Sentence processing in English natives and non-natives: Evidence from attachment resolution and agreement processing

A thesis submitted in fulfilment of requirements for the degree of

Doctor of Philosophy

School of Psychology and Clinical Language Sciences

University of Reading

Yesi Cheng

June 2021

Abstract

How similarly or differently native speakers (L1ers) and non-native speakers (L2ers) resolve attachment ambiguities and compute agreement during online reading has been widely examined to inform the theoretical debates on whether L1 and L2 acquisition and processing are qualitatively different. While L2 attachment resolution has been studied using either globally (e.g., We called the brother of the man who bought himself a book yesterday.) or temporarily ambiguous relative clauses (e.g., We called the brother of the woman who bought himself a book yesterday.), how L2ers resolve attachment ambiguities when reading both types of relative clauses is little known. While previous studies have examined the roles of linguistic information and participant-level individual differences in L1 and L2 attachment resolution separately, whether these factors modulate L1 and L2 processing to a similar extent has not been studied within the same study. Also, while studies that investigated agreement processing in L2ers with an L1 that does not have agreement (e.g., Chinese) have reported contradicting findings, what underlies the existing inconsistencies is unknown. Further, little is known about how number is specified, in terms of on nouns and other constituents such as demonstratives, may facilitate computation of non-local agreement in L1 and L2ers. Thus, this thesis reports three studies that address these issues, using both offline and online tasks. Study 1 showed that despite some quantitative differences, L1 and L2 attachment resolution are guided by the same parsing strategy, indexed by a low attachment preference, and modulated by similar factors, such as individual differences in lexical automaticity. Study 2 and Study 3 further demonstrated that L1 and L2 agreement computation are not fundamentally different and are modulated by additional number marking similarly, even when L2ers' L1 does not instantiate relevant features. Taken together, the findings from these three studies have lent support to the theories that suggest no qualitative differences between L1 and L2 acquisition and processing.

Acknowledgement

Firstly, I would really like to extend my most sincere gratitude to my fantastic supervisors, Ian Cunnings, Jason Rothman and Fang Liu, who have constantly offered me so much help and guidance during my whole PhD journey. I would surely not have been able to complete my PhD without them. Secondly, many thanks go to Arpita Bose for kindly involving me in her fantastic projects on aphasia and dementia where I have gained a range of analytic skills. I would also very much like to thank all the participants of my experiments for their generous help. Furthermore, many thanks go to The University of Reading and School of Psychology and Clinical Language Sciences as they have offered me the international studentship and laboratory facilities, which have allowed me to conduct my research smoothly. Last but not least, I would really like to thank my examiners for spending their time and energy on reading my thesis during this pandemic.

Declaration: I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Yesi Cheng

Declaration of Authorship

I declare that the three studies in this thesis report my original work in collaboration with my supervisors and have been either published or submitted, and that I am the first author on these papers.

Study 1

Cheng, Y., Rothman, J., & Cunnings, I. (2020). Parsing preferences and individual differences in nonnative sentence processing: Evidence from eye movements. *Applied Psycholinguistics*, *42*(1), 129-151. DOI: http://dx.doi.org/10.1017/S014271642000065X.

Study 2

Cheng, Y., Cunnings, I., Miller, D., & Rothman, J. (Under second round of review). Determiner-number specification matters for both L1 and L2 processing of non-local agreement similarly: An ERP investigation. Submitted to *Studies in Second Language Acquisition.*

Study 3

Cheng, Y., Rothman, J., & Cunnings, I. (Under review). Determiner-number specification and non-local agreement computation in L1 and L2 processing: A self-paced reading study. Submitted to *Journal of Psycholinguistic Research.*

Table of Contents

Chapter 1. General Introduction

Within the fields of second language acquisition and language comprehension, how native (L1) and non-native (L2) speakers process sentences has been a longstanding debate. What has remained strongly debated is whether L2ers acquire and compute syntactic structures in the same way as L1ers during online comprehension. While some predict qualitative differences between L1 and L2 acquisition and processing and claim that L2ers may not be able to acquire or process syntactic features in a native-like way (e.g., Clahsen & Felser, 2006, 2018; Hawkins & Chan,1997; Tsimpli & Dimitrakopoulou, 2007), others argue that, despite some differences, L1 and L2 acquisition and processing are qualitatively similar (e.g., Cunnings, 2017, Lardiere, 2009; McDonald, 2006; Schwartz & Sprouse, 1996). A growing body of research has examined how L2ers process syntactic structures when compared to L1ers during online reading. Attachment resolution for ambiguous relative clauses (RC) and agreement violations have been widely studied in L2 processing as findings from these two domains have been inconsistent. To tease apart the theoretical debate, this thesis examines L1 and L2 processing of ambiguous RCs (Study 1) and non-local agreement violations (Study 2 and 3) when various concerning factors are manipulated, using three different methods that have been extensively used in language comprehension research, which are eye-tracking (Study 1), EEG (Study 2) and self-paced reading (Study 3).

Study 1 investigates attachment preferences for ambiguous RCs in L1ers and L2ers. When a sentence allows multiple interpretations, readers tend to interpret the sentence in a particular way. In the sentence "We called the brother of the man who bought himself a book yesterday", the RC "who bought himself a book yesterday" is globally ambiguous as it can be interpreted as referring back to either "the brother" or "the man". L1ers of English have been reported to prefer attaching the RC to the second noun phrase (NP) "the man" (e.g., Cuetos & Mitchell, 1988). Such preference is termed "low attachment" and taken to indicate that L1

processing is guided by a syntactic parsing strategy called “Late Closure”. However, different attachment preferences have been attested across languages. For example, Spanish L1ers prefer high attachment, meaning that they prefer to attach the RC to the first noun phrase (NP) “the brother” (e.g., Cuetos & Mitchell, 1988). A growing body of research has been motivated to examine attachment preferences in L2ers whose L1 is reported to have an opposite preference to the target language as this can inform the theoretical debate of whether L2ers can employ native-like parsing strategies (e.g., Felser, Roberts, Marinis, & Gross, 2003). While attachment preferences in L2ers have been extensively investigated using either globally or temporarily ambiguous RCs (e.g., We called the brother of the woman who bought himself a book yesterday), how L2ers resolve RC ambiguities when both types of RCs are employed has not been systematically studied. Study 1 aims to address this issue by comparing attachment preferences in L1 and L2ers of English when testing globally and temporarily ambiguous RCs together.

Study 1 also probed how linguistic factors and participant-level individual differences interact with RC attachment preferences in L2 parsing. Previous findings have shown that attachment choices are modulated by discourse-level information in L1 processing (e.g., Hemforth, Fernandez, Clifton, Frazier, Konieczny, & Walter, 2015; Rohde, Levy & Keller, 2011), while little is known about whether L2 attachment resolution is influenced by such linguistic information. Study 1 addresses this question by manipulating the syntactic position of the constituent (subject/object) the RC modifies. While some recent findings suggest that native-like L2 parsing is achievable when individual differences are considered (e.g., Hopp, 2014), whether individual differences modulate L1 and L2 parsing to a similar extent remains unclear. Thus, Study 1 investigates the role of individual differences in L1 and L2 processing within the same study.

Another way of examining how syntactic information is utilised is to look at how readers respond to syntactic violations during processing. Local and non-local subject-verb agreement violations have been widely employed to answer this question (e.g., Dillon, Mishler, Sloggett, & Phillips, 2013; Tanner & Bulkes, 2015). Local agreement violations refer to contexts where the sentence subject does not match the verb (e.g., in number) that is directly adjacent to it (e.g., "The key are rusty"). Non-local agreement violations are more complicated as they contain an intervening noun phrase (intervening NP) that is embedded between the subject and the verb. For example, in the following sentence "The key to the cabinets are rusty", despite the intervening NP "the cabinets" being linearly closer to the verb and congruent with it on number, the singular subject "the key" is the only agreement controller and the verb has to agree with it on number properties according to English syntax. Therefore, the sentence contains a subject-verb agreement violation as the verb is plural as opposed to singular. If readers cannot successfully integrate syntactic information and are simply guided by linear relationship, they would not detect the syntactic error. While previous findings have suggested that L1ers of English are consistently sensitive to non-local agreement violations despite the fact that they sometimes may be influenced by interference from the linearly closer NP (e.g., Dillon et al., 2013; Jäger, Engelmann, & Vasishth, 2017), whether L2ers with a first language that does not license overt agreement ('non-agreement L2ers', e.g., Chinese, Korean) can successfully acquire and process this feature has been widely debated over the past decades due to contradicting results across studies (e.g., Chen, Shu, Liu, Zhao, & Li, 2007; Hawkins & Chan, 1997; Jiang, 2004, 2007; Lim & Christianson, 2015).

Also, recent findings have suggested that sensitivity to local agreement violations, such as "The/Many cookies tastes...", increases when the head noun is modified by a determiner that marks number ("e.g., many") in L1ers of English (Tanner & Bulkes, 2015) but decreases in Chinese L2ers of English under the same condition (Armstrong, Bulkes, & Tanner, 2018).

While this double number marking effect has been examined in the case of local agreement violations in both L1 and L2 processing, it has not been systematically investigated in the case of non-local agreement violations where memory retrieval interference could arise due to the intervening NP. Thus, Study 2 and Study 3 examine whether double marking from determiner-number specification modulates sensitivity to number violations in non-local dependencies. While Study 2 uses ERPs to investigate this effect, Study 3 addresses the same research questions using a self-paced reading task to examine if the same effect can be replicated in a reading time task given that prior research only examined this effect using ERPs (e.g., Armstrong et al., 2018; Tanner & Bulkes, 2015).

Study 2 and Study 3 also aim to help reconcile inconsistencies across L2 studies. While Lim & Christianson (2015) provided reading time evidence of native-like agreement processing in Korean L2ers of English, ERP evidence from Chen et al. (2007) suggested non-agreement L2ers of English may employ a different neural mechanism to process agreement. However, Chen et al. (2007) tested their L2ers in China whereas Lim & Christianson tested L2ers in the United States where immersive L2 input is provided. Thus, little is known about whether the neural substrates underlying agreement processing are different from L1ers under similar conditions to Lim & Christianson (2015). Study 2 can help shed light on this issue by testing a similar structure to Chen et al. (2007) on Chinese L2ers of English living in the United Kingdom, using ERP evidence. Furthermore, while Jiang (2004) suggested that immersed Chinese L2ers of English were insensitive to non-local agreement violations in a self-paced reading task (cf. Lim & Christianson, 2015), the conclusion was solely based on a significant grammaticality effect in the L1 group that was absent in the L2 group, despite numerical trends that were in the same direction as the L1 group. The fact that Jiang (2004) did not test for a group by grammaticality interaction makes it difficult to be certain that the L1 and L2 groups did indeed behave differently. By considering this critical methodological difference, Study 3

can serve as a good comparison to Jiang (2004) as it tests a similar syntactic structure and population, using a self-paced reading task.

The remainder of this chapter discusses relevant L1 and L2 models that are adopted to account for previous L1 and L2 findings with regards to attachment resolution of ambiguous RCs and processing of non-local agreement violations.

1.1 L1 sentence processing and attachment resolution

Attachment preferences have been examined in L1 sentence processing and various findings have been explained mainly using three different theoretical models, which are the garden path model, the competition models and the unrestricted race model.

1.1.1 Garden path model

One of the widely known models in the attachment resolution literature is the garden path model (Frazier,1987). It predicts that parsing is serial so that parsers only construct one syntactic analysis at a time. It also claims that initial parsing decisions are determined by syntactic information only and guided by two parsing principles, namely "Minimal Attachment" and "Late Closure". While Minimal Attachment predicts that the simplest analysis is initially adopted, Late Closure predicts that incoming material should be attached to the constituent that is currently in the parse. The latter is particularly relevant to attachment resolution of ambiguous RCs. For the sentence "We called the brother of the man who bought himself some books yesterday", the garden path model predicts that low attachment ("the man") should be favoured at "who" as the NP "the man" is the most recently processed material.

The garden path model has been supported by evidence from English L1 studies using various methods. For example, an early study conducted by Cuetos & Mitchell (1988) simply

examined offline attachment preferences in English L1ers using ambiguous RCs, as in (1). They reported that the L1ers of English preferred to attach the RC low, which was in accordance with "Late Closure".

(1) The journalist interviewed the daughter of the colonel who had had the accident.

Studies measuring online reading times also attested this low attachment preference (e.g., Felser, Roberts et al., 2003, Hopp, 2014). Unlike Cuetos & Mitchell (1988), Felser, Roberts et al. (2003) adopted temporarily ambiguous RCs that are disambiguated to attach either low or high via number agreement, as in (2).

(2a) The dean liked the secretary of the professors who was reading a letter.
(2b) The dean liked the secretary of the professors who were reading a letter.

The temporary ambiguity arises at the relative pronoun "who" as the relative pronoun can be interpreted as referring back to either the high attachment site "the secretary" or the low attachment site "the professors". The sentence is disambiguated later at the auxiliary verb ("was/were") and forced to attach high, as in (2a) or low, as in (2b), due to number congruency between the auxiliary verb and NPs. If attachment resolution is guided by "Late Closure" as predicted by the garden path model, English L1ers should only show processing difficulty and longer reading times at "was" in (2a) where low attachment is impossible. In line with the model, Felser, Roberts et al. found longer reading times in (2a) than in (2b) at the disambiguation region, which is taken to suggest that L1 initial parsing preferences are guided by syntactic information.

1.1.2 Competition models

Contrary to the garden path model, the competition models (also known as constraint-based theories) predict that parsing is parallel and interactive (e.g., MacDonald, 1994; Tabor & Tanenhaus, 1999; Taraban & McClelland, 1988; Trueswell, Tanenhaus, & Garnsey, 1994). To resolve syntactic ambiguity, all possible analyses are constructed simultaneously and compete for activation using different linguistic resources. Little competition should occur when linguistic constraints clearly favour one of the analyses. Competition becomes particularly strong when all the analyses are equally possible due to receipt of same amount of support, which causes processing difficulty. Also, parsing is highly interactive in this account, so that any relevant syntactic and non-syntactic linguistic information (e.g., semantics, discourse) can immediately influence the initial stage of syntactic ambiguity resolution.

The competition models have been tested in studies investigating syntactic ambiguity resolution. For example, MacDonald, Just, & Carpenter (1992) compared temporarily ambiguous sentences, such as (3a), and unambiguous sentences, such as (3b).

(3a) The experienced soldiers warned about the dangers before the midnight raid.
(3b) The experienced soldiers spoke about the dangers before the midnight raid.

(3a) contains temporary ambiguity at "warned" as it can be analysed either as a matrix verb (main clause analysis) or a past participle (reduced relative clause analysis). In contrast, the main verb "spoke" in (3b) can only render main clause analysis as it cannot be analysed as a participle. Hence, according to competition models, longer reading times should be expected in (3a) than in (3b) as competition between the two analyses in (3a) causes processing difficulty. Compatible with the models, MacDonald et al., (1992) reported longer reading times

in (3a) at the region “before the midnight”, which was taken to suggest that both analyses were equally activated in memory for competition before ambiguity can be resolved.

Nevertheless, evidence from attachment ambiguity resolution is taken against the competition models (e.g., Van Gompel, Pickering, & Traxler, 2001; Van Gompel et al., 2005). For example, Van Gompel et al. (2005) tested whether single or multiple analyses are computed by L1ers of English during attachment resolution of ambiguous RCs.

(4a) The governor of the province that will be retiring after the troubles is very rich.
(4b) The province of the governor that will be retiring after the troubles is very rich.
(4c) The bodyguard of the governor that will be retiring after the troubles is very rich.

Both (4a) and (4b) contain temporary ambiguity that is resolved via plausibility information at the verb “retiring” as only the animate NP “governor” can be the agent of the verb. Therefore, only the high attachment analysis in (4a) and the low attachment analysis in (4b) are possible. (4c) is globally ambiguous as both high attachment site “the bodyguard” and low attachment site “the governor” are animate and either can be the agent of the verb “retiring”. According to the competition models, (4c) should be more difficult to read as both analyses receive equal amount of support and hence lead to a stronger competition, in comparison to the other two conditions where only one analysis is plausible. However, the results showed that despite the reading times being equally long in (4a) and (4b), (4c) had shorter reading times than the other two conditions. Hence, the findings from Van Gompel et al. (2005) suggested that parsing is not parallel, and no competition occurs during attachment ambiguity resolution.

