-called “intrinsically reflexive” verbs, which require an object that is co-indexed with the subject; (ii) verbs of typically self-directed action, which can optionally be used intransitively or with an accompanying reflexive or pronoun; and (iii) a subset of syntactically transitive verbs, which require an object that can either be a reflexive or a pronoun.

[34] “Construction” is used here in an informal sense, in line with Oosthuizen (2013, p. 10), to refer to a collection of phenomena involving the syntactic distribution of a particular kind of linguistic item.

The verbs in (6) (examples from Oosthuizen, 2013, p. 12) are intrinsically reflexive: they require a complement, but this complement can only be an SE or a SELF-anaphor. It is the SE that is standardly used, but the complex reflexive form also occurs, particularly in colloquial language use. In the English equivalents of these sentences, no object is required, as is indicated in the translations.

(6a) *Die vrou$_i$ ontferm haar$_i$ / haarself$_i$ / *haar$_j$ oor die kinders.*

the woman pities her / herself / her over the children

"The woman takes pity on the children."

(6b) *Jan$_i$ het hom$_i$ / homself$_i$ / *hom$_j$ verset teen die aanval.*

Jan has him / himself / him resist against the attack

"Jan resisted the attack."

(6c) *Die seuns$_i$ moet hulle$_i$ / hulself$_i$ / *hulle$_j$ gedra.*

the boys must them / themselves / them behave

"The boys must behave (themselves)."

Verbs of typically self-directed action in Afrikaans can optionally be used intransitively, but can also co-occur with a SELF-anaphor, an SE or a pronoun. Note here that the SE and the pronoun are formally indistinguishable. In the English equivalents of the sentences in (7), a reflexive reading can only be obtained with the SELF-anaphor (*himself*).

(7a) *Jan$_i$ skeer (homself$_i$ / hom$_i$ / hom$_j$) elke oggend.*

John shaves himself / him / him every morning

"John shaves (himself/him) every morning."

(7b) *Jan_i was ($homself_i$ / hom_i / hom_j) elke oggend.*

John washes himself / him / him every morning

"John washes (himself/him) every morning."

(7c) *Jan_i trek ($homself_i$ / hom_i / hom_j) elke oggend aan.*

John pulls himself / him / him every morning on

"John dresses (himself/him) every morning."

Lastly, there are a number of syntactically transitive verbs in Afrikaans that can also take SEs as complements. These verbs differ from those of typically self-directed action in that they must have a direct object. Again, the equivalent constructions in English can only have a reflexive reading if they co-occur with a SELF-anaphor. Examples are provided in (8).

(8a) *Jan_i het $homself_i$ / hom_i / hom_j beseer.*

John has himself / him / him hurt

"John hurt himself/him."

(8b) *$Marie_i$ het $haarself_i$ / $haar_i$ / $haar_j$ na die venster toe gedraai.*

Marie has herself / her / her towards the window to turned

"Marie turned herself/her towards the window."

The other group of verbs of interest in the present paper are those that generally cannot have a reflexive interpretation.[35] Oosthuizen (2013, p. 15) describes these verbs as "inherently non-reflexive", as they can take neither a SELF-anaphor nor an SE, but only a pronoun as a complement. He divides these inherently non-reflexive verbs into two subclasses: one that expresses movement of one entity relative to another (e.g., *vergesel* 'accompany' and *volg* 'follow') and another that expresses the action of one entity on or directed towards another (e.g., *stamp* 'push' and *wink* 'beckon'). Examples are provided in (9) (from Oosthuizen, 2013, pp. 15–16).

(9a) *Jan$_i$ het *homself$_i$ / *hom$_i$ / hom$_j$ net voor die brug verbygesteek.*

John has himself / him / him just before the bridge by-passed

"John overtook him just before the bridge."

(9b) *Jan$_i$ het *homself$_i$ / *hom$_i$ / hom$_j$ huis toe gevolg.*

John has himself / him / him home to followed

"John followed him home."

(9c) *Marie$_i$ het *haarself$_i$ / *haar$_i$ / haar$_j$ omgestamp.*

Marie has herself / her / her over-pushed

"Marie pushed her over."

[35] It is possible for these verbs to be used reflexively under specific circumstances. For example, with respect to (9b), one could imagine a situation where Jan has refused to follow the leader on a group hike, opting instead to follow himself. Such interpretations, however, are dependent on a licensing context and are arguably unlikely to be accessed when the sentence is presented in isolation.

Thus, the class of verbs exemplified in (9) pattern identically in English and Afrikaans, in that they can generally only take an accompanying pronoun (or, of course, another NP) as a complement. In the verbs in (7) and (8), however, there is surface overlap between the Afrikaans pronoun and SE constructions on the one hand and the English pronoun construction on the other, but the binding constraints that apply differ. In both languages, the pronoun must take a non-local antecedent, but in Afrikaans, the SE takes a local antecedent.

5.6 Aims of the present study

In Reinhart and Reuland's (1993) analysis, verbs that can occur with SEs are marked as reflexive in the lexicon, as SEs themselves do not serve as reflexivizers. Adopting such an understanding, the point of interest in this paper is whether, amongst Afrikaans–English bilinguals, reflexivity information linked to verbs in Afrikaans might influence the interpretation of equivalent constructions in English. Such influence may result from lemmatic transfer – caused by the sharing of syntactic and semantic representations across languages at the lemma level – that subsequently shapes higher-level interpretive processes.

The paper reports on two experiments, which test for effects of cross-language activation in English by factorially manipulating the type of the English verbs' Afrikaans translation equivalents in a gender mismatch paradigm similar to that exemplified in (3) (section 5.3). We employ one group of "Potentially Reflexive" verbs, whose Afrikaans equivalents can occur with SEs (the verbs of typically self-directed action *dress*, *shave* and *wash* and the transitive verbs *injure*, *cut* and *turn*), and one group of

"Non-Reflexive" verbs, which in both English and Afrikaans are generally not used reflexively (*push*, *follow*, *overtake*, *accompany*, *lead* and *chase*).

In what follows, we first present the results of an offline experiment designed to test English pronoun resolution preferences in L1 English speakers and Afrikaans–English bilinguals across Potentially Reflexive and Non-Reflexive verbs. To foreshadow the results, both participant groups show a weaker preference for the non-local antecedent in the Potentially Reflexive condition, but the difference between conditions is significantly larger in the bilingual group. Having established an effect of the verb type manipulation on bilinguals' adherence to Binding Principle B, Experiment 2 investigates the time course of this effect during real-time antecedent search using eye-tracking-while-reading.

5.7 Experiment 1

The purpose of Experiment 1 was to examine L1 and bilingual speakers' pronoun antecedent choices across the two verb type conditions in the absence of time pressure.

5.7.1 Participants

The L1 English group consisted of 30 participants, who were recruited via word-of-mouth and social media. Although no data were collected on these participants' L2 knowledge, it is likely that they had some proficiency in Afrikaans, as Afrikaans is a compulsory subject in many schools in the region from which participants were recruited. The bilingual group consisted of 36 Afrikaans–English bilinguals (mean age 20.5 years, range 18–25). They were students at a university in South Africa's Western

Cape province. All bilingual participants were pre-pubescent acquirers of English (mean age of acquisition 2.2 years, SD = 3 years, range 0–10 years). Their English proficiency was assessed by means of a C-test consisting of three texts. In each text, 20 words had been removed, with the first half of the missing letters provided. Participants were tasked with filling in the remainder of each incomplete word. Two of the texts in the C-test were developed by Keijzer (2007) in order to assess L1 attrition in L1 English speakers, and so are relatively difficult. They were therefore deemed appropriate for assessing the L2 speakers in the population under study. The third text was a modified version of a Wikipedia article about a South African university. Scores were high (mean 71.39%, SD = 11.17%) and, compared to the average score (75%) of a group of age-matched South African L1 English speakers (n = 70) on the same task, indicate that the bilingual speakers were of a highly advanced proficiency level. Information on the bilingual participants' language background was collected using the Language History Questionnaire (Li et al., 2014).

5.7.2 Materials

The questionnaire contained 12 sentences testing each type of verb (Non-Reflexive, Potentially Reflexive). In addition to the 24 target sentences, 26 filler sentences were included, which probed the interpretations of referential expressions of various types (unambiguous pronouns, ambiguous pronouns and reflexives). Target and filler sentences were mixed and presented in random order. Examples are given in (10–14).

(10) Jill remembered that Hope had followed **her** to sit behind the speaker.

(Non-reflexive)

(11) Luke claimed that Paul had cut **him** badly during the fencing match.

(Potentially reflexive)

(12) While Mary chatted with Joan, **she** saw a coin on the floor.

(Ambiguous)

(13) Tim thought that Mark owed **him** far more than a hundred rand.

(Unambiguous)

(14) Peter watched as John bought a huge box of popcorn for **himself**.

(Reflexive)

The critical sentences are exemplified in (10) and (11). Each critical verb (six Non-Reflexive, six Potentially Reflexive) was tested in two sentences that differed both in content and in the gender of the names and the pronoun used: one sentence included male names and a male pronoun, and the other female names and a female pronoun.

5.7.3 Procedure

The questionnaire was administered online using Google forms. Participants were instructed to read each sentence and answer the question that followed. The question always asked *Who does [referential expression] refer to?*, and participants had to select their answer from the three options provided – the two names mentioned in the sentence, and an *Either* option. In (10), for example, the options provided were *Jill*, *Hope* and *Either*.

In the presentation of these options, the order of the two names was varied (such that the name that appeared first in the sentence appeared as the first possible answer half of the time), and the *Either* option was always provided last.

5.7.4 Analysis

The data were analyzed in the R environment for statistical computing (version 3.6; R Core Team, 2018). Mixed effects models were computed using the lme4 package (version 1.1.21; Bates et al., 2015), and subsequent pairwise comparisons were conducted using the emmeans package (version 1.3.5; Russell Lenth, 2019). *P*-values for the mixed effects models were obtained using the lmerTest package (version 3.1; Kuznetsova et al., 2017).

5.7.5 Predictions

Binding Principle B allows only for a non-local NP to be an antecedent of a pronoun. As such, if this principle is strictly adhered to, the non-local antecedent in the experimental items should be chosen without exception, regardless of the participants' language background or the verb type manipulation. If, however, the bilinguals are influenced by verb affordances in Afrikaans, they may opt for either the 'Local' or the 'Either' option in the Potentially Reflexive condition, as simplex pronominal forms accompanying the Afrikaans translation equivalents of these verbs can be interpreted either as reflexives (taking a local antecedent) or as pronouns (taking a non-local antecedent).

5.7.6 Results and discussion

The answers of one L1 English participant who answered *Either* to every question, including those following the filler sentences, were discarded, leaving 29 L1 English participants. The bilingual group's mean accuracy score on the unambiguous fillers was 93.2% (SD = 7.9%); the L1 group's mean score was 98.1% (SD = 4.1%).

Figure 5-1 illustrates the distribution of the responses to the experimental items across the two verb type conditions.

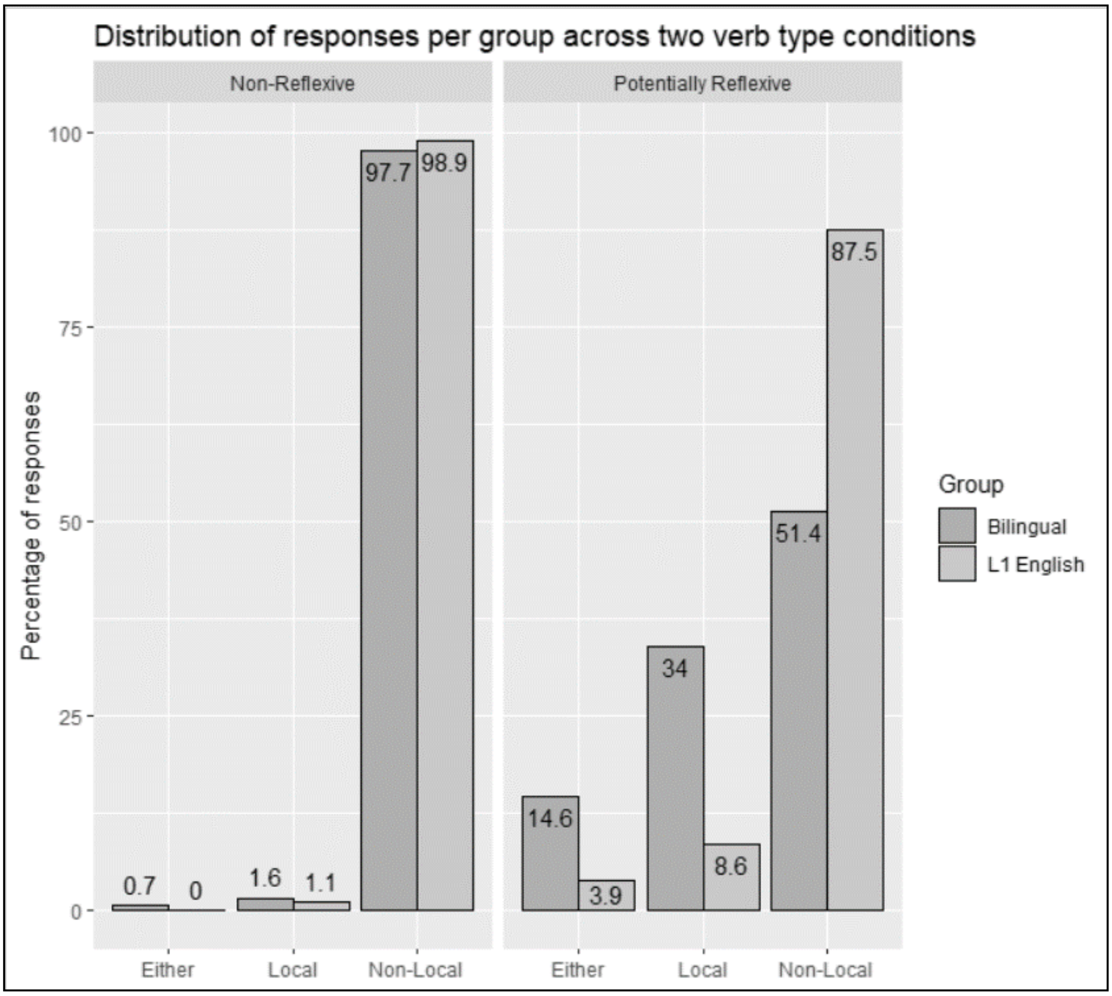

Figure 5-1: Questionnaire responses

Figure 5-1 shows that the L1 and bilingual participants responded similarly to sentences containing Non-Reflexive verbs, or those verbs that behave identically in English and Afrikaans in not allowing for an accompanying pronoun to be reflexively interpreted. However, the groups responded differently to sentences containing Potentially Reflexive verbs.

A generalized linear mixed effects model was fit, with fixed effects of Verb Type (Non-Reflexive or Potentially Reflexive, contrast coded as -1 and 1) and Group (L1 or Bilingual, contrast coded as -1 and 1), random intercepts for participant and item, and by-participants random slopes for Verb Type.[36] The model included a significant effect of Verb Type, such that participants were less accurate in the Potentially Reflexive than the Non-Reflexive condition (β = -1.6, SE = 0.4, p < .0001), and of Group, where the L1 English speakers were more accurate than the bilinguals (β = 1.05, SE = 0.3, p < .0001). There was also a significant Group x Verb Type interaction (β = 0.7, SE = 0.23, p = .002). Subsequent pairwise comparisons indicated that although the L1 English speakers were significantly more accurate in the Potentially Reflexive condition (β = 3.5, SE = 0.6, p < .0001), there was no difference in accuracy across the groups in the Non-Reflexive condition (β = 0.7, SE = 0.7, p = .3).

The responses to the Potentially Reflexive items were analyzed separately to further explore the group differences in response patterns to these items. Model results indicated that the L1 English group chose the non-local antecedent significantly more

[36] A model including both by-participants and by-items slopes did not converge. AIC comparisons indicated that the model with only by-participants slopes was superior to the model with only by-items slopes.

often than the bilingual group (Non-local model, Group: β = 1.8, SE = 0.3, $p < .0001$) and chose the 'Local' and 'Either' options significantly less often (Local model, Group: β = -1.4, SE = 0.3, $p < .0001$; Either model, Group: β = -1.18, SE = 0.4, $p = .003$). When the non-local antecedent was not selected, the L1 English and the bilingual group did not differ significantly in their preference for the 'Local' option over the 'Either' option (L1 Group: β = 0.3, SE = 0.56, $p = .5$).[37]

Overall, although the bilinguals' response patterns are nativelike in the Non-Reflexive condition, they deviate markedly from those of the L1 English speakers in the Potentially Reflexive condition. In this condition, the bilinguals are significantly more likely to violate Binding Principle B. This result aligns with the outcome that was predicted should the bilinguals have been influenced by the reflexivity information of the English verbs' Afrikaans translation equivalents.

The approximately 10% of non-local responses among the L1 English speakers in the Potentially Reflexive condition was not predicted. It is worth noting that other offline studies on L1 pronoun resolution (e.g., Bertenshaw, 2009; White, 1998) have also observed slight deviations from Binding Principle B in native speaker responses. Nonetheless, in light of this unpredicted finding, we checked whether particular aspects of the verbs within this condition might have driven the Principle B violations. For example, three of the six Potentially Reflexive verbs (*dress*, *wash* and *shave*) denote typically self-directed actions. As these verbs can also be interpreted reflexively in their

---

[37] The bilinguals' accuracy in the Potentially Reflexive condition was not correlated with their C-test scores ($r = -0.09$, $p = .56$). As can be expected, given the ceiling performance in the Non-Reflexive condition, there was also no correlation between participants' accuracy scores across verb types (Bilinguals: $r = 0.19$, $p = .27$; L1 English: $r = 0.06$, $p = .76$).

intransitive forms in English, they presumably also have a reflexive lexical entry in this language, which could conceivably influence antecedent choices. However, an examination of the L1 English speakers' responses to the Potentially Reflexive items revealed that self-directedness had no major effect on response patterns. Another possible driver of the L1 English speakers' behavior in the Potentially Reflexive condition is cross-linguistic influence, such that these individuals' responses were influenced by their knowledge of Afrikaans. This is at least plausible, given that the L1 English speakers are likely to have learned Afrikaans at school and were drawn from a population that receives relatively frequent exposure to Afrikaans, and given findings that L1 binding preferences can be affected by the L2 (Kim et al., 2010). However, strong claims cannot be made in this regard, since, as noted above, data on the L1 English speakers' Afrikaans proficiency and exposure were not collected.

The findings of Experiment 1 indicate that the verb type manipulation has a significant effect on bilinguals' adherence to Binding Principle B in English. The significant difference between response patterns across the two groups within the Potentially Reflexive condition suggests, further, that the verb type effect may be due to cross-language activation of the Afrikaans verbs' reflexivity information among the Afrikaans–English bilinguals. Experiment 1 leaves unaddressed the question of when this verb reflexivity information is accessed during real-time antecedent search and, more specifically, whether it is accessed before, after or in parallel to Binding Principle B. Thus, Experiment 2 made use of eye-tracking-while-reading to examine the interplay between verb-level and structural information during real-time pronoun resolution.

5.8 Experiment 2

Experiment 2 set out to answer two questions:

(i) To what extent is there evidence for cross-language activation of Afrikaans verb reflexivity information during online English pronoun resolution among Afrikaans–English bilinguals?

(ii) When, relative to Binding Principle B, do verb-level influences on antecedent search apply?

5.8.1 Participants

The same 36 Afrikaans–English bilinguals who participated in Experiment 1 first completed the eye-tracking experiment reported on below. All participants had normal or corrected-to-normal vision and were paid for their participation in the experiment.

5.8.2 Materials

Twenty-four experimental items were constructed. These were based on the items used in Patterson et al. (2014). Each experimental item consisted of a lead-in sentence, a critical sentence that included a pronoun and two potential antecedent NPs that were both proper names, and a wrap-up sentence. The gender match between the pronoun and the proper names in the critical sentence was manipulated (‘Match’ factor, levels Double Match, Local Mismatch and Non-Local Mismatch). Further, we manipulated the type of verb used in the critical sentence (‘Verb Type’ factor, levels Non-Reflexive and Potentially Reflexive). Examples are provided in (15) and (16) below.

(15) Non-Reflexive verb

Lead-in sentence: The presentation was being set up at the front of the room.

Double-Match Condition:

Jill remembered that Elle had followed her to sit at the back.

Local Mismatch Condition:

Jill remembered that John had followed her to sit at the back.

Non-Local Mismatch Condition:

John remembered that Jill had followed her to sit at the back.

Wrap-up sentence: They wanted to leave before the question session.

(16) Potentially Reflexive verb

Lead-in sentence: The hike up the mountain did not go well at all.

Double-Match Condition:

John claimed that Mark had injured him quite badly during the climb.

Local Mismatch Condition:

John claimed that Jane had injured him quite badly during the climb.

Non-Local Mismatch Condition:

Jane claimed that John had injured him quite badly during the climb.

Wrap-up sentence: Luckily there were no fatalities.

Six Non-Reflexive (*push*, *follow*, *overtake*, *accompany*, *lead* and *chase*) and six Potentially Reflexive verbs (*injure*, *dress*, *shave*, *turn*, *wash* and *cut*) were used. All critical sentences consisted of 12 words. The proper names in the critical sentences all consisted of four letters and one syllable, and all were either typically masculine or typically feminine (i.e., names used for both genders were avoided). The names were distributed equally across the conditions to control for effects of frequency.

In terms of structure, the first proper name in each critical sentence is the only accessible antecedent according to Binding Principle B, as it is outside the local domain of the pronoun. The second proper name was always the subject of an embedded complement clause, and as a co-argument of the pronoun was ruled out as an antecedent by Binding Principle B. The pronouns were always object pronouns, and an equal number of masculine and feminine pronouns were used.

The experimental items were divided across three lists in a Latin-square design and mixed and pseudo-randomized with 48 filler items. In eight of the filler items, the critical sentences contained reflexive anaphors (*himself*, *herself*) instead of pronouns. Eight further items included a pronoun that was the object of the main clause verb. As in Patterson et al. (2014), this was done to ensure that participants were exposed to items that were similar to the experimental items but differed in the factor under investigation (i.e., the type of referring expression and the position of the possible antecedent). The inclusion of such fillers should have helped to prevent participants from developing expectations about the type of referring expression that was under investigation. The remainder of the filler items included critical sentences that were similar to those in the experimental sentences, but included either two proper names or two proper names and a subject pronoun. To ensure that participants were paying attention, binary (yes/no) comprehension questions were presented after all the experimental items and 28 of the filler items (thus 52 out of the 72 items, 72.2%). Some of the questions following the filler items directly probed the interpretation of a referring expression in order to encourage participants to process these fully. The questions following the experimental items did not target pronoun interpretations, focusing rather on the scenario sketched as a background to the critical sentences. At the beginning of the experiment, participants were presented with six practice sentences to familiarize them with the procedure. Two of these practice sentences were followed by a comprehension question.

5.8.3 Procedure

Participants were tested in a quiet room using an Eyelink 1000 Plus eye-tracker, set to record monocularly (the right eye, barring tracking difficulties) at 1,000 Hz. The desktop setup was used, such that the Eyelink camera was positioned below a computer monitor of 1920 x 1080 resolution. The camera was 65 cm away from the participant, and the computer monitor 82 cm away. Participants rested their chin on a chin mount during the experiment. The chin mount was fixed to a height-adjustable table that could be electronically adjusted to allow each participant to sit comfortably.

Each experimental session began with calibration of the eye-tracker using a nine-point grid. The experiment was divided into three blocks, between which participants could opt to take a break. Calibration was performed again at the beginning of each new block.

All items were presented in Arial (font size 18) across three lines (one sentence per line), in black on a white background. Each trial began with a screen that was blank other than the presence of a fixation cross in the position where the first word of the first sentence of the item would appear. The experiment administrator performed drift correction at the beginning of each trial, such that the text of an item only appeared once the participant had fixated on the cross on the screen. Participants were instructed to read silently at their normal pace, and to push any key on a keyboard to continue to the next screen (which was sometimes a comprehension question relating to the previous item, and sometimes a blank screen with a fixation cross preceding the following item). Comprehension questions were answered using red and green keys on a keyboard (red for ‘no’, green for ‘yes’).

The eye-tracking experiment took 30 minutes on average. An entire testing session, including completion of the Language History Questionnaire (Li et al., 2014), the C-test and the offline task reported on in Experiment 1, lasted approximately 80 minutes.

5.8.4 Analysis

The reading times for four regions of the critical sentences are reported: the pronoun region, which includes the pronoun and the last three letters of the preceding word; the spillover region, which includes the two words following the pronoun; the prefinal region, which includes the two words after the spillover region; and the final region, which consists of the final word of the sentence. The pronoun region also includes part of the preceding word because short words are frequently skipped during reading, and expanding the region of interest around the pronoun made it more likely that fixations in this region would be captured.

All reading-time measures were log-transformed for the analyses, as is standard practice in reading-time studies (see e.g., Vasishth & Nicenboim, 2016). Six reading time measures are reported. First fixation duration is the duration of the participant's first fixation within a region of interest. First-pass reading time is the total duration of fixations made within a region of interest before that region is exited for the first time, either to the left or the right. Regression path duration is the total duration of fixations made within a region of interest before that region is exited to the right. Selective regression path duration only includes regressions made within the region of interest, thus excluding regressions to earlier regions of the text. Re-reading time is calculated by subtracting first-pass reading time from total reading time. Finally, total reading time is the sum of all fixations made within a region of interest. The paper adheres to the literature (see e.g., Sturt, 2003) in assuming that early and late measures in the eye-tracking data reflect processes that occur during the early and late stages of processing, respectively.

Prior to analysis, short fixations of 80 ms or less located within one degree of visual arc from another fixation were automatically merged with this longer fixation. Other extremely short (≤ 80 ms) or extremely long (> 1200 ms) fixations were deleted. Reading times for regions that were initially skipped were treated as missing data.

Analyses were conducted in the R environment for statistical computing (version 3.6; R Core Team, 2018). We ran linear mixed effects models using the lme4 package (version 1.1.21; Bates et al., 2015). *P*-values for these models were obtained using the lmerTest package (version 3.1; Kuznetsova et al., 2017). The emmeans package (version 1.3.5; Russell Lenth, 2019) was used to conduct pairwise comparisons.

5.8.5 Predictions

The theory on anaphora resolution makes provision for three predictions regarding the effect of the Match manipulation:

Binding as Initial Filter (BAIF) Hypothesis (Nicol & Swinney, 1989): This hypothesis posits that Binding Principle B will act immediately to rule out inaccessible antecedents from consideration. From this perspective, reading times should be longest in the Non-Local Mismatch condition, as there is no structurally accessible gender-matching antecedent available. Reading times in the other two conditions should be equivalent, as the presence of an inaccessible, gender-mismatching antecedent in the Local Mismatch condition should not affect processing. Importantly, the BAIF Hypothesis predicts that the differences between conditions will remain consistent throughout processing.

Defeasible Filter Hypothesis (Sturt, 2003): The Defeasible Filter Hypothesis puts forward that although the binding principles act immediately to rule out structurally inaccessible antecedents, these antecedents may be considered during a later stage of processing. This hypothesis, like the BAIF Hypothesis, also predicts longer reading times in the Non-Local Mismatch condition. However, in contrast to the BAIF Hypothesis, the Defeasible Filter Hypothesis allows for effects of the Local Mismatch to surface during later reading-time measures, such that processing difficulty may also be experienced in this condition. Moreover, it allows for a competition effect to surface in the Double Match condition, also leading to elevated reading times during later processing stages.

Cue-Based Retrieval Hypothesis (Badecker & Straub, 2002; Patil et al., 2016): According to the Cue-Based Retrieval Hypothesis, all possible antecedents are considered from the onset of processing, with the features of each being compared against those of the pronoun. This account, unlike the two reviewed above, thus predicts interference from a structurally illegitimate antecedent to occur from the earliest processing stages. Reading times are expected to be elevated in the Double Match condition, where there are two feature-matching NPs, compared to the other two conditions, where there is only one feature-matching antecedent.

A second set of predictions relates to the potential influence of the Verb Type manipulation. If this manipulation does have an effect, the following can be predicted to occur.

Firstly, processing difficulty may be attenuated in the Non-Local Mismatch condition with Potentially Reflexive verbs, as participants may settle on the linearly nearest NP as a possible antecedent for the pronoun (cf. the recovery strategy discussed in relation to Sturt, 2003). Shorter reading times in the Non-Local Mismatch condition for Potentially Reflexive items compared to Non-Reflexive items would indicate a facilitatory interference effect driven by the Verb Type manipulation.

Secondly, processing difficulty may be heightened in the Local Mismatch condition with Potentially Reflexive verbs, as participants may be more prone to consider the linearly nearest NP as a possible antecedent for the pronoun, and the feature mismatch between the pronoun and this NP may result in elevated reading times. This too would be a facilitatory interference effect, as per Parker and Phillips' (2017) understanding of facilitatory interference as marked by differences in processing ease across equally acceptable sentences.

Thirdly, processing difficulty may also be heightened in the Double Match condition with Potentially Reflexive verbs, as increased consideration of the linearly nearest NP may lead to increased competition effects, and thus increased inhibitory interference.

Taken together, these effects may reflect in a difference in reading times between the Non-Local Mismatch condition and the other two conditions for Non-Reflexive verbs, but similar reading times across all three Match conditions for Potentially Reflexive verbs. If the Verb Type manipulation acts on antecedent search before or in parallel to Binding Principle B, as per the Cue-Based Retrieval Hypothesis, these effects should be visible during early-stage reading measures. However, if the binding constraints act to rule out structurally inaccessible antecedents from consideration, the verb type effects may either surface in later processing measures (as per the Defeasible Filter Hypothesis) or not at all (as per the BAIF Hypothesis).

5.8.6 Results

Participants answered 91.2% of the comprehension questions correctly, indicating that they were paying attention during the task. Skipping rates for the four reported regions were as follows: pronoun region 22.8%, spillover region 8%, prefinal region 5% and final region 26.5%. Summaries of participants’ reading times are provided in Table 5-2.