1.1.3 Unrestricted race model

Even though the majority of prior research in English has lent support to a low attachment preference, some recent findings indicated that this preference may not be as strong as claimed as a variable attachment choice was attested in English L1ers when both globally and temporarily ambiguous RCs were tested together (e.g., Traxler, Pickering, & Clifton, 1998), which cannot be explained by the garden path model or the competition models. Thus, the unrestricted race model was adopted to account for such finding. The unrestricted race model is a hybrid model built upon the garden path model and the competition models (e.g., Traxler, Pickering, & Clifton, 1998; Van Gompel, Pickering, & Traxler, 2000). Specifically, similar to the garden path model, this model claims that parsing is serial so that the parser only computes one analysis at a time. Also, like the competition models, it suggests that parsing is interactive, meaning that the initial syntactic choice is not made using syntactic information only but also resources from other linguistic aspects. In this case, any attachment analysis could be favoured. Whichever analysis is favoured by the various linguistic information used during initial parsing will win the race and be adopted.

Several attachment resolution studies have reported findings that are in accordance with the unrestricted race model (e.g., Traxler et al., 1998; Van Gompel et al., 2001; Van Gompel, Pickering, Pearson, & Liversedge, 2005). Traxler et al. (1998) tested how RC attachment ambiguity is resolved in sentences that are globally ambiguous and those disambiguated using semantic information, as in (5).

(5a) The driver of the car that had the moustache was pretty cool.

(5b) The car of the driver that had the moustache was pretty cool.

(5c) The son of the driver that had the moustache was pretty cool.

(5a) and (5b) are temporarily ambiguous at the relative pronoun "that" as syntactically "that" can attach to either "the driver" or "the car". However, the sentence is disambiguated to high attachment in (5a) and low attachment in (5b) as the RC "that had the moustache" can only attach to "the driver" due to semantics. In contrast, the RC in (5c) can attach to either "the son" or "the driver" as either attachment will be plausible. The unrestricted race model predicts that only the analysis that wins the race is activated, and either low or high attachment could be chosen as attachment choices may vary due to the interplay between different sources of linguistic constraints. Therefore, (5c) should be easier to process compared to (5a) and (5b) since whichever analysis is activated will turn out to be plausible in the end. Also, the parser should encounter processing difficulty in (5a) and (5b) as there will be 50% of chance the initially chosen attachment turns out to be implausible at the disambiguation region "the moustache". Consistent with the unrestricted race model, Traxler et al. (1998) reported more processing difficulty for the sentences where high attachment is forced, like (5a), and for those where low attachment is forced, like (5b), in comparison to (5c) where either attachment is correct. In addition, the reading times did not statistically differ between the forced high and low attachment conditions. Hence, the results showed that the RC is variably attached to one of the attachment sites during initial parsing, which could imply that attachment resolution may also be guided by non-syntactic factors.

In summary, despite some evidence of a low attachment preference, as predicted by the garden path model, in English L1ers, it may not be as robust as indicated because other existing findings suggest that L1ers of English have a variable attachment choice that may be influenced by a range of factors, as predicted by the unrestricted race model.

1.2 L1 sentence processing and subject-verb number agreement

The mechanisms that underlie processing of subject-verb agreement have been widely investigated. It involves a series of complex processes including encoding words and phrases in working memory and retrieving relevant information later for successful comprehension, which can be modulated by other factors such as number marking.

1.2.1 Cue-based retrieval model

One of the influential models that has been used to explain findings from agreement research is the cue-based retrieval model (e.g., Lewis & Vasishth, 2005). This model claims that when memory retrieval is initiated, a set of retrieval cues are compared against all possible lexical items simultaneously in memory. The item that provides the best match to the retrieval cues will be retrieved as the target. To comprehend sentences such as "The key to the cabinets is rusty", memory retrieval is initiated at the verb "is" and the operation starts to search for a noun in memory that provides the best match to the retrieval cues to the verb "is". In this case, the retrieval cues are "subject" and "singular". Thus, "the key" is the most appropriate item and will be retrieved.

The cue-based retrieval model has also been used to account for how readers process agreement violations, and predicts that memory retrieval is subject to similarity-based interference (e.g., Jäger, Engelmann, & Vasishth, 2017; Jäger, Mertzen, Van Dyke, & Vasishth, 2020; Lewis, Vasishth, & Van Dyke, 2006; Van Dyke, 2007; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2011). Interference typically occurs in ungrammatical sentences where the distractor partially matches the cue while the target does not (e.g., "The key to the cabinets were rusty"). The intervening NP ("the cabinets") may be activated and mistakenly retrieved as the grammatical target at the verb ("are") as its number feature matches the retrieval cue "plural noun", despite the fact that the sentence subject ("the key") is the only grammatically

accessible target. Such retrieval interference could give rise to grammatical illusions (e.g., Phillips, Wagers, & Lau, 2011).

Interference has been extensively tested in studies investigating processing of number agreement in subject-verb dependencies (e.g., Chen, Shu, Liu, Zhao, & Li, 2007; Clifton, Frazier, & Deevy, 1999; Dillon et al., 2013; Eberhard, Cutting, & Bock, 2005; Pearlmutter, Garnsey, & Bock, 1999; Tanner, 2011). Previous findings have suggested that although L1ers of English are sensitive to number agreement violations during online reading, they are sometimes susceptible to interference, indexed by a cognitive ease-up when reading sentences containing agreement violations. For example, Wagers, Lau, & Phillips (2009) conducted a series of self-paced reading experiments to investigate whether readers are influenced by the number feature of the intervening NP in sentences containing non-local agreement, as in (6).

(6a) The key to the cell was rusty from many years of disuse.
(6b) The key to the cells was rusty from many years of disuse.
(6c) * The key to the cells were rusty from many years of disuse.
(6d) * The key to the cell were rusty from many years of disuse.

The results found that L1ers had slower reading times for (6c) and (6d), where subject-verb agreement violations occur, than the grammatical conditions (6a) and (6b). They also found that (6c), where the intervening NP matches the verb in number, elicited significantly shorter reading times than (6d) where it does not, even though they were both ungrammatical. This is taken as evidence of retrieval interference.

Studies using electrophysiological evidence also replicated this interference effect in sentences like (7) (e.g., Shen et al., 2013; Tanner, 2011).

(7a) The winner of the big trophy has proud parents.

(7b) *The winner of the big trophy have proud parents.

(7c) The winner of the big trophies has proud parents.

(7d) *The winner of the big trophies have proud parents.

Similar to Wagers et al. (2009), Tanner (2011) reported that even though both ungrammatical conditions (7b) and (7d) elicited a P600 effect (an index of detection of agreement violations) at the critical verb, the effect was smaller in the number match condition (7d) than the number mismatch condition (7b), suggesting retrieval interference from the intervening NP that matches the verb on number properties.

1.2.2 Number marking and agreement processing

Tanner, Nicol, & Brehm (2014) proposed that number agreement comprehension involves prediction upon number properties and cue-based retrieval when linguistic predictions fail. Under this approach, if the number feature of a noun becomes more pronounced, the prediction for the upcoming verb's number feature becomes stronger, giving rise to stronger sensitivity to agreement violations when the prediction fails at the verb. Subsequently, Tanner & Bulkes (2015) used ERPs to test how double number marking on the subject NP influences sensitivity to simple number agreement violations, as in (8).

(8a) The <u>cookies</u> taste/*tastes the best when dipped in milk.

(8b) Many <u>cookies</u> taste/*tastes the best when dipped in milk.

In (8a), the subject contains a determiner "the" that does not specify number as it can modify either singular or plural nouns. In (8b), the quantifier "many" marks plurality and thus

creates double number marking where number is more explicitly marked in combination with the plural marker "-s". According to Tanner et al. (2014), sensitivity to agreement violations should be stronger in (8b) where a stronger prediction is constructed due to double number marking, in comparison to in (8a) where number properties are comparatively less marked. In line with this account, Tanner & Bulkes (2015) reported that the P600 effect elicited by violations in (8b) was larger than that in (8a). Their findings suggested that how number is marked may modulate sensitivity to agreement errors, at least in L1 processing.

1.3 L2 acquisition and sentence processing

There has been a long-standing debate on whether L1 and L2 acquisition and processing are qualitatively different, and several theories have been proposed to address this issue (e.g., Cunnings, 2017; Clahsen & Felser, 2006, 2018; Hawkins & Chan,1997; Hawkins & Casillas, 2008; Hopp, 2014, 2016; Lardiere, 2009; Schwartz & Sprouse, 1996; Ullman, 2005). In L2 acquisition, while some hypotheses propose no or limited access to Universal Grammar in adult L2 learners and thus predict failure to acquire L2 grammatical features in a native-like way (e.g., Bley-Vroman, 1989; Hawkins & Chan,1997; Hawkins & Casillas, 2008; Tsimpli & Dimitrakopoulou, 2007), other approaches argue that despite some L1/L2 differences due to various factors, adult L2ers have full access to Universal Grammar (e.g., Lardiere, 2009; Schwartz & Sprouse, 1996; Prévost & White, 2000). In L2 processing, some models predict qualitative differences between L1 and L2 processing (e.g., Clahsen & Felser, 2006, 2018; Jiang, 2004). For example, Jiang (2004) argued that such differences arise due to deficit in L2ers' integrated linguistic competency, with difficulty integrating morphosyntactic information during real-time processing. Meanwhile, other models predict only quantitative L1/L2 differences, especially when individual differences are taken into consideration (e.g., Hopp, 2014; McDonald, 2006). Current findings from L2 acquisition and processing research

have not resolved these disputes as contradicting findings have been reported (e.g., Chen et al., 2007; Lardiere, 2007; Lim & Christianson, 2015; Felser, Roberts et al., 2003; Hopp, 2014).

The following will discuss several L2 models in detail that are particularly relevant to this thesis. This includes the Failed Functional Features Hypothesis (FFFH; Hawkins & Chan,1997), the Interpretability Hypothesis (Tsimpli & Dimitrakopoulou, 2007), the Full Transfer/Full Access Hypothesis (FT/FA; Schwartz & Sprouse, 1996), the Shallow Structure Hypothesis (SSH; Clahsen & Felser, 2006), the Lexical Bottleneck Hypothesis (Hopp, 2014), and the Interference model (Cunnings, 2017).

1.3.1 L2 acquisition models

1.3.1.1 Failed Functional Features Hypothesis and Interpretability Hypothesis

As a language acquisition theory, the FFFH (Hawkins & Chan,1997) differs from the SSH in terms of the source of L1/L2ers differences. Instead of claiming differences in the way syntactic information is utilised, the FFFH holds that the root of non-native-like performance lies in the parameter setting in the language faculty which is strongly influenced by one's L1. Specifically, during L1 acquisition, children adjust the parameters of L1 grammar including functional categories, such as agreement, as they learn the language. Once the parameters of these L1 functional features are set and fixed, they are impossible to reset after the critical period. Thus, when adult L2ers learn a novel L2 functional feature that has a different parametric setting from their L1, they cannot set a new parameter or readjust the relevant parameter for that feature in the language faculty, and as a result, this L2 grammar cannot be integrated into adult L2ers' linguistic representation.

As a more recent version of the FFFH, the Interpretability Hypothesis (Tsimpli & Dimitrakopoulou, 2007) maintains that the critical period effect only applies to uninterpretable features. While the adult L2ers have access to meaningful and interpretable features in

Universal Grammar, they do not have access to purely grammatical features that lack semantic content and are not instantiated in their L1. According to Carstens (2000), number is argued to be interpretable on nouns as it indicates semantic information about quantification, but uninterpretable on agreement targets. Thus, number on verbs is uninterpretable as it does not make a semantic contribution to the interpretation and only functions as a syntactic feature. Under these two accounts, L2ers are assumed to have difficulty with building a native-like mental representation of a syntactic structure that is absent in their L1 and so cannot fully acquire it. This has motivated researchers to investigate how agreement violations are processed by L2ers of English with an L1 that does not have agreement, such as Chinese.

Several studies have shown findings that may be taken as evidence for these theories (e.g., Jiang, 2004, 2007; Chen et al., 2007). For example, Lardiere (1998a, 1998b, 2007) conducted a case study with a Chinese L1er who had been living in the US for a long time and found that this participant struggled to produce the third-person singular "-s" in obligatory contexts in a production experiment. Jiang (2004) examined how Chinese L2 learners of English process long-distance subject-verb agreement in contexts like (9) and (10) during language comprehension.

(9a) The key to the cabinet was rusty from many years of disuse.
(9b) The key to the cabinets was rusty from many years of disuse.

For grammatical sentences like (9), unlike the L1ers, the Chinese L2ers did not show significantly longer reading times for (9b), where the intervening NP mismatches the verb on number, than the baseline (9a) where a number match occurs. This was taken to suggest that the Chinese L2ers were not sensitive to number agreement as otherwise they would have noticed the mismatch between the local NP and verb and shown significantly slower reading

times for (9b). In addition, Jiang (2004) argued that the Chinese L2ers were not sensitive to subject-verb agreement violations as otherwise they would have shown longer reading times for sentences containing agreement violations, such as (10b), than grammatical sentences as in (10a), but the relevant comparison was not significant.

(10a) The bridges to the island were about ten miles away.
(10b) *The bridge to the island were about ten miles away.

Chen et al. (2007) investigated comprehension of subject-verb agreement in Chinese L2ers of English using ERPs. Despite the fact that their Chinese L2ers detected the errors in the non-local agreement as in (11a) and (11b), indexed by high grammaticality judgement accuracy, the brain responses elicited by agreement violations in the L2ers were qualitatively different from those in the L1ers.

(11a) *The price of the cars were too high.
(11b) *The price of the car were too high.

As Chen et al. argued, since specific L1 experience shapes neural substrates of language processing, Chinese L2 learners cannot employ native-like processing mechanisms due to the fact that Chinese does not have agreement.

1.3.1.2 Full Transfer/Full Access Hypothesis

In contrast to the FFFH which claims unavailable parameter resetting after the critical period, the Full Transfer/Full Access (FT/FA) hypothesis (Schwartz & Sprouse, 1996) maintains that parameter resetting is possible during L2 acquisition and therefore, novel structures can be

acquired by adult L2ers. Specifically, this hypothesis proposes that when L2 acquisition commences, the L1grammar and its parameter settings are fully transferred to constitute the initial state of L2 acquisition. At this stage, as adult L2 learners use the L1 grammar to learn the L2, they may show errors, such as morphological inflection in agreement. However, L2 learners can continue to acquire the L2 grammar by modifying the initial grammar transferred from the L1 until it becomes target-like as they have full access to Universal Grammar and can reset linguistic parameters in the language faculty. Thus, learning a novel L2 structure that the L1 does not have is achievable as the new L2 parameter can be added.

Findings from several L2 studies are consistent with this hypothesis (e.g., Ågren, Michot, Granget, Gerolimich, Hadermann, & Stabarin, 2021; Sagarra & Herschensohn, 2011; White, Valenzuela, Kozlowska-Macgregor, & Leung, 2004). For example, White et al. (2004) examined acquisition of gender agreement by L2 learners of Spanish who had an L1 with (French) and without (English) gender agreement, using oral production and picture identification tasks. The findings showed that intermediate and advanced English L2ers of Spanish did not significantly differ from the L1 Spanish controls and the French L2ers of Spanish, despite the fact that English does not have gender agreement between determiners, nouns and adjectives. Similar findings have been reported in studies that examined gender agreement during language comprehension, suggesting that English L2 leaners of Spanish can acquire and process this feature in a native-like way as they, similar to L1ers, showed a P600 effect for gender agreement violations (e.g., Alemán Bañón, Fiorentino, & Gabriele, 2014; Alemán Bañón, Miller, & Rothman, 2017).