Table 5-2: Mean reading times for six measures at four sentence regions of the experimental items (standard deviations in parentheses)

	NON-REFLEXIVE CONDITION						POTENTIALLY REFLEXIVE CONDITION					
REGION	**FIRST FIXATION DURATION**	**FIRST-PASS READING TIME**	**REGRESSION PATH DURATION**	**SELECTIVE REGRESSION PATH DURATION**	**RE-READING TIME**	**TOTAL READING TIME**	**FIRST FIXATION DURATION**	**FIRST-PASS READING TIME**	**REGRESSION PATH DURATION**	**SELECTIVE REGRESSION PATH DURATION**	**RE-READING TIME**	**TOTAL READING TIME**
PRONOUN												
Double match	211.54 (80.9)	277.03 (131)	340.85 (227.7)	305.8 (170)	379.57 (324.4)	463.3 (313.6)	221.33 (80.3)	292.58 (178.9)	329.67 (213.3)	307.68 (186.9)	441.54 (249.2)	525.62 (339.8)
Local mismatch	201.49 (61.3)	280.67 (180.5)	437.36 (425.8)	329.85 (194.5)	413.07 (316.8)	500.23 (347)	236.6 (110.7)	323.38 (184.7)	444.23 (445.9)	357.64 (208.6)	499.73 (461.6)	635.72 (474.9)
Non-local mismatch	235.67 (100.3)	296.62 (159.8)	348.42 (225.9)	314.54 (167.3)	473.59 (313.3)	591.02 (388.1)	215.88 (70)	286.28 (171.8)	417.74 (506.1)	317.36 (182.5)	444.35 (267.6)	563.5 (355.9)
SPILLOVER												
Double match	229.79 (84.7)	320.51 (189.6)	380.88 (244.7)	347.24 (197.3)	377.64 (270.2)	506.56 (342.2)	236.55 (76.4)	357.55 (205.2)	603.55 (540)	457.18 (295)	424.8 (319.5)	598.92 (383.8)
Local mismatch	228.72 (86.5)	296.1 (149.7)	415.14 (282.8)	345.77 (175.5)	439.21 (300.9)	533.42 (392.8)	222.22 (83.7)	347.65 (235.2)	595.07 (570.6)	463.49 (328.8)	544.96 (442.9)	675.42 (521.9)
Non-local mismatch	234.78 (88.9)	337.63 (202.6)	555.69 (491.7)	403.49 (244.9)	472.34 (365.6)	667.17 (459.9)	232.74 (104.6)	358.3 (237.7)	562.38 (438.5)	447.41 (279.3)	482.68 (354.8)	649.33 (437.5)
PREFINAL												
Double match	228.02 (73.9)	342.6 (205.9)	586.65 (637.9)	394.6 (237.2)	403.34 (349.1)	505.1 (355.6)	227.65 (91.2)	379.58 (218.8)	626.68 (618.4)	448.79 (284.8)	427.81 (326.8)	560.21 (406.2)
Local mismatch	222.5 (66.5)	321.91 (160.7)	528.23 (442.7)	400.07 (256.4)	444.52 (303.5)	550.81 (342.9)	222.29 (85.7)	331.62 (179.5)	568.66 (573.7)	383.94 (214.7)	444.3 (407.5)	544.11 (417.5)
Non-local mismatch	209.16 (69.4)	364.23 (223.9)	752.31 (805.1)	454.53 (265.5)	445.84 (357.3)	627.24 (425.9)	221.03 (110.1)	347.4 (199.9)	699.72 (738.5)	443.07 (276.4)	391.45 (308)	532.98 (345.5)
FINAL												
Double match	208.35 (96.2)	245.12 (137.9)	501.46 (501.8)	261.89 (141.3)	220.47 (149.2)	287.87 (189.9)	218.81 (96.2)	269.97 (158.2)	621.92 (753)	311.97 (193.2)	350.65 (266.2)	336.76 (221)
Local mismatch	209.94 (92.4)	252.15 (140.9)	671.47 (756.1)	314.81 (218.5)	332.87 (268.8)	347.26 (254.4)	205.13 (74.4)	269.53 (184.9)	748.99 (874.2)	336.2 (335.7)	379.94 (501.9)	380.49 (393.4)
Non-local mismatch	203.25 (71.7)	261.76 (148.9)	878.58 (1196.8)	304.45 (188.9)	326.73 (217.9)	345.87 (246.1)	198.02 (119.5)	241.12 (175.3)	471.04 (480.2)	257.82 (184.4)	204.67 (113.4)	279.15 (212.1)

We examined first fixation times, first-pass reading times, and regression path times in the precritical region to check for effects of the two experimental manipulations on reading times prior to the regions of interest. The precritical region was defined as the verb that occurred before the pronoun (minus the verb's last three letters, which were included in the pronoun region) and the preceding auxiliary verb (either *had* or *was*). The skipping rate for this precritical region was 39%. No effects of the two experimental manipulations were observed for any of the reading time measures in this region (see Table 5-8 in the appendix for the model results). This indicates that the early regions of the critical sentences were read similarly regardless of the experimental manipulations.

In the analysis that follows, log-transformed reading times for each region were entered as the dependent variable in a linear mixed effects model, with the fixed effects of Match (Double Match, Local Mismatch or Non-Local Mismatch) and Verb Type (Non-Reflexive or Potentially Reflexive) and random intercepts for Participant and Item.[38] Model outputs are provided in Table 5-3. For all regions where interactions between the fixed effects were observed, the data for each verb type were analyzed separately. The results of these models are presented in separate tables for each sentence region. The results of pairwise comparisons are reported within parentheses in the text.

[38] The majority of models would not converge with any random slopes. As such, no random slopes have been included here.

Table 5-3: Results of the linear mixed models for the pronoun, spillover, prefinal and final regions for Experiment 2

	PRONOUN REGION	SPILLOVER REGION	PREFINAL REGION	FINAL REGION
FIRST FIXATION DURATION				
Intercept	7.70 (0.04)***	5.38 (0.03)***	5.34 (0.03)***	5.23 (0.03)***
DM vs NLM	-0.05 (0.04)	0.01 (0.03)	0.06 (0.03)*	0.05 (0.03)
LM vs NLM	-0.06 (0.04)	-0.02 (0.03)	0.05 (0.03)ₒ	0.04 (0.03)
Verb Type	-0.04 (0.05)	0.01 (0.03)	0.00 (0.03)	-0.01 (0.04)
DM vs NLM x Verb Type	-0.15 (0.09)ₒ	-0.07 (0.05)	0.05 (0.05)	-0.13 (0.07)ₒ
LM vs NLM x Verb Type	-0.25 (0.08)**	0.01 (0.05)	0.05 (0.05)	-0.07 (0.07)
FIRST-PASS READING TIME				
Intercept	5.53 (0.04)***	5.67 (0.05)***	5.70 (0.05)***	5.35 (0.06)***
DM vs NLM	-0.01 (0.04)	-0.01 (0.04)	0.00 (0.04)	0.02 (0.04)
LM vs NLM	0.03 (0.04)	-0.06 (0.04)	-0.05 (0.04)	0.04 (0.04)
Verb Type	-0.04 (0.06)	-0.08 (0.05)	-0.02 (0.08)	-0.02 (0.07)
DM vs NLM x Verb Type	-0.06 (0.08)	-0.07 (0.08)	-0.14 (0.08)ₒ	-0.21 (0.09)*
LM vs NLM x Verb Type	-0.17 (0.08)*	-0.05 (0.08)	-0.04 (0.08)	-0.16 (0.08)ₒ
REGRESSION PATH DURATION				
Intercept	5.71 (0.06)***	5.98 (0.07)***	6.12 (0.07)***	5.97 (0.09)***
DM vs NLM	-0.06 (0.05)	-0.11 (0.05)*	-0.15 (0.06)**	-0.10 (0.08)
LM vs NLM	0.09 (0.05)ₒ	-0.09 (0.05)ₒ	-0.20 (0.06)***	0.11 (0.08)
Verb Type	-0.03 (0.07)	-0.20 (0.10)*	-0.01 (0.13)	0.09 (0.11)
DM vs NLM x Verb Type	0.07 (0.11)	-0.30 (0.10)**	-0.14 (0.11)	-0.58 (0.16)***
LM vs NLM x Verb Type	0.05 (0.10)	-0.25 (0.10)*	-0.07 (0.11)	-0.42 (0.15)**
SELECTIVE REGRESSION PATH DURATION				
Intercept	5.61 (0.05)***	5.84 (0.06)***	5.87 (0.06)***	5.45 (0.07)***
DM vs NLM	-0.03 (0.04)	-0.05 (0.04)	-0.07 (0.04)ₒ	0.02 (0.05)
LM vs NLM	0.07 (0.04)ₒ	-0.04 (0.04)	-0.11 (0.04)**	0.09 (0.05)*
Verb Type	-0.03 (0.07)	-0.18 (0.08)*	-0.02 (0.11)	-0.02 (0.09)
DM vs NLM x Verb Type	0.00 (0.09)	-0.14 (0.08)ₒ	-0.15 (0.08)ₒ	-0.32 (0.09)***
LM vs NLM x Verb Type	-0.05 (0.08)	-0.14 (0.08)ₒ	0.01 (0.08)	-0.18 (0.09)*
RE-READING TIME				
Intercept	5.83 (0.05)***	5.82 (0.07)***	5.73 (0.08)***	5.39 (0.07)***
DM vs NLM	-0.13 (0.08)	-0.18 (0.07)*	-0.05 (0.07)	0.05 (0.13)
LM vs NLM	-0.06 (0.08)	0.03 (0.07)	0.03 (0.07)	0.20 (0.11)ₒ
Verb Type	-0.14 (0.07)*	-0.08 (0.09)	0.01 (0.10)	-0.00 (0.11)
DM vs NLM x Verb Type	-0.24 (0.16)	-0.13 (0.14)	-0.17 (0.14)	-0.76 (0.26)**
LM vs NLM x Verb Type	-0.14 (0.16)	-0.17 (0.14)	-0.08 (0.14)	-0.52 (0.23)*
TOTAL READING TIME				
Intercept	6.08 (0.07)***	6.18 (0.07)***	6.11 (0.07)***	5.50 (0.08)***
DM vs NLM	-0.14 (0.05)**	-0.15 (0.04)***	-0.11 (0.04)*	0.01 (0.05)
LM vs NLM	-0.01 (0.05)	-0.11 (0.04)*	-0.05 (0.04)	0.12 (0.05)*
Verb Type	-0.13 (0.09)	-0.13 (0.09)	0.03 (0.10)	-0.03 (0.09)
DM vs NLM x Verb Type	-0.12 (0.10)	-0.22 (0.09)*	-0.22 (0.09)*	-0.35 (0.10)***
LM vs NLM x Verb Type	-0.21 (0.10)*	-0.27 (0.09)**	-0.08 (0.09)	-0.23 (0.09)*
***p < 0.001, **p < 0.01, *p < 0.05, ₒp < 0.1				

Note: “DM” stands for Double Match; “LM” for Local Mismatch; “NLM” for Non-Local Mismatch.

Pronoun region

First fixation times differed significantly across the verb types only for the Local Mismatch condition, where fixations were longer for the Potentially Reflexive than the Non-Reflexive items (β = 0.16, SE = 0.07, *p* = .02). There was also a significant interaction between the Match and Verb Type factors for this measure, and so the verb types were analyzed separately. For **Non-Reflexive verbs**, first fixation times were significantly lower in the Double Match and Local Mismatch conditions compared to the Non-Local Mismatch condition. For **Reflexive verbs,** no difference in first fixation times across the Match conditions is observed (Table 5-4).

For regression path duration, a significant effect of Match is observed, in which regression path times are higher for the Local Mismatch condition than the Double Match condition across verb types (β = 0.15, SE = 0.05, *p* = .01). The same effect is evident in selective regression path durations (β = 0.1, SE = 0.04, *p* = .04).

For re-reading times, there is an effect of Verb Type, such that participants re-read less in the Non-Reflexive than in the Potentially Reflexive condition. Finally, total reading times in the Local Mismatch condition were again significantly longer for Potentially Reflexive items (β = 0.23, SE = 0.1, *p* = .04). An interaction between Match and Verb Type was also present at this region. For the **Non-Reflexive verbs**, total reading times were significantly lower in the Double Match than in the Non-Local Mismatch condition, and marginally lower in the Local Mismatch compared to the Non-Local Mismatch condition. For the **Reflexive verbs**, there is a significant difference between the Double Match and the Local Mismatch conditions, where total reading times were longer in the latter (β = 0.18, SE = 0.07, *p* = .02) (Table 5-4).

Table 5-4: Separate models for the two verb types at the pronoun region

	FIRST FIXATION TIMES		TOTAL READING TIMES	
	NREF β (SE)	**REF β (SE)**	**NREF β (SE)**	**REF β (SE)**
Intercept	7.67 (0.04)***	7.72 (0.05)***	6.02 (0.08)***	6.14 (0.09)***
DM vs NLM	-0.14 (0.06)*	0.02 (0.06)	-0.20 (0.07)**	-0.08 (0.07)
LM vs NLM	-0.18 (0.06)**	0.07 (0.06)	-0.12 (0.07)°	0.10 (0.07)
***p < 0.001, **p < 0.01, *p < 0.05, °p < 0.1				

Spillover region

No significant effects of Match or Verb Type are observed in the first fixation times or first-pass reading times at the spillover region. Regression path duration is higher for Potentially Reflexive compared to Non-Reflexive items in both the Double Match and the Local Mismatch condition (Double Match: β = -0.3, SE = 0.1, *p* = .007; Local Mismatch: β = -0.27, SE = 0.1, *p* = .02). For this measure, there is also a significant interaction between Match and Verb Type. For the **Non-Reflexive verbs**, reading times are significantly higher in the Non-Local Mismatch condition than in both the Double Match and Local Mismatch conditions. For **Reflexive verbs**, no significant differences across the Match conditions are observed (Table 5-5). Effects are similar for selective regression path duration, where for **Non-Reflexive verbs**, reading times are higher in the Non-Local Mismatch condition than in the Double Match condition and marginally higher in the Non-Local Mismatch than the Local Mismatch condition, and no effects of Match are observed for **Reflexive verbs** (Table 5-5).

In re-reading times, an effect of Match is observed, such that re-reading times in the Double Match condition are significantly lower than in the Local and the Non-Local Mismatch conditions for both verb types (Double Match vs Local Mismatch: β = -0.2, SE = 0.07, *p* = .01; Double Match vs Non-Local Mismatch: β = -0.18, SE = 0.07, *p* =

.03). Finally, total reading times are again longer in the Double Match and Local Mismatch conditions for Potentially Reflexive compared to Non-Reflexive verbs (Double Match: β = -0.18, SE = 0.1, p = .08; Local Mismatch: β = -0.23, SE = 0.1, p = .03), and there is again an interaction between Match and Verb Type for this measure. For the **Non-Reflexive verbs**, total reading times are significantly longer in the Non-Local Mismatch than in both the Double Match and Local Mismatch conditions. For **Reflexive verbs**, there are no significant differences across the Match conditions (Table 5-5).

Table 5-5: Separate models for the two verb types at the spillover region

	REGRESSION PATH DURATION		**SELECTIVE REGRESSION PATH DURATION**		**TOTAL READING TIMES**	
	NREF β (SE)	**REF β (SE)**	**NREF β (SE)**	**REF β (SE)**	**NREF β (SE)**	**REF β (SE)**
Intercept	5.89 (0.05)***	6.08 (0.10)***	5.76 (0.05)***	5.93 (0.09)***	6.12 (0.08)***	6.25 (0.09)***
DM vs NLM	-0.27 (0.07)***	0.03 (0.07)	-0.12 (0.06)*	0.01 (0.06)	-0.26 (0.06)***	-0.05 (0.06)
LM vs NLM	-0.20 (0.07)**	0.04 (0.07)	-0.11 (0.06)$^{\circ}$	0.03 (0.06)	-0.23 (0.07)***	0.03 (0.06)
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\circ}p < 0.1$						

Prefinal region

First fixation times at the prefinal region revealed an effect of Match, such that participants were slower in the Double Match condition than in the Non-Local Mismatch condition and marginally slower in the Local Mismatch compared to the Non-Local Mismatch condition. No significant effects are found for first-pass reading times.

Regression path durations were longer in the Non-Local Mismatch condition than in both the Double Match and the Local Mismatch conditions. For selective regression path durations, reading times are significantly lower in the Local Mismatch than in the Non-Local Mismatch condition, and marginally lower in the Double Match compared to the Non-Local Mismatch condition.

There were no significant effects of the experimental manipulations on re-reading times. Finally, total reading times were significantly lower in the Double Match compared to the Non-Local Mismatch condition. An interaction between Match and Verb Type is also observed for this measure. For the **Non-Reflexive verbs**, total reading times are significantly lower in the Double Match compared to the Non-Local Mismatch condition. For the **Reflexive verbs**, no significant differences between the Match conditions are observed (Table 5-6).

Table 5-6: Separate models for the two verb types at the prefinal region

	TOTAL READING TIMES	
	NREF **β (SE)**	**REF** **β (SE)**
Intercept	6.13 (0.09)***	6.10 (0.09)***
DM vs NLM	-0.22 (0.06)***	0.01 (0.06)
LM vs NLM	-0.09 (0.06)	-0.01 (0.06)
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\circ}p < 0.1$		

Final region

An interaction between the Match and Verb Type factors is observed for first fixation times at the final region. For the **Non-Reflexive verbs**, no significant differences between the Match conditions are observed. For the **Reflexive** verbs, first fixations are significantly longer in the Double Match than in the Non-Local Mismatch condition (Table 5-7).

For first-pass reading times, there is again a significant interaction between Match and Verb Type. For the **Non-Reflexive verbs**, no significant differences between the Match conditions are observed. For the **Reflexive** verbs, first-pass reading times are longer in the Double Match and Local Mismatch conditions compared to the Non-Local Mismatch condition (Table 5-7).

Regression path duration in the Non-Local Mismatch condition was significantly shorter for Potentially Reflexive items than for Non-Reflexive items ($\beta = -0.4$, SE = 0.14, $p = .003$). The model for this measure also includes a significant Match x Verb Type interaction. For the **Non-Reflexive verbs**, regression path duration is significantly lower in the Double Match condition than in the other two conditions (Double Match vs Local Mismatch: $\beta = -0.29$, SE = 0.1, $p = .02$). For the **Reflexive verbs**, regression path durations were significantly higher in the Local Mismatch than the Non-Local Mismatch condition and marginally higher in the Double Match than in the Non-Local Mismatch condition (Table 5-7). Similar results were found for selective regression path duration, where for **Non-Reflexive verbs**, reading times were shorter in the Double Match condition than in the Non-Local Mismatch condition, and for **Reflexive verbs**, reading times were shorter in the Non-Local Mismatch condition than in the other two conditions (Table 5-7).

A Match x Verb Type interaction is also present for re-reading times. For **Reflexive verbs**, re-reading times in the Non-Local mismatch condition are significantly shorter than those in the Double Match and Local Mismatch conditions. For **Non-Reflexive verbs**, re-reading times were significantly lower in the Double Match compared to the Non-Local Mismatch condition (Table 5-7).

Finally, total reading times were marginally longer in the Double Match condition when the pronoun followed a Potentially Reflexive verb (β = -0.19, SE = 0.1, p = .08), and there was a Match x Verb Type interaction for this measure. For **Non-Reflexive verbs**, total reading times are significantly shorter in the Double Match condition than in the Non-Local Mismatch condition. For **Reflexive verbs**, reading times are significantly shorter in the Non-Local Mismatch condition compared to the other two conditions (Table 5-7).

Table 5-7: Separate models for the two verb types at the final region

	FIRST FIXATION TIMES		**FIRST-PASS READING TIMES**		**REGRESSION PATH DURATION**		**SELECTIVE REGRESSION PATH DURATION**		**RE-READING TIMES**		**TOTAL READING TIMES**	
	NREF β (SE)	**REF β (SE)**	**NREF β (SE)**	**REF β (SE)**	**NREF β (SE)**	**REF β (SE)**	**NREF β (SE)**	**REF β (SE)**	**NREF β (SE)**	**REF β (SE)**	**NREF β (SE)**	**REF β (SE)**
Int.	7.56 (0.05)***	7.56 (0.05)***	5.36 (0.06)***	5.36 (0.07)***	6.04 (0.10)***	5.91 (0.11)***	5.45 (0.08)***	5.46 (0.08)***	5.45 (0.08)***	5.39 (0.09)***	5.51 (0.08)***	5.52 (0.09)***
DM vs NLM	-0.02 (0.07)	0.17 (0.07)*	-0.08 (0.06)	0.12 (0.06)*	-0.37 (0.11)**	0.19 (0.11)°	-0.15 (0.07)*	0.18 (0.07)**	-0.36 (0.18)*	0.42 (0.18)*	-0.16 (0.07)*	0.19 (0.06)**
LM vs NLM	0.01 (0.07)	0.11 (0.07)	-0.04 (0.06)	0.12 (0.06)*	-0.08 (0.11)	0.31 (0.10)**	0.01 (0.07)	0.19 (0.06)**	-0.04 (0.16)	0.48 (0.16)**	0.01 (0.07)	0.23 (0.06)***
***p < 0.001, **p < 0.01, *p < 0.05, °p < 0.1												

5.8.7 Discussion

The aims of Experiment 2 were to determine the extent to which there is evidence for cross-language activation of verb reflexivity information in Afrikaans during online bilingual pronoun resolution in English, and when, relative to Binding Principle B, verb-level influences on antecedent search apply.

There is a clear effect of Verb Type in the eye-tracking data. Contra the BAIF and Defeasible Filter Hypotheses, the results provided clear evidence that structural constraints were not respected from the outset of processing for sentences containing Potentially Reflexive verbs. From the earliest reading time measure – first fixation duration at the pronoun region – participants read significantly more slowly in the Non-Local Mismatch condition than in the other two Match conditions when the pronoun followed a Non-Reflexive verb. In contrast, when the pronoun followed a Potentially Reflexive verb, there was no difference in first fixation times across the three Match conditions. This effect is also present in the remaining regions of the sentence, indicating a persistent interaction between the Verb Type manipulation and the structural constraints on pronoun resolution.

The general pattern of results for the Potentially Reflexive items reflects participants' attempt, arguably prompted by cross-language activation of verb reflexivity information, to link the pronoun to the local antecedent. For the Double Match and Local Mismatch conditions, reading times are generally longer for the Potentially Reflexive than the Non-Reflexive items. This difference is significant for the Local Mismatch condition in both the pronoun and spillover regions, and for the Double Match condition at the spillover and final regions. As discussed in the predictions section, these elevated reading times may be the product of an increased competition effect in the Double Match condition and increased processing difficulty caused by the gender mismatch between the pronoun and nearest NP in the Local Mismatch condition. In the Non-Local Mismatch condition, however, the predicted effect – where reading times would be shorter for Potentially Reflexive items – only surfaces at the final region. This may indeed reflect a recovery strategy, as discussed in Sturt (2003), where participants would only settle on the structurally inaccessible antecedent in the absence of a feature-matching accessible antecedent.

Overall, the data provide clear evidence of both facilitatory and inhibitory interference. Facilitatory interference effects are evident for both the Local Mismatch and Non-Local Mismatch conditions, albeit in opposite directions. The elevated reading times in the Double Match condition for Potentially Reflexive items can plausibly be attributed to inhibitory interference. In all cases, these effects seem to be caused by the Verb Type manipulation.

5.9 General discussion and conclusion

This paper investigated whether cross-language activation of verb reflexivity information affects pronoun resolution in English among Afrikaans–English bilinguals. To this end, two experiments were conducted. Experiment 1 examined whether offline pronoun resolution preferences differed across English sentences containing verbs whose Afrikaans translation equivalents either require an accompanying pronominal form to be resolved in the English fashion (Non-Reflexive verbs) or allow this pronominal form to be interpreted as a reflexive, taking a local antecedent (Potentially Reflexive verbs). The results of this experiment revealed a clear effect of the Verb Type manipulation on reference assignment, as well as an interaction between Group and Verb Type, where the Verb Type effect was larger for the bilingual group than the L1 English group. The bilingual participants respected the structural constraints on English pronoun resolution almost without exception for the Non-Reflexive items, but violated this constraint approximately 50% of the time for the Potentially Reflexive items.

Experiment 2 used eye-tracking-while-reading to determine whether this Verb Type effect is also visible during online reference resolution and when this effect surfaces relative to the structural constraints that apply to reference resolution. Adherence to the structural constraints on pronoun resolution was tested by manipulating the gender of the potential antecedent NPs in the experimental items, creating three Match conditions (Double Match, Local Mismatch and Non-Local Mismatch). The results revealed a Verb Type x Match interaction from the earliest processing stages at the pronoun region of the experimental items, which generally persisted across the remainder of the sentence. Evidently, then, the Verb Type manipulation did affect participants' online processing patterns and was operative alongside the structural constraints on pronoun resolution from the onset of the reference resolution process.

Structural constraints and cross-language activation in bilingual pronoun resolution

The results of the present study are not compatible with either the BAIF Hypothesis (Nicol & Swinney, 1989) or the Defeasible Filter Hypothesis (Sturt, 2003), as participants only show evidence of having adhered to Binding Principle B – reflected in insensitivity to the gender of the structurally illegitimate antecedent – in the Non-Reflexive verb condition. The interference effects that were observed are consistent with the Cue-Based Retrieval Hypothesis (Badecker & Straub, 2002; Parker & Phillips, 2017; Patil et al., 2016), where the structural constraints operate in parallel to other cues that guide antecedent search. Previous studies on reflexive resolution have only sporadically observed interference effects (see e.g., Cunnings & Felser, 2013 (Experiment 2, high-working-memory-span readers); King et al., 2012; Parker & Phillips, 2017; Patil et al., 2016), although they seem to be somewhat more common in L2 speakers (e.g., Felser & Cunnings, 2012; Felser et al., 2009). The one reading-time study on L2 pronoun resolution (Patterson et al., 2014) remarks on a fleeting effect of a gender-matching inaccessible antecedent, but found this effect only in an analysis that was compromised by large amounts of missing data. A revised model revealed no interference effects among the L2 participants.

The present study contributes novel results regarding attraction effects in bilingual pronoun resolution, which to the best of our knowledge have not previously been reported on (see Chow et al., 2014, p. 3). In our data, interference effects were observed in sentences both with and without a feature mismatch between the pronoun and the target NP. Indeed, from the earliest reading time measures, participants were significantly slower when the structurally illegitimate antecedent mismatched the pronoun in gender (Local Mismatch condition) if the pronoun followed a Potentially Reflexive verb. Nonetheless, the results do not indicate that participants ignored Binding Principle B completely, as the facilitatory interference effect in the Non-Local Mismatch condition, predicted to surface if participants linked the pronoun to the local NP in the Potentially Reflexive items, only occurred at the final region of the sentence. It seems, then, that while the Verb Type manipulation may have led to the structural cue being demoted in importance, it did not override this cue completely.

Interpreting the implications of these results for theories of L2 processing requires consideration of the nature of the observed Verb Type effect. Following Reinhart and Reuland (1993), reflexivity is understood in this paper as a property of verbs that can receive a reflexive interpretation either in the absence of an accompanying reflexive (as, for example, in *John washed*) or when accompanied by an SE (as, for example, in the Afrikaans *Jan*$_i$ *was hom*$_i$ 'John washes him', where *hom* and *Jan* are interpreted as co-referential). Experiment 1 showed that the pronoun was almost never linked to the local antecedent when paired with a verb whose Afrikaans translation equivalent is Non-Reflexive. In contrast, sentences containing Potentially Reflexive verbs yielded significantly more local antecedent choices in the bilingual group than in the L1 English group. This finding suggests that the Afrikaans speakers' responses in the Potentially Reflexive verb condition may have been driven by cross-language activation of the Afrikaans translation equivalents.

A point that is often raised in relation to claims of cross-linguistic influence is whether the observed effects may not rather be due to general learner- or bilingualism effects. One way of addressing this concern is to include another bilingual group, whose two languages pattern in terms of verb reflexivity exactly like English. The design of the present study instead exploits what Jarvis (2010) terms an "intralingual contrast", where a single feature in one language is realized in two different ways in another language. In the case of the experimental items employed here, no English verbs allow for pronominal forms to be interpreted as reflexives, whereas in Afrikaans, some verbs do and some verbs do not. The effect that would suggest cross-linguistic influence, then, is an interaction between Group and Verb Type, where the effect of Verb Type differs across the groups (see Kim & Grüter, 2019, for a similar discussion). This is exactly the effect that was observed in Experiment 1. Thus, although no online data have yet been collected from an L1 English group, we have at least some indication that pronoun resolution preferences differ across L1 speakers and bilinguals for the two verb types employed. Given the parallels between the bilinguals' offline and online behavior, it is plausible that the L1 English speakers' online behavior would also reflect their offline responses.

Cross-language activation is thought to proceed as follows: upon reading the English verb, its Afrikaans counterpart in the lexicon, with its accompanying reflexivity information, is also activated. Subsequently, in a process of lemmatic transfer (Jarvis, 2009), the reflexivity information of the Afrikaans verb affects the interpretation of the English verb, which has consequences for the interpretation of the pronoun. As in Kim and Grüter (2019), then, verb-level information from the bilinguals' other language is shown here to have an effect on syntactic processing. However, a novel contribution of the present study is its indication that, for bilinguals, such verb-level information plays at least as important a role in processing as structural constraints on reference resolution.

The bilingual speakers' immediate sensitivity to the verb type manipulation is compatible with models of bilingual/L2 processing in which structural cues do not take priority over semantic/discourse cues (Clahsen & Felser 2006a, 2006b, 2006c, 2017; Cunnings, 2017a, 2017b). Cunnings (2017a, 2017b) in particular proposes that bilinguals show "increased susceptibility to interference during memory retrieval operations" (Cunnings, 2017b, p. 659); a claim which, as it applies to reference resolution, was initially based primarily on findings from reflexive processing. Considering that interference effects have not previously been observed in monolingual pronoun resolution (Chow et al., 2014), the present findings may lend tentative support to Cunnings' (2017a, 2017b) account. Here, it should be noted that interference effects were observed in our data across both Verb Types in the Double Match and Local Mismatch conditions (see e.g., regression path duration at the pronoun region), and so cross-language activation is likely not the only source of these effects. As such, the present findings encourage further examinations to determine the extent and possible underlying causes of interference effects in bilingual pronoun resolution.

In sum, this study has demonstrated that robust interference effects in bilingual reference resolution can be provoked by exploiting cross-linguistic differences in verb reflexivity. These interference effects appear during the earliest processing measures and thus constitute evidence against a “strictly syntactic” antecedent search procedure. The findings contribute to the growing body of results suggesting that all linguistic dependency resolution relies on the same error-prone cue-based retrieval mechanism (Lewis & Vasishth, 2005).

5.10 Appendix

Table 5-8: Linear mixed model results at the precritical region for Experiment 2

	First Fixation Duration	**First-Pass Reading Time**	**Regression Path Duration**
Intercept	5.33 (0.02)***	5.40 (0.03)***	5.54 (0.04)***
DM vs NLM	-0.00 (0.01)	0.00 (0.02)	-0.01 (0.02)
LM vs NLM	0.00 (0.01)	0.00 (0.02)	0.02 (0.02)
Verb Type	-0.01 (0.01)	0.01 (0.01)	0.01 (0.02)
DM vs NLM x Verb Type	-0.00 (0.01)	0.00 (0.02)	0.00 (0.02)
LM vs NLM x Verb Type	-0.01 (0.01)	-0.00 (0.02)	0.02 (0.02)
***p < 0.001, **p < 0.01, *p < 0.05, ∘p < 0.1			

6. L2 PROCESSING OF FILLER–GAP DEPENDENCIES ATTENUATED EFFECTS OF NATURALISTIC L2 EXPOSURE IN A MULTILINGUAL SETTING

6.1 Study abstract

In the online processing of long-distance wh-dependencies, native speakers have been found to make use of intermediate syntactic gaps, which has the effect of facilitating dependency resolution. This strategy has also been observed in L2 speakers living in an L2 immersion context, but not in classroom L2 learners. The aim of this paper is to investigate whether there is evidence of use of the intermediate gap among L2 speakers that have received considerable naturalistic exposure to the L2 from a young age, but do not live in a standard immersion context. Two groups of participants, one L1 English–L2 Afrikaans (n = 36) and one L1 Afrikaans–L2 English (n = 38), completed a self-paced reading task involving English sentences containing long-distance wh-dependencies. The data were analyzed using Bayesian regression. The results indicate that while the L1 English group made use of intermediate syntactic gaps, the L2 group did not. As such, the L2 group align with the classroom learners that have previously been studied. The findings shed light on the potential limits of non-immersive L2 experience in fostering nativelike sensitivity to abstract grammatical cues in L2 processing.

6.2 Introduction

It has been proposed that second language (L2) speakers, unlike first language (L1) speakers, have difficulty employing abstract syntactic information to guide real-time processing (Clahsen & Felser 2006a, 2006b, 2006c, 2017). While research has shown that this does not hold for all syntactic phenomena – see for example Boxell and Felser (2017) and Felser et al. (2012) for evidence that L2 speakers respect constraints on extraction – persistent differences have been observed between the L1 and L2 processing of sentences involving long-distance wh-dependencies, such as (1). As such, constructions of this type have received considerable attention in the discussion regarding whether, to what extent and under which circumstances L1 and L2 processing can converge fully (see Dallas & Kaan, 2008, for discussion).

(1a) The manager **who** the consultant's claim about the new proposal had pleased __ will hire five workers tomorrow.

(1b) The manager **who** the consultant claimed that the new proposal had pleased __ will hire five workers tomorrow.

To comprehend a sentence like (1), the reader or listener must reintegrate the moved element (or “filler”) into its original position (or “gap”) in the sentence so that it can be assigned a grammatical role. To do so, the filler must be held in working memory until the gap positon is reached, at which point reintegration can occur. Generative syntactic theories propose that in this process, the presence of a clause boundary (marked by *that* in 1b) allows the filler to be reactivated, thus accelerating reintegration at the upcoming gap in comparison to equivalent sentences in which no clause boundary is present (1a).

Such a facilitative effect has repeatedly been observed in L1 English speakers (Felser & Roberts, 2007; Gibson & Warren, 1999, 2004; Marinis et al., 2005; Pliatsikas & Marinis, 2013). However, amongst L2 speakers, findings have been less consistent. Classroom L2 learners seem to process sentences of the form in (1a) and (1b) identically, which has led to the suggestion that they do not make use of the clause boundary when processing dependencies of this type. On the other hand, naturalistic L2 exposure has been found to influence the L2 processing of these constructions, in that a facilitative effect of the clause boundary has been identified amongst L2 speakers who have spent time in L2 immersion contexts (Pliatsikas et al., 2017; Pliatsikas & Marinis, 2013). This result has been attributed to increased proceduralization of grammatical knowledge brought about by extensive L2 exposure, in line with an account in which L2 grammatical processing becomes more automatic as L2 proficiency and exposure increase (see e.g., Ullman, 2001b).