Following Tanner & Bulkes (2015) (see 8), Armstrong et al. (2018) tested how double number marking influences number agreement processing in Chinese L2ers of English. The results showed that Chinese L2ers showed a P600 effect to number agreement violations, which is expected under the FA/FT. The findings also showed that the effect was reduced when the

subject NP had double number marking (e.g., “many cookies”), which stands in contrast to the L1 finding in Tanner & Bulkes (2015). Armstrong et al. argued that this was due to an L1-transfer strategy employed by Chinese L2ers of English. Since Chinese, like English, can mark number using quantifiers (e.g., “many”), Chinese L2ers of English relied more on this lexical cue shared by the L1 and L2 than the morphological number cues for a more robust number encoding when such information is available. As a result, they reallocated more cognitive resources to quantification than morphological agreement marking, leading to a reduced P600 amplitude in the double marking condition.

1.3.2 L2 processing models

1.3.2.1 Shallow Structure Hypothesis

The SSH (Clahsen & Felser, 2006) is a processing theory which proposes that adult L2ers utilise structural information in a fundamentally different, or shallower way than L1ers during sentence processing. This difference is assumed to arise from L2ers’ reduced ability to use syntactic cues and overreliance on non-syntactic cues, such as lexical-semantic information. In other words, when processing a sentence, L2ers rely on their knowledge of the individual words in the sentence rather than the holistic syntactic relationship. Thus, they are predicted to show difficulties when processing syntactically complex structures, such as temporarily ambiguous RCs and non-local agreement.

A number of L2 studies have reported L2 difficulty with attachment resolution for temporarily ambiguous RCs (e.g., Dinçtopal-Deniz, 2010; Felser, Roberts et al., 2003; Omaki, 2005; Papadopoulou & Clahsen, 2003). For example, Felser Roberts et al., (2003) and Papadopoulou & Clahsen (2003) both investigated attachment preferences for RCs containing the lexical preposition “with” like (12a) and for those containing the structural preposition “of” like (12b), in a self-paced reading task.

(12a) The man liked the teacher with the boy who was smiling all the time.

(12b) The man liked the teacher of the boy who was smiling all the time.

They found that despite the fact that both the L1 and L2ers showed a low attachment preference when processing sentences containing the lexical preposition "with", the L2ers did not show any attachment preferences for RCs containing the structural preposition "of", unlike the L1ers who showed a low attachment preference. These findings were taken to suggest that L2ers are similar to L1ers in terms of using lexical-semantic information, but different when using structural information.

There are other studies indicating L2ers' overreliance on non-syntactic linguistic information over structural cues during attachment resolution. Dinçtopal-Deniz (2010) looked at whether attachment preferences would be influenced by animacy in Turkish speakers of English, using RCs disambiguated by plausibility. The results suggested that while the L1ers preferred low attachment irrespective of animacy, the L2ers' attachment preferences were influenced by animacy. In addition, Pan, Schimke, & Felser (2015) suggested that discourse-level information, such as preceding context, modulates attachment preferences in German and Chinese L2ers of English more strongly than in L1ers. They manipulated preceding contexts to bias towards either high or low attachment and found that the L2ers favoured high attachment in high attachment-biasing contexts and preferred low attachment in low attachment-biasing contexts.

Tanner et al., (2012) discussed how one instantiation of shallow L2 processing might be a reliance on linear order during the processing of non-local dependencies. As a result, L2ers may use the intervening NP to control agreement. In Chen et al. (2007), the Chinese L2ers of English showed a more positive brain response (P600 effect) for (13b) where the intervening

NP does not match the verb on number, compared to (13a), where it does, even though both conditions are grammatical. The P600 effect indicated that the Chinese L2ers interpreted grammatical sentences, as in (13b), as ungrammatical, showing that the L2ers might have been influenced by the local linear adjacency.

(13a) The price of the car was too high.
(13b) The price of the cars was too high.

In Tanner et al. (2012), the Spanish L2ers of English showed a P600 effect to syntactic violations in non-local agreement, as in (14a) and (14b). However, this effect observed in (14b) where the local NP and verb match on number properties was reduced compared to (14a) where a mismatch occurs.

(14a) *The winner of the big trophy have proud parents.
(14b) *The winner of the big trophies have proud parents.

This seems to be compatible with one of their initial assumptions with regards to L2ers' reliance on local linear adjacency. However, as Tanner et al. (2012) argued, if the attenuated sensitivity in (14b) resulted from L2ers' reliance on linear relationship, they would have observed a P600 effect for grammatical sentences like (13b) as the number mismatch between the local NP and verb would have been taken as ungrammatical. However, this was not observed. This implies that L2ers did not simply rely on linear order when processing non-local dependencies. Furthermore, the attenuated P600 effect was also found in the L1ers, indicating that L2ers weren't being any shallower than L1ers. Therefore, Tanner et al. argued, their findings cannot be explained by the SSH.

1.3.2.2 Alternative models of L2 processing

Unlike models that predict L2ers' reduced ability to utilise syntactic information, several L2 processing theories predict that L1 and L2 processing are qualitatively similar, but they may be quantitatively different due to various factors. The Lexical Bottleneck Hypothesis proposed by Hopp (2014) claims that the differences between L1 and L2ers do not lie in syntactic integration but individual differences in lexical access efficiency which reflects quantitative differences in resource allocation. Slowdown in lexical processing that subserves syntactic processing may give rise to parsing difficulty in L2ers, such as attenuated or delayed syntactic integration. Hopp (2014) provided evidence for this hypothesis by investigating attachment resolution in German L2ers of English. Specifically, the findings showed that while L2ers with slow lexical processing speed failed to exhibit any structural preferences, those with faster lexical access were more native-like by showing a low attachment preference. Therefore, it is compatible with the claim that native-like L2 syntactic processing is possible when quantitative differences are neutralised.

The Interference model (Cunnings, 2017) argues that L2ers can compute structures but may have difficulties due to memory retrieval interference. Under this account, L1/L2 differences arise from cue-based memory retrieval that involves retrieving information encoded in working memory for successful language comprehension. During retrieval process, competitors that partially match the retrieval cues may be activated in memory and cause retrieval interference. Even though L2ers can process complex structures like L1ers, they are predicted to be more susceptible to retrieval interference during sentence processing in comparison to L1ers.

Cunnings (2017) used this model to account for previous findings in L2 attachment resolution. Recall that a few studies did not attest a clear attachment preference in L2ers, which was taken as evidence of shallow parsing (e.g., Felser, Roberts et al., 2003; Omaki, 2005;

Papadopoulou & Clahsen, 2003). Cunnings argued that null preferences could mean variable attachment choices rather than indicating shallow parsing as null preferences have also been observed in L1ers (e.g., Traxler et al., 1998; Van Gompel et al., 2005). Secondly, prior research has shown that attachment preferences are modulated by a number of factors including discourse contexts (e.g., Pan et al., 2015; Pan & Felser, 2011), working memory span (e.g., Swets, Desmet, Hambrick, & Ferreira, 2007) and task demands (Swets, Desmet, Clifton & Ferreira, 2008), which could have contributed to variable attachment to either high or low attachment site. Finally, some recent studies have attested clear attachment preferences in L2ers (e.g., Hopp, 2014; Witzel, Witzel and Nicol, 2012), which clearly does not support the SSH. Further, Cunnings (2017) argued that the findings that L2 attachment preferences are influenced by discourse context (e.g., Pan et al., 2015; Pan & Felser, 2011) may not suggest that L2 processing rely more heavily on discourse-level information over syntactic cues. Rather, it may reflect L2ers increased sensitivity to adopting pragmatically appropriate interpretations and their increased reliance on discourse-based cues to memory retrieval compared to L1ers.

The Interference model is also used to account for findings from L2 agreement processing. Keating (2010) examined processing of gender agreement in L1 and L2ers of Spanish using sentences like (15) where grammaticality was manipulated via gender agreement between the head noun and adjective. Unlike (15a), (15b) contains an intervening NP that either matches or mismatches the adjective in gender.

(15a) La tienda está abierta/*abierto los sábados y domingos por la tarde.

'The store-FEM is open-FEM/*open-MASC Saturdays and Sundays in the afternoon.'

(15b) El vestido de la muchacha es rosado/*rosada y tiene lunares blanco.

'The dress-MASC of the girl-FEM is pink-MASC/*pink-FEM and has white polka dots.'

The results showed that while the L1 group had slower reading times for ungrammatical sentences where the adjective mismatched the subject on gender in both (15a) and (15b), the L2 group only showed this grammaticality effect for (15a). Cunnings (2017) argued that as (15b) has an intervening NP ("the girl") that may serve as a distractor, there is a greater chance of interference when the distractor partially matches the retrieval cues to the adjective. Therefore, the distractor may be retrieved as the target in the ungrammatical condition in (15b) as it matches the gender of the adjective. Such retrieval interference leads to a speedup in processing of sentences containing gender violations, which neutralised the reading time difference between the grammatical and ungrammatical conditions and cancelled out the grammaticality effect for (15b) in the L2 group. Thus, Cunnings suggested that L2ers are more sensitive to interference than L1ers.

1.4 This thesis

The literature discussed above has shown that even though L1ers of English prefer low attachment for temporarily ambiguous RCs, variable attachment may present when a globally ambiguous condition is provided to serve as a baseline (Van Gompel et al., 2005), which indicates that attachment choices may be influenced not only by syntactic information but also by a range of non-syntactic factors. While L2ers of English have exhibited either clear (e.g., Hopp, 2014) or null attachment preferences (e.g., Felser, Roberts et al., 2003), existing L2 studies, however, have not included a globally ambiguous condition. Including such a condition for L2 testing provides an important reference when compared with L1ers as results could inform us about the implications of the null preference observed in previous studies. If L2ers also show variable attachment like L1ers under the same condition, it may suggest that the variable attachment choice in L2ers does not imply shallow parsing and that L1 and L2

parsing are not fundamentally different. Since attachment preferences could be modulated by various kinds of factors (Cunnings, 2017), whether linguistic information such as syntactic position of the RC plays a role in L2 attachment resolution is little known. Additionally, it is currently unknown whether individual differences regulate L1 and L2 attachment resolution to a similar extent as no published research has examined this effect in both L1 and L2ers within the same study.

Previous findings from agreement research show that while L1ers can consistently detect non-local agreement violations, their sensitivity to syntactic violations may be reduced due to retrieval interference. Whether non-agreement L2ers can acquire and process non-local agreement in a native-like way still remains debated and the factors contributing to the inconsistencies across studies are unclear. Furthermore, while relevant findings suggest that sensitivity to violations is modulated by how number is marked in the case of local agreement (e.g., Armstrong et al., 2018; Tanner & Bulkes, 2015), this has not been examined in the context of non-local agreement where sensitivity to violations can be influenced by interference from the intervening NP.

This thesis reports three studies that aim to test the theoretical models and address those aforementioned issues. Study 1 investigates attachment resolution in L1 and L2ers of English and its interaction with discourse-level information and individual differences, using both globally and temporarily ambiguous RCs. Study 2 and 3 examine whether Chinese L2ers of English process non-local agreement differently from L1 controls and how number marking modulates detection of non-local agreement violations in L1 and L2ers.

In Study 1, offline data were collected using a comprehension task to tap into offline attachment preferences. Participants simply read globally ambiguous RCs and indicated their preferences by answering comprehension questions (e.g., “We called the brother of the man who bought himself a book yesterday. Question: Who bought himself a book yesterday?”).

Online attachment preferences were examined using an eye-tracking during reading paradigm. In Study 2, an offline judgement task was administered to test whether participants know explicit grammatical rules of subject-verb agreement. Online data were measured using EEG which examines implicit use of the grammar during incremental reading. In Study 3, online data were collected using a self-paced reading task with a concurrent offline judgement task.

To briefly summarise the results, Study 1 revealed that both L1 and L2 parsing are guided by "Late Closure" as L1ers and L2ers indicated a low attachment preference during offline and online reading, though the experimental effects were generally larger in L1ers. This preference was modulated by discourse-level information, reading span and lexical automaticity to a similar degree in L1 and L2 processing. The judgement data and ERP responses in Study 2 suggested that Chinese L2ers of English can detect non-local agreement violations and the neurocognitive processes underlying this operation were qualitatively similar to L1ers. However, the ERP effect was quantitatively smaller compared to L1ers', suggesting that Chinese L2ers may have experienced a stronger interference from intervening constituents (Cunnings, 2017). Also, double number marking was found to facilitate detection of non-local violations in L1ers and Chinese L2ers, indexed by increased judgement accuracy and enhanced neural sensitivity. This suggests that double number marking was processed similarly by L1ers and Chinese L2ers. Study 3 also demonstrated evidence that Chinese L2ers of English did not process non-local agreement and double number marking differently to L1ers. However, while double number marking facilitated detection of agreement violations, this effect was only found in the offline judgement data. In addition, the judgement results suggested that Chinese L2ers may be more sensitive to interference from non-local linguistic dependencies (Cunnings, 2017).

In summary, I argue that the findings from the three studies show that L2ers can acquire and parse syntactic structures in the same way as L1ers, even when the parsed L2 feature is

opposite to their L1 or entirely novel. However, the results suggest L2ers are more susceptible to interference during online reading and sometimes may show smaller effects. Therefore, the results from this thesis suggest quantitative rather than qualitative differences between L1 and L2 processing. The following three chapters report these three studies in more detail.

Chapter 2. Study 1: Parsing preferences and individual differences in non-native sentence processing: Evidence from eye-movements

Abstract

Using both offline and online measures, the present study investigates attachment resolution in relative clauses (RC) in English natives (L1) and non-natives (L2). We test how RC resolution interacts with linguistic factors and participant-level individual differences. Previous L1 English studies have demonstrated a low attachment preference and also an "ambiguity advantage", suggesting that L1ers may not have as strong a low attachment preference as is sometimes claimed. We employ a similar design to examine this effect in L1 and L2 comprehension. Offline results indicate that both groups exhibit a low attachment preference, positively correlated with reading span scores and with proficiency in the L2 group. Online results also suggest a low attachment preference in both groups. However, our data show that individual differences influence online attachment resolution for both native and non-natives; higher lexical processing efficiency correlates with quicker resolution of linguistic conflicts. We argue that the current findings suggest that attachment resolution during L1 and L2 processing share the same processing mechanisms and are modulated by similar individual differences.

Keywords: Non-native sentence processing; attachment resolution; eye-tracking during reading; individual differences

2.1 Introduction

The similarities and differences between native (L1) and non-native (L2) sentence processing are widely debated (e.g., Clahsen & Felser, 2006, 2018; Cunnings, 2017; Grüter & Rohde, 2013; Hopp, 2014; Kaan, 2014; McDonald, 2006). Some theories propose that L2 processing employs different parsing mechanisms from L1 processing (e.g., Clahsen & Felser, 2006; Jiang, 2004). Alternatively, others have argued that L1 and L2 processing employ similar mechanisms and that L2ers construct syntactic analyses in the same way as L1ers (e.g., Cunnings, 2017; Hopp, 2014; Kaan, 2014). Under such accounts, observable differences between groups are argued to be non-qualitative. Additionally, the extent to which L1 and L2 parsing are modulated by individual differences is also debated (e.g., Hopp, 2014; McDonald, 2006; Tanner et al., 2014). For example, Hopp (2014) argued that lexical automaticity plays a role in native-like performance by L2ers, with more efficient lexical processing leading L2ers to behave more like L1ers during sentence processing.

Relevant to this debate are studies that have examined how L1ers and L2ers resolve ambiguous relative clauses (RCs), using offline and online tasks with sentences like (1) (e.g., Carreiras & Clifton, 1999; Cuetos & Mitchell, 1988; Dussias, 2003; Felser, Roberts et al., 2003; Maia et al., 2007; Pan et al., 2015; Rothman, 2010; Scheepers et al., 2011).

(1a) The brother of the man who bought himself some books lived here.
(1b) We knew the brother of the man who bought himself some books.

In (1) the RC is embedded to modify a complex noun phrase (NP) ("the brother of the man"), which either serves as the syntactic subject (1a) or object (1b) of the sentence. In both sentences, the RC ("who bought himself…") is ambiguous, as it can be interpreted as referring back to either the local NP "the man" (low attachment) or the non-local NP "the brother" (high

attachment). English readers tend to prefer low attachment (e.g., Cuetos & Mitchell, 1988; Gibson et al., 1996). However, low attachment is not a universal tendency and attachment preferences vary across languages (e.g., Cuetos & Mitchell, 1988; Hemforth et al., 2015; Zagar et al., 1997). Additionally, whether L1 and L2 readers show similar attachment preferences has been contested, especially in the case where the L1 and L2 display distinct attachment preferences. While some studies suggest L2ers do not show any clear attachment preferences (e.g., Felser, Roberts et al., 2003; Omaki, 2005), others have argued that L2ers can behave similarly to L1ers, even when their L1 displays an opposite tendency and especially if individual differences in L2 processing are considered (e.g., Dekydtspotter et al., 2008; Hopp, 2014).