Findings such as the above have sparked interest in the effect of immersion on L2 acquisition and processing (see Pliatsikas & Chondrogianni, 2015, for an overview). In this regard, situations of societal multilingualism have yet to receive much scholarly attention. In such situations, L2 acquisition typically begins at an early age, with the learner receiving considerable naturalistic exposure to the L2 whilst never being fully immersed. Whether this level of naturalistic exposure is sufficient to bring about nativelike L2 processing of long-distance wh-dependencies is an open question.

The present paper reports on a self-paced reading study on the processing of long-distance wh-dependencies in English in a group of bilinguals drawn from the multilingual South African context, L1 Afrikaans–L2 English speakers. First-language English speakers serve as a control group. The paper investigates whether there is evidence of use of the intermediate gap in the L2 participants' processing patterns.

6.3 Naturalistic exposure in cases of asymmetric societal multilingualism

In the introduction to a collection of articles on naturalistic exposure, Pliatsikas and Chondrogianni (2015, p. 2) state that "experimental evidence suggests that one crucial factor for efficient native-like performance in the non-native language is the amount of naturalistic exposure, or immersion, that the learners receive to that language". Immersion here is defined as "the degree to which language learners use their non-native language outside the classroom and for their day-to-day activities", and is said to "usually presuppose that the learners live in an environment where their non-native language is exclusively or mostly used" (Pliatsikas & Chondrogianni, 2015, p. 2).

Immersion defined in this way applies to communities where the learner's non-native language is the dominant societal language, and would include L2 English speakers who are L1 speakers of a minority and/or immigrant language in the UK and the US, such as Turkish. In such cases, exposure to the L2 is pervasive, and the L2 is frequently used outside of the classroom, with use of the L1 perhaps being limited to the domestic environment.

Similar situations also apply in the case of early bilinguals. For example, heritage language speakers (see Montrul, 2016, for discussion) have exposure to a minority language in early childhood, but by late childhood have typically become dominant in the majority language of their community. Speakers of a heritage language have reduced exposure to this language and fewer opportunities to use it; exposure and use generally being restricted to the home. By the above definition, they are then immersed in the majority language, and typically they attain nativelike competence in this language.

However, another situation obtains in cases of societal bi- and multilingualism in which most if not all individuals have some proficiency in two or more languages. Such cases may involve one dominant language, but more than one language may be used in education, the media and government. South Africa is an example of such a society. The South African Constitution recognises 11 official languages; legislating that the governments of each of the nine provinces must promote these languages and use at least two of them in governmental business. In the Western Cape province, the research site of the present study, the three main languages are Afrikaans, English, and isiXhosa, which are the “first languages spoken at home” by 49.7%, 20.2% and 24.3% of the province’s population, respectively (Census, 2011). English however enjoys a privileged and even hegemonic status in the country: it is described as “the language of power and access – economically, politically, socially” (Probyn, 2001, p. 250), and consequently it is widespread as an additional language. Many individuals thus receive extensive exposure to and make extensive use of English in addition to their L1.

The L2 population under study in this paper are highly proficient early Afrikaans–English bilinguals who attend a bilingual Afrikaans–English university in the Western

Cape. They have all received exposure to both languages at both primary and secondary school, where typically one language is used as medium of instruction and the other is a compulsory school subject. Additionally, given the societal status of English and its public presence, the situation is such that, for the majority of their lives, these L2 speakers have had fairly intense contact with English, although they cannot be said to have been fully “immersed” in the language as per Pliatsikas and Chondrogianni’s (2015) definition, which foregrounds near-exclusive use of the L2.

The question arises as to whether this intense contact situation, like the full immersion contexts previously examined, is sufficient to foster nativelike L2 processing of long-distance wh-dependencies. The present paper aims to address this question.

6.4 L1 and L2 processing of long-distance wh-dependencies

A filler–gap dependency is a non-local relationship established between a moved element (the “filler”) and the base position in which it is lexically licensed (the “gap”, indicated by the underscore in 2). An example of a filler–gap dependency is given in (2):

(2) [$_{CP}$ The man [$_{CP}$ who the detective concluded [$_{CP}$ that the dangerous thief had distressed ___]] will buy a new alarm].

Sentences such as (2) pose a challenge to the human parser, as when the filler is first encountered, there is no indication of what thematic role or grammatical function it fulfils. To successfully interpret the sentence, the filler must be held in working memory until the gap is encountered and the moved element can be reintegrated into its canonical position in the sentence. Online studies have provided evidence of a related working memory cost, where a sentence with a filler–gap dependency is processed more slowly than an equivalent sentence without such a dependency (see e.g., Kluender & Kutas, 1993; Pliatsikas et al., 2017). Such studies have also shown that, in an attempt to minimize the demands on working memory, the parser actively looks for any legal gaps at which the filler can be reintegrated (see e.g., Stowe, 1986). This behavior is captured in Clifton and Frazier's (1989) Active Filler Hypothesis, and has been observed in both L1 and L2 speakers (see e g., Williams, 2006; Williams et al., 2001).

Another aspect of linguistic theory that is relevant in the parsing of sentences such as (2) is the notion of successive cyclic movement (Chomsky, 1986), which entails that a dislocated element moves through all intervening Spec,CP positions on the way to its landing site. Theoretically, this form of movement is motivated by the principle of subjacency (Chomsky, 1977), which prohibits a dislocated element from crossing a clause boundary without first moving to that clause boundary (for evidence of this constraint, see e.g., Ross, 1967). This successive cyclic movement leaves behind a phonologically unrealised copy of the filler at each clause boundary. Intervening clause boundaries thus constitute "intermediate gaps", which mediate between the filler's landing site and the 'true' gap at which it originated, thereby reducing the distance over which the filler–gap dependency extends.

In parsing, this reduced distance should be reflected in faster reintegration of the filler into its canonical position. In (2), the intermediate gap directly before *that* hosts a phonologically unrealised copy of *who*. In (3), there is also a filler–gap dependency, but no intermediate gap to host a copy of the filler.

(3) [$_{CP}$ The man [$_{CP}$ who the detective's conclusion about the dangerous thief had distressed ___] will buy a new alarm].

Successive cyclic movement predicts that (2) would be processed faster than (3), and this prediction is borne out in the case of L1 speakers. For example, Gibson and Warren (1999, 2004), in two self-paced reading tasks with monolingual English speakers, found shorter RTs at the final gap in sentences like (2) than in sentences like (3).

Research on L2 populations, in contrast, has shown that not all non-native speakers make use of intermediate gaps during processing (see Dallas & Kaan, 2008, for an overview). In Marinis et al. (2005), for example, native English speakers' RTs provided evidence of the facilitative effect of the intermediate gap. However, although they were equally able to comprehend these sentences, as suggested by their comparable performance on the comprehension questions, none of the L2 learner groups in this study showed the facilitative effect. Based on this finding, the authors propose that L2 speakers do not make use of intermediate gaps in the processing of long-distance filler–gap dependencies. Rather, they are said to maintain the filler in working memory until the subcategorizing verb is reached, at which point reintegration is triggered. One explanation that has been offered for this behavior is that non-native speakers may be less sensitive to the syntactic cue that identifies the clause boundary as a site for filler reactivation (Felser, 2015).

Marinis et al.'s (2005) participants had on average spent less than 2.5 years in an L2 immersion context. To explore the potential role of type of exposure on the L2 processing of long-distance filler–gap dependencies, Pliatsikas and Marinis (2013) examined two groups of L2 English speakers: one that had received only classroom exposure to English and another that had been immersed in an English-speaking environment for an average of nine years. Both the native speaker control group and the naturalistic exposure group showed faster RTs at the final gap position in the intermediate gap condition. In a follow-up fMRI study, Pliatsikas et al. (2017) also found an effect of naturalistic exposure: here, a group of L2 speakers with a mean duration of immersion of 5.37 years showed patterns that suggested they made use of the intermediate gap when processing the same set of stimuli used in Marinis et al. (2005) and Pliatsikas and Marinis (2013). It is against the background of these findings, which seem to indicate a central role of immersion in facilitating nativelike and syntactically driven L2 processing of long-distance wh-dependencies, that the present study is situated.

6.5 Aim of the present study

The results of the studies reviewed above suggest that L2 speakers that have spent a certain amount of time in an L2 immersion context can process long-distance wh-dependencies in a nativelike fashion. However, the precise nature of the linguistic setting that is required to bring about this processing shift remains undetermined. It is, for example, unclear whether full immersion in the L2 is necessary, or whether naturalistic L2 exposure received alongside considerable L1 exposure is sufficient. It is the aim of the present study to address this question by comparing long-distance wh-dependency processing across L1 speakers and a group of L2 speakers who have received divided exposure to their L1 and L2 for the majority of their lives. To enable comparisons with previous results to be drawn, the experiment uses the stimuli and experimental design employed in Marinis et al. (2005) and Pliatsikas and Marinis (2013).

6.6 Method

6.6.1 Participants

Two groups participated in this study: one group of 36 L1 English–L2 Afrikaans speakers (mean age: 20 years, SD: 1.2 years, range: 18–23) and one group of 38 L1 Afrikaans–L2 English speakers (mean age: 20 years, SD: 1 year, range: 18–22). All participants had lived in South Africa from birth with no significant amount of time spent outside of the country, and all L2 participants were early L2 acquirers (mean age of acquisition 5.16 years, SD 2.2 years). The Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian et al., 2007) was administered to assess participants' language backgrounds, and participants' English proficiency was evaluated by means of a C-test.

The participants' background data and C-test scores are summarised in Table 6-1.

Table 6-1: Background data on L1 English and L1 Afrikaans speakers (standard deviation)

	L1 ENG	**L1 AFR–L2 ENG**
Current English exposure %	74.02 (13.5)	47.8 (17.87)
Current Afrikaans exposure %	25.1 (16.3)	46.05 (19.17)
C-test score %	75.27 (13.27)	73.42 (13.21)
Self-rating English speaking (0-10)*	9 (0.9)	8.1 (1.4)
Self-rating English comp. spoken (0-10)	9.2 (0.89)	8.9 (1.07)
Self-rating English reading (0-10)	9 (0.85)	8.66 (1.18)

Note: *The self-rating scores in Tables 1 range from 0 (none) to 10 (perfect).

6.6.2 Materials

In the main experiment, participants read a total of 66 sentences. Six were practice items, 20 were experimental items, and the remaining 40 were fillers. The experimental sentences were identical to those used by Marinis et al. (2005) and Pliatsikas and Marinis (2013). Fewer fillers were used in this replication because participants completed this task as a part of a larger data collection procedure.

Each experimental sentence came in four versions. These were distributed in a 2 x 2 design, in which the conditions Extraction (Present or Absent) were crossed with Phrase Type (NP or VP). The four experimental conditions are illustrated below. The complete list of experimental items can be found in Marinis et al. (2005).

(4) a. [$_{CP}$ The nurse [$_{CP}$ who$_i$ the doctor argued [$_{CP}$ e$_i$ that the rude patient had angered e$_i$]] is refusing to work late.]

(*Extraction VP*)

b. [$_{CP}$ The nurse [$_{CP}$ who$_i$ the doctor's argument about the rude patient had angered e$_i$] is refusing to work late.]

(*Extraction NP*)

c. [$_{CP}$ The nurse thought [$_{CP}$ the doctor argued [$_{CP}$ that the rude patient had angered the staff at the hospital.]]]

(*Non-Extraction VP*)

d. [$_{CP}$ The nurse thought [$_{CP}$ the doctor's argument about the rude patient had angered the staff at the hospital.]]

(*Non-Extraction NP*)

In the two Extraction conditions, the initial NP (*the nurse*) is followed by a relative clause headed by *who*, which is the object of the embedded verb (*had angered*). The Extraction VP condition provides an intermediate gap for the filler at the embedded clause boundary. The verbs in the embedded clause were all transitive and biased towards taking a sentential object in order to prevent the filler *who* being construed as a direct object. The embedded verbs were also bridge verbs that permit *wh*-extraction from their complement clause. The Extraction NP sentences were identical in length to the Extraction VP sentences, but provided no intermediate landing site.

The Non-Extraction sentences also had the same number of words as their Extraction counterparts. No syntactic movement was involved in these sentences.

All sentences were divided into six segments as follows:

(5) The nurse who$_i$ / the doctor argued / e_i that / the rude patient / had angered e_i /

1 2 3 4 5

is refusing to work late.

6

The segments of interest are segments 3 and 5. Segment 3, which marks the beginning of the embedded clause, is the position of the intermediate gap in the Extraction VP sentences. Segment 5 contains the subcategorising verb, at which point reintegration of the filler must occur.

The experimental items were distributed across four lists in a Latin-square design and combined with the fillers in pseudorandom order. Each list was divided into three blocks of 20 sentences each. Participants took a short break between each block. All participants saw the same six practice items at the beginning of the experiment.

Comprehension questions were asked after all of the experimental sentences and 45% of the filler sentences in order to determine whether the participants were paying attention and whether they had understood the sentences. As in Marinis et al. (2005), the comprehension questions following the experimental items all asked which person mentioned in the sentence had committed a particular action; e.g., the question following (4a) was *Who is refusing to work late?*, with *The nurse* and *The doctor* offered as answers.

The C-test used to assess English proficiency consisted of three texts, each of which contained 20 incomplete words, with the first 50% of the letters of each word provided. Two of the texts in the C-test were developed by Keijzer (2007) in order to assess L1 attrition in L1 English speakers, and so are relatively difficult. They were therefore deemed appropriate for assessing the L2 speakers in the population under study. The third text was a modified version of a Wikipedia article about a South African university.

6.6.3 Procedure

The self-paced reading task was designed and administered in PsychoPy (Peirce et al., 2019). The accuracy of participants' question responses as well as their RTs for each segment were captured. The experiment made use of the non-cumulative moving window procedure (Just et al., 1982). The sentences were presented segment-by-segment in black letters (font: Consolas) on a light grey background, and the end of each sentence was marked by a full stop. The text was displayed in the centre of a 15-inch laptop screen (resolution: 1366 x 768)[39] and always ran over two lines, where only the final segment of each sentence appeared on the second line. Participants used the space bar on the keyboard to prompt the display of the next sentence segment.

[39] Due to a technical error, the RTs for the final segment were not captured, and they are therefore not reported here. However, segments at the end of a sentence are typically not analyzed in RT studies, as potential sentence wrap-up effects make these RTs difficult to interpret. Furthermore, neither Marinis et al. (2005) nor Pliatsikas and Marinis (2013) analyzed the results for this segment: the facilitative effect of the intermediate gap was always observed at the fifth segment. Thus, the analysis presented in this paper does not deviate from the practices of previous studies.

The task was explained to the participants by the experiment administrator. The oral explanation was accompanied by written instructions. Participants were given an opportunity to ask questions before and after completing the practice session.

Participants had two self-timed breaks during the task, where they could press any key whenever they were ready to continue with the experiment. The task took approximately 20 minutes to complete. Subsequently, the participants filled in the LEAP-Q and completed the C-test.

6.6.4 Analysis

The data were analyzed using Bayesian regression in order to allow conclusions to be drawn regarding the strength of the evidence in favour of the existence of processing differences across groups and construction types. Whilst conventional null hypothesis significance testing approaches do not make it possible to assess whether the lack of a significant effect is due to the non-existence of this effect or simply to a failure to detect it (due e.g., to a lack of statistical power), Bayesian analyses can provide support in favour of the null hypothesis, and thus allow for more nuanced interpretations of empirical data (Dienes 2014, 2016)[40].

Another advantage of the Bayesian approach is that it enables previous findings in a particular domain to be incorporated into an analysis. This is done by specifying so-called "priors", which in Bayesian regression provide an indication of the expected

[40] For an introduction to the use of Bayesian statistics in L2 research, the reader is referred to Norouzian et al. (2019).

sizes of the coefficients of particular effects. In the present analysis, we used informative priors for those effects that were of interest to the research question, and for which previous studies had returned converging results. These informative priors were specified in milliseconds, based on the averages of the raw reading times reported in Marinis et al. (2005) and Pliatsikas and Marinis (2013), with a standard deviation that captured the observed spread around the means.[41] For Pliatsikas and Marinis (2013), the differences between the L1 English group and the Naturalistic Exposure English group were employed, as this contrast is more representative of the L1–L2 contrast in the present study. The supplementary materials provide the details of the informative prior calculations.

For all effects that were not of direct relevance to the paper’s research question, or for which the two studies listed above returned conflicting results, vague priors were used. Following Nalborczyk, Batailler, Lœvenbruck, Vilain, and Bürkner (2019), these were specified as normally distributed around 0 for all terms except the model residuals and standard deviation, which were specified as having a Half-Cauchy distribution to restrict their values to being positive.

To fit each model, we utilized four chains with 5,000 samples per chain, a warm-up of 2,500 samples, and no thinning, resulting in 10,000 samples for each parameter estimate. Results are presented in the text with an indication of the evidence ratio (or Bayes factor), *P*(*b*), which provides an indication of the evidence regarding the

---

[41] Pliatsikas et al.’s (2017) findings were not used for the prior calculation because this study employed functional magnetic resonance imaging (fMRI), and it is difficult to relate results from this paradigm to those of reading-times studies such as Marinis et al. (2005) and Pliatsikas and Marinis (2013).

existence of a particular effect. Here, evidence ratios are interpreted in line with the guidelines in Jeffreys (1998), where ratios of 3 or greater indicate substantial evidence for an effect, and ratios of 1/3 or smaller indicate substantial evidence for the non-existence of this effect.

6.6.5 Reading time predictions

As noted above, the critical segments in this experiment are segments 3 and 5. At segment 3, regardless of their L1, both participant groups are expected to slow down in the Extraction conditions, as this reflects the cost of holding the filler in working memory. Here, no difference between the Extraction VP and Extraction NP conditions is expected, as the filler cost should be seen in both (Marinis et al., 2005; Pliatsikas & Marinis, 2013). The filler cost also entails that participants should be slower in segment 4 in the Extraction conditions compared to the Non-Extraction conditions (see Dekydtspotter, Schwartz, & Sprouse, 2006, for discussion of this point).

Segment 5 is where the subcategorizing verb is located. It is at this point that the filler must be reintegrated into the original gap position. This is also expected to reflect in a time cost, and so participants should be slower here in the Extraction than in the Non-Extraction conditions. At this point – on the basis of past findings with L1 speakers – a difference between the Extraction VP and the Extraction NP conditions is predicted to occur in the L1 English group. This is because the Extraction VP condition makes available a site for reactivation of the filler (segment 3, the location of the clause boundary). No such possibility is available in the Extraction NP condition. Consequently, retrieval of the filler in the Extraction NP condition must take place over a greater temporal distance than in the Extraction VP condition. This should result in a time cost for reintegration, leading to longer RTs in the Extraction NP than in the Extraction VP condition. This effect was also found for Pliatsikas and Marinis' (2013) immersed L2 speakers. If the naturalistic exposure to English received by the L1 Afrikaans participants in this study has been sufficient, then we expect RTs between the Extraction VP and Extraction NP conditions to differ in this group as well.

6.7 Results

6.7.1 Accuracy

For all comprehension questions, the mean accuracy score was 84.1%. For the questions following the experimental items, the mean accuracy score was 76.13%, with the L1 English group scoring 78.2% and the L1 Afrikaans group 74.12%. The data from two L1 English and two L1 Afrikaans participants who performed close to chance (below 60% correct) were removed from further analysis. This raised the overall accuracy score to 77.5% (range 60–100%). The L1 English participants' mean score was 79.7%, and that of the L1 Afrikaans participants was 75.3%. These relatively high scores, which are comparable to those obtained in Marinis et al. (2005) and Pliatsikas and Marinis (2013), indicate that the participants were paying attention during the task and that they understood these complex constructions.[42]

[42] As Pliatsikas and Marinis (2013) note, the fact that neither group performed at ceiling level can be attributed to the complexity of the sentences.

6.7.2 Reading times

In accordance with standard practice in sentence processing studies, only data from trials where the comprehension question was answered correctly were analyzed. Following Pliatsikas and Marinis (2013), we excluded the data of five participants (two L1 English, three L1 Afrikaans) whose mean RTs were more than two SDs above the group mean across all conditions. We then screened RTs for extreme values and outliers. Also in accordance with Pliatsikas and Marinis (2013), extreme values were defined as below 100 ms and above 4,000 ms, and outliers were defined as values greater than two SDs above or below the mean for each condition per subject and per item. Extreme values and outliers were replaced with the participant's mean RT for that condition and segment, unless this mean value itself was extreme, in which case the extreme value or outlier in question was eliminated. Following this procedure, 0.27% of the L1 Afrikaans data and 0.19% of the L1 English data were eliminated. Mean values replaced 3.59% of the L1 Afrikaans data and 4% of the L1 English data.

The mean reading times (standard deviations) per group per condition are shown in Table 6-2.

Table 6-2: Mean reading times in milliseconds (standard deviations) per group per condition

L1	Condition	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5
Afrikaans	**Extraction NP**	1121 (683)	1928 (806)	964 (434)	1321 (708)	1237 (692)
	Extraction VP	994 (566)	1577 (649)	949 (523)	1230 (544)	1272 (726)
	Non-extraction NP	1006 (593)	1577 (626)	831 (457)	1182 (609)	919 (483)
	Non-extraction VP	1108 (494)	1664 (735)	919 (486)	1163 (519)	958 (409)
English	**Extraction NP**	1047 (542)	1743 (698)	1060 (531)	1340 (634)	1147 (754)
	Extraction VP	1052 (532)	1646 (698)	987 (538)	1406 (772)	1165 (725)
	Non-extraction NP	1064 (554)	1667 (792)	855 (462)	1040 (518)	878 (383)
	Non-extraction VP	1184 (641)	1597 (642)	893 (448)	1264 (716)	859 (346)

Reading times were log-transformed to meet the normality requirement for linear regression. Bayesian regression models were fit using the brms package (version 2.9.0; Bürkner 2017, 2018) in R (version 3.6; R Core Team, 2018). In all models, L1 Group (L1 Afrikaans or L1 English, sum coded as -1 and 1), Extraction (Absent or Present; sum coded as -1 and 1) and Phrase Type (NP or VP; sum coded as -1 and 1) were fixed effects. We employed the maximal random effects structure that would converge (Barr, Levy, Scheepers, & Tily, 2013), which for all models included random intercepts for Participants and Items and by-participants and by-items random slopes for Extraction, Phrase Type and their interaction.

Recall that there are three segments of interest for the analysis: segment 3, the clause boundary and location of the intermediate gap in the Extraction VP condition; segment 4, where possible spill-over effects from segment 3 in the Extraction conditions may arise; and segment 5, the location of the subcategorising verb where the filler must be reintegrated in the Extraction conditions. The RT results for segments 3–5 are presented below, with the overall pattern illustrated in Figure 6-1. The results of the Bayesian regression models for each segment are presented in Table 6-3.

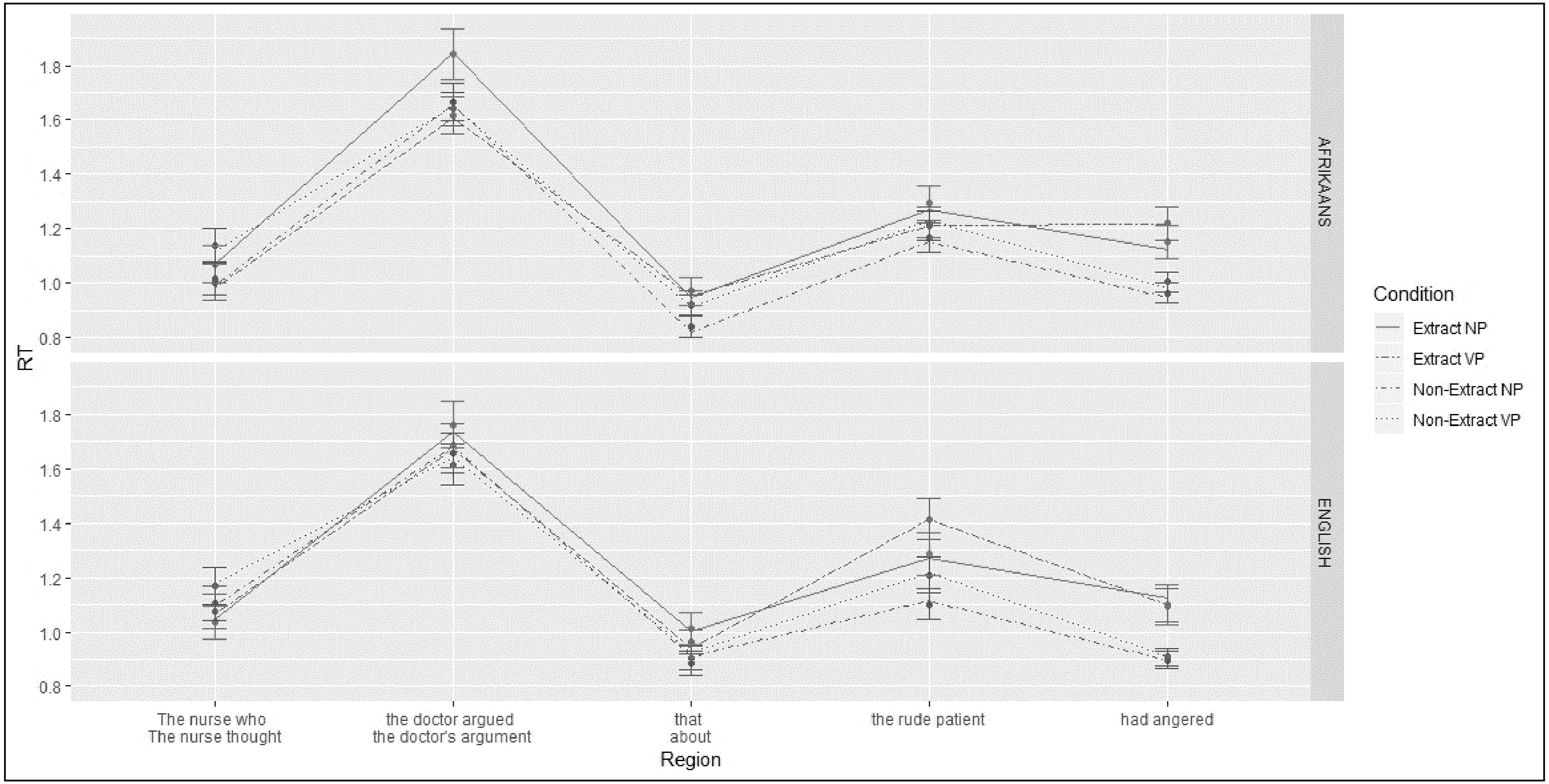

Figure 6-1: Reading times in seconds for regions of interest. Error bars represent standard deviations. Extraction sentences start with *The nurse who*. VP sentences continue with *the doctor argued that*; NP sentences continue with *the doctor's argument about*.

Table 6-3: Results of Bayesian regression models for the segment 3–5 reading times

	SEGMENT 3				SEGMENT 4				SEGMENT 5			
Term	**Est.**	**SE**	**CI L.**	**CI U.**	**Est.**	**SE**	**CI L.**	**CI U.**	**Est.**	**SE**	**CI L.**	**CI U.**
Intercept	-0.21	0.04	-0.27	-0.15	0.07	0.04	0.00	0.14	-0.10	0.03	-0.16	-0.04
L1	0.01	0.03	-0.05	0.06	0.00	0.04	-0.07	0.06	-0.04	0.03	-0.09	0.02
Extraction	0.06	0.02	0.03	0.09	0.07	0.02	0.03	0.10	0.07	0.02	0.03	0.11
Phrase Type	0.02	0.02	-0.01	0.05	0.02	0.02	-0.01	0.05	0.01	0.02	-0.01	0.04
L1:Extraction	0.01	0.02	-0.02	0.03	0.03	0.02	0.00	0.05	-0.01	0.02	-0.04	0.03
L1:Phrase Type	-0.01	0.01	-0.04	0.01	0.02	0.01	-0.01	0.04	-0.01	0.01	-0.03	0.01
Extraction:Phrase Type	-0.04	0.02	-0.07	-0.01	-0.01	0.02	-0.04	0.02	0.00	0.02	-0.03	0.02
L1:Extraction:Phrase Type	0.01	0.01	-0.01	0.03	0.01	0.01	-0.01	0.03	-0.02	0.01	-0.04	0.01

Note: *Est.* represents the parameter estimate; *SE* the standard error; *CI L.* the lower end of the 95% credible interval, and *CI U.* the upper end of the 95% credible interval.

Segment 3: The model revealed two robust effects at segment 3. The first was that of Extraction, where RTs were higher in the Extraction than the Non-Extraction conditions, indicating the cost of holding the moved element in working memory ($P(b > 0) = 713.29$). The second was the Extraction x Phrase Type interaction ($P(b < 0) = 187.68$), where participants were faster in the Extraction VP than in the Extraction NP condition. The latter effect was also found in Pliatsikas and Marinis' (2013) L2 groups, and is not related to intermediate gap processing. Further, a reliable interaction between L1 and Phrase Type was also observed, such that the L1 speakers were faster in the Verb Phrase condition relative to the Noun Phrase condition compared to the L2 speakers ($P(b < 0) = 5.91$). This effect is also not related to the processing of intermediate gaps.

Segment 4: At this segment, there was again a robust effect of Extraction ($P(b > 0) = 2499$), indicating that RTs were higher in the Extraction than the Non-Extraction conditions. There was also a reliable interaction between L1 and Extraction ($P(b > 0)$

= 30.75), where the L1 speakers were slower in the Extraction conditions than the L2 speakers. No other effects were robust at this segment.

Segment 5: At this segment, there is substantial evidence that the L1 speakers were faster than the L2 speakers ($P(b < 0) = 6.28$). There is also a robust effect of Extraction, where RTs were higher in the Extraction than the Non-Extraction conditions ($P(b > 0) = 453.55$). Then, there is substantial evidence that the L1 speakers were faster in the Verb Phrase condition relative to the Noun Phrase condition compared to the L2 speakers ($P(b < 0) = 3.35$).

At segment 5, shorter RTs in the Extraction VP compared to the Extraction NP condition would indicate that participants had made use of the intermediate gap in processing the experimental items. The model provides only anecdotal evidence of such an effect ($P(b < 0) = 1.31$). There is however substantial evidence that the L1 speakers' RTs across the two Extraction conditions differed to a larger extent than those of the L2 speakers ($P(b < 0) = 7.86$). As such, we conducted separate analyses for the groups at this segment. The results are presented in Table 6-4.

Table 6-4: Results of the Bayesian regression analyses for the L1 English and L1 Afrikaans groups at segment 5

Term	**Est.**	**SE**	**CI L.**	**CI U.**
L1 English				
Intercept	-0.14	0.05	-0.23	-0.05
Extraction	0.08	0.03	0.03	0.13
Phrase Type	0.00	0.02	-0.03	0.04
Extraction:Phrase Type	-0.02	0.02	-0.05	0.02
L1 Afrikaans				
Intercept	-0.06	0.05	-0.14	0.02
Extraction	0.08	0.04	0.01	0.14
Phrase Type	0.02	0.02	-0.01	0.06
Extraction:Phrase Type	0.01	0.02	-0.02	0.05

For the L1 English group, RTs in the Extraction conditions were higher than in the Non-Extraction conditions ($P(b > 0) = 89.91$). Further, there is substantial evidence that RTs were lower in the Extraction VP than the Extraction NP condition ($P(b < 0) = 3.6$).

The L1 Afrikaans groups' RTs were also higher in the Extraction than the Non-Extraction conditions ($P(b > 0) = 37.31$). Reaction times in this group were slower in the Verb Phrase than the Noun Phrase condition ($P(b > 0) = 6.49$). Finally, the model results provide borderline substantial evidence that the L1 Afrikaans speakers were not faster in the Extraction VP than the Extraction NP condition ($P(b < 0) = 0.35$). Follow-up analyses to check for an effect of proficiency revealed that participants with higher C-test scores were slower in the Extraction conditions ($b = 0.04$, 95% credible interval [-0.01, 0.1], $P(b > 0) = 11.27$), but there was no reliable proficiency effect on RTs in the Extraction VP relative to the Extraction NP condition.

6.8 Discussion

This study's aim was to investigate whether previously observed L1–L2 differences in the processing of long-distance wh-dependencies are found with L2 speakers who have received extensive naturalistic L2 exposure in a non-immersion environment. This aim relates to a larger question regarding the circumstances under which L1 and L2 processing can converge.