Linguistic factors also influence attachment preferences (e.g., Desmet et al., 2002; Desmet et al., 2006; Fodor, 2002). Most importantly for present purposes, some L1 studies suggest that the syntactic position of the constituent that the RC modifies influences attachment preferences of sentences like (1) (e.g., Hemforth et al., 2000; Hemforth et al., 2015). Whether L2ers are sensitive to such subtle differences, is, however yet to be systematically explored.

Against the above background, we conducted a study on relative cause (RC) attachment in L1and L2 English speakers, testing the syntactic position of the RC. To tease apart different accounts of L1 and L2 processing, we aimed to test the extent to which L2ers can process and interpret RCs as in (1) in a nativelike way. To examine whether individual differences influence how nativelike L2 processing can become in this domain (Hopp, 2014), we also investigated how individual differences in working memory, lexical processing and for L2ers proficiency, influence L1 and L2 RC resolution offline and during processing.

2.1.1 Relative clauses in L1 processing

A large literature has contested how parsing preferences influence the processing of RCs (e.g., Cuetos &Mitchell, 1988, and much subsequent literature), and two competing principles are believed to influence low vs. high attachment. *Late Closure* (Frazier, 1979) or *Recency* (Gibson et al., 1996) predicts that new material is attached to the most recently processed constituent. On the other hand, *Predicate Proximity* (Gibson et al., 1996) holds that incoming material is preferably attached as close as possible to the head of a predicate. Therefore, readers who follow *Late Closure* or *Recency* favour low attachment of the RC to the local NP. Alternatively, readers who are guided by *Predicate Proximity* prefer high attachment to the non-local NP. As mentioned above, English L1ers generally have a low attachment preference for ambiguous RCs as in (1) in offline tasks (e.g., Cuetos & Mitchell, 1988). In tasks that measure online processing, researchers have manipulated agreement features to force either high or low attachment, as in (2a) and (2b) respectively. These studies have typically shown shorter reading times at the disambiguating word ("was/were" in (2a/b)) for RCs that attach low than those that attach high in English (e.g., Felser, Roberts et al., 2003; Hopp, 2014; Omaki, 2005). However, the low attachment preference of English is not universal, and a high attachment preference has been attested in offline and online tasks in L1 studies of various other languages (e.g., Bidaoui et al., 2016; Brysbaert & Mitchell, 1996; Carreiras et al., 2004; Chernova & Chernigovskaya, 2015; De Vincenzi & Job, 1993; Maia et al., 2004; Papadopoulou & Clahsen, 2003).

(2a) The man blamed the brothers of the boy who were smiling all the time.

(2b) The man blamed the brothers of the boy who was smiling all the time.

Determining the robustness of the low attachment preference in L1 English is complicated by studies testing three different types of RCs together. For example, Van Gompel et al. (2005) compared reading times for sentences that force high attachment (3a), low attachment (3b) and a globally ambiguous condition (3c).

(3a) The governor of the province that will be retiring after the troubles is very rich.
(3b) The province of the governor that will be retiring after the troubles is very rich.
(3c) The bodyguard of the governor that will be retiring after the troubles is very rich.

Van Gompel et al.'s results suggested that, instead of showing a low attachment reading time advantage, the English L1ers they tested demonstrated an "ambiguity advantage", with shorter reading times for ambiguous sentences like (3c) than sentences where the RC was forced to attach either high (3a) or low (3b). Furthermore, the forced low attachment sentences were not significantly different from the high attachment ones in terms of reading times. These findings were interpreted as supporting an "unrestricted race model" of ambiguity resolution (Traxler et al., 1998; Van Gompel et al., 2000). The faster reading times for (3c) were taken to indicate that readers variably attached the RC either high or low across trials. (3c) is thus easiest because whichever attachment was initially computed at "that" will turn out to be plausible once the verb "retiring" is encountered. For both (3a) and (3b), whichever attachment is initially computed at "that" will be incorrect 50% of the time at the verb, leading to (3a/b) having equally longer reading times compared to (3c). Therefore, these results suggest that L1ers of English may not have as strong a low attachment preference as claimed, variably attaching instead to either available site. The fact that the ambiguous condition had faster reading times compared to the disambiguated conditions was also taken as evidence against the idea that the two possible interpretations in the ambiguous condition competed for

activation in parallel as otherwise, Van Gompel et al. (2005) reasoned, the ambiguous condition should have had longer reading times due to competition.

A number of studies have examined how various linguistic factors influence attachment preferences, including whether attachment choices in L1ers are modulated by whether the complex NP the RC modifies is in subject position as in (1a) or object position as in (1b) (e.g., Hemforth et al., 2015; Kim & Christianson, 2017). Hemforth et al. (2015) found a stronger high attachment preference for NPs in object position in German and Spanish. Alternatively, the syntactic position did not influence attachment preferences in English and French, with English showing a general low attachment preference. However, this study tested offline preferences only, and did not test online processing.

Researchers have also examined how participant-level individual differences influence attachment resolution. This includes work examining the role of working memory, as measured by reading span tasks, although results have been mixed (e.g., Kim & Christianson, 2013; Kim & Christianson, 2017; Payne et al., 2014; Swets et al., 2007). Some studies reported that the low attachment preference in English increases as a function of higher reading span scores in either offline (Kim & Christianson, 2013; Omaki, 2005; Swets et al., 2007) or online measures (Kim & Christianson, 2013; Payne et al., 2014), suggesting that high-span individuals prefer attaching low more often than low-span individuals. However, results have been interpreted differently across these studies. Some researchers (e.g., Omaki, 2005) interpret these results from a resource limitation perspective whereas others (e.g., Kim & Christianson, 2013; Swets et al., 2007) see this as evidence for a chunking based account in which high-span and low-span readers chunk the relative clause differently, leading to different NPs being more salient.

Other studies observed the reverse pattern online (Felser, Marinis et al., 2003; Traxler, 2007) or no such effects offline (Felser, Marinis et al., 2003). For example, Traxler (2007) found that high-span individuals preferred high attachment while low-span readers preferred

low attachment in an eye-tracking study. Traxler claimed that high-span individuals are more sensitive to discourse-salience and are drawn to the head NP rather than the subordinate NP, while low-span individuals may simply rely on linear distance.

Finally, Kim & Christianson (2017) and Payne et al. (2014) found that L1 readers with higher reading spans have greater difficulty processing globally ambiguous RCs than readers with lower spans in an online self-paced reading task. Kim & Christianson (2017) interpreted this as indicating that high-span readers can hold the two potential interpretations in mind at the same time, which leads to competition between high and low attachment that is not found in lower span readers.

2.1.2 Relative clauses in L2 processing

The question of whether L2 learners show the same parsing preferences as L1 speakers has been widely examined. Some L2 studies found non-nativelike attachment preferences in L2ers (e.g., Dinçtopal-Deniz, 2010; Felser, Roberts et al., 2003; Fernandez, 2003; Omaki, 2005; Papadopoulou & Clahsen, 2003). For example, Felser, Roberts et al. (2003) tested sentences like (2) in an offline task and online self-paced reading task. They found that in both measures, the L2ers did not exhibit any structural preferences. These findings led to the conclusion that L2 processing is different from L1 processing when processing is guided by structural information.

Conversely, other studies have lent support to the claim that L2 parsing is guided by structure-based information in either offline or online tasks, or both (e.g., Bidaoui et al., 2016; Hopp, 2014; Witzel et al., 2012). For example, Hopp (2014) investigated online attachment preferences of advanced German speakers of L2 English using sentences like (2) during eye-tracking and found that the L2ers preferred to attach low, which was in line with the L1 results.

Less is known about how RC position modulates L2ers' attachment choices, and we are aware of only one study that has examined this issue. Kim and Christianson (2017) conducted a study with Korean learners of English using sentences like (1a/b) and found the L2ers' offline attachment preferences were not influenced by RC position. Syntactic position also was not found to influence online processing in self-paced reading. However, Kim and Christianson tested globally ambiguous sentences only, and did not test sentences disambiguated to high or low attachment.

Individual differences have also been examined in L2 attachment resolution. For example, Dekydtspotter et al. (2008) reported more nativelike attachment preferences as proficiency increased in an online study of L2 French. Hopp (2014) investigated the role of working memory, as measured by a reading span task, and lexical efficiency, tested using a lexical decision task, in both offline and online attachment resolution. Offline, increased reading span scores correlated with an increased low attachment preference, replicating some previous L1 English findings (e.g., Swets et al., 2007). The results from an eye-tracking experiment suggested that even though the L2 group as a whole did not show any clear attachment preferences for sentences like (4), where the ambiguity is not resolved immediately, a low attachment preference emerged when individual differences in lexical automaticity were considered, with the L2 individuals with high lexical automaticity preferring low attachment. Reading span scores, however, were not significantly correlated with online processing.

(4) The doctor examined the mother of the boy who had badly injured herself with the knife.

Finally, Kim & Christianson (2017) tested individual differences in English RCs during self-paced reading in Korean learners of English. They found that high span L2ers had longer reading times for sentences containing globally ambiguous RCs relative to those with low span

scores. As similar patterns were observed when these L2ers processed their L1, Korean, Kim and Christianson concluded that high-span readers are more likely to consider both interpretations of ambiguous RCs in parallel than low-span readers, in both their L1 and L2.

2.1.3 The present study

The mixed findings in previous L2 studies suggest that L2 attachment resolution is influenced by a wide variety of factors. However, the role that the syntactic position of the RC may play in influencing L2 attachment resolution has not been systematically examined. Furthermore, while existing studies have examined whether or not L1ers consider different potential attachment choices in parallel (e.g., Traxler et al., 1998), the extent to which L2 comprehension may involve parallel competition between multiple possible analyses has not been examined extensively. Although Kim and Christianson (2017) examined this issue in globally ambiguous RCs, to our knowledge, no existing published study has tested this issue by directly comparing the processing of ambiguous RCs to those disambiguated low and high.

As such, we employed a similar design to Van Gompel et al.'s (2005) L1 study. We tested RCs in subject and object position, as in (5) and (6) respectively. We tested globally ambiguous sentences, as in (5a/6a), and compared them to sentences that forced low attachment, as in (5b/6b), and sentences that forced high attachment, as in (5c/6c). Given that some previous L2 studies have reported no differences between high and low attachment conditions during processing (e.g., Felser, Roberts et al., 2003; Papadopoulou & Clahsen, 2003), our approach of including a globally ambiguous condition provides an important control for comparing low and high attachment sentences.

(5a) Subject-Modifying RC, Ambiguous

The brother of the man who accidentally hurt <u>himself</u> <u>yesterday afternoon</u> lived in town.

(5b) Subject-Modifying RC, Low Attachment

The sister of the man who accidentally hurt himself yesterday afternoon lived in town.

(5c) Subject-Modifying RC, High Attachment

The brother of the woman who accidentally hurt himself yesterday afternoon lived in town.

(6a) Object-Modifying RC, Ambiguous

We saw the brother of the man who accidentally hurt himself yesterday afternoon.

(6b) Object-Modifying RC, Low Attachment

We saw the sister of the man who accidentally hurt himself yesterday afternoon.

(6c) Object-Modifying RC, High Attachment

We saw the brother of the woman who accidentally hurt himself yesterday afternoon.

… Luckily it wasn't serious in the end.

If L1ers demonstrate a low attachment preference, as predicted by the garden path model (Frazier,1987), reading times should be longer in (5c/6c), where low attachment isn't possible, compared to conditions (5a/5b/6a/6b), where it is. Alternatively, if L1ers randomly attach either low or high, as predicted by the unrestricted race model (e.g., Van Gompel et al., 2000), reading times should be longer in disambiguating conditions (5b/5c/6b/6c) compared to ambiguous (5a/6a). Such findings would support the "ambiguity advantage". If L1ers consider both low and high attachment in parallel, longer reading times should be expected in (5a/6a) than (5b/5c/6b/6c), due to the competition between the two possible interpretations (e.g., MacDonald, 1994; Tabor & Tanenhaus, 1999).

If L1 and L2 processing are qualitatively similar, the L2ers should behave like the L1ers. On the contrary, if they are different, L2ers should show different reading time patterns to the L1ers. For example, if L2ers do not show any clear preferences between low and high

attachment (e.g., Felser, Roberts et al., 2003), then (5b/5c) and (6b/6c) should not differ. If this effect results from L2ers exhibiting variable attachment, then (5b/5c/6b/6c) should have longer reading times than ambiguous (5a/6a). Alternatively, if L2ers do not clearly resolve the RC at all during online processing, they may show no significant differences between any conditions.

Additionally, our study addresses another important gap in the literature by examining individual differences in this domain. To our knowledge, all other existing studies examining individual differences have tested either L1 speakers or L2 speakers but not both within the same study. If L1 and L2 processing are similar, we would expect that individual differences should impact L1 and L2 processing similarly. As such, we examined how individual differences in reading span and lexical automaticity affect both offline and online ambiguity resolution in L1ers and L2ers. Based on prior findings, we predicted both L1/L2 individuals with higher reading spans should prefer low attachment more than those with low span scores for globally ambiguous sentences in offline comprehension (e.g., Hopp, 2014; Swets et al., 2007). For online processing, we expected L2ers with higher levels of lexical automaticity to behave more nativelike, and should show a stronger low attachment preference (Hopp, 2014). If L1 and L2 processing are qualitatively similar, lexical automaticity may influence L1 and L2 processing in a similar way. Finally, we also expected L2ers with higher proficiency to exhibit a more native-like pattern during attachment resolution (Dekydtspotter et al., 2008).

2.2 Method

Participants

The experiment was conducted with 66 English L1 speakers and 66 English L2 speakers with a first language that is reported to have a high attachment preference. This included Spanish, Italian, German, Dutch, French, Russian, Portuguese, Greek and Arabic. All participants were recruited from the University of Reading and surrounding areas. L1 participants were

university students who participated in the study for course credit or a small payment. The L2 speakers were either students or were working in the local community at the time of testing. Their English proficiency was measured by a short version of the Oxford Placement Test. Their proficiency scores ranged from 30-60 out of 60 (mean = 48, SD = 0.93), representing intermediate to advanced level. All participants had normal or corrected to normal vision.

Materials

We combined an offline judgement and an online reading task to test participant's attachment preferences. The offline task consisted of 20 experimental items as in (7) and 40 fillers, which were pseudo-randomised in a latin-square design. We tested participants' offline attachment preferences for ambiguous RCs like (7a/b) in cases where the RC is in subject or object position.

(7a) Subject-Modifying RC

The brother of the man who often bought himself some books got married yesterday.

(7b) Object-Modifying RC

We met the brother of the man who often bought himself some books.

Question: Who bought himself some books?

Answer options: The brother/The man

Participants only saw one condition of each pair and, therefore, read 10 sentences like (7a) and 10 sentences like (7b). To indicate their attachment preferences, participants needed to answer comprehension questions such as "Who bought some books?" by choosing one of the two individuals shown in those sentences (i.e., either "the brother" or "the man"). The order

of the options was counterbalanced. The choice of high attachment was coded as value 0 and low attachment was coded as 1. As such, a value towards 1 indicates a low attachment preference.

In the online reading task, we monitored participant's eye-movements as they read a series of texts. The materials for this task consisted of 36 experimental items like (5/6) and 80 fillers, randomised in a latin-square design. The experimental items contained a critical first sentence and a wrap-up sentence. The experimental items manipulated the position of the RC such that the RC was either in subject position (5a-c) or object position (6a-c). Half of the items contained a masculine reflexive ("himself") and half a feminine reflexive ("herself"). The temporarily ambiguous RCs were disambiguated at the reflexive via gender match between the reflexive and the local or non-local NP. (5a/6a) are globally ambiguous as the reflexive matches the gender of both NPs. (5b/6b) are forced to attach low as the reflexive only matches the gender of the local NP whereas (5c/6c) are forced to attach high due to the gender match between the reflexive and the non-local NP. A full list of experimental materials for the offline and online tasks can be found in the supplementary materials.

We ran two additional tasks with both the L1 and L2 participants to investigate potential individual differences. To tap into working memory, a reading span task adapted from Daneman & Carpenter (1980) was administered to both groups. The participants read aloud sets of sentences presented one by one on a computer screen, as in (8) and (9). Half the sentences were grammatical and half ungrammatical. After each sentence participants judged whether it was grammatical or ungrammatical by pressing "1" or "2" on the keyboard. In the meantime, they had to memorise the final word of each sentence, underlined in (8) and (9). After the last sentence of a set, "RECALL" appeared onscreen and participants had to recall the to-be-remembered words, which were recorded by the experimenter. The set size increased from 2 sentences to 5 sentences as the experiment progressed. There were three sets at each set

size, and participants completed all sets. The reading span score was calculated by dividing the total number of correctly recalled words by the total number of words that needed to be recalled.

(8) The young boy listened to music in his bedroom for hours. (Grammatical)

(9) *The old man picked up the phone his to spoke and daughter. (Ungrammatical)

A second individual differences task measured levels of lexical automaticity, using a lexical decision task adapted from Hopp (2014). There were 80 words in total, half of which were real English words and the other half were non-words following English phonotactics rules. The order in which the words were presented was randomised. Participants needed to decide as quickly as possible whether the word they saw was a real English word or not. They were instructed to rest two fingers from their preferred hand on the "1" and "2" keys, pressing "1" for real English words and "2" for non-words. Reaction times and accuracy were recorded. Following Hopp (2014), lexical automaticity was calculated by dividing the standard deviation of the reaction times to the real English words judged correctly by their average reaction times. For the L2 learners, a vocabulary screening task was administered at the end of the second session to test if they were familiar with the meaning (and gender) of the critical vocabulary (i.e. the noun phrases in the RCs).

Procedure

The study was conducted in two sessions at least 3 days apart. In the first session, participants completed the background questionnaire which provided information on language experience. This was followed by the main reading experiment where eye-movements were monitored by an SR Research Eyelink 1000 eye-tracker. Although viewing was binocular, the eye-movement

record was recorded for the right eye only. The experiment began with a calibration procedure on a 9-point grid, and calibration was adjusted as needed between trials. The stimuli were presented onscreen in black letters. Before each sentence, a fixation marker appeared onscreen above the first word to be displayed. Upon fixating the marker, the sentence appeared. Participants were told to read as naturally as possible and to make sure they understood the sentences. All sentences were followed by a yes/no comprehension question that was answered by a push-button response. Comprehension questions did not probe attachment of the RC in the critical sentences. Participants familiarised themselves with the procedure by first completing some practice trials before the main experiment. In the main experiment, experimental and filler items were pseudo-randomised in a latin-square design across six presentation lists that were completed by the same number of participants.

In the second session, participants first completed the reading span task and then the lexical decision task. Following that, they completed the offline attachment preference task. In addition, L2 speakers completed the proficiency test.

Data analysis

For the eye-tracking data, reading times were calculated at two regions of text. The reflexive region consisted of the critical reflexive, while the spillover region contained the rest of the clause (as exemplified in (5) & (6) using underlining). We calculated three reading time measures at each region. First-pass times summed the duration of fixations in a region entered from the left up until it was exited for the first time. Regression path times summed the duration of fixations starting when a region was first entered and up until but not including the first fixation in a region to the right. Total reading times refer to the duration of all fixations in a region regardless of when they occurred.

Trials in which a region was not fixated were treated as missing data. Fixations less than 80ms were combined with any neighbouring fixations if they were within one character of each other. All other fixations less than 80ms, as well as all those over 800ms, were removed. Due to a typographic error, the responses from one item in condition (6c) (object modifying RC, high attachment) were removed before analysis. Trials with excessive track loss were also removed, accounting for less than 1% of the data. Based on the vocabulary screening task, trials with critical vocabulary that the L2ers did not know were removed before analysis, accounting for less than 1% of the data.

Analysis was conducted using mixed-effects models (Baayen et al., 2008). For the offline task, a generalised mixed model was used containing sum coded (-1/1) fixed effects of Group (L1/L2), Position (subject RC/object RC) and their interaction. For the eye-movement data, reading times were log-transformed to minimise skew (see Vasishth & Nicenboim, 2016). Mixed models included sum coded fixed effects of Group (L1/L2) and Position (subject RC/object RC). The three-condition ambiguity manipulation involved two treatment-coded contrasts. One contrast (low attachment (LA)) compared the low attachment condition to the globally ambiguous condition, while the second contrast (high attachment (HA)) compared the high attachment condition to the globally ambiguous condition. In the case of any interactions between Position and the LA and/or the HA contrast, follow-up comparisons were conducted at the two levels of Position. For interactions with Group, additional analyses were conducted for each group separately.

All models were fit using the maximal random effects structure that converged (Barr, 2013; Barr et al., 2013). When the maximal model failed to converge, the random correlations were removed first. If the model still failed to converge, the random effect with the least variance was iteratively removed until the model converged.

We first conducted a main analysis as above to test for between-groups effects. We then conducted a series of additional analyses with three individual differences measures (working memory, lexical automaticity, and L2 proficiency) by adding each predictor separately into the maximal model, using the same method to achieve convergence. We analysed each individual differences measure separately to avoid issues related to multicollinearity. The three individual differences measures were included as centred, continuous predictors in each model, along with relevant interactions. The data and analysis code for our experiments is available at the Open Science Framework (OSF) website (https://osf.io/tvakf/).

2.3 Results

Individual differences measures and offline results

For reading span scores, the L1 (mean = 0.71, SD = 0.107) and L2 (mean = 0.72, SD = 0.117) groups did not differ significantly (estimate = 0.001, SE = 0.002, t = 0.801, p = .423), despite the fact that the task was presented in English. L1ers did however have significantly faster lexical automaticity than L2ers (L1 mean = 0.25, SD = 0.116; L2 mean = 0.34, SD = 0.155; estimate = 0.04, SE = 0.002, t = 16.41, p < .001).

The results from the offline task are shown in Table 1. The proportions here are descriptively all above 0.70 indicating a low attachment preference. Analysis revealed a significant effect of Position (estimate = 0.329, SE = 0.065, z = 4.99, p < .001), with the low attachment preference being stronger in object-modifying RCs than subject-modifying RCs in both groups. Neither the effect of group or interaction were significant (both z < 0.18, both p > 0.23).

	Native Speakers	Non-Native Speakers
Subject Modifying RC	0.73 (0.01)	0.77 (0.01)
Object Modifying RC	0.82 (0.01)	0.82 (0.01)

Table 1. *Low attachment preferences in the offline task (standard errors in parentheses), Study 1*

Eye-tracking results

Accuracy to the comprehension questions was 95% for both groups (all participants scored above 82%), indicating all participants paid attention. A summary of the reading time data and statistical analysis is shown in Tables 2 and 3 respectively. For brevity, main effects of Group were found in each measure at each region, indicating slower reading times for the L2ers. We also do not discuss main effects of Position, or Position by Group interactions, below, as these are difficult to interpret on their own, but further interactions between Position, Group and LA or HA are informative about attachment preferences.

At the reflexive region, in first pass reading times we observed a significant interaction between Group and HA. Separate analyses on each group revealed that the L1 group showed significantly longer reading times for the high attachment than globally ambiguous condition (estimate = 0.061, SE = 0.021, t = 2.92, p = .005), whereas the L2 group showed no significant differences (estimate = -0.007, SE = 0.02, t = -0.35, p = .728).

Moving onto regression path time, the HA effect was significant with longer reading times for high attachment RCs relative to ambiguous RCs. There were also numerical trends for an LA effect that differed across groups, with L1ers, but not L2ers, tending to show longer reading times in LA than ambiguous conditions. However, neither the LA effect nor the Group by LA interaction was significant.

	First Pass Time		Regression Path Time		Total Viewing Time	
	Native Speakers	Non-Native Speakers	Native Speakers	Non-Native Speakers	Native Speakers	Non-Native Speakers
Reflexive Region						
Subject Modifying RC, Globally Ambiguous	226 (5)	282 (7)	266 (10)	366 (18)	389 (13)	481 (15)
Subject Modifying RC, Low Attachment	241 (5)	275 (6)	311 (16)	351 (19)	397 (15)	470 (15)
Subject Modifying RC, High Attachment	251 (7)	275 (7)	326 (17)	387 (17)	501 (19)	578 (20)
Object Modifying RC, Globally Ambiguous	235 (6)	271 (6)	327 (20)	339 (13)	380 (12)	491 (17)
Object Modifying RC, Low Attachment	239 (5)	286 (7)	382 (45)	352 (13)	406 (15)	491 (17)
Object Modifying RC, High Attachment	251 (7)	276 (6)	379 (29)	355 (13)	523 (19)	543 (20)
Spillover Region						
Subject Modifying RC, Globally Ambiguous	413 (11)	485 (12)	496 (25)	555 (16)	757 (30)	849 (28)
Subject Modifying RC, Low Attachment	397 (11)	488 (13)	522 (27)	587 (22)	735 (27)	818 (25)
Subject Modifying RC, High Attachment	398 (11)	488 (12)	598 (35)	614 (25)	803 (28)	924 (27)
Object Modifying RC, Globally Ambiguous	428 (14)	523 (15)	990 (55)	1292 (70)	654 (23)	809 (25)
Object Modifying RC, Low Attachment	412 (14)	507 (14)	997 (67)	1226 (63)	612 (23)	765 (25)
Object Modifying RC, High Attachment	418 (14)	504 (14)	1628 (103)	1532 (88)	753 (29)	815 (24)

Table 2. *Summary of reading times in milliseconds (standard errors in parentheses), Study 1*

	Reflexive Region			*Spillover Region*		
	Estimate (SE)	*t*	*p*	*Estimate (SE)*	*t*	*p*
First Pass Time						
Group	**0.085 (0.017)**	**5.19**	**< .001**	**0.094 (0.023)**	**4.08**	**< .001**
Position	0.001 (0.009)	0.18	.856	0.009 (0.014)	0.68	.494
LA	0.023 (0.013)	1.73	.083	-0.026 (0.015)	-1.68	.092
HA	0.026 (0.013)	1.92	.055	-0.017 (0.016)	-1.08	.277
Group * Position	-0.011 (0.010)	-1.06	.290	0.014 (0.014)	0.99	.321
Group * LA	-0.017 (0.013)	-1.26	.206	0.024 (0.016)	1.48	.139
Group * HA	**-0.034 (0.014)**	**-2.45**	**.016**	0.017 (0.016)	1.04	.299
Position * LA	0.001 (0.013)	0.03	.970	-0.004 (0.015)	-0.25	.798
Position * HA	0.001 (0.013)	0.04	.967	0.002 (0.016)	0.14	.883
Group * Position * LA	0.017 (0.013)	1.25	.209	-0.003 (0.015)	0.20	.835
Group * Position * HA	0.013 (0.013)	0.98	.324	-0.012 (0.016)	0.75	.447
Regression Path Time						
Group	**0.092 (0.025)**	**4.21**	**< .001**	**0.108 (0.028)**	**3.79**	**< .001**
Position	0.016 (0.015)	1.10	.271	**0.277 (0.026)**	**10.57**	**< .001**
LA	0.033 (0.019)	1.66	.097	-0.001 (0.022)	-0.03	.975
HA	**0.071 (0.020)**	**3.42**	**<.001**	**0.159 (0.023)**	**6.90**	**< .001**
Group * Position	**-0.028 (0.013)**	**-2.09**	**.035**	0.021 (0.025)	0.84	.402
Group * LA	-0.039 (0.019)	-1.96	.0504	0.003 (0.022)	0.17	.868
Group * HA	-0.028 (0.020)	-1.34	.179	**-0.064 (0.023)**	**-2.82**	**.004**
Position * LA	0.002 (0.019)	0.11	.907	-0.023 (0.022)	-1.04	.299
Position * HA	-0.012 (0.019)	-0.65	.513	**0.074 (0.024)**	**3.04**	**.003**
Group * Position * LA	0.024 (0.020)	1.20	.229	0.004 (0.022)	0.18	.855
Group * Position * HA	0.005 (0.022)	0.23	.816	-0.036 (0.023)	-1.61	.108
Total Viewing Time						
Group	**0.119 (0.027)**	**4.30**	**< .001**	**0.102 (0.029)**	**3.50**	**< .001**
Position	-0.002 (0.013)	-1.15	.881	**-0.041 (0.014)**	**-2.84**	**.004**
LA	0.007 (0.019)	0.39	.694	**-0.039 (0.017)**	**-2.23**	**.025**
HA	**0.167 (0.022)**	**7.61**	**< .001**	**0.084 (0.020)**	**4.10**	**< .001**
Group * Position	-0.001 (0.013)	-0.08	.935	0.026 (0.014)	1.83	.066
Group * LA	-0.025 (0.019)	-1.29	.197	0.007 (0.017)	0.40	.683
Group * HA	**-0.052 (0.022)**	**-2.37**	**.019**	-0.015 (0.019)	-0.82	.408
Position * LA	0.019 (0.019)	0.99	.845	-0.020 (0.017)	-1.17	.240
Position * HA	0.004 (0.022)	0.20	.828	0.001 (0.017)	0.05	.958
Group * Position * LA	-0.004 (0.019)	-0.22	.166	0.004 (0.017)	0.22	.818
Group * Position * HA	-0.028 (0.021)	-1.40	.324	**-0.036 (0.017)**	**-2.05**	**.040**

Note: LA = Low Attachment, HA = High Attachment

Table 3. *Summary of the statistical analysis, Study 1*

In terms of total reading times, the HA effect was statistically significant with longer reading times for high attachment RCs than ambiguous ones. There was also a significant Group by HA interaction. Follow-up analyses revealed that the HA effect was present in both groups but with a larger effect in the L1 group (estimate = 0.221, SE = 0.034, t = 6.48, p < .001) compared to the L2 group (estimate = 0.12, SE = 0.03, t = 4.27, p < .001). This is illustrated in Figure 1.

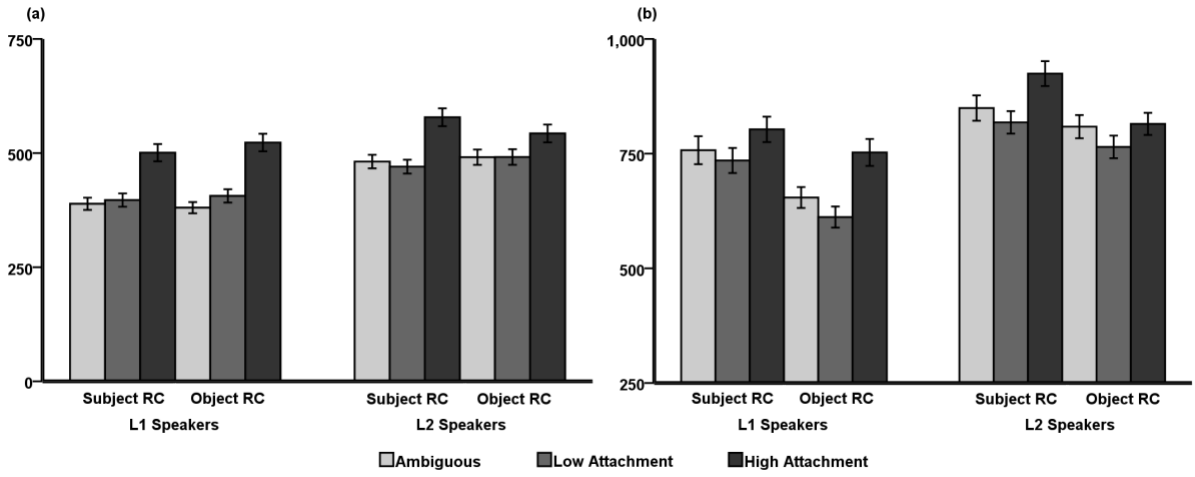

Figure 1. *Total viewing times at the reflexive region (a) and spillover region (b), Study 1*

At the spillover region, no significant effects were found for first pass reading time. For regression path time, there were three significant effects: the HA effect, the Group by HA interaction and the Position by HA interaction. To test the HA effect that was modulated by an interaction with Group, we conducted separate analyses for each group. These results found that both L1ers and L2ers had significantly longer reading times for high attachment RCs than the ambiguous baseline, but the effect was larger in the L1 group (estimate = 0.21, SE = 0.05, t = 4.63, p < .001) than the L2 group (estimate = 0.09, SE = 0.04, t = 2.18, p = .036). To examine the Position by HA interaction we tested HA effects in each position, collapsed across groups. These indicated that the HA effect was significant in both positions but with a smaller

effect for subject position (estimate = 0.086, SE = 0.03, t = 3.33, $p < .001$) than object position (estimate = 0.231, SE = 0.045, t = 5.15, $p < .001$).