The analyses revealed some group differences in processing across the three segments of interest. The finding of relevance to the research question was the interaction at segment 5 – the segment at which reintegration of the filler must occur – between L1 Group, Extraction and Phrase Type. Here, the separate group analyses revealed a reliable Extraction x Phrase Type interaction for the L1 group, but not for the L2 group. In the L1 group, this interaction effect reflected that RTs at segment 5 were lower in the Extraction VP than the Extraction NP condition, in line with the findings of previous studies with L1 speakers (Gibson & Warren, 1999, 2004; Marinis et al., 2005; Pliatsikas & Marinis, 2013). This difference between the two Extraction conditions suggests that in the Extraction VP condition, the availability of an intermediate gap at which the filler could be reactivated facilitated the filler's later reintegration.

The absence of such a facilitatory effect in the L2 group, who patterned with the classroom English learners in Marinis et al. (2005) and Pliatsikas and Marinis (2013), was not expected, given that these participants have received extensive naturalistic exposure to English for most of their lives. As they still scored highly on the comprehension questions, the L2 participants clearly understood these complex constructions. However, they seemingly did not make use of abstract syntactic elements to do so. The processing pattern observed in this group suggests rather that they made use of a strategy in which the moved element is held in working memory until the reintegration site is reached.

It has been proposed that non-use of intermediate gaps when parsing long-distance wh-dependencies is due to delayed or absent sensitivity to the abstract syntactic cue – that is, the clause boundary – that marks a potential site for filler reactivation (Felser, 2015). This reduced sensitivity, furthermore, does not seem to be linked to L2 proficiency, as suggested by the present findings, as well as those of Pliatsikas and Marinis (2013, see the footnote on p. 179). It has been proposed that reduced sensitivity to syntactic cues is a general characteristic of L2 processing, which is said to rely to a greater extent on "semantic, pragmatic, probabilistic and surface-level information" (Clahsen & Felser, 2017, p. 2). This argument is formalized as the so-called Shallow Structure Hypothesis (SSH) (Clahsen & Felser 2006a, 2006b, 2006c, 2017). Importantly, the SSH does not propose that L2 learners never make use of syntactic cues during parsing. It accommodates processing behavior such as that of Pliatsikas and Marinis' (2013) (see also Pliatsikas et al., 2017) naturalistic exposure group by attributing qualitative changes in L2 processing routines to exposure-induced automatization of L2 processing (Clahsen & Felser, 2017, p. 6).

Indeed, based on their findings, Pliatsikas and Marinis (2013, p. 180) suggest that “the SSH applies only to L2 learners with limited or no naturalistic exposure”. This suggestion should be qualified in light of the present data. Our L1 Afrikaans participants are on average exposed to English 47.8% of the time, and for the majority of them, this exposure began in early childhood. Nonetheless, they still employ so-called “shallow” processing, indicating that the scope of the SSH may extend beyond classroom L2 learners.

The present findings call for a more careful consideration of the nature of naturalistic L2 exposure across different linguistic settings. Our L2 group differs from Pliatsikas and Marinis’ (2013) – which consisted of L1 Greek speakers living in the UK – in that they receive substantial L1 exposure alongside their L2 exposure. Almost 50% of the inhabitants of the province in which they live are L1 speakers of Afrikaans, and they attend a university where teaching and learning are also partly conducted in this language. Therefore, these L2 participants differ from those in ‘true’ L2 immersion environments, in that the use possibilities for their L1 are substantially greater.

Divided exposure to two or more languages is known to have consequences for language processing, leading, for example, to reduced fluency in reading (Whitford & Titone, 2012, 2015) and delayed lexical access (Gollan, Montoya, Cera, & Sandoval, 2008). Speculatively, this divided exposure could act on language processing by reducing the automaticity of processing in the L2, in turn reducing sensitivity to abstract syntactic information. Consequently, although our L2 participants’ English exposure may facilitate their processing of this language, their divided L1–L2 exposure may still cause their processing behavior to differ from that of an immersed L2 group. A comparison across L2 groups with varying levels of L2 exposure would be one way to gain further insight into the workings of the naturalistic exposure effect on L2 processing.

Ultimately, the present study highlights the value and importance of considering new contexts in investigations of bilingual language acquisition and processing. If we only investigate foreign-language or full immersion settings, we overlook the wide range of situations in which multiple languages coexist alongside one another. An understanding of the consequences of such environments for the bi- and multilingual mind is crucial to developing a full understanding of the human capacity for language (for a similar argument, see Bylund & Athanasopoulous, 2014).

6.9 Conclusion

This paper examined the online processing of long-distance wh-dependencies in L1 English–L2 Afrikaans speakers and L1 Afrikaans–L2 English speakers who have received considerable naturalistic exposure to the L2 from a young age. The paper's aims were to determine whether these groups made use of intermediate syntactic gaps during processing, as has previously been observed in monolingual English speakers and immersed L2 speakers. Evidence of use of the intermediate gap was found among the L1 speakers but not among the L2 speakers, despite the latter group's considerable naturalistic exposure to English. The paper provides an indication of the limits of non-immersive L2 experience in fostering nativelike sensitivity to abstract grammatical cues in L2 speakers.

6.10 Supplementary materials I: Working memory effects

Processing long-distance dependencies relies to some extent on working memory (e.g., Kluender & Kutas, 1993). Capacity-based models of L2 processing (e.g., Hopp, 2010; McDonald, 2006) therefore predict that individual differences in working memory may affect the L2 processing of constructions such as the experimental items in the present study. Although capacity limitations are not predicted to affect L2 processing in early L2 acquirers like those tested in the present study, we nonetheless investigated whether working memory capacity had any effects on our L2 participants' processing.

To test for possible effects of working memory capacity on processing, we collected working memory data from a subset of our L2 participants (n = 19). These participants completed a reading span task on a computer (Stone & Towse, 2015; von Bastian, Locher, & Ruflin, 2013). The reading span task presented participants with a set of sentences, each of which had to be judged as either 'makes sense' or 'does not make sense'. Each sentence was followed by a number that had to be remembered until the end of the set of sentences, at which point all the numbers encountered in that set had to be provided, in order of appearance. The sets varied in size from two to five sentences. Scoring was done according to the proportion of numbers that were correctly recalled. The participants had a mean proportion correct score of 0.51 (SD: 0.15, range: 0.26–0.83).

Possible effects of working memory on the processing of the experimental stimuli were examined using Bayesian regression analyses, computed with the brms package (version 2.9.0; Bürkner 2017, 2018) in R (version 3.6; R Core Team, 2018). In all models, Extraction (Absent or Present; sum coded as -1 and 1) and Phrase Type (NP or VP; sum coded as -1 and 1) were fixed effects. Working memory (centred around the mean) was entered as a continuous predictor alongside Extraction and Phrase Type. Three-way interactions were allowed between all factors. All models included random intercepts for Participant and Number. The model results are presented in Table 6-5.

Table 6-5: Working memory models

	SEGMENT 3				SEGMENT 4				SEGMENT 5			
Term	**Est.**	**SE**	**CI L.**	**CI U.**	**Est.**	**SE**	**CI L.**	**CI U.**	**Est.**	**SE**	**CI L.**	**CI U.**
Intercept	-0.19	0.06	-0.29	-0.08	0.15	0.06	0.06	0.24	0.06	0.05	-0.02	0.14
Working Memory (WM)	-0.06	0.06	-0.17	0.04	-0.03	0.05	-0.11	0.06	0.02	0.05	-0.05	0.10
Extraction	0.06	0.03	0.02	0.10	0.06	0.03	0.01	0.11	0.16	0.03	0.12	0.21
Phrase Type	0.01	0.03	-0.03	0.06	0.01	0.03	-0.04	0.06	0.01	0.03	-0.04	0.05
WM:Extraction	-0.02	0.02	-0.04	0.01	0.01	0.02	-0.02	0.04	0.04	0.02	0.01	0.07
WM:Phrase Type	-0.03	0.02	-0.05	0.00	-0.01	0.02	-0.04	0.02	0.00	0.02	-0.03	0.03
Extraction:Phrase Type	-0.04	0.03	-0.08	0.00	0.02	0.03	-0.03	0.07	0.01	0.03	-0.04	0.06
WM:Extraction:Phrase Type	-0.02	0.02	-0.05	0.01	0.00	0.02	-0.03	0.03	0.02	0.02	-0.01	0.05

Note: *Est.* represents the parameter estimate; *SE* the standard error; *CI L.* the lower end of the 95% credible interval, and *CI U.* the upper end of the 95% credible interval.

The effect of working memory is robust only at segment 3 ($P(b < 0) = 5.64$), where participants with higher working memory scores had lower RTs. This effect was receives only anecdotal support at segments 4 ($P(b < 0) = 2.32$) and 5 ($P(b > 0) = 2.45$).

There is a Working Memory x Extraction effect at each segment. At segment 3, participants with higher working memory scores have lower RTs ($P(b < 0) = 5.23$). The reverse is true at the following two segments (segment 4: $P(b > 0) = 3.35$; segment 5: $P(b > 0) = 35.63$).

Finally, there is substantial evidence of the absence of the Extraction x Phrase Type interaction indicating use of the intermediate gap ($P(b < 0) = 0.17$). Higher working memory scores, instead of facilitating processing in the Extraction VP condition at segment 5, led to higher RTs in this condition ($P(b > 0) = 6.05$).

Overall, the results of this analysis indicate that participants with higher working memory scores did not show more nativelike processing of the experimental items. Rather, higher working memory scores are associated with longer RTs in the Extraction conditions. Speculatively, it may be that participants with greater working memory capacity are more likely to maintain the moved element in working memory during processing and to execute reintegration upon reaching the subcategorizing verb at segment 5. Participants with lower working memory capacity may wait until the end of the sentence is encountered before deriving a global interpretation of the input (see e.g., the discussion in Roberts et al., 2007). While the relationship between these results and those of the main experiment must be considered with caution, given that working memory scores were not obtained from the full set of L2 participants, the present findings suggest that capacity-based limitations were not responsible for the L2 speakers' non-use of the intermediate gap.

6.11 Supplementary materials II: Bayesian analyses

Setting the priors

Informative priors were only specified for effects observed in both Marinis et al. (2005) and Pliatsikas and Marinis (2013). These informative priors were specified in milliseconds, based on the averages of the raw reading times reported in each of these studies, with a standard deviation that captured the observed spread around the means. For Pliatsikas and Marinis (2013), the differences between the L1 English group and the Naturalistic Exposure English group are employed, as this contrast is more representative of the L1–L2 contrast in the present study.

Vague priors were specified for those model terms that either (i) were not significant in both of the two studies mentioned above or (ii) were not of interest to the research question. These were specified as normally distributed around 0 for all terms except the model residuals and standard deviation, which were restricted to be positive by means of a Half-Cauchy distribution (see e.g., Nalborczyk et al., 2019). Calculations of the informative priors are provided below.

Segment 3

- Extraction vs Non-Extraction: Marinis et al. (2005): 272 ms.

  Pliatsikas and Marinis (2013): 106 ms.

  Prior: Cauchy, mean = 180, sd = 90.

- L1 vs L2 : Marinis et al. (2005): -135 ms,

  Pliatsikas and Marinis (2013): -14 ms.

  Prior: Cauchy, mean = -80, sd = 65.

Segment 4

- Extraction vs Non-Extraction: Marinis et al. (2005): 331 ms.

  Pliatsikas and Marinis (2013): 108 ms.

  Prior: Cauchy, mean = 200, sd = 100.

- L1 vs L2 : Marinis et al. (2005): -409 ms.

  Pliatsikas and Marinis (2013): -509 ms.

  Prior: Cauchy, mean = -450, sd = 50.

Segment 5

- Extraction vs Non-Extraction: Marinis et al. (2005): 751 ms.

  Pliatsikas and Marinis (2013): 180 ms.

  Prior: Cauchy, mean = 500, sd = 300.

- L1 vs L2 : Marinis et al. (2005): -348 ms.

  Pliatsikas and Marinis (2013): -276 ms.

  Prior: Cauchy, mean = -300, sd = 50.

Main analyses

Model 3

```
## Set priors specified above for model 3
priors.3 <- c(set_prior("cauchy(-80,65)", class = "b", coef = "L11"),
    set_prior("cauchy(180,90)", class = "b", coef = "EXTRACT1"),
    set_prior("normal(0,10)", class = "b", coef = "L11:EXTRACT1:VP1"),
    set_prior("normal(0,10)", class = "b", coef = "L11:EXTRACT1"),
    set_prior("normal(0,10)", class = "b", coef = "L11:VP1"),
    set_prior("cauchy(0,10)", class = "sigma"), set_prior("cauchy(0,10)",
        class = "sd"))
model.3 <- brm(IRI_3 ~ L1 * EXTRACT * VP + (1 + EXTRACT * VP |
    PARTICIPANT) + (1 + EXTRACT * VP | NUMBER), family = lognormal(),
    prior = priors.3, data = spr.df.model, chains = 4, iter = 5000,
    warmup = 2500, sample_prior = TRUE, cores = 4)
```

Hypothesis testing

```
## L1 speakers are faster than L2 speakers:
(hypL1 <- hypothesis(model.3, hypothesis = "L11<0"))

## Hypothesis Tests for class b:
##   Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob
## 1  (L11) < 0     0.01      0.03    -0.05     0.06        0.7      0.41
##   Star
## 1
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## CI.Lower and CI.Upper are the 95% credible intervals within
## which the true value of the parameter is likely to fall.
## Evid.Ratio is the Bayes Factor for the specified
## hypothesis.
## Bayes Factor = 0.7, anecdotal evidence that
## L1 speakers are not faster than L2 speakers.

## Extraction conditions are slower:
(hypEXTRACT <- hypothesis(model.3, hypothesis = "EXTRACT1>0"))

## Hypothesis Tests for class b:
##       Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob
## 1 (EXTRACT1) > 0     0.06      0.02     0.03     0.09     713.29         1
##   Star
## 1    *
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 713.29, decisive evidence that RTs are
## slower in the Extraction conditions than the Non-Extraction
## conditions.

## Extraction VP condition is faster than Extraction NP
## condition:
(hypEXTRACT.VP <- hypothesis(model.3, hypothesis = "EXTRACT1:VP1<0"))

## Hypothesis Tests for class b:
##           Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio
## 1 (EXTRACT1:VP1) < 0    -0.04      0.02    -0.07    -0.01     187.68
##   Post.Prob Star
## 1      0.99    *
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 187.68, decisive evidence that RTs are lower
## in the Extraction VP than the Extraction NP condition.
```

```
## L1 speakers are slower than L2 speakers in the Extraction
## condition:
(hypL1.EXTRACT <- hypothesis(model.3, hypothesis = "L11:EXTRACT1>0"))

## Hypothesis Tests for class b:
##           Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio
## 1 (L11:EXTRACT1) > 0     0.01      0.02    -0.02     0.03       1.76
##   Post.Prob Star
## 1      0.64
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 1.76, anecdotal evidence that L1 speakers
## were slower in the Extraction conditions than L2 speakers.

## L1 speakers are faster than L2 speakers in the Verb Phrase
## condition:
(hypL1.VP <- hypothesis(model.3, hypothesis = "L11:VP1<0"))

## Hypothesis Tests for class b:
##      Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob
## 1 (L11:VP1) < 0    -0.01      0.01    -0.04     0.01       5.91      0.86
##   Star
## 1
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 5.91, substantial evidence that L1 speakers
## were faster in the Verb Phrase condition than L2 speakers.

## The difference between the two Extraction conditions is
## larger for L1 than for L2 speakers:
(hypL1.EXTRACT.VP <- hypothesis(model.3, hypothesis = "L11:EXTRACT1:VP1>0"))

## Hypothesis Tests for class b:
##               Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio
## 1 (L11:EXTRACT1:VP1) > 0     0.01      0.01    -0.01     0.03      2.57
##   Post.Prob Star
## 1      0.72
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 2.57, anecdotal evidence that L1 speakers
## were slower in the Extraction VP condition than L2
## speakers.
```

Model 4

```
## Set priors specified above for model 4
priors.4 <- c(set_prior("cauchy(-450,50)", class = "b", coef = "L11"),
    set_prior("cauchy(200,100)", class = "b", coef = "EXTRACT1"),
    set_prior("normal(0,10)", class = "b", coef = "L11:EXTRACT1:VP1"),
    set_prior("normal(0,10)", class = "b", coef = "L11:EXTRACT1"),
    set_prior("normal(0,10)", class = "b", coef = "L11:VP1"),
    set_prior("cauchy(0,10)", class = "sigma"), set_prior("cauchy(0,10)",
        class = "sd"))
model.4 <- brm(IRI_4 ~ L1 * EXTRACT * VP + (1 + EXTRACT * VP |
    PARTICIPANT) + (1 + EXTRACT * VP | NUMBER), family = lognormal(),
    prior = priors.4, data = spr.df.model, chains = 4, iter = 5000,
    warmup = 2500, sample_prior = TRUE, cores = 4)
```

Hypothesis testing

```
## L1 speakers are faster than L2 speakers:
(hypL1 <- hypothesis(model.4, hypothesis = "L11<0"))

## Hypothesis Tests for class b:
##   Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob
## 1  (L11) < 0        0      0.04    -0.07     0.06       1.07      0.52
##   Star
## 1
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 1.07, anecdotal evidence that L1 speakers
## are faster than L2 speakers.

## Extraction conditions are slower:
(hypEXTRACT <- hypothesis(model.4, hypothesis = "EXTRACT1>0"))

## Hypothesis Tests for class b:
##       Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob
## 1 (EXTRACT1) > 0     0.07      0.02     0.03      0.1       2499         1
##   Star
## 1    *
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 2499, decisive evidence that RTs are slower
## in the Extraction conditions than the Non-Extraction
## conditions.
```

```
## Extraction VP condition is faster than Extraction NP
## condition:
(hypEXTRACT.VP <- hypothesis(model.4, hypothesis = "EXTRACT1:VP1<0"))

## Hypothesis Tests for class b:
##            Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio
## 1 (EXTRACT1:VP1) < 0    -0.01      0.02    -0.04     0.02       1.84
##   Post.Prob Star
## 1      0.65
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 1.84, anecdotal evidence that RTs are lower
## in the Extraction VP than the Extraction NP condition.

## L1 speakers are slower than L2 speakers in the Extraction
## condition:
(hypL1.EXTRACT <- hypothesis(model.4, hypothesis = "L11:EXTRACT1>0"))

## Hypothesis Tests for class b:
##            Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio
## 1 (L11:EXTRACT1) > 0     0.03      0.02        0     0.05      30.75
##   Post.Prob Star
## 1      0.97    *
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 30.75, very strong evidence that L1 speakers
## were slower in the Extraction conditions than L2 speakers.

## L1 speakers are faster than L2 speakers in the Verb Phrase
## condition:
(hypL1.VP <- hypothesis(model.4, hypothesis = "L11:VP1<0"))

## Hypothesis Tests for class b:
##       Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob
## 1 (L11:VP1) < 0     0.02      0.01    -0.01     0.04       0.14      0.12
##   Star
## 1
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 0.14, substantial evidence that L1 speakers
## were not faster in the Verb Phrase condition than L2
## speakers.
```

```
## The difference between the two Extraction conditions is
## larger for L1 than for L2 speakers:
(hypL1.EXTRACT.VP <- hypothesis(model.4, hypothesis = "L11:EXTRACT1:VP1>0"))

## Hypothesis Tests for class b:
##                Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio
## 1 (L11:EXTRACT1:VP1) > 0     0.01      0.01    -0.01     0.03      2.59
##   Post.Prob Star
## 1      0.72
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 2.59, anecdotal evidence that L1 speakers
## were slower in the Extraction VP condition than L2
## speakers.
```

Model 5

```
## Set priors specified above for model 5
priors.5 <- c(set_prior("cauchy(-300,50)", class = "b", coef = "L11"),
    set_prior("cauchy(500,300)", class = "b", coef = "EXTRACT1"),
    set_prior("normal(0,10)", class = "b", coef = "L11:EXTRACT1:VP1"),
    set_prior("normal(0,10)", class = "b", coef = "L11:EXTRACT1"),
    set_prior("normal(0,10)", class = "b", coef = "L11:VP1"),
    set_prior("cauchy(0,10)", class = "sigma"), set_prior("cauchy(0,10)",
        class = "sd"))
model.5 <- brm(IRI_5 ~ L1 * EXTRACT * VP + (1 + EXTRACT * VP |
    PARTICIPANT) + (1 + EXTRACT * VP | NUMBER), family = lognormal(),
    prior = priors.5, data = spr.df.model, chains = 4, iter = 5000,
    warmup = 2500, sample_prior = TRUE, cores = 4)
```

Hypothesis testing

```
## L1 speakers are faster than L2 speakers:
(hypL1 <- hypothesis(model.5, hypothesis = "L11<0"))

## Hypothesis Tests for class b:
##   Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob
## 1  (L11) < 0    -0.04      0.03    -0.09     0.02       6.28      0.86
##   Star
## 1
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 6.28, substantial evidence that L1 speakers
## are faster than L2 speakers.
```

```
## Extraction conditions are slower:
(hypEXTRACT <- hypothesis(model.5, hypothesis = "EXTRACT1>0"))

## Hypothesis Tests for class b:
##       Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob
## 1 (EXTRACT1) > 0     0.07      0.02     0.03     0.11     453.55         1
##   Star
## 1    *
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 453.55, decisive evidence that RTs are
## slower in the Extraction conditions than the Non-Extraction
## conditions.

## Extraction VP condition is faster than Extraction NP
## condition:
(hypEXTRACT.VP <- hypothesis(model.5, hypothesis = "EXTRACT1:VP1<0"))

## Hypothesis Tests for class b:
##           Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio
## 1 (EXTRACT1:VP1) < 0        0      0.02    -0.03     0.02       1.31
##   Post.Prob Star
## 1      0.57
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 1.31, anecdotal evidence that RTs are lower
## in the Extraction VP than the Extraction NP condition.

## L1 speakers are slower than L2 speakers in the Extraction
## condition:
(hypL1.EXTRACT <- hypothesis(model.5, hypothesis = "L11:EXTRACT1>0"))

## Hypothesis Tests for class b:
##           Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio
## 1 (L11:EXTRACT1) > 0    -0.01      0.02    -0.04     0.03       0.52
##   Post.Prob Star
## 1      0.34
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 0.52, anecdotal evidence that L1 speakers
## were not slower in the Extraction conditions than L2
## speakers.
```

```
## L1 speakers are faster than L2 speakers in the
## Verb Phrase condition:
(hypL1.VP <- hypothesis(model.5, hypothesis = "L11:VP1<0"))

## Hypothesis Tests for class b:
##      Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob
## 1 (L11:VP1) < 0    -0.01      0.01    -0.03     0.01       3.35      0.77
##   Star
## 1
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 3.35, substantial evidence that L1 speakers
## were faster in the Verb Phrase condition than L2 speakers.

## The difference between the two Extraction conditions is
## larger for L1 than for L2 speakers:
(hypL1.EXTRACT.VP <- hypothesis(model.5, hypothesis = "L11:EXTRACT1:VP1<0"))

## Hypothesis Tests for class b:
##               Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio
## 1 (L11:EXTRACT1:VP1) < 0    -0.02      0.01    -0.04     0.01      7.86
##   Post.Prob Star
## 1      0.89
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 7.86, substantial evidence that L1 speakers
## were faster in the Extraction VP condition than L2
## speakers.
```

Separate group analyses at segment 5

```
eng.model <- dplyr::filter(spr.df.model, L1 == "ENGLISH")
afr.model <- dplyr::filter(spr.df.model, L1 == "AFRIKAANS")
```

English group

```
priors.eng <- c(set_prior("cauchy(500,300)", class = "b", coef = "EXTRACT1"),
    set_prior("normal(180,60)", class = "b", coef = "EXTRACT1:VP1"),
    set_prior("normal(0,10)", class = "b", coef = "VP1"), set_prior("cauchy(0,10)",
        class = "sigma"), set_prior("cauchy(0,10)", class = "sd"))
model.5.eng <- brm(IRI_5 ~ EXTRACT * VP + (1 + EXTRACT * VP |
    PARTICIPANT) + (1 + EXTRACT * VP | NUMBER), family = lognormal(),
    prior = priors.eng, data = eng.model, chains = 4, iter = 5000,
    warmup = 2500, sample_prior = TRUE, cores = 4)
```

Hypothesis testing

```
## Extraction conditions are slower than Non-Extraction
## conditions:
(hypEng <- hypothesis(model.5.eng, hypothesis = "EXTRACT1>0"))

## Hypothesis Tests for class b:
##       Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob
## 1 (EXTRACT1) > 0     0.08      0.03     0.03     0.13      89.91      0.99
##   Star
## 1    *
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 89.91, decisive evidence that RTs are
## slower in the Extraction conditions than the Non-Extraction
## conditions.
```

```
# Verb Phrase condition is slower than Noun Phrase condition:
(hypEng <- hypothesis(model.5.eng, hypothesis = "VP1>0"))

## Hypothesis Tests for class b:
##   Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob
## 1  (VP1) > 0        0      0.02    -0.03     0.04       1.05      0.51
##   Star
## 1
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 1.05, anecdotal evidence that RTs were
## higher in the VP condition than the NP condition.

## Extraction VP condition is faster than Extraction NP
## condition:
(hypEng <- hypothesis(model.5.eng, hypothesis = "EXTRACT1:VP1<0"))

## Hypothesis Tests for class b:
##           Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio
## 1 (EXTRACT1:VP1) < 0    -0.02      0.02    -0.05     0.02        3.6
##   Post.Prob Star
## 1      0.78
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 3.6, substantial evidence of a facilitatory
## effect in the Extraction VP condition relative to the
## Extraction NP condition.
```

Afrikaans group

```
priors.afr <- c(set_prior("cauchy(500,300)", class = "b", coef = "EXTRACT1"),
    set_prior("normal(180,60)", class = "b", coef = "EXTRACT1:VP1"),
    set_prior("normal(0,10)", class = "b", coef = "VP1"), set_prior("cauchy(0,10)",
        class = "sigma"), set_prior("cauchy(0,10)", class = "sd"))
model.5.afr <- brm(IRI_5 ~ EXTRACT * VP + (1 + EXTRACT * VP |
    PARTICIPANT) + (1 + EXTRACT * VP | NUMBER), family = lognormal(),
    prior = priors.afr, data = afr.model, chains = 4, iter = 5000,
    warmup = 2500, sample_prior = TRUE, cores = 4)
```

Hypothesis testing

```
## Extraction conditions are slower than Non-Extraction
## conditions:
(hypAfr <- hypothesis(model.5.afr, hypothesis = "EXTRACT1>0"))

## Hypothesis Tests for class b:
##        Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob
## 1 (EXTRACT1) > 0     0.08      0.04     0.01     0.14      37.31      0.97
##   Star
## 1    *
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 37.31, very strong evidence that RTs were
## higher in the Extraction than the Non-Extraction
## conditions.

## Verb Phrase condition is slower than Noun
## Phrase condition:
(hypAfr <- hypothesis(model.5.afr, hypothesis = "VP1>0"))

## Hypothesis Tests for class b:
##   Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob
## 1  (VP1) > 0     0.02      0.02    -0.01     0.06       6.49      0.87
##   Star
## 1
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 6.49, substantial evidence that RTs were
## higher in the VP than the NP condition.
```

```
## Extraction VP condition is faster than Extraction NP
## condition:
(hypAfr <- hypothesis(model.5.afr, hypothesis = "EXTRACT1:VP1 < 0"))

## Hypothesis Tests for class b:
##           Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio
## 1 (EXTRACT1:VP1) < 0     0.01      0.02    -0.02     0.05       0.35
##   Post.Prob Star
## 1      0.26
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 0.35, borderline substantial evidence that there was no
## facilitatory effect in this group.
```

C-test models

```
afr.model <- dplyr::filter(spr.df.model, L1 == "AFRIKAANS")
afr.model$CTEST_SCALE <- as.vector(scale(afr.model$CTEST, center = TRUE,
    scale = TRUE))
## Segment 3
priors.3.afr <- c(set_prior("cauchy(180,90)", class = "b", coef = "EXTRACT1"),
    set_prior("cauchy(0,10)", class = "sigma"), set_prior("cauchy(0,10)",
        class = "sd"))
model.3.afrc <- brm(IRI_3 ~ EXTRACT * VP * CTEST_SCALE + (1 +
    EXTRACT * VP | PARTICIPANT) + (1 + EXTRACT * VP | NUMBER),
    family = lognormal(), prior = priors.3.afr, data = afr.model,
    chains = 4, iter = 5000, warmup = 2500, sample_prior = TRUE,
    cores = 4)

## Segment 4
priors.4.afr <- c(set_prior("cauchy(200,100)", class = "b", coef = "EXTRACT1"),
    set_prior("cauchy(0,10)", class = "sigma"), set_prior("cauchy(0,10)",
        class = "sd"))
model.4.afrc <- brm(IRI_4 ~ EXTRACT * VP * CTEST_SCALE + (1 +
    EXTRACT * VP | PARTICIPANT) + (1 + EXTRACT * VP | NUMBER),
    family = lognormal(), prior = priors.4.afr, data = afr.model,
    chains = 4, iter = 5000, warmup = 2500, sample_prior = TRUE,
    cores = 4)

## Segment 5
priors.5.afr <- c(set_prior("normal(180,60)", class = "b", coef = "EXTRACT1:VP1"),
    set_prior("normal(0,10)", class = "b", coef = "VP1"), set_prior("cauchy(0,10)",
        class = "sigma"), set_prior("cauchy(0,10)", class = "sd"))
model.5.afrc <- brm(IRI_5 ~ EXTRACT * VP * CTEST_SCALE + (1 +
    EXTRACT * VP | PARTICIPANT) + (1 + EXTRACT * VP | NUMBER),
    family = lognormal(), prior = priors.5.afr, data = afr.model,
    chains = 4, iter = 5000, warmup = 2500, sample_prior = TRUE,
    cores = 4)
```

Hypothesis testing

```
## Model 3:
## C-test x Extraction effect:
(hypAfr <- hypothesis(model.3.afrc, hypothesis = "EXTRACT1:CTEST_SCALE<0"))

## Hypothesis Tests for class b:
##                 Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio
## 1 (EXTRACT1:CTEST_S... < 0    -0.02      0.02    -0.05     0.02    3.36
##   Post.Prob Star
## 1      0.77
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 3.36, substantial evidence that RTs were
## lower in the Extraction condition for participants with
## higher C-test scores.

## C-test x Phrase Type effect:
(hypAfr <- hypothesis(model.3.afrc, hypothesis = "VP1:CTEST_SCALE<0"))

## Hypothesis Tests for class b:
##              Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio
## 1 (VP1:CTEST_SCALE) < 0        0      0.02    -0.03     0.03       0.91
##   Post.Prob Star
## 1      0.48
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 0.91, anecdotal evidence that RTs were not
## lower in the Verb Phrase condition for participants with
## higher C-test scores.

## C-test x Phrase Type x Extraction effect:
(hypAfr <- hypothesis(model.3.afrc, hypothesis = "EXTRACT1:VP1:CTEST_SCALE<0"))

## Hypothesis Tests for class b:
##                 Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio
## 1 (EXTRACT1:VP1:CTE... < 0    -0.02      0.02    -0.05     0.01    5.86
##   Post.Prob Star
## 1      0.85
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 5.86, substantial evidence that RTs were
## lower in the Extraction VP relative to the Extraction NP
## condition for participants with higher C-test scores.
```

```
## Model 4:
## C-test x Extraction effect:
(hypAfr <- hypothesis(model.4.afrc, hypothesis = "EXTRACT1:CTEST_SCALE<0"))

## Hypothesis Tests for class b:
##                 Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio
## 1 (EXTRACT1:CTEST_S... < 0    -0.01      0.02    -0.05     0.03    1.74
##   Post.Prob Star
## 1      0.64
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 1.74, anecdotal evidence that RTs were lower
## in the Extraction condition for participants with higher
## C-test scores.

## C-test x Phrase Type effect:
(hypAfr <- hypothesis(model.4.afrc, hypothesis = "VP1:CTEST_SCALE<0"))

## Hypothesis Tests for class b:
##              Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio
## 1 (VP1:CTEST_SCALE) < 0    -0.01      0.02    -0.04     0.02       1.81
##   Post.Prob Star
## 1      0.64
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 1.81, anecdotal evidence that RTs were
## lower in the Verb Phrase condition for participants with
## higher C-test scores.

## C-test x Phrase Type x Extraction effect:
(hypAfr <- hypothesis(model.4.afrc, hypothesis = "EXTRACT1:VP1:CTEST_SCALE<0"))

## Hypothesis Tests for class b:
##                 Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio
## 1 (EXTRACT1:VP1:CTE... < 0        0      0.02    -0.03     0.03    1.26
##   Post.Prob Star
## 1      0.56
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 1.26, anecdotal evidence that RTs were lower
## in the Extraction VP relative to the Extraction NP
## condition for participants with higher C-test scores.
```

```
## Model 5:
## C-test x Extraction effect:
(hypAfr <- hypothesis(model.5.afrc, hypothesis = "EXTRACT1:CTEST_SCALE>0"))

## Hypothesis Tests for class b:
##                 Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio
## 1 (EXTRACT1:CTEST_S... > 0     0.04      0.03    -0.01      0.1   11.27
##   Post.Prob Star
## 1      0.92
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 11.27, strong evidence that RTs were higher
## in the Extraction condition for participants with higher
## C-test scores.

## C-test x Phrase Type effect:
(hypAfr <- hypothesis(model.5.afrc, hypothesis = "VP1:CTEST_SCALE>0"))

## Hypothesis Tests for class b:
##              Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio
## 1 (VP1:CTEST_SCALE) > 0        0      0.02    -0.03     0.03       1.15
##   Post.Prob Star
## 1      0.54
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 1.15, anecdotal evidence that RTs were
## higher in the Verb Phrase condition for participants with
## higher C-test scores.

## C-test x Phrase Type x Extraction effect:
(hypAfr <- hypothesis(model.5.afrc, hypothesis = "EXTRACT1:VP1:CTEST_SCALE<0"))

## Hypothesis Tests for class b:
##                 Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio
## 1 (EXTRACT1:VP1:CTE... < 0     0.01      0.02    -0.02     0.04    0.59
##   Post.Prob Star
## 1      0.37
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 0.59, anecdotal evidence that RTs were not
## lower in the Extraction VP relative to the Extraction NP
## condition for participants with higher C-test scores.
```

Working memory models

```
spr.wm.afr <- filter(spr.wm, L1 == "AFRIKAANS")
priors.3.wm <- c(set_prior("cauchy(180,90)", class = "b", coef = "EXTRACT1"),
    set_prior("cauchy(0,10)", class = "sigma"), set_prior("cauchy(0,10)",
        class = "sd"))
model.3.wm <- brm(IRI_3 ~ (prop.score_scale + EXTRACT + VP)^3 +
    (1 | PARTICIPANT) + (1 | NUMBER), family = lognormal(), prior = priors.3.wm,
    data = spr.wm.afr, chains = 4, iter = 5000, warmup = 2500,
    sample_prior = TRUE, cores = 4)

priors.4.wm <- c(set_prior("cauchy(200,100)", class = "b", coef = "EXTRACT1"),
    set_prior("cauchy(0,10)", class = "sigma"), set_prior("cauchy(0,10)",
        class = "sd"))
model.4.wm <- brm(IRI_4 ~ (prop.score_scale + EXTRACT + VP)^3 +
    (1 | PARTICIPANT) + (1 | NUMBER), family = lognormal(), prior = priors.4.wm,
    data = spr.wm.afr, chains = 4, iter = 5000, warmup = 2500,
    sample_prior = TRUE, cores = 4)

priors.5.wm <- c(set_prior("cauchy(500,300)", class = "b", coef = "EXTRACT1"),
    set_prior("cauchy(0,10)", class = "sigma"), set_prior("cauchy(0,10)",
        class = "sd"))
model.5.wm <- brm(IRI_5 ~ (prop.score_scale + EXTRACT + VP)^3 +
    (1 | PARTICIPANT) + (1 | NUMBER), family = lognormal(), prior = priors.5.wm,
    data = spr.wm.afr, chains = 4, iter = 5000, warmup = 2500,
    sample_prior = TRUE, cores = 4)
```

Hypothesis testing

```
## Main effects of Working Memory:

## Model 3:
(hypAfrwm <- hypothesis(model.3.wm, hypothesis = "prop.score_scale<0"))

## Hypothesis Tests for class b:
##                Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob Star
## 1 (prop.score_scale) < 0    -0.06      0.06    -0.17     0.04       5.46      0.85
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 5.46, substantial evidence that RTs were
## lower for participants with higher working memory.

## Model 4:
(hypAfrwm <- hypothesis(model.4.wm, hypothesis = "prop.score_scale<0"))

## Hypothesis Tests for class b:
##                Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob Star
## 1 (prop.score_scale) < 0    -0.03      0.05    -0.11     0.06       2.52      0.72
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 2.52, anecdotal evidence that RTs were lower
## for participants with higher working memory.
```

```
## Model 5:
(hypAfrwm <- hypothesis(model.5.wm, hypothesis = "prop.score_scale>0"))

## Hypothesis Tests for class b:
##                 Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob Star
## 1 (prop.score_scale) > 0     0.02      0.05    -0.05      0.1       2.42      0.71
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 2.42, anecdotal evidence that RTs were
## higher for participants with higher working memory.

## Interaction between Working Memory and Extraction:

## Model 3:
(hypAfrwm_ext <- hypothesis(model.3.wm, hypothesis = "prop.score_scale:EXTRACT1<0"))

## Hypothesis Tests for class b:
##                   Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob Sta
r
## 1 (prop.score_scale... < 0    -0.02      0.02    -0.04     0.01       4.66      0.82
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 4.66, substantial evidence that RTs in the
## Extraction conditions were lower for participants with
## higher working memory.

## Model 4:
(hypAfrwm_ext <- hypothesis(model.4.wm, hypothesis = "prop.score_scale:EXTRACT1>0"))

## Hypothesis Tests for class b:
##                   Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob Sta
r
## 1 (prop.score_scale... > 0     0.01      0.02    -0.02     0.04       3.26      0.76
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 3.26, substantial evidence that RTs in the
## Extraction conditions were higher for participants with
## higher working memory.
```

```r
## Model 5:
(hypAfrwm_ext <- hypothesis(model.5.wm, hypothesis = "prop.score_scale:EXTRACT1>0"))

## Hypothesis Tests for class b:
##                 Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob Sta
r
## 1 (prop.score_scale... > 0     0.04      0.02     0.01     0.07      39.16      0.98
*
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 39.16, very strong evidence that RTs in the
## Extraction conditions were higher for participants with
## higher working memory.


## Interaction between Working Memory, Extraction and Phrase
## Type:
(hypAfrwm_extVP <- hypothesis(model.5.wm, hypothesis = "prop.score_scale:EXTRACT1:VP1>0"))

## Hypothesis Tests for class b:
##                 Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob Sta
r
## 1 (prop.score_scale... > 0     0.02      0.02    -0.01     0.05       5.63      0.85
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.

## Bayes Factor = 5.63, substantial evidence that RTs in the
## Extraction VP condition were higher relative to the
## Extraction NP condition for participants with higher
## working memory.
```

7. GENERAL DISCUSSION

7.1 Overview of the study

In recent years, online psycholinguistic techniques have increasingly been used to study L2 processing as part of an investigation into the potential limits (or lack thereof) on ultimate attainment in a non-native language. Section 1.3 of this dissertation identified three phenomena that have been found to be processed differently in L2 compared to L1 speakers: garden-path sentences, referential dependencies, and long-distance wh-dependencies. Section 1.4 then introduced a number of factors that seem to contribute to L2 processing nativelikeness. Among these are an early age of L2 acquisition, extensive naturalistic L2 exposure, and high L2 proficiency. Second-language speakers with this combination of characteristics are understudied in the L2 processing literature, and therefore the extent to which their processing behavior aligns with or diverges from that of native speakers is currently unclear. In order to address this knowledge gap, the dissertation examined an L2 population of exactly this profile, namely L1 Afrikaans–L2 English speakers in South Africa; a country in which, despite considerable linguistic diversity, English holds a privileged position as the primary language of education, commerce and public office (Posel & Zeller, 2016).

Three main research questions were posed in section 1.7. The research articles in chapters 4–6 each set out to answer one of these three questions, repeated for ease of reference below:

1. To what extent are there differences between L1 and L2 speakers' abilities to reanalyze temporarily ambiguous input during real-time processing?

2. To what extent is bilingual processing of English pronouns guided by English structural constraints on pronoun reference assignment?
3. Do L1 and L2 speakers rely equally on abstract syntactic structures to process sentences involving long-distance wh-dependencies?

This concluding chapter begins by summarizing the answers to each of these questions. Subsequently, the findings are considered in concert in order to draw out general implications for the study of L2 processing. The limitations of the study are then discussed, and recommendations for future research are provided.

7.1.1 To what extent are there differences between L1 and L2 speakers' abilities to reanalyze temporarily ambiguous input during real-time processing?

The results of the experiment presented in chapter 4 revealed L1–L2 convergence in reanalysis behavior when the revision of the initial incorrect parse allowed the temporarily ambiguous NP to remain within the thematic domain of the preceding verb, as is the case, for example, in *The boy wrote the report about the budget would start an important debate*. For these so-called complement-clause constructions, both L1 and L2 speakers showed only a mild processing disruption at the region of the sentence (beginning with *would* in the example above) that signaled that reanalysis was necessary.

In the other type of construction that was considered, the temporarily ambiguous NP had to be reanalyzed as the subject of a separate main clause, as, for example, in *While the men drank the beer imported from Europe pleased all the customers*. For these so-called adjunct-clause constructions, reanalysis behavior converged across the L1 speakers and early L2 acquirers (AoA < 5), but the later L2 acquirers (AoA ≥ 5) showed a different processing pattern. Specifically, the L1 speakers and early L2 acquirers clearly initiated reanalysis during real-time processing, but there is no evidence that this was the case for the later L2 acquirers. Notably, however, comprehension question accuracy was equivalent across all participants, indicating that the later L2 acquirers were still able to recover a correct interpretation of the input. Overall, the findings suggest that when revisions to an initial parse require significant structural modifications, L2 speakers' ability to initiate reanalysis in real-time may drop off following a certain, relatively early AoA.

7.1.2 To what extent is bilingual processing of English pronouns guided by English structural constraints on pronoun reference assignment?

The findings presented in chapter 5 revealed that the Afrikaans–English bilinguals did not adhere to Binding Principle B in English when the pronoun was a co-argument of a verb whose Afrikaans translation equivalent allows an accompanying pronominal form to be reflexively interpreted (i.e., what is termed here a "Potentially Reflexive" verb). However, when the verb's translation equivalent did not allow this possibility, they did show evidence of having adhered to the English structural constraints. It was proposed that this verb type effect arises through cross-language activation, where the English verb activates its Afrikaans counterpart in the lexicon, and the reflexivity information linked to the Afrikaans verb is shared with the English equivalent by means of lemmatic transfer. The resulting influence of the Afrikaans verb information then has consequences for processing at the sentence level.

The effects of the verb type manipulation were found in both an offline antecedent choice task and an online eye-tracking-while-reading task. The eye-tracking results indicated that Binding Principle B was violated immediately for items containing Potentially Reflexive verbs, suggesting that the Afrikaans verb type information was recruited from the outset of the reference resolution process.

7.1.3 Do L1 and L2 speakers rely equally on abstract syntactic structures to process sentences involving long-distance wh-dependencies?

The analysis of the self-paced reading data presented in chapter 6 provided evidence that the L1 English group did make use of abstract syntactic structures in the processing of the experimental items, but suggested that this processing strategy was not employed by the L2 group. The findings for the L2 group therefore align with those previously obtained for classroom L2 learners (e.g., Marinis et al., 2005; Pliatsikas & Marinis, 2013), rather than those obtained for immersed L2 speakers (e.g., Pliatsikas et al., 2017; Pliatsikas & Marinis, 2013). Instead of making use of the clause boundary as a site for reactivation of the moved element, the data suggest that the L2 participants maintained the moved element in working memory until they reached the subcategorizing verb, at which point they performed reintegration.

7.2 Implications

A consideration of the implications of the study's findings as a whole should begin with an indication of how the three sets of experimental findings presented in chapters 4–6 relate to one another. The processing phenomena that were examined were either forwards-looking – requiring the maintenance of a moved element in working memory until its reintegration site is reached (chapter 6) – or backwards-looking, requiring previously constructed parts of a syntactic representation to be accessed and/or manipulated (chapters 4 and 5). Aspects of the L2 group's behavior on each of these tasks can be classified as non-nativelike. Firstly, while they successfully processed the long-distance wh-dependencies, they did so without making use of the so-called intermediate gap provided by the clause boundary, unlike the L1 speakers tested here and in previous studies. Secondly, a subset of the relatively later L2 acquirers (AoA ≥ 5), unlike the L1 speakers and earlier L2 acquirers, did not initiate extensive structural revisions to an incorrect parse during real-time processing. Thirdly, the pronoun resolution strategies of the bilinguals in the experiment reported on in chapter 5 show clear evidence of CLI, both offline and online. Further, even aside from the CLI manipulation, the results of this experiment show interference effects that have not previously been observed in monolingual pronoun resolution.

One of the study's central findings, then, is the indication that the particular characteristics of the L2 population studied here – namely an early age of L2 acquisition, extensive naturalistic L2 exposure, and high L2 proficiency – are themselves not sufficient to engender L1-like processing behavior (chapters 4 and 6) or to give rise to the prioritization of L2 structural constraints on reference resolution (chapter 5). This finding speaks to the models of L2 processing that were discussed in section 1.5 in a number of ways, which are considered in turn below. Before proceeding, though, it should be reiterated that the dissertation did not aim to produce evidence for or against any particular model, and so the discussion that is offered here confines itself to a consideration of how the present findings relate to the models that were introduced in chapter 1.

Beginning with the role of AoA: the fact that the early L2 acquirers examined here did not pattern across the board with the L1 group in chapter 4 or chapter 6, and showed susceptibility to CLI and interference in chapter 5, prompts some consideration of the role AoA may have played in the extent of their processing (non)nativelikeness. Models such as the SSH (Clahsen and Felser 2006a, 2006b, 2006c, 2017) and the declarative/procedural model (Ullman, 2001b) predict nativelike processing behavior among earlier L2 acquirers. The present findings may raise questions about this prediction and/or about the nature of early acquisition in the South African context. Regarding the former, one possibility, already raised in recent work from Clahsen and Felser's lab (e.g., Bosch et al., 2019, Veríssimo, 2018, Veríssimo et al., 2018; see also Long, 1990), is that multiple sensitive periods may exist for the various systems and sub-systems involved in grammatical processing (see e.g., the discussion of AoA effects on inflectional priming versus stem priming in Veríssimo, 2018). It may therefore be that the average AoA of the L2 speakers studied here would reveal age effects on the processing of some syntactic phenomena, but not others. For example, as discussed above, the processes involved in revising an incorrect parse are not identical to those involved in resolving long-distance wh-dependencies, and so the duration of sensitive periods for the development of nativelikeness in these domains may differ.

Alternatively, or additionally, there is a consideration regarding the nature of the onset of L2 acquisition in the South African setting. Age of acquisition – especially for younger learners – is generally understood as "the age at which immersion in the L2 begins in earnest" (Birdsong, 2014, p. 374). That is, in the vast majority of studies on ultimate attainment to date, AoA marks the point at which a childhood L2 learner moves to a country or region where the L2 is the primary language and thus becomes immersed in the L2. Such a situation does not apply in the South African context, where for the majority of English learners, exposure to this language commences and persists alongside exposure to the L1. This is certainly the case for the L2 group in the present study: because Afrikaans is spoken as L1 by approximately 50% of the Western Cape's population (Census, 2011), it has a strong societal presence alongside English. Of course, an early start does not guarantee nativelike L2 outcomes (Birdsong, 2014; Munoz, 2008) – even when this early start occurs in naturalistic settings and not the classroom – and the quantity of exposure received to the language in question plays a role in levels of ultimate attainment. This is perhaps most clearly illustrated in heritage language acquisition, where insufficient L1 input in the early years leads to divergent L1 outcomes (see Montrul, 2008; Polinsky & Scontras, 2019 for discussion). In the present study, the same applies to L2 outcomes: insufficient L2 input is likely to have consequences for the development of the L2 system, again with different aspects of processing potentially being affected differently, as has been found to be the case for heritage language speakers (see the references in Jacob, Şafak, Demir, & Kırkıcı, 2018, p. 175).

Relatedly, and aside from potential effects of AoA, we also did not observe nativelike L2 processing of long-distance wh-dependencies in chapter 6, despite previous findings (Pliatsikas et al., 2017; Pliatsikas & Marinis, 2013) that extensive naturalistic L2 exposure is sufficient for the development of nativelike processing in this domain among later L2 acquirers. Plausibly, the fact that our L2 speakers' naturalistic L2 exposure is not immersive may account for this as well; for example by limiting processing automaticity in the L2 (Caffarra et al., 2015) and consequently reducing sensitivity to parsing cues that are not overtly marked, such as clause boundaries

(Felser, 2015). Another consideration, raised in section 1.6 of the dissertation, relates to the nature of the naturalistic exposure that is received. Given that L1 English speakers make up a minority of the country's population, the majority of interactions in English occur between non-native speakers of the language. While lingua francae have received little attention in the acquisition and processing literature (see Canagarajah, 2007, and Ortega, 2018, for discussion), the syntax of such varieties has been described as characterized by "transfer phenomena, developmental patterns and nativised forms and ... simplification, regularisation and levelling processes" (Meierkord, 2004, p. 128). While it is not obvious that exposure to lingua franca English would affect use of the intermediate gap in long-distance wh-dependency processing, the probabilistic nature of processing in general – where predictions about upcoming input are shaped by previous language exposure (see Gibson, Bergen, & Piantadosi, 2013; MacDonald et al., 1994) – entails that the nature of the input received will have consequences for processing (see also Bylund, in press).

The present findings align with previous literature (see section 1.4.2) in indicating that L1–L2 processing differences may persist even at high levels of L2 proficiency. Aside from AoA and an insufficient amount of L2 exposure, which may prevent proceduralization of L2 grammatical knowledge (Ullman, 2001b), this fact has also been explained on the basis of the increased cognitive demands imposed by L2 processing (e.g., Hopp, 2010; McDonald, 2006). The studies presented in this dissertation do not address such accounts directly, given that early L2 acquirers were tested. However, the findings presented in the supplementary materials to chapter 6 suggest that even among such L2 speakers, cognitive capacity may have some effect on L2 processing. In this study, L2 participants with higher WMC were found to maintain the filler in working memory as parsing continued, while lower-WMC participants seemed to wait until the entire sentence had been read before deriving a global interpretation thereof. Again, that there does seem to be an effect of WMC on L2 processing among our L2 participants may be a consequence of the acquisition context: perhaps non-immersive early L2 exposure prompts greater reliance on general cognitive mechanisms during L2 processing than would be found among L2 speakers who received extensive exposure during early childhood.

Finally, the interference effects observed in chapter 5 align with Cunnings' (2017a, 2017b) account of bilingual processing, which explicitly predicts that bilinguals should show susceptibility to interference during memory retrieval. The findings in this chapter also lend support to the claim that not only late L2 learners but also early bilinguals may diverge from monolinguals in their antecedent search behavior. The CLI effect evident in these data indicates that one source of this divergence may be differences in verb argument structure across the bilingual's two languages, which, as a result of cross-language activation, may have subsequent effects on higher-level syntactic processing operations. This finding is compatible with Cunnings' (2017b, p. 674) proposal that retrieval processes in one of a bilingual speaker's languages may influence these processes in their other language.

In sum, a selection of this dissertation's findings did bear out certain predictions in the literature about the conditions under which L2 processing should be more (or less) nativelike. The dissertation identified one processing phenomenon where relatively earlier L2 acquirers show more nativelike behavior than relatively later L2 acquirers (chapter 4) and found that reference resolution in early bilinguals may indeed be affected by cross-linguistic differences in reference resolution strategies (chapter 5). That an early AoA and extensive naturalistic L2 exposure did not engender nativelike processing across the board should inform both our knowledge of language acquisition and processing in the South African context and our understanding of the interactions between individual-level and environmental-level effects on language development.

As such, the dissertation's overarching contribution is two-tiered. Firstly, it contributes by virtue of its examination of a context and an L2 population that have not received previous attention in the syntactic processing literature. The results provide an initial indication of the consequences of the South African linguistic setting for L2 processing, when the L2 in question is the societally dominant language. The findings can be considered in light of calls – which have recently increased in urgency (see e.g., Andringa & Godfroid, 2019; Ortega, 2019, and, for the cognitive sciences more generally, Henrich et al., 2010a, 2010b) – to broaden the contexts in which research into language acquisition and processing is conducted, so that our theories can account for variation across settings, social groups, and language combinations.

Secondly, it provides insight into the potential differential effects of AoA and L2 exposure across different environments and learner groups. The majority of research regarding the effects of such factors is conducted in WEIRDer, more linguistically homogeneous settings. As noted above, AoA studies conducted in such settings typically focus on the age at which the learner was first immersed in the L2 environment. Similarly, studies of the effects of naturalistic exposure on processing, few as these may be, deal with immersive naturalistic exposure. Diversifying our research sites – specifically to less WEIRD, more linguistically heterogeneous locales

– not only allows for the generalizability of our conclusions to be put to the test, but indeed enables entirely new research questions to be posed.

In this regard, the present study suggests that naturalistic exposure received in multilingual settings may not yield the same L2 processing outcomes as that received in monolingual settings. Relatedly, the age of first exposure to a language that is one component of a multilingual ecosystem may not necessarily mark the onset of exposure that is sufficient to lead to acquisition. From these observations spring new research questions, for example regarding the potential benefits, if any, of early, non-immersive naturalistic L2 exposure for ultimate attainment in L2 syntactic processing. Expanding our investigations to other multilingual populations within non-WEIRD settings – for example, African language speakers, who tend to have considerably more diverse linguistic repertoires (see e.g., Banda, 2012; Peirce & Ridge, 1997) – would undoubtedly further enrich existing understandings of the role of such factors.

Ultimately, examining language acquisition and processing in multilingual settings such as South Africa calls for careful consideration of how best to characterize both bi-/multilingual participant groups and multilingual contexts. Terms such as ‘early bilingual’ and ‘naturalistic exposure’ may require qualification, and we may need to adopt novel approaches to select and group participants for data collection and analysis. These are challenges that should not be shied away from: rigorous investigations of L2 and bilingual development in such settings are long overdue (see e.g., Ortega, 2018), and are indeed indispensable in arriving at a complete understanding of the human capacity for language.

7.3 Limitations and recommendations for future research

This section discusses the limitations of the studies presented in this dissertation. Interspersed are recommendations for future research that would address these limitations and expand on the current findings.

A first limitation relates to the kinds of processing phenomena that were investigated in the dissertation. At the outset of the project, it was assumed that, given the nature of the L2 population under study, they would almost inevitably show nativelike sensitivity to phenomena such as morphosyntactic agreement violations. In hindsight, considering the findings that were obtained, this assumption may have been unfounded. It would have been useful to have, as a first study, an examination of an aspect of the L2 that has generally been found to be processed in a nativelike fashion by L2 learners with other L1s in other contexts. This would have made it possible to determine whether the aspects of non-nativelikeness observed here are tied only to relatively more complex phenomena, or whether they occur even, for example, in the processing of local dependencies. This is a question worth examining in future studies. As Afrikaans does not mark person and number features on the verb, an investigation of subject–verb agreement processing in the L2 would provide an indication of whether these L2 speakers show nativelike sensitivity to a feature that is absent in their L1.

A second limitation relates to the entanglement of AoA and L2 exposure in our L2 sample. It would be desirable to be able to isolate the effects of these two factors on L2 processing. For future studies, one possibility would be to compare two groups, matched in current L2 exposure and general proficiency, where one group had early childhood exposure to limited quantities of input and the other group's L2 exposure commenced only later in life.

A third limitation relates to the lack of an objective measure of the L2 participants' language dominance, which is "a measure of relative frequency of use and proficiency in each language" (Wang, 2013, p. 768, cited in Treffers-Daller & Korybski, 2015). In this regard, while the dissertation's L2 participants on average report approximately equal exposure to both of their languages, contained within this average are a number of individuals who report considerably higher L2 than L1 exposure. Although this fact does not compromise the conclusions drawn here in any way, it does leave unexplored the relationship between relative L2 dominance and L2 processing nativelikeness. Research indicates that dominance shifts do have consequences for processing (see e.g., Caffarra, Zimnukhova, & Mancini, 2016; Dussias & Sagarra, 2007; Kasparian & Steinhauer, 2017; Kasparian, Vespignani, & Steinhauer, 2017), and so the possibility of shifts in dominance should be taken into account in future processing studies. One way to do this is by relying on self-reports; certain language background questionnaires (e.g., Li et al., 2014), for example, provide an estimate of dominance based on self-reported proficiency, exposure and use across a bilingual's two languages. Alternatively, an objective measure of dominance could be obtained based on a comparison of the bilingual individual's relative performance in their two languages. Naming tasks (e.g., the Boston Naming Test, Kaplan, Goodglass, & Weintraub, 1978, or a version thereof adapted for the South African context, the Groote Schuur Naming Test, Mosdell, Balchin, & Ameen, 2010) provide one relatively hassle-free way of doing this, as they can provide a measure of naming fluency and accuracy in one language relative to the other.

The question of dominance relates to a more general consideration about the classification of bilinguals in quantitative research. The relevant point, which was raised as early as the 1960s (Fishman & Cooper, 1969), is made succinctly in the title of Luk and Bialystok's (2013) paper: "Bilingualism is not a categorical variable". This statement reiterates that within bilingual populations, there is inevitably variation in AoA, proficiency, and frequency of exposure and use.

The extent to which such variation affects processing outcomes within a predefined participant group can be investigated in multiple ways. One approach, for example, is to incorporate background variables such as AoA and proficiency into the data analysis procedure (as is done, for example, in chapters 4 and 6, respectively). An alternative is to seek out sub-groups within a predefined group that pattern similarly to one another. Quantitative approaches such as cluster analysis can be used to achieve this goal. The primary aim of cluster analysis, given a set of data consisting of objects with specific characteristics, is to identify structure in the data set such that the objects can be grouped into clusters whose members are maximally similar to one another, but maximally dissimilar to the members of other clusters (Gries, 2010, p. 285).

Clusters can be identified on the basis of either participants' background characteristics, or their performance on one or more dependent measures. An example of the former is found in von Bastian et al. (2013), where groups of participants with different levels of bilingualism were identified based on AoA, language usage and language proficiency. The performance of these groups on a battery of cognitive tests was then assessed in order to determine the extent to which varying levels of bilingualism affected cognitive performance. Similarly, Sparks, Patton and Ganschow

(2012) (see also Rysiewicz, 2008; Skehan, 1986) formed clusters of learners on the basis of L1 literacy, intelligence and L2 aptitude in order to investigate whether they would pattern differently in terms of L2 outcomes. An example of forming clusters based on a dependent measure is found in Diaz, Mitterer, Broersma, Escera, and Sebastián-Gallés (2016), who used performance on three phoneme perception tasks to identify sub-groups of bilinguals who were either ‘good’ or ‘poor’ perceivers.

The bottom-up, data-driven nature of cluster analysis and other similar approaches can be especially useful for data exploration and hypothesis generation in relation to relatively understudied research contexts and participant groups. As such, tools of this type may be useful in future explorations of bilingual and L2 populations in linguistically diverse settings such as South Africa.

7.4 Concluding remarks

This dissertation examined L2 syntactic processing in a multilingual setting that has gone unexplored in the psycholinguistic literature. The primary aim was to determine the extent to which the L2 group, made up of high-proficiency L2 speakers with an early AoA and extensive naturalistic L2 exposure, showed nativelike processing of a selection of syntactic phenomena. The findings revealed L1–L2 convergence for a subset of the L2 participants on one processing task, but non-nativelike L2 processing elsewhere. These results contribute to our current knowledge of the conditions under which L2 processing nativelikeness is most likely and, conversely, of the conditions under which non-nativelikeness is likely to persist, despite the presence of certain factors generally deemed beneficial for levels of ultimate L2 attainment. More generally, the results provide insight into the nature of L2 processing in one non-

WEIRD, highly linguistically diverse environment. It is hoped that this will be the first of many such investigations conducted in South Africa and other naturally multilingual settings.

REFERENCES

Altmann, G., Garnham, A., & Dennis, Y. (1992). Avoiding the garden path: Eye movements in context. *Journal of Memory and Language, 31*(5), 685–712.

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

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

Andringa, S., & Godfroid, A. (2019). SLA for all?: Reproducing second language acquisition research in non-academic samples. *Language Learning,* 5–9.

Avery, N., & Marsden, E. (2019). A meta-analysis of sensitivity to grammatical information during self-paced reading: Towards a framework of reference for reading time effect sizes. *Studies in Second Language Acquisition*, 1–33.

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. doi:10.1016/j.jml.2007.12.005

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Recent Advances in Learning and Motivation* (pp. 47–89). New York: Academic Press.

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(4), 748–769. doi:10.1037//0278-7393.28.4.748

Bancroft, T. D., Jones, J. A., Ensor, T. M., Hockley, W. E., & Servos, P. (2016). Overwriting and intrusion in short-term memory. *Memory & Cognition, 44*(3), 435–443. doi:10.3758/s13421-015-0570-y

Banda, F. (2012). Critical perspectives on language planning and policy in Africa: Accounting for the notion of multilingualism. *Stellenbosch Papers in Linguistics Plus, 38*(0). doi:10.5842/38-0-50

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. doi:10.1016/j.jml.2012.11.001

von Bastian, Claudia C., Locher, A., & Ruflin, M. (2013). Tatool: A Java-based open-source programming framework for psychological studies. *Behavior research methods, 45*(1), 108–115. doi:10.3758/s13428-012-0224-y

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models Using lme4. *Journal of Statistical Software, 67*(1). doi:10.18637/jss.v067.i01

Bernolet, S., & Hartsuiker, R. J. (2018). Syntactic representations in late learners of a second language. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Bilingual Cognition and Language* (pp. 205–224). Amsterdam: John Benjamins Publishing Company.

Bertenshaw, N. (2009). *The Application of Binding Constraints by Japanese L2 Learners of English.* (PhD). University of Essex, Essex.

Bigelow, M., & Tarone, E. (2004). The role of literacy level in second language acquisition: Doesn't who we study determine what we know? *TESOL Quarterly, 38*(4), 689. doi:10.2307/3588285

Birdsong, D. (2014). Dominance and age in bilingualism. *Applied Linguistics, 35*(4), 374–392. doi:10.1093/applin/amu031

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning, 56*(Suppl. 1), 9–49. doi:10.1111/j.1467-9922.2006.00353.x

Birdsong, D. (2018). Plasticity, variability and age in second language acquisition and bilingualism. *Frontiers in Psychology, 9,* 81. doi:10.3389/fpsyg.2018.00081

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. doi:10.1075/lia.4.2.01bir

Birdsong, D., & Vanhove, J. (2016). Age of second language acquisition: Critical periods and social concerns. In E. Nicoladis & S. Montanari (Eds.), *Bilingualism across the lifespan: Factors moderating language proficiency* (pp. 163–181). Washington: American Psychological Association.

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

Bley-Vroman, R. (2009). The evolving context of the fundamental difference hypothesis. *Studies in Second Language Acquisition, 31*(2), 175–198. doi:10.1017/s0272263109090275

Bolhuis, J. J., Okanoya, K., & Scharff, C. (2010). Twitter evolution: Converging mechanisms in birdsong and human speech. *Nature Reviews Neuroscience, 11*(11), 747.

Bosch, S., Veríssimo, J., & Clahsen, H. (2019). Inflectional morphology in bilingual language processing: An age-of-acquisition study. *Language Acquisition, 26*(3), 339–360. doi:10.1080/10489223.2019.1570204

Boxell, O., & Felser, C. (2017). Sensitivity to parasitic gaps inside subject islands in native and non-native sentence processing. *Bilingualism: Language and Cognition, 20*(03), 494–511. doi:10.1017/S1366728915000942

Breen, M. (2014). Empirical investigations of the role of implicit prosody in sentence processing. *Language and Linguistics Compass, 8*(2), 37–50. doi:10.1111/lnc3.12061

Britt, M. A. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language*, *33*(2), 251–283.

Britt, M. A., Perfetti, C. A., Garrod, S., & Rayner, K. (1992). Parsing in discourse: Context effects and their limits. *Journal of Memory and Language, 31*(3), 293–314.