With respect to total reading times, a significant LA main effect was observed across the groups. Here, however, shorter reading times were found for low attachment RCs compared to globally ambiguous ones. The main effect of HA was also significant, with longer reading times for the high attachment condition than the baseline condition. This was, however, modulated by a significant three-way interaction between Group, Position and HA, which is illustrated in Figure 1. The follow-up analyses indicated that in subject position, The HA effect was significant in the L2 group (estimate = 0.105, SE = 0.038, t = 2.79, $p = .009$), while the same numerical trend was not significant in the L1 group (estimate = 0.065, SE = 0.037, t = 1.73, $p = .084$). In object position, the HA effect was significant in the L1 group (estimate = 0.138, SE = 0.039, t = 3.53, $p < .001$), but not in the L2 group (estimate = 0.04, SE = 0.033, t = 1.19, $p = .234$).

Individual differences analysis: Offline task

Reading span scores, lexical automaticity and proficiency were included in separate models to examine their correlation with attachment preferences. In the reading span model, the main effect of Position was still significant (estimate = 0.337, SE = 0.071, z = 4.73, $p < .001$), as was the effect of reading span (estimate = 6.600, SE =1.553, z = 4.25, $p < .001$), in the absence of any further significant effects or interactions (all $z < 0.12$, all $p > .24$). The results here indicated that reading span scores were positively correlated with the low attachment preference in both groups (see Figure 2).

In the lexical automaticity model, apart from the significant effect of Position (estimate = 0.355, SE = 0.061, z = 5.76, $p < .001$), there was no significant effects or interactions (all $z < 0.51$, all $p > 0.14$). In the proficiency model, the effect of Position (estimate = 0.273, SE =

0.082, z = 3.31, $p < .001$) and the effect of proficiency (estimate = 0.085, SE = 0.028, z = 2.96, $p < .01$) were significant, but the Position by proficiency interaction was not significant (estimate = 0.016, SE = 0.011, z = 1.54, $p = .123$). The proficiency model suggested language proficiency was positively correlated with the low attachment preference in the L2 group.

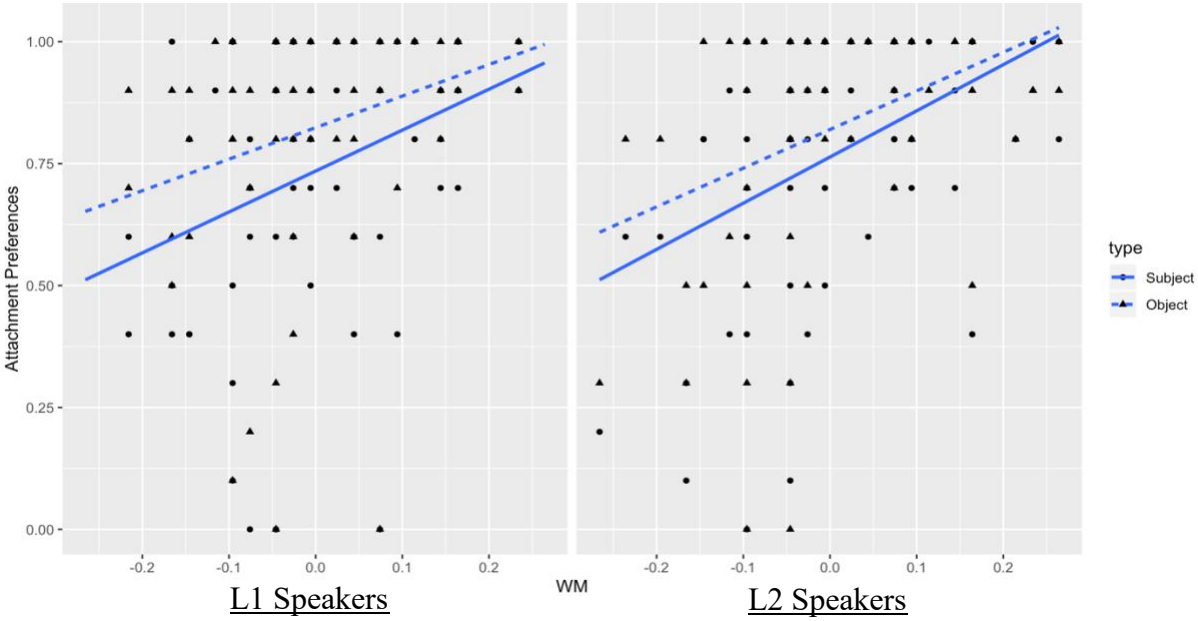

Figure 2. *Interaction between reading span and attachment preferences for each position and group, Study 1*

Individual differences analysis: Reading times

Lexical automaticity

At the reflexive region, there were numerical trends in first pass reading time for readers with high lexical automaticity to have shorter reading times, but the main effect of lexical automaticity was not significant (estimate = 0.237, SE = 0.121, t = 1.94, $p = .052$). We also observed numerical trends that the HA effect in both positions found in the main analysis in the L1 group was largely driven by those with high lexical efficiency, while the L2 group did not differ. However, the interaction between Group, Lexical Automaticity, Position and HA was not significant (estimate = -0.200, SE =0.103, t = -1.92, $p = .054$).

A significant three-way interaction between Lexical Automaticity, Position and HA was observed for total reading times at the reflexive region (estimate = 0.314, SE = 0.153, t = 2.04, p = .041). This is illustrated in Figure 3. We can see that in subject position, individuals with high and with low levels of lexical automaticity behaved quite similarly, whereas in object position, the HA effect seemed larger for individuals with lower levels of lexical automaticity compared to the highly automatized participants.

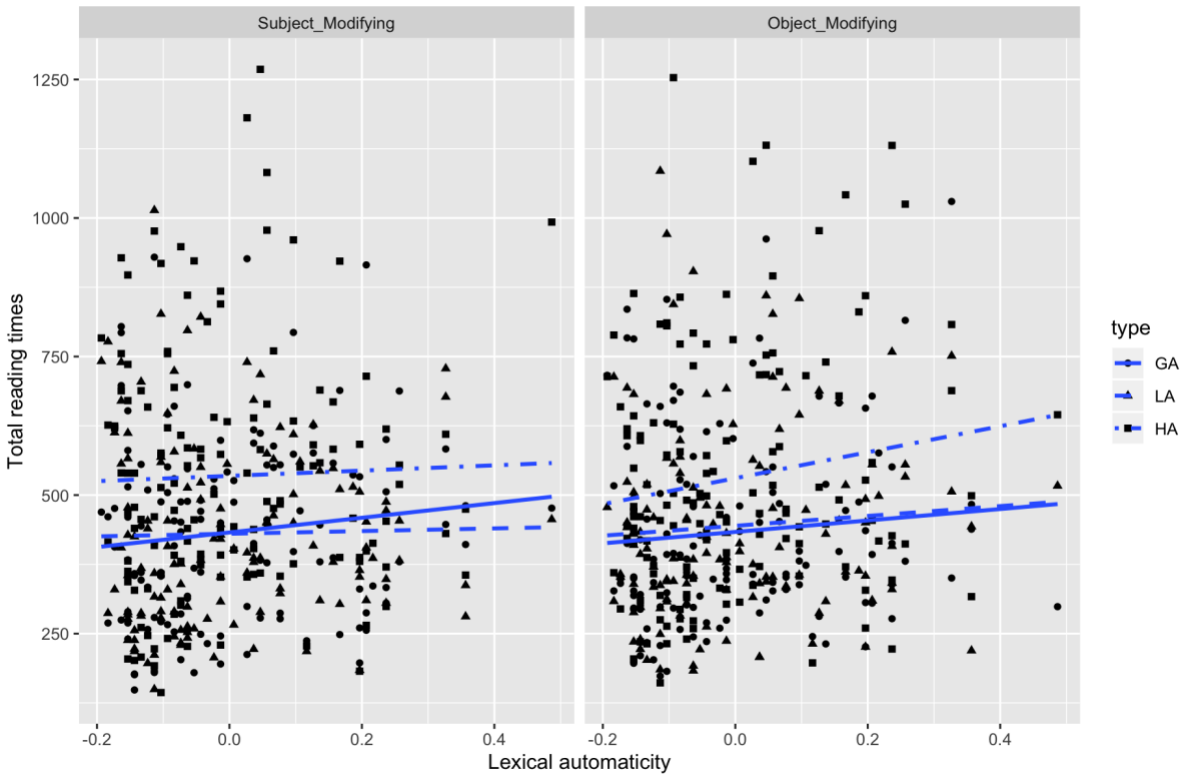

Figure 3. *Interaction between lexical automaticity and total reading times for each position across groups, Study 1*

At the spillover region, we only found a numerical trend for first pass reading times to be longer in less automatized individuals. However, the main effect of Lexical Automaticity was not significant (estimate = 0.308, SE = 0.175, t = 1.76, p = .079).

Reading span and proficiency

Regarding individual differences in reading span, we did not observe any statistically significant main effects or interactions of theoretical interest (all $t < 1.83$, all $p > .07$). In terms of proficiency, no effects of theoretical interest were significant (all $t < 1.65$, all $p > .101$).

2.4 General Discussion

This study aimed to investigate parsing strategies in L1 and L2 RC attachment, more specifically whether attachment preferences were influenced by RC position and/or individual differences. The results showed that both L1 and L2 groups demonstrated a clear low attachment preference, modulated by the syntactic position of the RC in both offline and online tasks. We also observed some interactions between individual differences and offline/online attachment preferences. The implications of our results are discussed in more detail below.

Attachment preferences and sentence processing in native and non-native speakers

In the offline task, both L1 and L2 groups preferred low attachment over high attachment for both types of RCs. The L1 result is generally in line with the L1 literature. The L2 results, however, differ from some previous findings where L2ers from an L1 that prefers high attachment demonstrated null preferences (e.g., Felser, Roberts et al., 2003; Kim & Christianson, 2017; Omaki, 2005; Papadopoulou & Clahsen, 2003) in offline measures. Even though some previous studies (e.g., Felser, Roberts et al., 2003; Omaki, 2005) tested L2ers in an English-speaking country, their participants might differ from those included in this study regarding the amount of exposure, the degree of and opportunity for usage of English, or English proficiency. Regardless, our results indicated that L2ers did not significantly differ from the L1ers in their offline attachment preferences.

For online processing, both groups exhibited a low attachment preference even though they differed during certain stages of processing. An absent LA effect and a clear HA effect were observed in both groups across several measures, especially total reading times, suggesting a low attachment preference in L1ers and L2ers (see Figure 1). However, there is some evidence showing that the HA effect was delayed and smaller in the L2ers, even though the effect was in the same direction for both groups. During first pass reading time, only L1ers showed the HA effect, suggesting that they preferred low attachment whereas L2ers did not show any preferences in this measure. Both groups showed the HA effect in total reading time at the reflexive region, regression path time and total reading time at the spillover region, although the effect was larger for the L1ers in some measures. We argue that this suggests slower but not qualitatively different processing for L2ers, who showed the same HA effect as L1ers, albeit delayed and numerically smaller.

The online results are generally in line with accounts which predict a low attachment preference in L1 English (e.g., Felser, Roberts et al., 2003; Hopp, 2014). Our data do not fully replicate the "ambiguity advantage" observed in previous studies (e.g., Traxler et al., 1998; Van Gompel et al., 2005). In regression path times at the reflexive, there were numerical trends for the ambiguous condition to have shorter reading times than both disambiguated conditions, though the LA comparisons were not fully reliable here. Across the majority of measures, although longer reading times were observed for high attachment conditions, the globally ambiguous conditions did not have significantly longer reading times than the low attachment conditions. This might be partially supportive of the "ambiguity advantage", in that this pattern of results does not suggest competition in the ambiguous conditions. However, we did find some evidence of competition in one measure, namely total viewing times at the spillover region, where the ambiguous condition had longer reading times than the low attachment condition. This might indicate an initial preference for low attachment with delayed

competition between the two attachment sites at the spillover region. We do not draw any strong conclusions about the ambiguity advantage here, but note that we did not find any L1/L2 differences in this regard. Thus, our clearest finding across measures was the low attachment preference for both L1 and L2 readers.

The L2 online results corroborate findings that L2ers can show online structural preferences (e.g., Bidaoui et al., 2016; Hopp, 2014). However, in Hopp (2014), the L2 group (as a whole) only showed the low attachment preference for sentences that are immediately disambiguated by a copula but not for sentences that are disambiguated later, as used in our study. Our study thus also indicates L2ers can exhibit online attachment preferences in sentences with later disambiguation. Our results stand in stark contrast with other studies showing that L2ers failed to show structural preferences (e.g., Felser, Roberts et al., 2003; Omaki, 2005; Papadopoulou & Clahsen, 2003), which could be due to methodological differences. Most L2 studies that failed to attest attachment preferences employed self-paced reading, which does not allow information from rereading and later processing stages to be captured. Also, with self-paced reading, individuals with poorer memory might have impoverished representations of the previous text when they reach the critical region, while eye-tracking allows for more naturalistic reading. Future research is required here to examine how these methodological issues may influence attachment resolution in L1 and L2 processing.

Recall that syntactic position interacted with offline attachment resolution as both groups preferred attaching low more often when the RC modified the object of the main verb compared to when it modified the subject. This suggests that the low attachment preference was stronger for object-modifying RCs in L1 and L2ers alike. It could be the case that since the high attachment site is the subject and always first-mentioned in the subject-modifying RC, it receives more discourse salience than the low attachment site, leading to a slightly weaker low attachment bias in the subject-modifying RC conditions. The offline findings are

inconsistent with Hemforth et al. (2015) and Kim & Christianson (2017), who did not find effects of position in English RCs. However, both these studies had smaller samples than the current study (48 and 34 participants respectively, compared to 132 in our experiment).

The position effect was also attested during online processing. In regression path times at the spillover region, both groups showed larger HA effects in object-modifying RCs compared to subject-modifying RCs, suggesting a stronger low attachment preference for object-modifying RCs in both groups. We propose that high attachment in object-modifying RCs was more demanding due to the attenuated prominence of the high attachment site in these sentences. This interplay of syntactic position and attachment choices across the groups suggests that discourse-level information was processed similarly in L1 and L2 readers.

Individual differences in attachment preferences

Working memory as measured by reading span interacted with offline attachment preferences in both the L1 and L2 groups. The offline results suggested that participants with high span preferred low attachment more than those with low span, which is consistent with most previous offline studies (e.g., Kim & Christianson, 2012; Omaki, 2005; Swets et al., 2007; but see Felser, Marinis et al., 2003). However, our study did not aim to tease apart different accounts of this correlation, and we cannot distinguish between whether this suggests different prosodic processing strategies (e.g., Swets et al., 2007) or resource limitations (e.g., Omaki, 2005). Importantly for present purposes, we did not find significant differences between L1ers and L2ers in this regard, suggesting that individual differences in reading span influence L1 and L2 processing in a similar way. We also found L2 proficiency interacted with offline attachment preferences, suggesting L2ers with higher proficiency attached low more often than those with lower proficiency. The finding on proficiency is compatible with the online finding

from Dekydtspotter et al. (2008) that participants with higher levels of proficiency showed a native-like attachment preference.

For online processing at the reflexive region, there was a trend during first-pass times that L1ers with fast lexical access exhibited the HA effect, but not L1ers with slower lexical efficiency or L2ers. This might suggest that lexical processing efficiency influences whether readers show attachment effects in early stages of processing, and it is compatible with claims that some aspects of lexical processing need to be complete before syntactic integration begins (e.g., Hopp, 2016; Tily et al., 2010).