Brown, R., & Fish, D. (1983). The psychological causality implicit in language. *Cognition, 14*(3), 237–273.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1). doi:10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal, 10*(1), 395. doi:10.32614/RJ-2018-017

Bylund, E. (in press). WEIRD psycholinguistics. In A. Deumert, Q. WIlliams, & T. Milani (Eds.), *Multilingualism and Linguistic Citizenship. Education, Narrative and Episteme*. Clevedon: Multilingual Matters.

Bylund, E., Abrahamsson, N., & Hyltenstam, K. (2012). Does first language maintenance hamper nativelikeness in a second language?: A study of ultimate attainment in early bilinguals. *Studies in Second Language Acquisition, 34*(2), 215–241.

Bylund, E., Abrahamsson, N., Hyltenstam, K., & Norrman, G. (2019). Revisiting the bilingual lexical deficit: The impact of age of acquisition. *Cognition, 182,* 45–49. doi:10.1016/j.cognition.2018.08.020

Bylund, E., & Athanasopoulous, P. (2014). Language and thought in a multilingual context: The case of isiXhosa. *Bilingualism: Language and Cognition, 17*(02), 431–441. doi:10.1017/S1366728913000503

Bylund, E., Hyltenstam, K., & Abrahamsson, N. (2013). Age of acquisition effects or effects of bilingualism in second language ultimate attainment? In G. Granena & M. Long (Eds.), *Sensitive Periods, Language Aptitude, and Ultimate L2 Attainment* (pp. 69–102). Amsterdam: John Benjamins Publishing Company.

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

Caffarra, S., Zimnukhova, S., & Mancini, S. (2016). What usage can do: The effect of language dominance on simultaneous bilinguals' morphosyntactic processing. *Linguistics Vanguard, 2*(s1), 127. doi:10.1515/lingvan-2016-0020

Canagarajah, S. (2007). Lingua franca English, multilingual communities, and language acquisition. *The Modern Language Journal, 91,* 923–939. doi:10.1111/j.0026-7902.2007.00678.x

Carminati, M. N. (2002). *The processing of Italian subject pronouns* (PhD). University of Massachusetts, Amherst.

Carpenter, P. A., & Just, M. A. (1983). What your eyes do while your mind is reading. In K. Rayner (Ed.), *Eye Movements in Reading: Perceptual and Language Processes* (pp. 275–307). New York: Academic Press.

Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 49*(3), 639–663. doi:10.1080/713755647

Census. (2011). *Census in brief*. Pretoria: Statistics South Africa.

Centre for Lexical Information. (1993). *The Celex lexical database*. Nijmegen: Max Planck Institute for Psycholinguistics.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, Mass.: MIT Press.

Chomsky, N. (1977). *Essays on form and interpretation*. New York: North-Holland.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht, Holland, Cinnaminson, [N.J.]: Foris Publications.

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

Chow, W.-Y., Lewis, S., & Phillips, C. (2014). Immediate sensitivity to structural constraints in pronoun resolution. *Frontiers in Psychology, 5,* 630. doi:10.3389/fpsyg.2014.00630

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology, 42*(4), 368–407. doi:10.1006/cogp.2001.0752

Clackson, K., Felser, C., & Clahsen, H. (2011). Children's processing of reflexives and pronouns in English: Evidence from eye-movements during listening. *Journal of Memory and Language, 65*(2), 128–144. doi:10.1016/j.jml.2011.04.007

Clackson, K., & Heyer, V. (2014). Reflexive anaphor resolution in spoken language comprehension: Structural constraints and beyond. *Frontiers in Psychology, 5,* 904. doi:10.3389/fpsyg.2014.00904

Clahsen, H., & Felser, C. (2006a). Continuity and shallow structures in language processing. *Applied Psycholinguistics, 27*(1), 107–126. doi:10.1017/S0142716406060206

Clahsen, H., & Felser, C. (2006b). Grammatical processing in language learners. *Applied Psycholinguistics, 27*(01), 3–42. doi:10.1017/S0142716406060024

Clahsen, H., & Felser, C. (2006c). How native-like is non-native language processing? *Trends in Cognitive Sciences, 10*(12), 564–570. doi:10.1016/j.tics.2006.10.002

Clahsen, H., & Felser, C. (2017). Some notes on the Shallow Structure Hypothesis. *Studies in Second Language Acquisition, 24,* 1–14. doi:10.1017/S0272263117000250

Clahsen, H., Felser, C., Neubauer, K., Sato, M., & Silva, R. (2010). Morphological structure in native and nonnative language processing. *Language Learning, 60*(1), 21–43. doi:10.1111/j.1467-9922.2009.00550.x

Clahsen, H., & Veríssimo, J. (2016). Investigating grammatical processing in bilinguals. *Linguistic Approaches to Bilingualism, 6*(5), 685–698.

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

Clifton, C., Frazier, L., & Deevy, P. (1999). Feature manipulation in sentence comprehension. *Italian Journal of Linguistics, 11*(1), 11–40.

Clifton, C., Kennison, S. M., & Albrecht, J. E. (1997). Reading the words her, his, him: Implications for parsing principles based on frequency and on structure. *Journal of Memory and Language, 36*(2), 276–292. doi:10.1006/jmla.1996.2499

Clifton Jr, C., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language, 49*(3), 317–334.

Coetzee-Van Rooy, S. (2012). Flourishing functional multilingualism: Evidence from language repertoires in the Vaal Triangle region. *International Journal of the Sociology of Language*, 2012(218), 87–119. doi:10.1515/ijsl-2012-0060

Conklin, K., & Pellicer-Sánchez, A. (2016). Using eye-tracking in applied linguistics and second language research. *Second Language Research, 32*(3), 453–467. doi:10.1177/0267658316637401

Contemori, C., Asiri, O., & Perea Irigoyen, Elva Deida. (2019). Anaphora resolution in L2 English. *Studies in Second Language Acquisition, 1,* 1–28. doi:10.1017/S0272263119000111

Conti-Ramsden, G., Ullman, M. T., & Lum, Jarrad A G. (2015). The relation between receptive grammar and procedural, declarative, and working memory in specific language impairment. *Frontiers in Psychology, 6,* 1090. doi:10.3389/fpsyg.2015.01090

Cook, V. J. (1990). Timed comprehension of binding in advanced L2 learners of English. *Language Learning, 40*(4), 557–599.

Cook, V. (1999). Going beyond the native speaker in language teaching. *TESOL Quarterly,* 33, 185.

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the syntactic processor. In D. R. Dowty, L. Kartunnen, & A. M. Zwicky (Eds.), *Natural Language parsing: Psychological, Computational and Theoretical Perspectives* (pp. 320–358). Cambridge: Cambridge University Press.

Crawley, R. A., Stevenson, R. J., & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research, 19*(4), 245–264.

Cuetos, F. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition, 30*(1), 73–105. doi:10.1016/0010-0277(88)90004-2

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

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

Cunnings, I., & Felser, C. (2013). The role of working memory in the processing of reflexives. *Language and Cognitive Processes, 28*(1-2), 188–219. doi:10.1080/01690965.2010.548391

Cunnings, I., Fotiadou, G., & Tsimpli, I. (2017). Anaphora resolution and reanalysis during L2 sentence processing. *Studies in Second Language Acquisition, 39*(04), 621–652. doi:10.1017/S0272263116000292

Curtiss, S. (1988). Abnormal language acquisition and the modularity of language. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge Survey* (pp. 96–116). Cambridge: Cambridge University Press.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech, 40*(2), 141–201.

Dallas, A., & Kaan, E. (2008). Second language processing of filler–gap dependencies by late learners. *Language and Linguistics Compass, 2*(3), 372–388. doi:10.1111/j.1749-818X.2008.00056.x

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*(4), 422–433. doi:10.3758/BF03214546

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450–466.

Dekydtspotter, L., Schwartz, B. D., & Sprouse, R. A. (2006). The comparative fallacy in L2 processing research. In M. Grantham O'Brien, C. Shea, & J. Archibald (Eds.), *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference* (pp. 33–40). Somerville, MA: Cascadilla Press.

Department of Basic Education. *The status of the language of learning and teaching (LoLT) in South African public schools: A quantitative overview.* Pretoria: Department of Basic Education.

Diaz, B., Mitterer, H., Broersma, M., Escera, C., & Sebastián-Gallés, N. (2016). Variability in L2 phonemic learning originates from speech-specific capabilities: An MMN study on late bilinguals. *Bilingualism: Language and Cognition, 19*(5), 955–970. doi:10.1017/S1366728915000450

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology, 5,* 781. doi:10.3389/fpsyg.2014.00781

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology, 72,* 78–89. doi:10.1016/j.jmp.2015.10.003

Dijkstra, T. (2005). Bilingual visual word recognition and lexical access. In J. F. Kroll & A. de Groot (Eds.), *Handbook of Bilingualism. Psycholinguistic Approaches* (pp. 179–201). Oxford: Oxford Univ. Press.

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language, 69*(2), 85–103. doi:10.1016/j.jml.2013.04.003

Dillon, B. W. (2011). *Structured access in sentence comprehension* (PhD). University of Maryland, Maryland.

Dimitriadis, A., Everaert, M., & Reuland, E. (2017). The world of reflexives: An introduction. *Studia Linguistica: A Journal of General Linguistics, 71*(1), 1–11.

Dowens, M. G., Vergara, M., Barber, H. A., & Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience, 22*(8), 1870–1887.

Drummer, J.-D., & Felser, C. (2018). Cataphoric pronoun resolution in native and non-native sentence comprehension. *Journal of Memory and Language, 101,* 97–113. doi:10.1016/j.jml.2018.04.001

Dussias, P. E. (2003). Syntactic ambiguity resolution in L2 learners. *Studies in Second Language Acquisition, 25*(04). doi:10.1017/S0272263103000238

Dussias, P. E., & Cramer Scaltz, Tracy 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. doi:10.1016/j.actpsy.2007.09.004

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

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish–English bilinguals. *Bilingualism: Language and Cognition, 10*(01), 101. doi:10.1017/S1366728906002847

Eberhard, D. M., Simons, G. F., Fennig, C. D. (2019). *Ethnologue: Languages of the World*. Retrieved from http://www.ethnologue.com/

Ellert, M. (2013). Resolving ambiguous pronouns in a second language: A visual-world eye-tracking study with Dutch learners of German. *International Review of Applied Linguistics in Language Teaching, 51*(2), 171–197. doi:10.1515/iral-2013-0008

Ellis, R., & Roever, C. (2018). The measurement of implicit and explicit knowledge. *The Language Learning Journal, 21*(5), 1–16. doi:10.1080/09571736.2018.1504229

Fabbro, F. (2015). *The neurolinguistics of bilingualism: An introduction*. Hove, East Sussex: Psychology Press.

Felser, C. (2015). Native vs. non-native processing of discontinuous dependencies. *Second Language*, (14), 5–19.

Felser, C., & Cunnings, I. (2012). Processing reflexives in a second language: The timing of structural and discourse-level constraints. *Applied Psycholinguistics, 33*(03), 571–603. doi: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. doi:10.1017/s0272263111000507

Felser, C., Marinis, T., & Clahsen, H. (2003). Children's processing of ambiguous sentences: A study of relative clause attachment. *Language Acquisition, 11*(3), 127–163. doi:10.1207/s15327817la1103_1

Felser, C., & Roberts, L. (2007). Processing wh -dependencies in a second language: A cross-modal priming study. *Second Language Research, 23*(1), 9–36. doi:10.1177/0267658307071600

Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics, 24*(3), 453–489. doi:10.1017/S0142716403000237

Felser, C., Sato, M., & Bertenshaw, N. (2009). The on-line application of binding Principle A in English as a second language. *Bilingualism: Language and Cognition, 12*(04), 485. doi:10.1017/S1366728909990228

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47*(2), 164–203.

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*(1), 11–15. doi:10.1111/1467-8721.00158

Ferreira, F., & Clifton Jr, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*(3), 348–368.

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language, 30*(6), 725–745.

Ferreira, F., & Patson, N. D. (2007). The "good enough" approach to language comprehension. *Language and Linguistics Compass, 1*(1-2), 71–83. doi:10.1111/j.1749-818X.2007.00007.x

Fishman, J. A., & Cooper, R. L. (1969). Alternative measures of bilingualism. *Journal of Verbal Learning and Verbal Behavior, 8*(2), 276–282.

Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition, 23*(4), 527–552.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language, 41*(1), 78–104.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge: MIT Press.

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies*. West Bend, Indiana: Indiana University Linguistics Club.

Frazier, L., Carlson, K., & Clifton, C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences, 10*(6), 244–249. doi:10.1016/j.tics.2006.04.002

Frazier, L., & Clifton, C. (1998). Sentence reanalysis, and visibility. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in Sentence Processing* (pp. 143–176). Dordrecht: Springer Netherlands. doi:10.1007/978-94-015-9070-9_5

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

Frazier, L., & d'Arcais, Giovanni B Flores. (1989). Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language, 28*(3), 331–344.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6*(4), 291–325.

Frazier, L., & Fodor, J. D. (1980). Is the human sentence parsing mechanism an ATN? *Cognition, 8,* 411–459.

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.

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. doi:10.1080/027249897392251

Frenck-Mestre, C. (2002). An on-line look at sentence processing in the second language. In R. R. Heredia & J. Altarriba (Eds.), *Bilingual Sentence Processing* (pp. 217–236). Amsterdam: North-Holland.

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. doi:10.1006/jmla.1997.2512

Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry, 5*(3), 459–464.

Garvey, C., Caramazza, A., & Yates, J. (1974). Factors influencing assignment of pronoun antecedents. *Cognition, 3*(3), 227–243. doi:10.1016/0010-0277%2874%2990010-9

Gernsbacher, M. A., Hargreaves, D. J., & Beeman, M. (1989). Building and accessing clausal representations: The advantage of first mention versus the advantage of clause recency. *Journal of Memory and Language, 28*(6), 735–755.

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. doi:10.1177/1367006915604401

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

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences of the United States of America, 110*(20), 8051–8056. doi:10.1073/pnas.1216438110

Gibson, E., & Pearlmutter, N. J. (2000). Distinguishing serial and parallel parsing. *Journal of Psycholinguistic Research, 29*(2), 231–240.

Gibson, E., & Warren, T. (1999). *The psychological reality of intermediate linguistic structure in long-distance dependencies*. Unpublished manuscript.

Gibson, E., & Warren, T. (2004). Reading-time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax, 7*(1), 55–78. doi:10.1111/j.1368-0005.2004.00065.x

Goldrick, M., Putnam, M. T., & Schwarz, L. (2016). Coactivation in bilingual grammars: A computational account of code mixing. *Bilingualism: Language and Cognition, 19*(5), 857–876. doi:10.1017/S1366728915000802

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language, 58*(3), 787–814. doi:10.1016/j.jml.2007.07.001

Gollan, T. H., Slattery, T. J., Goldenberg, D., van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General, 140*(2), 186–209. doi:10.1037/a0022256

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition, 33*(7), 1220–1234. doi:10.3758/BF03193224

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(6), 1411.

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

Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(6), 1304–1321. doi:10.1037/0278-7393.32.6.1304

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research, 29*(3), 311–343. doi:10.1177/0267658312461497

Green, D. W., Crinion, J., & Price, C. J. (2006). Convergence, degeneracy, and control. *Language Learning, 56,* 99–125. doi:10.1111/j.1467-9922.2006.00357.x

Greenberg, J. H. (1956). The measurement of linguistic diversity. *Language, 32*(1), 109. doi:10.2307/410659

Gries, S. T. (2010). Useful statistics for corpus linguistics. In A. Sánchez (Ed.), *A mosaic of corpus linguistics. Selected approaches* (pp. 269–291). Frankfurt am Main: Lang.

Grüter, T., Rohde, H., & Schafer, A. (2014). The role of discourse-level expectations in non-native speakers' referential choices. In W. Orman & M. J. Valleau (Eds.), *Proceedings of the 38th Annual Boston University Conference on Language Development* (pp. 179–191). Somerville, MA: Cascadilla Press.

Grüter, T., Rohde, H., & Schafer, A. J. (2017). Coreference and discourse coherence in L2. *Linguistic Approaches to Bilingualism, 7*(2), 199–229.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In NAACL '01, *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies* (pp. 1–8). Stroudsburg, PA, USA: Association for Computational Linguistics.

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science, 30*(4), 643–672. doi:10.1207/s15516709cog0000_64

Hale, J. (2016). Information-theoretical complexity metrics. *Language and Linguistics Compass, 10*(9), 397–412. doi:10.1111/lnc3.12196

Hale, J. T. (2011). What a rational parser would do. *Cognitive Science, 35*(3), 399–443. doi:10.1111/j.1551-6709.2010.01145.x

Hamrick, P., Lum, Jarrad A G, & Ullman, M. T. (2018). Child first language and adult second language are both tied to general-purpose learning systems. *Proceedings of the National Academy of Sciences of the United States of America, 115*(7), 1487–1492. doi:10.1073/pnas.1713975115

Harald Baayen, R., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research, 3*(2), 12. doi:10.21500/20112084.807

Hartshorne, J. K., & Snedeker, J. (2013). Verb argument structure predicts implicit causality: The advantages of finer-grained semantics. *Language and Cognitive Processes, 28*(10), 1474–1508. doi:10.1080/01690965.2012.689305

Hartshorne, J. K., Sudo, Y., & Uruwashi, M. (2013). Are implicit causality pronoun resolution biases consistent across languages and cultures? *Experimental Psychology*, 60(3), 179–196. doi:10.1027/1618-3169/a000187

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition, 177,* 263–277.

Havik, E., Roberts, L., van Hout, R., Schreuder, R., & Haverkort, M. (2009). Processing subject-object ambiguities in the L2: A self-paced reading study with German L2 learners of Dutch. *Language Learning, 59*(1), 73–112.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010a). Most people are not WEIRD. *Nature, 466*(7302), 29. doi:10.1038/466029a

Henrich, J., Heine, S. J., & Norenzayan, A. (2010b). The weirdest people in the world? *The Behavioral and Brain Sciences, 33*(2-3), 61-83; discussion 83-135. doi:10.1017/S0140525X0999152X

Hill, L., & Bekker, S. (2014). Language, residential space and inequality in Cape Town: Broad-brush profiles and trends. *African Population Studies, 28*(1), 661–680.

Holmes, V. M., Kennedy, A., & Murray, W. S. (1987). Syntactic structure and the garden path. *The Quarterly Journal of Experimental Psychology, 39*(2), 277–293.

Holmes, V. M., Stowe, L., & Cupples, L. (1989). Lexical expectations in parsing complement-verb sentences. *Journal of Memory and Language, 28*(6), 668–689.

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

Hopp, H. (2015). Individual differences in the second language processing of object–subject ambiguities. *Applied Psycholinguistics, 36*(02), 129–173. doi:10.1017/S0142716413000180

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

Hopp, H. (2018). The bilingual mental lexicon in L2 sentence processing. *Second Language, 17,* 5–27.

Hsiao, F., & Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition, 90*(1), 3–27.

Huang, B. H. (2015). A synthesis of empirical research on the linguistic outcomes of early foreign language instruction. *International Journal of Multilingualism, 13*(3), 257–273. doi:10.1080/14790718.2015.1066792

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica, 137*(2), 151–171. doi:10.1016/j.actpsy.2010.11.003

Hummel, K. M. (2009). Aptitude, phonological memory, and second language proficiency in nonnovice adult learners. *Applied Psycholinguistics, 30*(2), 225–249. doi:10.1017/S0142716409090109

Hyltenstam, K. (1992). Non-native features of near-native speakers: On the ultimate attainment of childhood L2 learners. In R. J. Harris (Ed.), *Cognitive Processing in Bilinguals* (pp. 351–368). Elsevier.

Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 538–588). Oxford, UK: Blackwell Publishing Ltd.

Hyltenstam, K., Bylund, E., Abrahamsson, N., & Park, H.-S. (2009). Dominant-language replacement: The case of international adoptees. *Bilingualism: Language and Cognition, 12*(2), 121–140.

Jackson, C. (2008). Proficiency level and the interaction of lexical and morphosyntactic information during L2 sentence processing. *Language Learning, 58*(4), 875–909.

Jacob, G., & Felser, C. (2016). Reanalysis and semantic persistence in native and non-native garden-path recovery. *Quarterly Journal of Experimental Psychology (2006), 69*(5), 907–925. doi:10.1080/17470218.2014.984231

Jacob, G., Şafak, D. F., Demir, O., & Kırkıcı, B. (2018). Preserved morphological processing in heritage speakers: A masked priming study on Turkish. *Second Language Research, 35*(2), 173–194. doi:10.1177/0267658318764535

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434–446. doi:10.1016/j.jml.2007.11.007.

Jäger, L. A., Benz, L., Roeser, J., Dillon, B. W., & Vasishth, S. (2015). Teasing apart retrieval and encoding interference in the processing of anaphors. *Frontiers in Psychology, 6,* 506. doi:10.3389/fpsyg.2015.00506

Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language, 94,* 316–339. doi:10.1016/j.jml.2017.01.004

Janet Choy, J., & Thompson, C. K. (2010). Binding in agrammatic aphasia: Processing to comprehension. *Aphasiology, 24*(5), 551–579. doi:10.1080/02687030802634025

Järvikivi, J., Van Gompel, Roger PG, Hyönä, J., & Bertram, R. (2005). Ambiguous pronoun resolution: Contrasting the first-mention and subject-preference accounts. *Psychological Science, 16*(4), 260–264.

Jarvis, S. (2009). Lexical transfer. In A. Pavlenko (Ed.), *The bilingual mental lexicon. Interdisciplinary approaches* (pp. 99–124). Clevedon: Multilingual Matters.

Jarvis, S. (2010). Comparison-based and detection-based approaches to transfer research. *Eurosla Yearbook, 10*(1), 169–192.

Jarvis, S., & Pavlenko, A. (2010). *Crosslinguistic influence in language and cognition*. London, New York: Routledge.

Jeffreys, H. (1998). *Theory and probability*. Oxford: Clarendon Press.

Jegerski, J. (2012). The processing of subject–object ambiguities in native and near-native Mexican Spanish. *Bilingualism: Language and Cognition, 15*(04), 721–735. doi:10.1017/S1366728911000654

Jessen, A., & Felser, C. (2018). Reanalysing object gaps during non-native sentence processing: Evidence from ERPs. *Second Language Research, 35*(2), 285–300. doi:10.1177/0267658317753030

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*(1), 60–99.

Juffs, A. (1998a). Main verb versus reduced relative clause ambiguity resolution in L2 sentence processing. *Language Learning, 48*(1), 107–147.

Juffs, A. (1998b). Some effects of first language argument structure and morphosyntax on second language sentence processing. *Second Language Research, 14*(4), 406–424.

Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society, 102*(2), 199–225.

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

Juffs, A. (2006). Processing reduced relative versus main verb ambiguity in English as a second language. In R. Slabakova, S. Montrul, & P. Prevost (Eds.), *Inquiries in linguistic development in honor of Lydia White* (pp. 213–232). Amsterdam: John Benjamins Amsterdam.

Juffs, A., & Harrington, M. (1995). Parsing effects in second language sentence processing. *Studies in Second Language Acquisition, 17*(4), 483–516. doi:10.1017/S027226310001442X

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

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching, 44*(2), 137–166. doi:10.1017/S0261444810000509

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111*(2), 228.

Just, M. A., & Carpenter, P. A. (1984). Using eye fixations to study reading comprehension. In D. E. Kieras & M. A. Just (Eds.), *New Methods in Reading Comprehension Research* (pp. 151–182). London: Routledge.

Kaan, E., Ballatyne, J. C., & Wijnen, F. (2015). Effects of reading speed on second-language sentence processing. *Applied Psycholinguistics, 36*(04), 799–830. doi:10.1017/S0142716413000519

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*(1), 133–156. doi:10.1016/S0749-596X(03)00023-8

Kaplan, E., Goodglass, H., & Weintraub, S. (1978). *The Boston Naming Test*. Philadelphia: Lea & Febiger.

Kasparian, K., & Steinhauer, K. (2017). When the second language takes the lead: Neurocognitive processing changes in the first language of adult attriters. *Frontiers in Psychology, 8,* 389. doi:10.3389/fpsyg.2017.00389

Kasparian, K., Vespignani, F., & Steinhauer, K. (2017). First language attrition induces changes in online morphosyntactic processing and re-analysis: An ERP study of number agreement in complex Italian sentences. *Cognitive Science, 41*(7), 1760–1803. doi:10.1111/cogs.12450

Katamba, F. (2003). Bantu nominal morphology. In D. Nurse & G. Philippson (Eds.), *The Bantu Languages* (pp. 103–120). New York: Routledge.

Keating, G. D. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning, 59*(3), 503–535. doi:10.1111/j.1467-9922.2009.00516.x

Keating, G. D., Jegerski, J., & VanPatten, B. (2016). Online processing of subject pronouns in monolingual and heritage bilingual speakers of Mexican Spanish. *Bilingualism: Language and Cognition, 19*(01), 36–49. doi:10.1017/S1366728914000418

Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford, Calif.: CSLI Publications.

Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics, 25*(1), 1–44. doi:10.1093/jos/ffm018

Keijzer, M. (2007). *Last in first out? An investigation of the regression hypothesis in Dutch emigrants in Anglophone Canada* (PhD dissertation). Vrije Universiteit Amsterdam.

Kennison, S. (2003). Comprehending the pronouns her, him, and his: Implications for theories of referential processing. *Journal of Memory and Language, 49*(3), 335–352. doi:10.1016/S0749-596X(03)00071-8

Kennison, S. M. (2001). Limitations on the use of verb information during sentence comprehension. *Psychonomic Bulletin & Review, 8*(1), 132–138.

Kentner, G. (2019). Rhythmic parsing. *The Linguistic Review, 34*(1), 123–155. doi:10.1515/tlr-2016-0005

Kentner, G., & Vasishth, S. (2016). Prosodic focus marking in silent reading: Effects of discourse context and rhythm. *Frontiers in Psychology, 7,* 319. doi:10.3389/fpsyg.2016.00319

Kim, E., Montrul, S., & Yoon, J. (2015). The on-line processing of binding principles in second language acquisition: Evidence from eye tracking. *Applied Psycholinguistics, 36*(06), 1317–1374. doi:10.1017/S0142716414000307

Kim, H., & Grüter, T. (2019). Cross-linguistic activation of implicit causality biases in Korean learners of English. *Bilingualism: Language and Cognition, 22*(3), 441–455. doi:10.1017/S1366728918000561

Kim, J. H., & Christianson, K. (2016). Working memory effects on L1 and L2 processing of ambiguous relative clauses by Korean L2 learners of English. *Second Language Research, 33*(3), 365–388. doi:10.1177/0267658315623322

Kim, J.-H., Montrul, S., & Yoon, J. (2010). Dominant language influence in acquisition and attrition of binding: Interpretation of the Korean reflexive caki. *Bilingualism: Language and Cognition, 13*(1), 73–84. doi:10.1017/S136672890999037X

King, J., Andrews, C., & Wagers, M. Do reflexives always find a grammatical antecedent for themselves? In *25th Annual CUNY Conference on Human Sentence Processing*. New York: The CUNY Graduate Center.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language, 40*(2), 153–194. doi:10.1006/jmla.1998.2620

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

Koornneef, A., & Van Berkum, J. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language, 54*(4), 445–465. doi:10.1016/j.jml.2005.12.003

Krashen, S. D. (1992). *The input hypothesis: Issues and implications*. Torrance, CA: Laredo Pub. Co.

Kuznetsova, A., Brockhoff, P. B., & Christensen, Rune H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13). doi:10.18637/jss.v082.i13

Kwon, N., Cunnings, I., & Lesmana, M. (2013). *Time course of reference resolution by early and late bilinguals*. Poster presented at *AMLAP*, Aix-Marseille Université.

Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., . . . Tsai, C.-C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review, 10,* 90–115. doi:10.1016/j.edurev.2013.10.001

Lee, E.-K., Lu, D. H.-Y., & Garnsey, S. M. (2013). L1 word order and sensitivity to verb bias in L2 processing. *Bilingualism: Language and Cognition, 16*(04), 761–775. doi:10.1017/S1366728912000776

Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwitserlood, P. (2008). Native language influences on word recognition in a second language: A megastudy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(1), 12–31. doi:10.1037/0278-7393.34.1.12

Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley.

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

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

Lewis, R. L., Vasishth, S., & Van Dyke, Julie A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences, 10*(10), 447–454. doi:10.1016/j.tics.2006.08.007

Lewis, S., & Phillips, C. (2015). Aligning grammatical theories and language processing models. *Journal of Psycholinguistic Research, 44*(1), 27–46. doi:10.1007/s10936-014-9329-z

Li, P., Zhang, F. A., Tsai, E., & Puls, B. (2014). Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition, 17*(3), 673–680. doi:10.1017/S1366728913000606

Lim, J. H., & Christianson, K. (2013). Second language sentence processing in reading for comprehension and translation. *Bilingualism: Language and Cognition, 16*(03), 518–537. doi:10.1017/S1366728912000351

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: a meta-analysis. *Psychonomic Bulletin & Review, 21*(4), 861–883. doi:10.3758/s13423-013-0565-2

Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics, 4*(2), 126–141. doi:10.1093/applin/4.2.126

Long, M. H. (2013). Maturational constraints on child and adult SLA. In G. Granena & M. Long (Eds.), Sensitive Periods, Language Aptitude, and Ultimate L2 Attainment (pp. 3–41). Amsterdam: John Benjamins Publishing Company.

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition, 12*(03), 251–285. doi:10.1017/s0272263100009165

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.

MacDonald, M. C., & Hsiao, F. (2018). Sentence comprehension. In S.-A. Rueschemeyer & M. G. Gaskell (Eds.), *The Oxford Handbook of Psycholinguistics* (pp. 171–196). Oxford: Oxford University Press.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 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. doi:10.1044/1092-4388(2007/067)

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language processing. *Studies in Second Language Acquisition, 27*(01), 483. doi:10.1017/S0272263105050035

Marsden, E., Thompson, S., & Plonsky, L. (2018). A methodological synthesis of self-paced reading in second language research. *Applied Psycholinguistics, 39*(5), 861–904. doi:10.1017/S0142716418000036

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature, 244*(5417), 522.

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

Matsumura, M. (1994). Japanese learners' acquisition of the locality requirement of English reflexives: Evidence for retreat from overgeneralization. *Studies in Second Language Acquisition, 16*(1), 19–42.

Mayberry, R. I., & Kluender, R. (2018). Rethinking the critical period for language: New insights into an old question from American Sign Language. *Bilingualism: Language and Cognition, 21*(5), 886–905. doi:10.1017/S1366728917000724

Mayberry, R. I., Lock, E., & Kazmi, H. (2002). Linguistic ability and early language exposure. *Nature, 417*(6884), 38. doi:10.1038/417038a

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics, 17*(6), 578–586.

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. doi:10.1016/j.jml.2006.06.006

McDonald, J. L., & MacWhinney, B. (1995). The time course of anaphor resolution: Effects of implicit verb causality and gender. *Journal of Memory and Language, 34*(4), 543–566.

McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), The Psychology of Learning and Motivation: Advances in Research and Theory (pp. 155–200). San Diego: Elsevier.

McNamara, T. F. (2000). *Language testing*. Oxford: Oxford University Press.

Meierkord, C. (2004). Syntactic variation in interactions across international Englishes. *English World-Wide, 25*(1), 109–132.

Meisel, J. M. (1997). The acquisition of the syntax of negation in French and German: Contrasting first and second language development. *Second Language Research, 13*(3), 227–263.

Mesthrie, R. (2009). *Language in South Africa*. Cambridge: Cambridge University Press.

Mitchell, D. C., Cuetos, F., Corley, Martin M. B., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research, 24*(6), 469–488. doi:10.1007/bf02143162

Montrul, S. (2008). *Incomplete Acquisition in Bilingualism: Re-examining the Age Factor.* Amsterdam: John Benjamins Publishing Company.

Montrul, S. (2016). *The acquisition of heritage languages*. Cambridge: Cambridge University Press.

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience, 24*(4), 933–947. doi:10.1162/jocn_a_00119

Mosdell, J., Balchin, R., & Ameen, O. (2010). Adaptation of aphasia tests for neurocognitive screening in South Africa. *South African Journal of Psychology, 40*(3), 250–261.

Munoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics, 29*(4), 578–596. doi:10.1093/applin/amm056

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18*(3), 251–269.