During later stages of processing, lexical automaticity modulated L1 and L2 processing to a similar extent. Both highly and less automatized L1/L2 individuals preferred low attachment. However, differences appeared with regards to the high attachment condition particularly in object-modifying RCs, where the difficulty associated with high attachment was reduced in those individuals with higher lexical automaticity (see Figure 3). As discussed above, forcing high attachment RCs in object position seemed particularly difficult to parse, indexed by the stronger low attachment preference in object position from our offline and online findings. Hence, it could the case that individuals with faster lexical access overcame the linguistic conflict and completed reanalysis more quickly compared to individuals with slower access, when processing difficulty increased. Taken together, our results suggest that faster lexical access can facilitate attachment resolution in both L1 and L2 processing.

Our findings in terms of lexical automaticity are not entirely consistent with Hopp (2014). For sentences that were not immediately disambiguated, as tested in our study, Hopp (2014) found that attachment preferences only emerged for L2ers with high levels of lexical automaticity, while L2ers with low lexical automaticity didn't show any significant differences between conditions. However, our results present a different pattern where the individuals with lower levels of lexical automaticity exhibited larger effects. These apparently opposing

findings might be due to differences at the group level. Specifically, the L2 group as a whole in Hopp (2014) did not show any attachment preferences for non-locally disambiguated sentences, while the L2 group in our study did. It could thus be that the L2ers in Hopp (2014) did not have as efficient lexical processing as those tested in our study at the group level. This would not be so odd when one considers the context of testing. Whereas Hopp's participants were tested in a non-native English environment (Germany), our participants were tested in native immersion (the United Kingdom).

Our online findings did not replicate our offline working memory and proficiency effects or any previously reported effects in relation to working memory or proficiency. The null effects of reading span during online processing fail to replicate some previous findings (e.g., Payne et al., 2014; Traxler, 2007), but they are consistent with Hopp (2014), who also did not observe effects of reading span during online reading. One possibility is that the spread of reading span scores in our study was less than in Payne et al., (2014) and Traxler (2007), potentially making reading span effects more difficult to observe. Note however that our combined L1/L2 sample is larger than previous studies. Therefore, sample size may also play a role in these mixed findings. Further replications with large samples are required to tease these issues apart.

Finally, although we tested only L2ers whose L1 has previously been reported to have a high attachment preference, our L2 group did comprise speakers from different L1 backgrounds. The two largest subgroups included L1 Italian speakers (n = 16) and L1 Spanish speakers (n = 11), while smaller subgroups of below 10 participants included L1 Dutch, German, French, Portuguese, Russian, Greek and Arabic speakers. Since these groups are not large enough for inferential analysis, especially for the eye-tracking data, we only descriptively examined how each subgroup behaved in the offline task. Here, all subgroups descriptively indicated a low attachment preference by choosing low attachment more than 65% of the time

in the offline task. Thus, we do not believe a single subgroup was responsible for our findings. However, given differences in the number of participants across the subgroups and in the strength of attachment biases across languages, this may have affected our results. Hence, future research will need to examine how specific L1 backgrounds may influence L2 attachment resolution.

2.5 Conclusion

We argue that our results suggest quantitative rather than qualitative differences between L1ers and L2ers in attachment resolution. Both groups showed similar parsing preferences online, though the effects were smaller and delayed in L2ers. Also, our results suggest that the syntactic position of the RC influences ambiguity resolution, which we argued results from differences in discourse salience between RCs in subject and object position. Importantly, this effect of discourse salience appeared to modulate L1 and L2 processing in a similar way. L1 and L2 attachment resolution also interacted with individual differences in a largely similar way across groups, with better lexical processing efficiency being related to quicker linguistic conflict resolution in both L1 and L2 processing. In sum, our results suggest similar processing strategies in ambiguity resolution are possible in L1 and L2 comprehension, and highlight the importance of considering both linguistic factors and individual differences when examining the similarities and differences between L1 and L2 sentence processing.

Chapter 3. Study 2: Determiner-number specification matters for both L1 and L2 processing of non-local agreement similarly: An ERP investigation

Abstract

The present study uses ERP responses to examine non-local agreement processing between native (L1) English speakers and Chinese-English second language (L2) learners, whose L1 lacks number agreement. We manipulated number marking with determiners (*the* versus *that/these*) to see how determiner-specification influences both native and non-native processing downstream for verbal number agreement. Behavioural and ERP results suggest both groups detected non-local agreement violations, indexed by a P600 effect. Moreover, the manipulation of determiner-number specification revealed a facilitation effect across the board in both grammaticality judgment and ERP responses for both groups: increased judgement accuracy and higher amplitude in the P600 time window for sentences containing violations with demonstratives rather than bare determiners. Contrary to some claims regarding the potential for non-native processing, the present data suggest that L1 and L2 agreement processing employ similar strategies, even when the L1 lacks the relevant distinction.

Keywords: ERPs; non-local agreement; determiner-number specification; bilingualism

3.1 Introduction

Whether adult second language (L2) processing is fundamentally different from native language (L1) processing remains a topic of debate. A broad range of theories, such as the Failed Functional Feature Hypothesis (Hawkins & Chan,1997; Hawkins & Liszka, 2003) and Shallow Structure Hypothesis (Clahsen & Felser, 2006, 2018), predict qualitative differences between L1 and L2 acquisition and processing, such that L2 speakers are predicted to have difficulty when acquiring a novel L2 grammatical feature absent in their L1 grammar or underutilise syntactic information during real-time processing. Alternatively, other approaches maintain that L2 speakers can acquire novel L2 features or construct syntactic representations in a similar way to L1ers (e.g., Cunnings, 2017; Hopp, 2014; Lardiere, 2009; Lim & Christianson, 2015; McDonald, 2006; Kaan, 2014; Schwartz & Sprouse, 1996). In sufficiently proficient L2 speakers, when quantitative differences between L1 and L2 arise, such theories predict that differences might be best explained by issues pertaining to task effects (e.g., Lim & Christianson, 2015), difficulties in anticipating upcoming information (Grüter, Rohde & Schafer, 2014) or with lexical processing efficiency (e.g., Hopp, 2014), memory retrieval interference (e.g., Cunnings, 2017) or participant-level individual differences (e.g., Lempert, 2016; Sagarra & Herchensohn, 2010; Tanner, Inoue, & Osterhout, 2014).

Situated within the above discussion, a growing body of research has investigated processing of subject-verb agreement as in (1) and (2) during comprehension in L2 populations with an L1 that does not license overt agreement (e.g., Chinese, Japanese, Korean) to inform these longstanding theoretical debates in L2 acquisition and processing (e.g., Clahsen & Felser, 2006; Cunnings, 2017; Hawkins & Chan,1997; Schwartz & Sprouse, 1996).

(1) *Turtles moves slowly (local agreement violation)

(2) *The key to the cabinet/s are rusty (non-local agreement violation)

In English, present-tense verbs agree with their subjects in number and person, which is seen reliably in the third person singular -s marking or with greater distinction in the copula *to be*. Both (1) and (2) contain an agreement violation as the subject does not match the verb in number. Contexts like (1) are local as the sentence subject and verb are directly adjacent to each other. However, contexts as in (2) are more complicated as they contain a non-local linguistic dependency where an intervening noun phrase (intervening NP; "the cabinets") is embedded between the subject "the key" and the verb "are". Research has tested how L2 speakers whose L1 does not have agreement process local and non-local agreement (violations) in juxtaposition to L1 speakers during real-time comprehension, using a variety of techniques including event-related potentials (ERPs), although results have been mixed (e.g., Armstrong, Bulkes & Tanner, 2018; Chen, Shu, Liu, Zhao, & Li, 2007; Jiang, 2004; Lim & Christianson, 2015; Ojima, Nakata, & Kakigi, 2005).

Recent ERP research has also examined how double number marking on a subject NP ("Many cookies"), as in (3b), influences sensitivity to agreement violations in comparison to sentences like (3a), without additional marking ("The cookies") in both L1 and L2 speakers (Armstrong et al., 2018; Tanner & Bulkes, 2015). To date however, how double number marking influences agreement processing has been tested in local agreement contexts only, and how double number marking regulates processing of non-local number agreement violations is yet to be explored.

(3a) *The <u>cookies</u> tastes the best when dipped in milk.

(3b) *<u>Many</u> <u>cookies</u> tastes the best when dipped in milk.

Extending previous research, we aim to gauge: (i) ERP responses to non-local agreement violations like (2) in English L1 speakers and Chinese speakers of L2 English in an immersion setting and (ii) how double number marking influences non-local agreement violation processing in L1 and L2 comprehension. Findings will weigh in on different L1/L2 processing accounts, providing evidence to inform theoretical debates regarding potential L2 computational capacity for non-local linguistic dependencies and whether linguistic features absent in an L2 speaker's L1 are indeed particularly problematic.

3.1.1 Agreement processing in L1

A significant amount of L1 research on language comprehension has tested processing of local and non-local agreement violations using behavioural and ERP experimentations (e.g., Alemán Bañón & Rothman, 2019; Dillon, Mishler, Sloggett, & Phillips, 2013; Osterhout & Mobley, 1995; Pearlmutter, Garnsey, & Bock, 1999; Shen, Staub, & Sanders, 2013; Tanner, 2011; Tanner & Bulkes, 2015; Tanner, Grey, & van Hell, 2017; Tanner, Nicol, & Brehm, 2014; Tanner, Nicol, Herschensohn, & Osterhout, 2012; Wagers, Lau, & Phillips, 2009). It has been consistently shown that agreement violations are detected by L1 speakers, reflected by either longer reading times or a P600 effect for sentences containing agreement violations like (1) and (2) compared to grammatical controls (e.g., Chen et al., 2007; Dillon et al., 2013; Lim & Christianson, 2015; Osterhout & Mobley, 1995; Osterhout, McKinnon, Bersick, & Corey, 1996; Tanner et al., 2012; Tanner & Bulkes, 2015), although other ERP responses have also been reported in some studies (e.g., Osterhout, 1997; Tanner, 2019; Tanner, Goldshtein & Weissman, 2018; Tanner & Van Hell, 2014). With non-local violations like (2), English natives show shorter reading times or reduced P600 effects when the intervening NP matches the number properties of the verb compared to when it does not (e.g., Dillon et al.,2013; Lim & Christianson, 2015; Pearlmutter et al., 1999; Shen et al., 2013; Tanner et al., 2012; Tanner et

al., 2017; Tanner, Nicol et al., 2014). This is hypothesized to occur due to a so-called "attraction effect": matching agreement between the intervening NP and verb in number, even though the head noun of the complex NP is the actual agreement controller (e.g., Dillon et al.,2013; Lim & Christianson, 2015; Pearlmutter et al., 1999; Shen et al., 2013; Tanner et al., 2012; Tanner et al., 2017).

How number is marked can also influence agreement violations in L1 processing. In English, number can be marked morphologically (e.g. "cookies") and also lexically, with words such as "many", as in "many cookies". With a quantifier like "many", which itself indicates plurality, the following nominal head must also be overtly marked by the plural marker "-s" ("cookies"). This combination forms a case of double number marking. Tanner & Bulkes (2015) manipulated this factor in a design that tested sentences like (4) to investigate whether double marking using quantifiers, as in (4c/d), would facilitate perception of local agreement violations compared to cases like (4a/b), without double marking.

(4a) The cookies taste the best when dipped in milk. (Grammatical, Unquantified)

(4b) *The cookies tastes the best when dipped in milk. (Ungrammatical, Unquantified)

(4c) Many cookies taste the best when dipped in milk. (Grammatical, Quantified)

(4d) *Many cookies tastes the best when dipped in milk. (Ungrammatical, Quantified)

Indeed, they found the difference in the amplitude of the P600 effect was larger between (4c) and (4d), where the plural subject NP was preceded by a number-marked quantifier, compared to between (4a) and (4b), where it was preceded by a number-unspecified determiner. This suggests that double number marking from the quantifier makes agreement errors more salient in L1 processing.

3.1.2 Agreement processing in L2

While mixed findings have been reported in L2 processing, most research on L1-L2 pairs of typologically similar languages reveals that L2 speakers can be native-like when processing local and non-local agreement violations (e.g., Alemán Bañón, Miller & Rothman, 2017; Frenck-Mestre, Osterhout, McLaughlin, Foucart, 2008; Sagarra & Herchensohn, 2010; Tanner, McLaughlin, Herschensohn, & Osterhout, 2013; Tanner, Inoue et al., 2014). Conversely, with language pairs where morphological agreement is not present in the L1, the picture is less clear, with some prior behavioural studies suggesting similarity to L1 speakers and others differences (e.g., Jiang, 2004; Jiang, Novokshanova, Masuda, & Wang, 2011; Lempert, 2016; Lim & Christianson, 2015). The existing evidence from ERP studies, both in and outside of a native-English immersion context, also paints an ambiguous picture, with studies reporting either similar or different neural responses to relevant agreement violations from L2 speakers of these languages compared to L1 speakers' (e.g., Armstrong et al., 2018; Chen et al., 2007; Ojima et al., 2005).

Cross-study divergence seems, at least in part, to be driven by methodological differences (i.e. materials, whether participants were tested in an immersion setting or not). Processing of local agreement violations like (1) was tested in Japanese speakers of English (Ojima et al., 2005) and Chinese speakers of English (Armstrong et al., 2018) and contradictory results were found. Ojima et al (2005) found that whilst L1 English controls demonstrated both a left-lateralised negativity and a P600 component at the verb for sentences containing local violations (e.g., "Turtles move slowly" vs "*Turtles moves slowly"), the highest proficiency L2 individuals tested who were living in Japan only showed the left-lateralised negativity and absence of the P600 effect. Hence, they claimed qualitative differences between L1 and L2 processing. Conversely, Armstrong et al. (2018) showed that both English controls and Chinese L2-English participants living in the United States exhibited a P600 effect to local violations

(e.g., “The cookies taste…” vs “*The cookies tastes..”), demonstrating that L2 speakers whose L1 lacks the relevant morphological agreement, at least under certain conditions such as in immersion, can demonstrate native-like neural responses to agreement violations.

Armstrong et al. (2018) employed the design and materials as in (4) from Tanner & Bulkes (2015), testing whether a stronger response to local agreement violations would be evoked following double number marking in Chinese speakers of L2 English. Recall that unlike English, number is not morphologically marked on nouns in Chinese but can be marked on determiners using quantifiers (e.g., “Many cookie”) and demonstratives (e.g., “Those cookie”). As such, double number marking is not possible in Chinese. The results found that, unlike English L1 speakers who showed an enhanced P600 effect for violations following double marking, L2 participants showed a reduced P600 effect for (4c)-(4d) relative to (4a)-(4b), suggesting that double marking decreased sensitivity to local violations. Armstrong et al. proposed that the L2 speakers’ failure to utilise double number marking in a native-like way was due to an L1 processing strategy that arises from the overlap in quantification between Chinese and English. Specifically, Armstrong et al. argued that once the Chinese L2 speakers parsed the number marked quantifier “many”, the way number marking happens in their L1, they paid less attention to the morphosyntactic cues on the noun. However, as they only tested one group of L2 speakers, Armstrong et al. acknowledged this could also be a general L2 processing strategy. Nevertheless, there is also a potential confound in their materials, as some of the quantifiers used (e.g., “some”) are number-ambiguous. Given that “some” can also occur with singular nouns (e.g., “Some bread is on the table”), this could have contributed to the L2 speakers’ apparent reduced sensitivity.

As to non-local agreement, Chen et al. (2007) tested a group of Chinese speakers of English in China using a design as in (5), which manipulated sentence grammaticality and the number properties of the intervening noun (car/s).