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

Nalborczyk, L., Batailler, C., Lœvenbruck, H., Vilain, A., & Bürkner, P.-C. (2019). An introduction to Bayesian multilevel models using brms: A case study of gender effects on vowel variability in standard Indonesian. *Journal of Speech, Language, and Hearing Research, 62*(5), 1225–1242.

Narayanan, S., & Jurafsky, D. (1998). Bayesian models of human sentence processing. In *Proceedings of the Twelfth Annual Meeting of the Cognitive Science Society* (pp. 752–757). Cambridge, MA: Erlbaum.

Newport, E. L., Bavelier, D., & Neville, H. J. (2001). Critical thinking about critical periods: Perspectives on a critical period for language acquisition. In E. DuPoux (Ed.), *Language, Brain, and Cognitive Development: Essays in Honor of Jacques Mehler* (pp. 481–502). Cambridge, MA: MIT Press.

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

Norouzian, R., Miranda, M. de, & Plonsky, L. (2018). The Bayesian revolution in second language research: An applied approach. *Language Learning, 68*(4), 1032–1075. doi:10.1111/lang.12310

Norouzian, R., Miranda, M. D. E., & Plonsky, L. (2019). A Bayesian approach to measuring evidence in L2 research: An empirical investigation. *The Modern Language Journal, 25,* 1. doi:10.1111/modl.12543

Norris, J. M. (2015). Statistical significance testing in second language research: Basic problems and suggestions for reform. *Language Learning, 65*(S1), 97–126. doi:10.1111/lang.12114

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience, 17*(8), 1212–1228. doi:10.1162/0898929055002436

Oller, J. W. (2014). Milestones in Language Development. In P. Brooks & V. Kempe (Eds.), *Encyclopedia of Language Development*. Thousand Oaks: SAGE Publications, Inc.

Oosthuizen, J. (2013). *Obligatory reflexivity in Afrikaans. A minimalist approach* (PhD). Stellenbosch University, Stellenbosch.

Ortega, L. (2005). For what and for whom is our research? The ethical as transformative lens in instructed SLA. *The Modern Language Journal, 89*(3), 427–443. doi:10.1111/j.1540-4781.2005.00315.x

Ortega, L. (2012). Epistemological diversity and moral ends of research in instructed SLA. *Language Teaching Research, 16*(2), 206–226. doi:10.1177/0267658311431373

Ortega, L. (2013). SLA for the 21st century: Disciplinary progress, transdisciplinary relevance, and the bi/multilingual turn. *Language Learning, 63*(1), 1–24. doi:10.1111/j.1467-9922.2012.00735.x

Ortega, L. (2018). Ontologies of language, Second Language Acquisition, and world Englishes. *World Englishes, 37*(1), 64–79. doi:10.1111/weng.12303

Ortega, L. (2019). SLA and the study of equitable multilingualism. *The Modern Language Journal, 103*(Supplement 2016), 23–38. doi:10.1111/modl.12525

Osaka, M., & Osaka, N. (1992). Language-independent working memory as measured by Japanese and English reading span tests. *Bulletin of the Psychonomic Society, 30*(4), 287–289.

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A means for exploring the neurocognition of second language processing. *Language Learning, 56,* 199–230. doi:10.1111/j.1467-9922.2006.00361.x

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language, 34*(6), 739–773. doi:10.1006/jmla.1995.1033

Pan, H.-Y., & Felser, C. (2011). Referential context effects in L2 ambiguity resolution: Evidence from self-paced reading. *Lingua, 121*(2), 221–236. doi:10.1016/j.lingua.2010.08.003

Pan, H.-Y., Schimke, S., & Felser, C. (2013). Referential context effects in non-native relative clause ambiguity resolution. *International Journal of Bilingualism, 19*(3), 298–313. doi:10.1177/1367006913515769

Papadopoulou, D. (2005). Reading-time studies of second language ambiguity resolution. *Second Language Research, 21*(2), 98–120. doi:10.1191/0267658305sr254oa

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing. *Studies in Second Language Acquisition, 25*(4), 501–528. doi:10.1017/s0272263103000214

Parker, D., & Phillips, C. (2017). Reflexive attraction in comprehension is selective. *Journal of Memory and Language, 94,* 272–290. doi:10.1016/j.jml.2017.01.002

Patil, U., Vasishth, S., & Lewis, R. L. (2016). Retrieval interference in syntactic processing: The case of reflexive binding in English. *Frontiers in Psychology, 7,* 329. doi:10.3389/fpsyg.2016.00329

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. doi:10.1037/a0014276

Patson, N. D., & Husband, E. M. (2016). Misinterpretations in agreement and agreement attraction. *The Quarterly Journal of Experimental Psychology, 69*(5), 950–971.

Patterson, C., & Felser, C. (2019). Delayed application of binding condition C during cataphoric pronoun resolution. *Journal of Psycholinguistic Research, 48*(2), 453–475. doi:10.1007/s10936-018-9613-4

Patterson, C., Trompelt, H., & Felser, C. (2014). The online application of binding condition B in native and non-native pronoun resolution. *Frontiers in Psychology, 5,* 147. doi:10.3389/fpsyg.2014.00147

Peirce, B. N., & Ridge, S. G. M. (1997). Multilingualism in southern Africa. *Annual Review of Applied Linguistics, 17,* 170–190.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., . . . Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. doi:10.3758/s13428-018-01193-y

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1-2), 8–13. doi:10.1016/j.jneumeth.2006.11.017

Penfield, W., & Roberts, L. (1959). *Speech and brain mechanisms*. Princeton, NJ: Princeton University Press.

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. doi: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.

Pliatsikas, C., & Chondrogianni, V. (2015). Editorial: Learning a non-native language in a naturalistic environment: Insights from behavioral and neuroimaging research. *Frontiers in Psychology, 6*(888), 240. doi:10.3389/fpsyg.2015.01009

Pliatsikas, C., Johnstone, T., & Marinis, T. (2017). An fMRI study on the processing of long-distance wh-movement in a second language. *Glossa: A journal of general linguistics, 2*(1), 1052. doi:10.5334/gjgl.95

Pliatsikas, C., & Marinis, T. (2013). Processing empty categories in a second language: When naturalistic exposure fills the (intermediate) gap. *Bilingualism: Language and Cognition, 16*(01), 167–182. doi:10.1017/S136672891200017X

Plüddemann, P. (2015). Unlocking the grid: Language-in-education policy realisation in post-apartheid South Africa. *Language and Education, 29*(3), 186–199. doi:10.1080/09500782.2014.994523

Polinsky, M. (2018). *Heritage languages and their speakers*. Cambridge: Cambridge University Press.

Polinsky, M., & Scontras, G. (2019). Understanding heritage languages. *Bilingualism: Language and Cognition, 58,* 1–17. doi:10.1017/S1366728919000245

Posel, D., & Zeller, J. (2016). Language shift or increased bilingualism in South Africa: Evidence from census data. *Journal of Multilingual and Multicultural Development, 37*(4), 357–370. doi:10.1080/01434632.2015.1072206

Potgieter, A. (2014). *The role of input in the early trilingual acquisition of English, Afrikaans and isiXhosa* (PhD). Stellenbosch University, Stellenbosch.

Prévost, P., & White, L. (2000). Missing Surface Inflection or Impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research, 16*(2), 103–133. doi:10.1191/026765800677556046

Pritchett, B. L. (1988). Garden path phenomena and the grammatical basis of language processing. *Language,* 539–576.

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

Probyn, M. (2001). Teachers' voices: Teachers' reflections on learning and teaching through the medium of English as an additional language in South Africa. *International Journal of Bilingual Education and Bilingualism, 4*(4), 249–266. doi:10.1080/13670050108667731

Qian, Z., Lee, E.-K., Lu, D. H.-Y., & Garnsey, S. M. (2018). Native and non-native (L1-Mandarin) speakers of English differ in online use of verb-based cues about sentence structure. *Bilingualism: Language and Cognition, 16,* 1–15. doi:10.1017/S1366728918001001

R Core Team. (2018). R: A language and environment for statistical computing. Austria, Vienna: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rampton, M. (1990). Displacing the 'native speaker': Expertise, affiliation, and inheritance. *ELT Journal,* 44, 97–101.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior, 22*(3), 358–374.

Reinhart, T., & Reuland, E. (1993). Reflexivity. *Linguistic Inquiry, 24*(4), 657–720.

Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics, 32*(02), 299–331. doi:10.1017/S0142716410000421

Roberts, L., Gullberg, M., & Indefrey, P. (2008). Online pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition, 30*(3), 333–357. doi:10.1017/s0272263108080480

Roberts, L., Marinis, T., Felser, C., & Clahsen, H. (2007). Antecedent priming at trace positions in children's sentence processing. *Journal of Psycholinguistic Research, 36*(2), 175–188. doi:10.1007/s10936-006-9038-3

Roberts, L., & Siyanova-Chanturia, A. (2013). Using eye-tracking to investigate topics in L2 acquisition and L2 processing. *Studies in Second Language Acquisition, 35*(02), 213–235. doi:10.1017/S0272263112000861

Ross, J. R. (1967). *Constraints on variables in syntax* (PhD). MIT, Massachusetts.

Ross, S. J., & Mackey, B. (2015). Bayesian approaches to imputation, hypothesis testing, and parameter estimation. *Language Learning, 65*(S1), 208–227. doi:10.1111/lang.12118

Rossi, E., Kroll, J. F., & Dussias, P. E. (2014). Clitic pronouns reveal the time course of processing gender and number in a second language. *Neuropsychologia, 62,* 11–25. doi:10.1016/j.neuropsychologia.2014.07.002

Rossi, S., Gugler, M. F., Friederici, A. D., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience, 18*(12), 2030–2048. doi:10.1162/jocn.2006.18.12.2030

Russell Lenth. (2019). *emmeans: Estimated marginal means, aka least-squares means*. Retrieved from https://CRAN.R-project.org/package=emmeans

Rysiewicz, J. (2008). Cognitive profiles of (un) successful FL learners: A cluster analytical study. *The Modern Language Journal, 92*(1), 87–99.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics, 2*(9), 437–442.

Sagarra, N. (2016). Longitudinal effects of working memory on L2 grammar and reading abilities. *Second Language Research, 33*(3), 341–363. doi:10.1177/0267658317690577

Sagarra, N., & Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua, 120*(8), 2022–2039. doi:10.1016/j.lingua.2010.02.004

Sagarra, N., & Herschensohn, J. (2011). Proficiency and animacy effects on L2 gender agreement processes during comprehension. *Language Learning, 61*(1), 80–116. doi:10.1111/j.1467-9922.2010.00588.x

Schimke, S., de la Fuente, Israel, Hemforth, B., & Colonna, S. (2018). First language influence on second language offline and online ambiguous pronoun resolution. *Language Learning, 68*(3), 744–779. doi:10.1111/lang.12293

Serafini, E. J., & Sanz, C. (2016). Evidence for the decreasing impact of cognitive ability on second language development as proficiency increases. *Studies in Second Language Acquisition, 38*(4), 607–646. doi:10.1017/S0272263115000327

Serratrice, L. (2007). Cross-linguistic influence in the interpretation of anaphoric and cataphoric pronouns in English–Italian bilingual children. *Bilingualism: Language and Cognition, 10*(3), 225–238. doi:10.1017/S1366728907003045

Serratrice, L., Sorace, A., & PAOLI, S. (2004). Crosslinguistic influence at the syntax–pragmatics interface: Subjects and objects in English–Italian bilingual and monolingual acquisition. *Bilingualism: Language and Cognition, 7*(3), 183–205. doi:10.1017/S1366728904001610

Skehan, P. (1986). Cluster analysis and the identification of learner types. In V. Cook (Ed.), *Experimental approaches to second language acquisition* (pp. 81–94). Oxford: Pergamon.

Slabakova, R., White, L., & Brambatti Guzzo, N. (2017). Pronoun interpretation in the second language: Effects of computational complexity. *Frontiers in Psychology, 8,* 1236. doi:10.3389/fpsyg.2017.01236

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology, 49*(3), 238–299. doi:10.1016/j.cogpsych.2004.03.001

Sorace, A., & Filiaci, F. (2016). Anaphora resolution in near-native speakers of Italian. *Second Language Research, 22*(3), 339–368. doi:10.1191/0267658306sr271oa

Sparks, R. L., Patton, J., & Ganschow, L. (2012). Profiles of more and less successful L2 learners: A cluster analysis study. *Learning and Individual Differences, 22*(4), 463–472. doi:10.1016/j.lindif.2012.03.009

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(6), 1521.

Spivey-Knowlton, M. J., Trueswell, J., & Tanenhaus, M. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology, 47*(2), 276–309.

SR Research (2017). *Experiment Builder*. Mississauga, Ontario, Canada: SR Research Ltd.

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, 2(2), 191–196.

Stone, J. M., & Towse, J. N. (2015). A working memory test battery: Java-based collection of seven working memory tasks. *Journal of Open Research Software, 3*(2), 92. doi:10.5334/jors.br

Stowe, L. A. (1986). Parsing wh-constructions: Evidence for on-line gap location. *Language and Cognitive Processes, 1*(3), 227–245. doi: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. doi:10.1016/S0749-596X(02)00536-3

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.

Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General, 136*(1), 64–81. doi:10.1037/0096-3445.136.1.64

Tabor, W., & Hutchins, S. (2004). Evidence for self-organized sentence processing: Digging-in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(2), 431.

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. doi:10.1016/j.jml.2014.07.003

Tarone, E., & Bigelow, M. (2012). A research agenda for second language acquisition of pre-literate and low-literate adult and adolescent learners. In P. Vinogradov & M. Bigelow (Eds.), *Proceedings from the 7th Annual Low Educated Second Language and Literacy Acquisition Symposium* (pp. 5–26). Minneapolis: University of Minnesota Printing Services.

Thomas, M. (1991). Universal Grammar and the interpretation of reflexives in a second language. *Language,* 211–239.

Thompson, C. K., & Choy, J. J. (2009). Pronominal resolution and gap filling in agrammatic aphasia: Evidence from eye movements. *Journal of Psycholinguistic Research, 38*(3), 255–283. doi:10.1007/s10936-009-9105-7

Tolentino, L. C., & Tokowicz, N. (2011). Across languages, space, and time. *Studies in Second Language Acquisition, 33*(1), 91–125. doi:10.1017/S0272263110000549

Traxler, M. J. (2002). Plausibility and subcategorization preference in children's processing of temporarily ambiguous sentences: Evidence from self-paced reading. *The Quarterly Journal of Experimental Psychology: Section A, 55*(1), 75–96.

Traxler, M. J. (2005). Plausibility and verb subcategorization in temporarily ambiguous sentences: Evidence from self-paced reading. *Journal of Psycholinguistic Research, 34*(1), 1–30. doi:10.1007/s10936-005-3629-2

Traxler, M. J. (2014). Trends in syntactic parsing: Anticipation, Bayesian estimation, and good-enough parsing. *Trends in Cognitive Sciences, 18*(11), 605–611. doi:10.1016/j.tics.2014.08.001

Traxler, M. J., Hoversten, L. J., & Brothers, T. A. (2018). Sentence processing and interpretation in monolinguals and bilinguals. In E. M. Fernández & H. S. Cairns (Eds.), *The Handbook of Psycholinguistics* (pp. 320–344).

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language, 35*(3), 454–475.

Traxler, M. J., Pickering, M. J., & Clifton Jr, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language, 39*(4), 558–592.

Treffers-Daller, J., & Korybski, T. (2015). Using lexical diversity measures to operationalise language dominance in bilinguals. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language dominance in bilinguals. Issues of measurement and operationalization* (pp. 106–123). Cambridge: Cambridge University Press.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33*(3), 285–318.

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. doi:10.1037//0278-7393.19.3.528

Tsimpli, I., Sorace, A., Heycock, C., & Filiaci, F. (2004). First language attrition and syntactic subjects: A study of Greek and Italian near-native speakers of English. *International Journal of Bilingualism, 8*(3), 257–277.

Ullman, M. T. (1999). Acceptability ratings of regular and irregular past-tense forms: Evidence for a dual-system model of language from word frequency and phonological neighbourhood effects. *Language and Cognitive Processes, 14*(1), 47–67. doi:10.1080/016909699386374

Ullman, M. T. (2001a). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews: Neuroscience, 2*(10), 717–726. doi:10.1038/35094573

Ullman, M. T. (2001b). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition, 4*(2), 105–122. doi:10.1017/s1366728901000220

Ullman, M. T. (2016). The declarative/procedural model. In G. Hickok & L. Small (Eds.), *Neurobiology of Language* (pp. 953–968). Amsterdam: Elsevier.

Ullman, M. T., & Gopnik, M. (1999). Inflectional morphology in a family with inherited specific language impairment. *Applied Psycholinguistics, 20*(1), 51–117. doi:10.1017/s0142716499001034

van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., & Hartsuiker, R. J. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language, 64*(1), 88–107. doi:10.1016/j.jml.2010.08.006

van Assche, E., Duyck, W., Hartsuiker, R. J., & Diependaele, K. (2009). Does bilingualism change native-language reading? *Psychological Science, 20*(8), 923–927. doi:10.1111/j.1467-9280.2009.02389.x

Van Berkum, J., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language, 41*(2), 147–182.

van Dyke, J. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language, 49*(3), 285–316. doi:10.1016/S0749-596X(03)00081-0

Van Gompel, R. P. G., & Pickering, M. J. (2007). Syntactic parsing. In M. G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistic*s (pp. 289–307). Oxford: Oxford University Press.

Van Gompel, R. P. G., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language, 45*(2), 225–258.

van Hell, J. G., & de Groot, A. M. B. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychologica, 128*(3), 431–451. doi:10.1016/j.actpsy.2008.03.010

van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review, 9*(4), 780–789. doi:10.3758/BF03196335

van Heuven, Walter J.B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language, 39*(3), 458–483. doi:10.1006/jmla.1998.2584

Vanhove, J. (2013). The critical period hypothesis in second language acquisition: A statistical critique and a reanalysis. *PloS One, 8*(7), e69172. doi:10.1371/journal.pone.0069172

Vasishth, S., Chen, Z., Li, Q., & Guo, G. (2013). Processing Chinese relative clauses: Evidence for the subject-relative advantage. *PloS One, 8*(10), e77006. doi:10.1371/journal.pone.0077006

Vasishth, S., & Nicenboim, B. (2016). Statistical methods for linguistic research: Foundational ideas - Part I. *Language and Linguistics Compass, 10*(8), 349–369. doi:10.1111/lnc3.12201

Veríssimo, J. (2016). Extending a Gradient Symbolic approach to the native versus non-native contrast: The case of plurals in compounds. *Bilingualism: Language and Cognition, 19*(5), 900–902. doi:10.1017/S1366728916000134

Veríssimo, J. (2018). Sensitive periods in both L1 and L2: Some conceptual and methodological suggestions. *Bilingualism: Language and Cognition, 21*(5), 932–933. doi:10.1017/S1366728918000275

Veríssimo, J., Heyer, V., Jacob, G., & Clahsen, H. (2018). Selective effects of age of acquisition on morphological priming: Evidence for a sensitive period. *Language Acquisition, 25*(3), 315–326.

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. doi:10.1016/j.jml.2009.04.002

Wang, X. (2013). Language dominance in translation priming: Evidence from balanced and unbalanced Chinese–English bilinguals. *Quarterly Journal of Experimental Psychology, 66*(4), 727–743. doi:10.1080/17470218.2012.716072

Warner, J., & Glass, A. L. (1987). Context and distance-to-disambiguation effects in ambiguity resolution: Evidence from grammaticality judgments of garden path sentences. *Journal of Memory and Language, 26*(6), 714–738.

Warren, P. (2016). Prosody and language processing. In S. C. Garrod & M. J. Pickering (Eds.), *Language Processing* (pp. 155–188). Hove: Psychology Press.

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition, 85*(1), 79–112.

Weber-Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience, 8*(3), 231–256. doi:10.1162/jocn.1996.8.3.231

Werker, J. F., & Hensch, T. K. (2015). Critical periods in speech perception: New directions. *Annual Review of Psychology, 66,* 173–196. doi:10.1146/annurev-psych-010814-015104

Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology, 46*(3), 233–251. doi:10.1002/dev.20060

White, L. (1998). Second language acquisition and Binding Principle B: Child/adult differences. *Second Language Research, 14*(4), 425–439. doi:10.1191/026765898675690231

White, L., Hirakawa, M., & Kawasaki, T. (1996). Effects of instruction on second language acquisition of the Japanese long-distance reflexive zibun. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique, 41*(3), 235–254.

Whitford, V., & Titone, D. (2012). Second-language experience modulates first- and second-language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin & Review, 19*(1), 73–80. doi:10.3758/s13423-011-0179-5

Whitford, V., & Titone, D. (2015). Second-language experience modulates eye movements during first- and second-language sentence reading: Evidence from a gaze-contingent moving window paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(4), 1118–1129. doi:10.1037/xlm0000093

Williams, J. N. (2006). Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition, 9*(01), 71–88. doi:10.1017/S1366728905002385

Williams, J. N., Möbius, P., & Kim, C. (2001). Native and non-native processing of English wh-questions: Parsing strategies and plausibility constraints. *Applied Psycholinguistics, 22*(04). doi:10.1017/S0142716401004027

Wilson, M. P., & Garnsey, S. M. (2009). Making simple sentences hard: Verb bias effects in simple direct object sentences. *Journal of Memory and Language, 60*(3), 368–392. doi:10.1016/j.jml.2008.09.005

Xiang, M., Dillon, B., & Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain and Language, 108*(1), 40–55. doi:10.1016/j.bandl.2008.10.002

Young, R. F., & Astarita, A. C. (2013). Practice theory in language learning. *Language Learning, 63,* 171–189. doi:10.1111/j.1467-9922.2012.00743.x

APPENDICES

APPENDIX A: Ethics approval

UNIVERSITEIT
STELLENBOSCH
UNIVERSITY

NOTICE OF APPROVAL

REC Humanities New Application Form

25 July 2017

Project number: GENL-2017-0382-284

Project Title: The influence of language dominance on second language syntactic processing in successive English-Afrikaans bilinguals

Dear Ms Robyn Berghoff

Your REC Humanities New Application Form submitted on 5 July 2017 was reviewed and approved by the REC: Humanities.

Please note the following about your approved submission:

Ethics approval period: 25 July 2017 - 24 July 2020

Please take note of the General Investigator Responsibilities attached to this letter. You may commence with your research after complying fully with these guidelines.

If the researcher deviates in any way from the proposal approved by the REC: Humanities, the researcher must notify the REC of these changes.

Please use your SU project number (GENL-2017-0382-284) on any documents or correspondence with the REC concerning your project.

Please note that the REC has the prerogative and authority to ask further questions, seek additional information, require further modifications, or monitor the conduct of your research and the consent process.

FOR CONTINUATION OF PROJECTS AFTER REC APPROVAL PERIOD

Please note that a progress report should be submitted to the Research Ethics Committee: Humanities before the approval period has expired if a continuation of ethics approval is required. The Committee will then consider the continuation of the project for a further year (if necessary)

Included Documents:

Document Type	File Name	Date	Version
Informed Consent Form	ConsentForm	23/06/2017	
Recruitment material	Afrikaans advertisement	23/06/2017	
Recruitment material	English advertisement	23/06/2017	
Research Protocol/Proposal	Berghoff_ResearchProposalJune	23/06/2017	
Request for permission	RBerghoff_Application Letter for Institutional Permission	04/07/2017	
Data collection tool	LEAPQ_Final	05/07/2017	

If you have any questions or need further help, please contact the REC office at cgraham@sun.ac.za.

Sincerely,

Clarissa Graham

REC Coordinator: Research Ethics Committee: Human Research (Humanities)

National Health Research Ethics Committee (NHREC) registration number: REC-050411-032. The Research Ethics Committee: Humanities complies with the SA National Health Act No.61 2003 as it pertains to health research. In addition, this committee abides by the ethical norms and principles for research established by the Declaration of Helsinki (2013) and the Department of Health Guidelines for Ethical Research: Principles Structures and Processes (2nd Ed.) 2015. Annually a number of projects may be selected randomly for an external audit.

Page 1 of 2

APPENDIX B: Ethics amendment approval

UNIVERSITEIT
STELLENBOSCH
UNIVERSITY

NOTICE OF APPROVAL

REC Humanities Amendment Form

23 April 2019

Project number: 0382

Project Title: The influence of language dominance on second language syntactic processing in successive English-Afrikaans bilinguals

Dear Ms. Robyn Berghoff

Your REC Humanities Amendment Form submitted on 20 March 2019 was reviewed and approved by the REC: Humanities.

Please note the following for your approved submission:

Ethics approval period:

Protocol approval date (Humanities)	Protocol expiration date (Humanities)
25 July 2017	24 July 2020

GENERAL COMMENTS:

Please take note of the General Investigator Responsibilities attached to this letter. You may commence with your research after complying fully with these guidelines.

If the researcher deviates in any way from the proposal approved by the REC: Humanities, the researcher must notify the REC of these changes.

Please use your SU project number (0382) on any documents or correspondence with the REC concerning your project.

Please note that the REC has the prerogative and authority to ask further questions, seek additional information, require further modifications, or monitor the conduct of your research and the consent process.

FOR CONTINUATION OF PROJECTS AFTER REC APPROVAL PERIOD

Please note that a progress report should be submitted to the Research Ethics Committee: Humanities before the approval period has expired if a continuation of ethics approval is required. The Committee will then consider the continuation of the project for a further year (if necessary)

Included Documents:

Document Type	File Name	Date	Version
Default	Berghoff_ResearchProposal_AMENDED	15/03/2019	2
Default	Project description REC	15/03/2019	1
Default	ET LIST Example	15/03/2019	1

If you have any questions or need further help, please contact the REC office at cgraham@sun.ac.za.

Sincerely,

Clarissa Graham

REC Coordinator: Research Ethics Committee: Human Research (Humanities)

*National Health Research Ethics Committee (NHREC) registration number: REC-050411-032.
The Research Ethics Committee: Humanities complies with the SA National Health Act No.61 2003 as it pertains to health research. In addition, this committee abides by the ethical norms and principles for research established by the Declaration of Helsinki (2013) and the Department of Health Guidelines for Ethical Research: Principles Structures and Processes (2nd Ed.) 2015. Annually a number of projects may be selected randomly for an external audit.*

Page 1 of 2

APPENDIX C: Institutional permission

UNIVERSITEIT • STELLENBOSCH • UNIVERSITY
jou kennisvennoot • your knowledge partner

INSTITUTIONAL PERMISSION:

AGREEMENT ON USE OF PERSONAL INFORMATION IN RESEARCH

Name of Researcher: Robyn Berghoff

Name of Research Project: The influence of language dominance on second language syntactic processing in successive English-Afrikaans bilinguals

Service Desk ID: IRPSD 539

Date of Issue: 3 August 2017

You have received institutional permission to proceed with this project as stipulated in the institutional permission application and within the conditions set out in this agreement.

1 WHAT THIS AGREEMENT IS ABOUT	
What is POPI?	1.1 POPI is the Protection of Personal Information Act 4 of 2013. 1.2 POPI regulates the entire information life cycle from collection, through use and storage and even the destruction of personal information.
Why is this important to us?	1.3 Even though POPI is important, it is not the primary motivation for this agreement. The privacy of our students and employees are important to us. We want to ensure that no research project poses any risks to their privacy. 1.4 However, you are required to familiarise yourself with, and comply with POPI in its entirety.
What is considered to be personal information?	1.5 'Personal information' means information relating to an identifiable, living, individual or company, including, but not limited to: 1.5.1 information relating to the race, gender, sex, pregnancy, marital status, national, ethnic or social origin, colour, sexual orientation, age, physical or mental health, well-being, disability, religion, conscience, belief, culture, language and birth of the person; 1.5.2 information relating to the education or the medical, financial, criminal or employment history of the person;

APPENDIX D: Consent form

UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvennoot • your knowledge partner

STELLENBOSCH UNIVERSITY
CONSENT TO PARTICIPATE IN RESEARCH

Reading of English sentences

You are asked to participate in a research study conducted by Robyn Berghoff, MA, from the Department of General Linguistics at Stellenbosch University. The study forms part of the investigator's doctoral research. You were selected as a possible participant in this study because of your language background.

1. PURPOSE OF THE STUDY

The study investigates how individuals read English sentences.

2. PROCEDURES

If you volunteer to participate in this study, we would ask you to do the following things:

- Complete a short language background questionnaire (±10 mins).
- Complete a so-called "cloze" test of English proficiency (±10 mins).
- Complete an experiment. In the experiment, you will read a number of sentences. Sometimes you will be required to answer questions about the content of these sentences (±30 mins).

The total length of time for participation should be approximately 50 minutes. All tasks will be completed in the laboratory space of the Department of General Linguistics on the fifth floor of the Arts and Social Sciences building. All tasks will be performed on a computer.

3. POTENTIAL RISKS AND DISCOMFORTS

This is a low-risk study, but subjects may feel fatigued during the procedure, as it is relatively time-consuming. If at any point the subject wishes to take a short break, s/he should please let the investigator know and s/he will be welcome to do so.

4. POTENTIAL BENEFITS TO SUBJECTS AND/OR TO SOCIETY

The subject is not expected to benefit directly from participation in the study.

Scientifically, the results of this study are relevant to our understanding of how individuals read in English.

5. PAYMENT FOR PARTICIPATION

The subject will receive a payment of R50 for participating in the study. This amount will be paid in cash at the end of the data collection procedure. Should the participant not complete the entire procedure, s/he will be paid pro-rata for his/her time at a rate of R20 per 30 mins.

6. CONFIDENTIALITY

Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law. Confidentiality will be maintained by use of participant numbers instead of names in the data coding. In addition, the data will be stored securely on an external hard drive to which only the researcher has access.

If results from the study are submitted for publication, participants will be referred to generally in terms of their language background and not by either name or participant number.

7. PARTICIPATION AND WITHDRAWAL

You can choose whether to be in this study or not. If you volunteer to be in this study, you may withdraw at any time without consequences of any kind. You may also refuse to answer any questions you don't want to answer and still remain in the study. The investigator may withdraw you from this research if circumstances arise which warrant doing so; for example, if more than a specified percentage of questions are answered incorrectly.

8. IDENTIFICATION OF INVESTIGATORS

If you have any questions or concerns about the research, please feel free to contact the principal investigator, Robyn Berghoff (robynberghoff@gmail.com; 062 634 2520), or the study's supervisor, Prof Manne Bylund (mbylund@sun.ac.za).

9. RIGHTS OF RESEARCH SUBJECTS

You may withdraw your consent at any time and discontinue participation without penalty. You are not waiving any legal claims, rights or remedies because of your participation in this research study. If you have questions regarding your rights as a research subject, contact Ms Maléne Fouché (mfouche@sun.ac.za; 021 808 4622) at the Division for Research Development.

SIGNATURE OF RESEARCH SUBJECT OR LEGAL REPRESENTATIVE

The information above was described to me by [*Robyn Berghoff/an authorized research assistant*] in [*Afrikaans/English/Xhosa/other]* and I am in command of this language or it was satisfactorily translated to me. I was given the opportunity to ask questions and these questions were answered to my satisfaction.

I hereby consent voluntarily to participate in this study. I have been given a copy of this form.

____

Name of Subject/Participant

____

Date

SIGNATURE OF INVESTIGATOR

I declare that I explained the information given in this document to __________________ [*name of the subject/participant]. [He/she]* was encouraged and given ample time to ask me any questions. This conversation was conducted in [*Afrikaans/*English/*Xhosa/*Other*] and [*no translator was used/this conversation was translated into* ___________ by _______________________].

____ ______________

Signature of Investigator **Date**

APPENDIX E: C-test

Fill in the gap

On each of the following three pages you will find a short text. Each text contains gaps where parts of some words have been left out. Please try to fill in the gaps. In many cases there are several possible answers.

1.

We all live with other people's expectations of us. These are a **refl** of **th** trying to **under** us; **the** are **predic** of **wh** they **th** we will think, **d** and feel .

Gene , we **acc** the **sta** quo, but these **expec** can be **ha** to **han** when they **co** from our **fami** and can be **diff** to **ign** , especially **wh** they come from our **par** .

2.

The BBC's core **purp** is broadcasting. Since the **lau** of Radio Times in 1923 it **h** also **eng** in **comme** activities. If **pur** properly, **su** commercial activities **he** to **rea** the **va** of **lic** payers' **ass** and **gene** income to be **plou** back **in** the public **ser** programming. The commercial Policy Guidelines **s** out the **fram** which **ens** that the BBC's commercial activities **supp** its public purpose.

3.

The University of the Witwatersrand, Johannesburg, is a multi-campus South African public research university situated in the northern areas of central Johannesburg. The university's **fir** library, **ho** at the **ti** in what was **me** to be a **tem** construction, was **des** in a **fi** on Christmas Eve in 1931. **Fol** this, an **app** was made to the **pu** for £80,000 to **p** for the **cons** of a new **li** , and the **acq** of books. This **res** in the fairly **rap** construction of the William Cullen Library; completed in 1935. During **th** period, as the Great Depression hit South Africa, the university was **fac** with **sev** financial **rest** .

APPENDIX F: LEAP-Q

Northwestern Bilingualism & Psycholinguistics Research Laboratory
Please cite Marian, Blumenfeld, & Kaushanskaya (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.

Language Experience and Proficiency Questionnaire (LEAP-Q)

Last Name		First Name		Today's Date	
Age		Date of Birth		Male ☐	Female ☐

(1) Please list all the languages you know **in order of dominance**:

1	2	3	4	5

(2) Please list all the languages you know **in order of acquisition** (your native language first):

1	2	3	4	5

(3) Please list what percentage of the time you are *currently* and *on average* exposed to each language.
(*Your percentages should add up to 100%*):

List language here:					
List percentage here:					

(4) When choosing to read a text available in all your languages, in what percentage of cases would you choose to read it in each of your languages? Assume that the original was written in another language, which is unknown to you.
(*Your percentages should add up to 100%*):

List language here					
List percentage here:					

(5) When choosing a language to speak with a person who is equally fluent in all your languages, what percentage of time would you choose to speak each language? Please report percent of total time.
(*Your percentages should add up to 100%*):

List language here					
List percentage here:					

(6) How many years of formal education do you have? ______ ______________________________
Please check your highest education level (or the approximate South African equivalent to a degree obtained in another country):

☐ Less than high school	☐ Some undergraduate university	☐ Masters
☐ High school	☐ Undergraduate degree	☐ Ph.D./M.D.
☐ Professional Training	☐ Some postgraduate university	☐ Other:

(7) Date of immigration to South Africa, if applicable ___ ______________________________
If you have ever immigrated to another country, please provide name of country and date of immigration here. ________________ ____

(8) Have you ever had a vision problem ☐, hearing impairment ☐, language disability ☐, or learning disability ☐ ? (Check all applicable). If yes, please explain (including any corrections):
______________________________ ______________________________

Language:

This is my Native language.

All questions below refer to your knowledge of .

(1) Age when you…:

began acquiring :	*became fluent* in :	*began reading* in :	*became fluent reading* in :

(2) Please list the number of years and months you spent in each language environment:

	Years	Months
A country where is spoken		
A family where is spoken		
A school and/or working environment where is spoken		

(3) On a scale from zero to ten, please select your *level of **proficiency*** in speaking, understanding, and reading from the scroll-down menus:

Speaking	(click here for scale	Understanding spoken language	(click here for scale)	Reading	(click here for scale

(4) On a scale from zero to ten, please select how much the following factors contributed to you learning :

Interacting with friends	(click here for pull-down scale)	Language tapes/self instruction	(click here for pull-down scale
Interacting with family	(click here for pull-down scale)	Watching TV	(click here for pull-down scale
Reading	(click here for pull-down scale)	Listening to the radio	(click here for pull-down scale

(5) Please rate to what extent you are currently exposed to in the following contexts:

Interacting with friends	(click here for pull-down scale)	Listening to radio/music	(click here for pull-down scale
Interacting with family	10 - always	Reading	(click here for pull-down scale
Watching TV	(click here for pull-down scale)	Language-lab/self-instruction	(click here for pull-down scale

(6) In your perception, how much of a foreign accent do you have in ?

10 - pervasive

(7) Please rate how frequently others identify you as a non-native speaker based on your <u>accent</u> in :

(click here for pull-down scale)

Language:

This is my (please select from pull-down menu) language.

All questions below refer to your knowledge of .

(1) Age when you…:

began acquiring :	*became fluent* in :	*began reading* in :	*became fluent reading* in :

(2) Please list the number of years and months you spent in each language environment:

	Years	Months
A country where is spoken		
A family where is spoken		
A school and/or working environment where is spoken		

(3) On a scale from zero to ten please select your *level of* ***proficiency*** in speaking, understanding, and reading from the scroll-down menus:

Speaking	(click here for scale)	Understanding spoken language	(click here for scale)	Reading	(click here for scale

(4) On a scale from zero to ten, please select how much the following factors contributed to you learning :

Interacting with friends	(click here for pull-down scale)	Language tapes/self instruction	(click here for pull-down scale
Interacting with family	(click here for pull-down scale)	Watching TV	(click here for pull-down scale
Reading	(click here for pull-down scale)	Listening to the radio	(click here for pull-down scale

(5) Please rate to what extent you are currently exposed to in the following contexts:

Interacting with friends	(click here for pull-down scale)	Listening to radio/music	(click here for pull-down scale
Interacting with family	(click here for pull-down scale)	Reading	(click here for pull-down scale
Watching TV	(click here for pull-down scale)	Language-lab/self-instruction	(click here for pull-down scale

(6) In your perception, how much of a foreign accent do you have in ?

(click here for pull-down scale)

(7) Please rate how frequently others identify you as a non-native speaker based on your accent in :

(click here for pull-down scale)

APPENDIX G: Language History Questionnaire (Li et al., 2014)

L2 Language History Questionnaire (Version 3.0, 2015)

See http://blclab.org/ for online use and credit

Participant ID: ____________________

1. Age (in years): __________

2. Sex (Circle one): Male / Female

3. Education (your current or most recent educational level, even you have not finished the degree) (Circle one):

- Graduate school (PhD/MD/JD)
- Graduate school (Masters)
- College (BA/BS)
- High school
- Middle school
- Other (specify): ___________

4. Have you ever studied or learned a second language in terms of listening, speaking, reading, or writing? (Circle one):
Yes / No

5. Indicate your native language(s) and any other languages you have studied or learned, the age at which you started using each language in terms of listening, speaking, reading, and writing, and the total number of years you have spent using each language.

Language	Listening	Speaking	Reading	Writing	Years of use[a]

a. You may have learned a language, stopped using it, and then started using it again. Please give the total number of years.

6a. Country of residence: ____________

6b. Country of origin: ____________

6c. If 6a and 6b are different, then when did you first move to the country where you currently live? ________________

7. If you have lived or travelled in countries other than your country of residence or country of origin for three or more months, then indicate the name of the country, your length of stay, the language you used, and the frequency of your use of the language for each country.

Country	Length of stay[a] [month(s)]	Language	Frequency of use[b]
			1 2 3 4 5 6 7
			1 2 3 4 5 6 7
			1 2 3 4 5 6 7
			1 2 3 4 5 6 7

a. You may have been to the country on multiple occasions, each for a different length of time. Add all the trips together.

b. Please rate according to the following scale (circle the number in the table)

Never	*Rarely*	*Sometimes*	*Regularly*	*Often*	*Usually*	*Always*
1	2	3	4	5	6	7

8. Indicate the age at which you started using each of the languages you have studied or learned in the following environments.

Language	At home	With friends	At school	At work	Language software	Online games

9. Indicate the language used by your teachers for instruction at each educational level. If the instructional language switched during any educational level, then also indicate the "Switched to" language.

	Language	(Switched to)
Elementary school		
Middle school		
High school		
College/university		

10. Rate your language learning skill. In other words, how good do you feel you are at learning new languages, relative to your friends or other people you know? (circle one)

Very poor	Poor	Limited	Average	Good	Very good	excellent
1	2	3	4	5	6	7

11. Rate your current ability in terms of listening, speaking, reading, and writing in each of the languages you have studied or learned. Please rate according to the following scale (circle the number in the table):

Very poor	Poor	Limited	Functional	Good	Very good	Native-like
1	2	3	4	5	6	7

Language	Listening	Speaking	Reading	Writing
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7

12. If you have taken any standardized language proficiency tests (e.g., TOEFL), then indicate the name of the test, the language assessed, and the score you received for each. If you do not remember the exact score, then indicate an "Approximate score" instead.

Test	Language	Score	(Approximate score)

13. Rate the strength of your foreign accent for each of the languages you have studied or learned. Please rate the strength of your accent according to the following scale (circle the number in the table):

None	Very weak	Weak	Moderate	Strong	Very strong	Extreme
1	2	3	4	5	6	7

Language	Strength of accent
	1 2 3 4 5 6 7
	1 2 3 4 5 6 7
	1 2 3 4 5 6 7
	1 2 3 4 5 6 7

14. Estimate how many hours per day you spend engaged in the following activities in each of the languages you have studied or learned.

	Language: ________	Language: ________	Language: ________
Watching television:	________(hrs)	________(hrs)	________(hrs)
Listening to radio:	________(hrs)	________(hrs)	________(hrs)
Reading for fun:	________(hrs)	________(hrs)	________(hrs)
Reading for school/work:	________(hrs)	________(hrs)	________(hrs)
Writing emails to friends:	________(hrs)	________(hrs)	________(hrs)
Writing for school/work:	________(hrs)	________(hrs)	________(hrs)

15. Estimate how many hours per day you spend speaking with the following groups of people in each of the languages you have studied or learned.

	Language: ________	Language: ________	Language: ________
Family members:	________(hrs)	________(hrs)	________(hrs)
Friends[a]:	________(hrs)	________(hrs)	________(hrs)
Classmates:	________(hrs)	________(hrs)	________(hrs)
Coworkers[b]:	________(hrs)	________(hrs)	________(hrs)

a. Include significant others in this category if you did not include them as family members (e.g., married partners).

b. Include anyone in the work environment in this category (e.g., if you are a teacher, include students as co-workers).

16a. Do you mix words or sentences from different languages when you speak? *(This includes, for example, starting a sentence in one language but using a word or phrase from another language in the middle of the sentence.)* (Circle one)

Yes / No

16b. If you answered "Yes" to 16a, then indicate the languages that you mix and estimate the frequency of mixing in normal conversation with the following groups of people. Please estimate the frequency of mixing according to the following scale (circle the number in the table):

	Language 1	Language 2	Frequency of mixing
Family members			1 2 3 4 5 6 7
Friends			1 2 3 4 5 6 7
Classmates			1 2 3 4 5 6 7
Coworkers			1 2 3 4 5 6 7

17. In which language do you communicate best or feel most comfortable in terms of listening, speaking, reading, and writing in each of the following environments?

	Listening	Speaking	Reading	Writing
At home				
With friends				
At school				
At work				

18. How often do you use each of the languages you have studied or learned for the following activities? Please circle the number in the table according to the scale below.

Never	Rarely	Sometimes	Regularly	Often	Usually	Always
1	2	3	4	5	6	7

Language	Thinking	Talking to yourself	Expression emotion[a]	Dreaming	Arithmetic[b]	Remembering numbers[c]
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7

a. This includes shouting, cursing, showing affection, etc.
b. This includes counting, calculating tips, etc.
c. This includes telephone numbers, ID numbers, etc.

19. What percentage of your friends speaks each of the languages you have studied or learned? *(The total percentage should add up to 100%.)*

Language	Percentage
	%
	%
	%
	%

20a. Do you feel that you are bicultural or multicultural? *(This includes, for example, growing up with parents or relatives from different cultures or living in different cultures for extensive periods of time.)* (Circle one)

Yes / No

20b. If you answered "Yes" to 20a, then which cultures/languages do you identify with more strongly? Rate the strength of your connection in the following categories for each culture/language. Circle the number in the table according to the following scale.

None	Very weak	Weak	Moderate	Strong	Very strong	Extreme
1	2	3	4	5	6	7

Culture/Language	Way of life	Food	Music	Art	Cities/ towns	Sports teams
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7

21. Please comment below to indicate any additional answers to any of the questions above that you feel better describe your language background or usage.

22. Please comment below to provide any other information about your language background or usage.

APPENDIX H: Stimuli

Chapter 4 – Experimental items

Adjunct-clause items

While the captain sailed the boat carrying the tourists passed by very quickly.
While the captain sailed the horse in the field passed by very quickly.

While the child climbed the ladder against the wall fell to the ground.
While the child climbed the housewife in the kitchen fell to the ground.

While the boy climbed the tree in the garden looked very beautiful indeed.
While the boy climbed the fish that was grilled looked very beautiful indeed.

As the girl painted the flowers in the vase smelled very fresh indeed.
As the girl painted the coffee in the mug smelled very fresh indeed.

As the girl drank the milk in the bottle disappeared from the kitchen.
As the girl drank the dog with thick fur disappeared from the kitchen.
As the men drank the beer imported from Europe pleased everybody very much.
As the men drank the song that was popular pleased everybody very much.

While the woman drank the coffee in the mug looked very nice indeed.
While the woman drank the flowers in the vase looked very nice indeed.

While the band played the beer imported from Europe pleased all the customers.
While the band played the song that was popular pleased all the customers.

As the mother baked the cake for the tea disappeared mysteriously from sight.
As the mother baked the truck in the driveway disappeared mysteriously from sight.

As the driver parked the car with broken windows made a loud noise.
As the driver parked the boy living next door made a loud noise.

As the window-cleaner called the ladder against the wall started to shake alarmingly.
As the window-cleaner called the housewife in the kitchen started to shake alarmingly.

As the woman called the ice-cream with the sprinkles fell to the ground.
As the woman called the puppy in the garden fell to the ground.

While the boy walked the milk in the bottle got hot and smelly.
While the boy walked the dog with thick fur got hot and smelly.

While the neighbor visited the car with broken windows passed by the house.
While the neighbor visited the boy living next door passed by the house.

As the woman rode the boat carrying the tourists raced under the bridge.
As the woman rode the horse in the field raced under the bridge.

While the cleaner polished the cake for the tea was stolen by someone.
While the cleaner polished the truck in the driveway was stolen by someone.

While the child ate the ice-cream with the sprinkles dropped to the floor.
While the child ate the puppy in the garden dropped to the floor.

As the woman ate the tree in the garden shone in the sun.
As the woman ate the fish that was grilled shone in the sun.

As the pilot flew the song that was popular sounded far too loud.
As the pilot flew the plane over the ocean sounded far too loud.

As the choir sang the song that was popular could be heard everywhere.
As the choir sang the plane over the ocean could be heard everywhere.

Complement clause items

The manager resolved the report about the budget would be discussed very soon.
The manager resolved the issue with the budget would be discussed very soon.

The judge resolved the problem with the case would be described in court.
The judge resolved the mother of the child would be described in court.
The man believed the book about the war had upset very many people.
The man believed the girl who was upset had upset very many people.

The spokesman confirmed the story about poor hygiene had surprised the president yesterday.
The spokesman confirmed the doctor who was fired had surprised the president yesterday.

The inspector warned the crimes committed last year would destroy very many lives.
The inspector warned the boss who was corrupt would destroy very many lives.

The man warned the story about poor hygiene would embarrass the hospital managers.
The man warned the doctor who was fired would embarrass the hospital managers.

The dean warned the magazine about academic research would never get published again.
The dean warned the professor who had lied would never get published again.

The headmaster cautioned the poem that was controversial was unsuitable for the class.
The headmaster cautioned the tutor with the tattoo was unsuitable for the class.

The headmaster cautioned the teacher who was angry would not be clearly understood.

The headmaster cautioned the answer to the problem would not be clearly understood.

The teacher cautioned the problem with the case would not go away easily.
The teacher cautioned the mother of the child would not go away easily.

The man read the poem that was controversial had excited the young children.
The man read the tutor with the tattoo had excited the young children.

The woman read the magazine about academic research had shocked the university staff.
The woman read the professor who had lied had shocked the university staff.

The scientist proved the theory about the universe could solve the difficult problem.
The scientist proved the lady who was intelligent could solve the difficult problem.

The student suggested the teacher who was angry was not actually correct today.
The student suggested the answer to the problem was not actually correct today.

The man confessed the truth about the crime was not believed by anyone.
The man confessed the king who was worried was not believed by anyone.

The criminal confessed the crimes committed last year had been discovered too late.
The criminal confessed the boss who was corrupt had been discovered too late.

The student wrote the report about the budget would start an important debate.
The student wrote the issue with the budget would start an important debate.

The journalist wrote the book about the war had amazed all the judges.
The journalist wrote the girl who was upset had amazed all the judges.

The minister advised the truth about the crime should be discussed in parliament.
The minister advised the king who was worried should be discussed in parliament.

The psychologist advised the theory about the universe should be accepted by everyone.
The psychologist advised the lady who was intelligent should be accepted by everyone.

Chapter 5 – Experimental items

Potentially reflexive verbs

The hike up the mountain did not go well at all.
John claimed that Mark had injured him quite badly during the climb.
Luckily there were no fatalities.

The hike up the mountain did not go well at all.
John claimed that Jane had injured him quite badly during the climb.
Luckily there were no fatalities.

The hike up the mountain did not go well at all.
Jane claimed that John had injured him quite badly during the climb.
Luckily there were no fatalities.

Today’s workout at the gym was tougher than usual.
Joan thought that Nick had injured her during the race before class.
The doctor recommended a period of rest.

Today’s workout at the gym was tougher than usual.
Nick thought that Joan had injured her during the race before class.
The doctor recommended a period of rest.

Today’s workout at the gym was tougher than usual.
Joan thought that Lynn had injured her during the race before class.
The doctor recommended a period of rest.

The employees had to look neat for the staff photo.
Sean thought that Jill had dressed her very well for the occasion.
The photo would look nice on the office wall.

The employees had to look neat for the staff photo.
Jill thought that Gwen had dressed her very well for the occasion.
The photo would look nice on the office wall.

The employees had to look neat for the staff photo.
Jill thought that Sean had dressed her very well for the occasion.
The photo would look nice on the office wall.

The end-of-year dinner was at a fancy restaurant.
Matt claimed that Jack had dressed him in the nicest Italian suit.
Everyone had a wonderful evening!

The end-of-year dinner was at a fancy restaurant.
Matt claimed that Jane had dressed him in the nicest Italian suit.
Everyone had a wonderful evening!

The end-of-year dinner was at a fancy restaurant.
Jane claimed that Matt had dressed him in the nicest Italian suit.
Everyone had a wonderful evening!

At the end of November, all the men got rid of their beards together.
Mark thought that Jane had shaved him quite well around his jawline.
Smooth skin is better for the summer months.

At the end of November, all the men got rid of their beards together.
Jane thought that Mark had shaved him quite well around his jawline.
Smooth skin is better for the summer months.

At the end of November, all the men got rid of their beards together.
Mark thought that John had shaved him quite well around his jawline.
Smooth skin is better for the summer months.

Everybody looked unfamiliar at the dress-up party.
John thought that Elle had shaved her to make the costume work.
With a bald head she could be Britney Spears.

Everybody looked unfamiliar at the dress-up party.
Elle thought that Hope had shaved her to make the costume work.
With a bald head she could be Britney Spears.

Everybody looked unfamiliar at the dress-up party.
Elle thought that John had shaved her to make the costume work.
With a bald head she could be Britney Spears.

After reaching the top of the tower, everyone paused to look at the view.
Jack remembered that Matt had turned him to see the beautiful sunset.
It was a great time to take a photo.

After reaching the top of the tower, everyone paused to look at the view.
Jack remembered that Anne had turned him to see the beautiful sunset.
It was a great time to take a photo.

After reaching the top of the tower, everyone paused to see the view.
Anne remembered that Jack had turned him to see the beautiful sunset.
It was a great time to take a photo.

It was hard to fit everybody into the photo.
Joan remembered that Nick had turned her so that everybody fit comfortably.
Now they could take the picture.

It was hard to fit everybody into the photo.
Nick remembered that Joan had turned her so that everybody fit comfortably.
Now they could take the picture.

It was hard to fit everybody into the photo.
Joan remembered that Lynn had turned her so that everybody fit comfortably.
Now they could take the picture.

Everyone had to shower using the hose because of the drought.
Nick claimed that Lynn had washed her without spilling very much water.
It would be nice to take a long bath.

Everyone had to shower using the hose because of the drought.
Lynn claimed that Joan had washed her without worrying about wasting water.
It would be nice to take a long bath.

Everyone had to shower using the hose because of the drought.
Lynn claimed that Nick had washed her without spilling very much water.
It would be nice to take a long bath.

Things got chaotic in the kitchen on Sunday afternoon.
Jack claimed that Matt had washed him after the oil had splattered.
They were glad they didn't mess on the white carpet!

Things got chaotic in the kitchen on Sunday afternoon.
Jack claimed that Anne had washed him after the oil had splattered.
They were glad they didn't mess on the white carpet!

Things got chaotic in the kitchen on Sunday afternoon.
Anne claimed that Jack had washed him after the oil had splattered.
They were glad they didn't mess on the white carpet!

People were upset after the accident at the factory.
Luke claimed that Elle had cut him badly on the broken glass.
The doctor put a bandage on to stop the bleeding.

People were upset after the accident at the factory.
Elle claimed that Luke had cut him badly on the broken glass.
The doctor put a bandage on to stop the bleeding.

People were upset after the accident at the factory.
Luke claimed that Paul had cut him badly on the broken glass.
The doctor put a bandage on to stop the bleeding.

Needlework can sometimes be a dangerous hobby.
Sean remembered that Gwen had cut her with the scissors last summer.
This time they would pay more attention to their hands.

Needlework can sometimes be a dangerous hobby.
Gwen remembered that Jill had cut her with the scissors last summer.
This time they would pay more attention to their hands.

Needlework can sometimes be a very dangerous hobby.
Gwen remembered that Sean had cut her with the scissors last summer.
This time they would pay more attention to their hands.

Non-reflexive verbs

Home improvement projects are always stressful.
Paul claimed that Luke had pushed him into the freshly painted wall.
Now the wall needed to be redone.

Home improvement projects are always stressful.
Paul claimed that Beth had pushed him into the freshly painted wall.
Now the wall needed to be redone.

Home improvement projects are always stressful for people.
Beth claimed that Paul had pushed him into the freshly painted wall.
Now the wall needed to be redone.

There was water all over the floor of the kitchen.
Lynn thought that John had pushed her into the puddle on purpose.
Someone had left the tap running.

There was water all over the floor of the kitchen.
John thought that Lynn had pushed her into the puddle on purpose.
Someone had left the tap running.

There was water all over the floor of the kitchen.
Lynn thought that Elle had pushed her into the puddle on purpose.
Someone had left the tap running.

The presentation was being set up at the front of the room.
John remembered that Jill had followed her to sit at the back.
They wanted to leave before the question session.

The presentation was being set up at the front of the room.
Jill remembered that Elle had followed her to sit at the back.
They wanted to leave before the question session.

The presentation was being set up at the front of the room.
Jill remembered that John had followed her to sit at the back.
They wanted to leave before the question session.

Sometimes it’s stressful to meet new people.
Will claimed that Nick had followed him around the room all night.
He avoided talking to strangers.

Sometimes it’s stressful to meet new people.
Will claimed that Hope had followed him around the function all night.
He avoided talking to strangers.

Sometimes it is stressful to meet new people.
Hope claimed that Will had followed him around the function all night.
He avoided talking to strangers.

The day of the big race was approaching.
Jill remembered that Sean had overtaken her at the race last year.
She would try harder this year to stay ahead.

The day of the big race was approaching.
Sean remembered that Jill had overtaken her at the race last year.
She would try harder this year to stay ahead.

The day of the big race was approaching.
Jill remembered that Gwen had overtaken her at the race last year.
She would try harder this year to stay ahead.

The students skated quickly on the ice rink.
Joan remembered that Nick had overtaken him after the second full lap.
He was out of practice on ice skates.

The students skated quickly on the ice rink.
Nick remembered that Sean had overtaken him after the second full lap.
He was out of practice on ice skates.

The students skated quickly on the ice rink.
Nick remembered that Joan had overtaken him after the second full lap.
He was out of practice on ice skates.

The students found this week’s homework quite difficult.
Luke claimed that Paul had accompanied him to ask the teacher’s advice.
Hopefully there would be a revision session before the exam.

The students found this week’s homework quite difficult.
Luke claimed that Beth had accompanied him to ask the teacher’s advice.
Hopefully there would be a revision session before the exam.

The students found this week’s homework quite difficult.
Beth claimed that Luke had accompanied him to ask the teacher’s advice.
Hopefully there would be a revision session before the exam.

Keeping track of deadlines can be hard for students.
Elle remembered that Will had accompanied her to submit the paper late.
More than one student had been confused.

Keeping track of deadlines can be hard for students.
Will remembered that Elle had accompanied her to submit the paper late.
More than one student had been confused.

Keeping track of deadlines can be hard for students.
Elle remembered that Hope had accompanied her to submit the paper late.
More than one student had been confused.

There were many important people at the business lunch.
Sean claimed that Jane had led her to seats near the CEO.
That was where important decisions were made.

There were many important people at the business lunch.
Jane claimed that Gwen had led her to seats near the CEO.
That was where important decisions were made.

There were many important people at the business lunch.
Jane claimed that Sean had led her to seats near the CEO.
That was where important decisions were made.

Not everyone knew the way back home.
Paul claimed that Jack had led him using a map and compass.
Thankfully they made it back before dark.

Not everyone knew the way back home.
Paul claimed that Gwen had led him using a map and compass.
Thankfully they made it back before dark.

Not everyone knew the way back home from the beach.
Gwen claimed Paul had led him using a map and compass.
Thankfully they made it back before dark.

Everyone was happy for the winners.
Hope remembered that Will had chased her right to the finish line.
It had been a close race.

Everyone was happy for the winners.
Will remembered that Hope had chased her right to the finish line.
It had been a close race.

Everyone was happy for the winners of the race.
Hope remembered that Elle had chased her right to the finish line.
It had been a close race.

The children put the music at the party too loud.
Beth claimed that Matt had chased him to get back the remote.
The parents didn't want the neighbours to complain.

The children put the music at the party too loud.
Matt claimed that Will had chased him to get back the remote.
The parents didn't want the neighbours to complain.

The children put the music at the party too loud.
Matt claimed that Beth had chased him to get back the remote.
The parents didn't want the neighbours to complain.

Chapter 6 – Experimental items

1a The manager who the secretary claimed that the new salesman had pleased will raise company salaries.

1b The manager who the secretary's claim about the new salesman had pleased will raise company salaries.

1c The manager thought the secretary claimed that the new salesman had pleased the boss in the meeting.

1d The manager thought the secretary's claim about the new salesman had pleased the boss in the meeting.

2a The student who the headmaster thought that the clever teacher had surprised does not like doing homework.

2b The student who the headmaster's thoughts about the clever teacher had surprised does not like doing homework.

2c The student believed the headmaster thought that the clever teacher had surprised everybody at school last week.

2d The student believed the headmaster's thoughts about the clever teacher had surprised everybody at school last week.

3a The nurse who the doctor argued that the rude patient had angered is refusing to work late.

3b The nurse who the doctor's argument about the rude patient had angered is refusing to work late.

3c The nurse thought the doctor argued that the rude patient had angered the staff at the hospital.

3d The nurse thought the doctor's argument about the rude patient had angered the staff at the hospital.

4a The witness who the lawyer proved that the evil criminal had confused does not want to testify.

4b The witness who the lawyer's proof about the evil criminal had confused does not want to testify.

4c The witness said the lawyer proved that the evil criminal had confused the judge during the trial.
4d The witness said the lawyer's proof about the evil criminal had confused the judge during the trial.

5a The actress who the journalist suggested that the talented writer had inspired will go on stage tonight.
5b The actress who the journalist's suggestion about the talented writer had inspired will go on stage tonight.
5c The actress thought the journalist suggested that the talented writer had inspired everybody with the new play.
5d The actress thought the journalist's suggestion about the talented writer had inspired everybody with the new play.

6a The customer who the receptionist stated that the lazy cleaner had annoyed will not pay his bill.
6b The customer who the receptionist's statement about the lazy cleaner had annoyed will not pay his bill.
6c The customer thought the receptionist stated that the lazy cleaner had annoyed the manager of the hotel.
6d The customer thought the receptionist's statement about the lazy cleaner had annoyed the manager of the hotel.

7a The farmer who the builder thought that the dedicated worker had amazed will give everybody extra money.
7b The farmer who the builder's thoughts about the dedicated worker had amazed will give everybody extra money.
7c The farmer said the builder thought that the dedicated worker had amazed the new boss last week.
7d The farmer said the builder's thoughts about the dedicated worker had amazed the new boss last week.

8a The singer who the musician stated that the drunken guitarist had offended will not perform this evening.
8b The singer who the musician's statement about the drunken guitarist had offended will not perform this evening.
8c The singer thought the musician stated that the drunken guitarist had offended the drummer after the performance.
8d The singer thought the musician's statement about the drunken guitarist had offended the drummer after the performance.

9a The schoolboy who the teacher proved that the aggressive child had distressed will complain at the meeting.
9b The schoolboy who the teacher's proof about the aggressive child had distressed will complain at the meeting.
9c The schoolboy said the teacher proved that the aggressive child had distressed the class at school yesterday.
9d The schoolboy said the teacher's proof about the aggressive child had distressed the class at school yesterday.

10a The girl who the policeman concluded that the nasty boy had frightened has stopped going to school.
10b The girl who the policeman's conclusion about the nasty boy had frightened has stopped going to school.
10c The girl said the policeman concluded that the nasty boy had frightened the children at the school.
10d The girl said the policeman's conclusion about the nasty boy had frightened the children at the school.

11a The coach who the manager decided that the violent footballer had annoyed will cancel the match today.
11b The coach who the manager's decision about the violent footballer had annoyed will cancel the match today.

11c The coach said the manager decided that the violent footballer had annoyed his fans at the match.
11d The coach said the manager's decision about the violent footballer had annoyed his fans at the match.

12a The politician who the minister stated that the TV journalist had upset will not give an interview.
12b The politician who the minister's statement about the TV journalist had upset will not give an interview.
12c The politician thought the minister stated that the TV journalist had upset the president on the programme.
12d The politician thought the minister's statement about the TV journalist had upset the president on the programme.

13a The chef who the cook argued that the head waitress had bothered wants to find another job.
13b The chef who the cook's argument about the head waitress had bothered wants to find another job.
13c The chef said the cook argued that the head waitress had bothered the manager of the restaurant.
13d The chef said the cook's argument about the head waitress had bothered the manager of the restaurant.

14a The director who the agent suggested that the unpleasant dancer had disappointed will cancel the performance tonight.
14b The director who the agent's suggestion about the unpleasant dancer had disappointed will cancel the performance tonight.
14c The director said the agent suggested that the unpleasant dancer had disappointed the others in the ballet.
14d The director said the agent's suggestion about the unpleasant dancer had disappointed the others in the ballet.

15a The film star who the interviewer suggested that the horrible photographer had embarrassed will not answer any questions.
15b The film star who the interviewer's suggestion about the horrible photographer had embarrassed will not answer any questions.
15c The film star said the interviewer suggested that the horrible photographer had embarrassed the editor of the newspaper.
15d The film star said the interviewer's suggestion about the horrible photographer had embarrassed the editor of the newspaper.

16a The man who the customer thought that the shop assistant had amused was trying not to laugh.
16b The man who the customer's thoughts about the shop assistant had amused was trying not to laugh.
16c The man believed the customer thought that the shop assistant had amused everybody in the store yesterday.
16d The man believed the customer's thoughts about the shop assistant had amused everybody in the store yesterday.

17a The therapist who the patient dreamed that the strange woman had fascinated is writing a new book.
17b The therapist who the patient's dream about the strange woman had fascinated is writing a new book.
17c The therapist said the patient dreamed that the strange woman had fascinated the members of the group.
17d The therapist said the patient's dream about the strange woman had fascinated the members of the group.

18a The man who the detective concluded that the dangerous thief had distressed will buy a new alarm.
18b The man who the detective's conclusion about the dangerous thief had distressed will buy a new alarm.

18c The man thought the detective concluded that the dangerous thief had distressed the people in the neighbourhood.
18d The man thought the detective's conclusion about the dangerous thief had distressed the people in the neighbourhood.

19a The captain who the officer decided that the young soldier had displeased will write a formal report.
19b The captain who the officer's decision about the young soldier had displeased will write a formal report.
19c The captain said the officer decided that the young soldier had displeased the colonel at training today.
19d The captain said the officer's decision about the young soldier had displeased the colonel at training today.

20a The tourist who the guide claimed that the hotel manager had angered wants to return home now.
20b The tourist who the guide's claim about the hotel manager had angered wants to return home now.
20c The tourist believed the guide claimed that the hotel manager had angered everybody in the holiday party.
20d The tourist believed the guide's claim about the hotel manager had angered everybody in the holiday party.