(5a) The price of the car was too high. (Grammatical, Singular Intervening Noun)

(5b) The price of the cars was too high. (Grammatical, Plural Intervening Noun)

(5c) *The price of the car were too high. (Ungrammatical, Singular Intervening Noun)

(5d) *The price of the cars were too high. (Ungrammatical, Plural Intervening Noun)

For grammatical sentences, a P600 effect was elicited for (5b), where the intervening NP does not match the verb in number, compared to (5a), where it does, in the Chinese speakers of English. Chen et al. interpreted this as indicating that the L2 speakers focused on the incongruency between the local noun and verb. For ungrammatical sentences, even though the L2 speakers detected non-local agreement violations in both (5c) and (5d), irrespective of the intervening NP's number, they showed a distinct neural response, a late negative shift, from the L1 speakers who showed a P600 component. Hence, Chen et al. (2007) sustained the claims of Ojima et al. (2005), concluding the neural underpinnings of L2 processing are qualitatively different from L1 processing when the processed features are absent in the L1. However, the Chen et al. (2007) study, like Ojima, was conducted outside of an immersion setting. Whether immersed L2 speakers' neural responses to non-local violations can be native-like and how double marking regulates non-local agreement processing has not been examined. The present study aims to address these questions.

3.1.3 The present study

In summary, to our knowledge, no existing published studies have used ERPs to examine the processing of non-local agreement in Chinese speakers of English in an immersion setting where both quantity and quality of native input exposure is increased. Related work has shown an association between naturalistic or immersion-like L2 exposure and native-like grammatical

processing (e.g., Dussias, 2003; Morgan-Short, Sanz, Steinhauer, & Ullman, 2010; Morgan-Short, Steinhauer, Sanz, & Ullman, 2012; Pliatsikas & Marinis, 2013). Thus, by testing the same domain of grammar, non-local agreement, as in Chen et al. (2007) in the context of immersion, we will be able to test for further evidence of this inference. Whilst prior research has tested local agreement and double number marking in immersion, no existing published studies have examined the case of non-local agreement and double marking. Also, the relevant previous research has examined double marking using quantifiers, some of which can be number-ambiguous (e.g. *Some* bread is on the table; *Some* breads are made of corn). Thus, the present study employed demonstratives (e.g., these, those) that more clearly mark number unambiguously. Moreover, by testing cases of double number marking using demonstratives, as opposed to quantifiers as in Tanner & Bulkes (2015) and Armstrong et al. (2018), the data will shed light on whether the previously reported effect was from quantification itself or more generalizable to all instances of double marking.

Therefore, our study fills a number of gaps in the literature with three interrelated goals: (a) reconciling some of the inconsistencies found across the above reviewed studies, (b) understanding more specifically what role lexical and morphological cues play in L2 parsing via examining the case of double marking and (c) interpreting what our results can add to debates within L2 acquisition and processing more generally. With this in mind, we addressed the following research questions:

(1) Will a P600 effect be elicited by non-local violations in Chinese speakers of English living in an immersion setting where increased exposure to native English is afforded?

(2) Does double number marking from determiner-number specification using demonstratives elicit a larger or smaller P600 effect in English L1 speakers and Chinese speakers of L2 English?

Based on previous L1/L2 findings, the following hypotheses are proposed. In this immersion setting, if native-like processing is attainable, Chinese speakers of L2 English and native L1 speakers alike should exhibit a P600 effect to agreement violations in non-local dependencies. Furthermore, double marking should enhance neural sensitivity to non-local violations in L1 speakers, leading to a larger P600 effect for violations following double marking (Tanner & Bulkes, 2015). If we replicate Armstrong et al's. (2018) results for L2 speakers, we should observe a reduced sensitivity to non-local violations following double marking in Chinese speakers of English, i.e., a smaller P600 effect. Alternatively, if L2 speakers are able to utilise double number marking like L1 speakers, the P600 effect should be larger in cases of double marking in both groups.

3.2 Method

Participants

The experiment was conducted in an English immersion setting with 32 English L1 speakers (mean age = 21.4) and 32 Chinese - English L2 speakers who learned English in school settings in China (mean age = 25.3). All participants were recruited from the University of Reading and were enrolled in either an undergraduate or postgraduate course. They received a small payment or course credit upon completion of the study. The L2 speakers were born and raised in China and came to the UK for higher education. They were living in the UK at the time of testing and reported their lengths of immersion experience which ranged from 2 to 48 months (mean = 17.7 months, SD = 13.18). Their English proficiency was measured by a short version of the Oxford Quick Placement Test (Oxford University Press, 2004). The proficiency scores ranged from 24-54 out of 60 (mean = 40, SD = 7.87). All participants were right-handed and had normal or corrected to normal vision.

Materials

We recorded EEG with ERP time locking concurrent with a grammaticality judgement task (GJT) to test participants' online processing and comprehension of non-local subject-verb agreement. Following that, we also administered a whole sentence GJT which was slightly different from the EEG concurrent GJT in terms of stimuli presentation, as described below. For the EEG task, 160 critical items like (6) were created, with four target conditions (40 trials per condition) that were distributed across four separate lists so that participants only saw one condition of each item. Each experimental sentence contained a critical verb (either "is" or "has") and manipulated sentence grammaticality (grammatical vs. ungrammatical). The subject was either singular or plural such that half the sentences were grammatical, as in (6a/c), and half were ungrammatical, as in (6b/d). The intervening noun was always singular so that it matched the number properties of the verb. Number specification on the determiner (number-specified vs. number-unspecified) was also manipulated using demonstratives. Conditions (6a/b) had a number-unspecified determiner ("The") while conditions (6c/d) had a demonstrative that specified number ("This/These" or "That/Those"). Across items, these two sets of demonstratives were used an equal number of times. Within the critical sentences, half had "is" as the verb and half had "has". Another 160 fillers were created with half being grammatical and half being ungrammatical. Some of the fillers contained a similar structure to the critical items but had a plural verb (i.e., The biscuits on the table are tasty) to minimise the possibility of participants expecting that the verb would always be singular given that all critical items contained singular verbs. All the sentences were displayed word by word.

(6a) The window of the house is really clean. (Grammatical, Number-Unspecified (NU))

(6b) The windows of the house is really clean. (Ungrammatical, Number-Unspecified (NU))

(6c) That window of the house is really clean. (Grammatical, Number-Specified (NS))

(6d) Those windows of the house is really clean. (Ungrammatical, Number-Specified (NS))

The whole sentence GJT task consisted of a different set of 24 experimental items, that manipulated the same four conditions as in (6), and 30 fillers, using a slightly different procedure from the EEG concurrent GJT. Instead of showing one word at a time, a whole sentence was presented at once, during which the participants made their response. The items were pseudo-randomised in a latin-square design so that each participant saw a different list. Participants only saw one condition of each item and, therefore, read 6 sentences for each condition. Participants were asked to judge whether the sentence they read was grammatical or not by pressing 1 (grammatical) and 2 (ungrammatical) on the keyboard. Correct answers were coded as 1 and incorrect answers were coded as 0. As such, a value closer to 1 indicates higher accuracy. The materials for the EEG and whole sentence GJT experiments can be found in the online Supplementary Materials.

Procedure

The study was conducted in one session. All participants were first asked to provide information on their language experiences by completing a participant form, followed by the main EEG experiment presented in rapid serial visual presentation (RSVP) while participants' EEG activity was recorded. Participants were told to read as naturally as possible and to make sure they understood the sentences. Before each sentence, a fixation marker appeared in the middle of the screen. Following that, the words of each sentence were displayed one at a time for 450 ms with interstimulus intervals of 200 ms. After each sentence, a happy face and a sad face that represented "grammatical" and "ungrammatical" respectively appeared onscreen. Even though it was untimed, participants were asked to make a judgment as quickly and

accurately as possible about whether the sentence they read was grammatical or not using the mouse clicking with their right hands. After that, a 1000 ms blank screen appeared before the presentation of the next sentence. Participants familiarised themselves with the procedure by first completing some practice trials before the experiment. After the EEG task, all participants completed the whole sentence GJT. Finally, the L2 speakers completed the proficiency test.

Data acquisition and analysis

The EEG activity was recorded by a 64-channel active cap system using Brain Vision Recorder and a BrainAmpDC amplifier system (Brain Products, Germany). Eye movements were monitored with Fp1 and Fp2. The data were recorded with a reference to FCz and re-referenced offline to the average of the mastoids. Impedances were maintained below 5 Ω for all channels. The EEG signals were digitized at a sampling rate of 1000 Hz with a bandpass filter of 0.016 to 200 Hz. Data pre-processing was done by Brain Vision Analyzer (Brain Products, Germany). The data were filtered offline at 0.1–30 Hz. Epochs of 1500 ms were segmented around the critical verb with 300 ms before the onset of the critical stimulus and 1200 ms post-onset. The baseline (300 ms pre-stimulus) was corrected for all epochs. Using similar parameters found in Spychalska, Kontinen & Werning (2016), semi-automatic artefact rejection was applied to help spot any trials with the absolute amplitude difference over 200 mV /200 ms, or with the amplitude lower than -130 mV or higher than 130 mV, or with the activity lower than 0.5 mV in intervals of 100 ms, or with a voltage step higher than 50 mV/ms. Trials with blinks, eye movements, excessive amplifier drift or noisy electrodes were removed, which kept at least 63% of the trials in any of the four experimental conditions for each participant in the L1 group and 75% in the L2 group. After the pre-processing procedure, 7% and 5% of the total data were excluded in the L1 and L2 groups respectively prior to averaging and grand averaging. ERPs were time-locked to the onset of the critical verb and averaged offline for each condition at

each electrode for each participant. For each participant, mean amplitudes were computed in the 500–1000 ms post-stimulus window which covers the P600 time window.

Nine regions of interest were included in the analysis: left anterior (F1, F3, F5, FC3, FC5, FC1), left medial (C1, C3, C5, CP1, CP3, CP5), left posterior (P1, P3, P5, P7, PO3, PO7), midline anterior (Fz and FCz), midline medial (Cz and CPz), midline posterior (Pz and POz), right anterior (F2, F4, F6, FC2, FC4, FC6), right medial (C2, C4, C6, CP2, CP4, CP6), and right posterior (P2, P4, P6, P8, PO4, PO8). Repeated measures ANOVAs were conducted separately for the midline and lateral electrode sites due to the different numbers of electrodes these sites had, with Group (L1 and L2) as a between-subject variable, Grammaticality (grammatical and ungrammatical), Number Specification (number-specified and number-unspecified), Caudality (anterior, medial and posterior) and Hemisphere (left and right (only for lateral analysis)) as within-subject variables. Following Armstrong et al. (2018), we only report effects and interactions relevant to Grammaticality and Number Specification effects. For any main effects and interactions involving a variable with more than two levels (caudality), we report the results based on the Mauchly's test for sphericity and sphericity corrections. Post hoc analyses were conducted for any further interactions.

The GJT data from the EEG recording were analysed using generalised (binomial) mixed effects logistic regression (Jaeger, 2008). A generalised mixed model was conducted including sum coded (-1/1) fixed effects of Group, Grammaticality and Number Specification and their interactions. One Chinese participant was removed due to the loss of data. The whole sentence GJT data were analysed using the same methods. The maximal models were computed and fit using the maximal random effects model that converged (Barr, 2013; Barr, Levy, Scheepers, & Tily, 2013). Random intercepts for subjects and items were included. By-subject random slopes included grammaticality*number specification and by-item random slopes included group*grammaticality*number specification. When the maximal model failed

to converge, we refitted the model by first removing the random correlation parameters. If the model still failed to converge, the random effect that accounted for the least variance was iteratively removed until convergence was achieved. The experimental materials and data for our experiments are available at the Open Science Framework (OSF) website (https://osf.io/5stmk/).

3.3 Results

Whole sentence GJT

The descriptive results from the four conditions in the whole sentence GJT are shown in Table 1. The overall judgement score across all four experimental conditions was 0.91 in the L1 group (range = 0.63 to1, SD = 0.1) and 0.92 in the L2 group (range = 0.58 to 1, SD = 0.09). The statistical results (all estimates are in logits) revealed neither main effects of Group nor any interactions by Group (all $z < 1.02$, $p > .3$). The main effect of Grammaticality and Grammaticality by Number Specification interaction were significant (Grammaticality: estimate = -0.55, SE = 0.18, z = -3.13, $p = .002$; Grammaticality by Number Specification: estimate = -0.3, SE = 0.13, z = -2.4, $p = .02$). The Grammaticality effect showed both groups made more incorrect judgements for the ungrammatical sentences relative to the grammatical ones. For the two-way interaction, follow up analyses indicated that while there was no difference between the two grammatical conditions (estimate = 0.22, SE = 0.28, z = 0.79, $p = .428$), participants made significantly more correct judgements on the ungrammatical sentences with a number-specified determiner than those with a number-unspecified determiner (estimate = -0.36, SE = 0.13, z = -2.81, $p = .005$). Also, ungrammatical sentences were judged significantly more poorly than the grammatical ones only when the sentences had a number-unspecified determiner (estimate = -0.86, SE = 0.21, z = -4.14, $p < .001$), but not for number-specified determiner sentences (estimate = -0.29, SE = 0.25, z = -1.16, $p = .25$).

	L1 Speakers	L2 Speakers
Grammatical, NU	0.97 (0.01)	0.96 (0.01)
Ungrammatical, NU	0.82 (0.03)	0.87 (0.02)
Grammatical, NS	0.95 (0.02)	0.94 (0.02)
Ungrammatical, NS	0.90 (0.02)	0.92 (0.02)

Table 1. *Judgement accuracy in the whole sentence GJT (standard errors in parentheses), Study 2*

GJT during EEG

The descriptive results from the four experimental conditions in the EEG concurrent GJT are shown in Table 2. The overall score across all the experimental conditions was 0.86 (range = 0.53 to 0.98, SD = 0.09) in the L1 group and 0.85 (range = 0.51 to 0.99, SD = 0.1) in the L2 group. The results suggest no main effects of Group or any interactions by Group (all $z < -1.5$, $p > .13$). There was a significant main effect of Number Specification (estimate = -0.11, SE = 0.03, z = -3.13, $p = .002$), which was qualified by a significant Grammaticality by Number Specification interaction (estimate = -0.21, SE = 0.03, z = -6.44, $p < .001$). The follow-up analyses showed within the ungrammatical conditions, number-specified determiners elicited more correct judgements than the number-unspecified ones (estimate = -0.31, SE = 0.05, z = -6.57, $p < .001$). Also, within the grammatical conditions, the number-unspecified determiners elicited more correct judgements than the number-specified ones (estimate = 0.11, SE = 0.05, z = 2.23, $p = .03$). Additionally, both groups made better judgments on the grammatical sentences compared to the ungrammatical counterparts for sentences with a number-unspecified determiner (estimate = -0.41, SE = 0.12, z = -3.35, $p < .001$) but did not exhibit such difference for sentences with a number-specified determiner (estimate = -0.002, SE = 0.12, z = -0.01, $p = .989$).

	L1 Speakers	L2 Speakers
Grammatical, NU	0.89 (0.01)	0.90 (0.01)
Ungrammatical, NU	0.83 (0.01)	0.76 (0.01)
Grammatical, NS	0.86 (0.01)	0.90 (0.01)
Ungrammatical, NS	0.88 (0.01)	0.84 (0.01)

Table 2. *Judgement accuracy in the concurrent EEG GJT (standard errors in parentheses), Study 2*

ERPs

Figure 1 illustrates the voltage deflections elicited by (6a-d) at 19 electrodes in both groups.

Lateral analysis results

The ANOVA results for the mean voltage measured along the lateral electrodes during the 500–1000 ms time window indicated a significant main effect of Grammaticality showing the ERP responses were more positive for the ungrammatical sentences than the grammatical ones ($F (1, 62) = 29.72$, $p < .001$), which reflects a P600 effect. The Group by Number Specification interaction was significant ($F (1, 62) = 4.39$, $p = .04$). Follow-up tests showed the voltage was more positive for sentences with a number-unspecified determiner than those with a number-specified determiner in the L1 speakers ($t = 3.59$, $p < .001$), but the opposite in the L2 speakers ($t = -7.23$, $p < .001$). The Grammaticality by Caudality interaction was also significant ($F (2, 124) = 56.96$, $p < .001$). Follow-up t-tests demonstrated that the brain responses elicited by the ungrammatical sentences were more positive than the grammatical ones in both medial ($t = 17.05$, $p < .001$) and posterior ($t = 28.31$, $p < .001$) areas but less positive than the grammatical sentences in the anterior region ($t = -4.99$, $p < .001$).

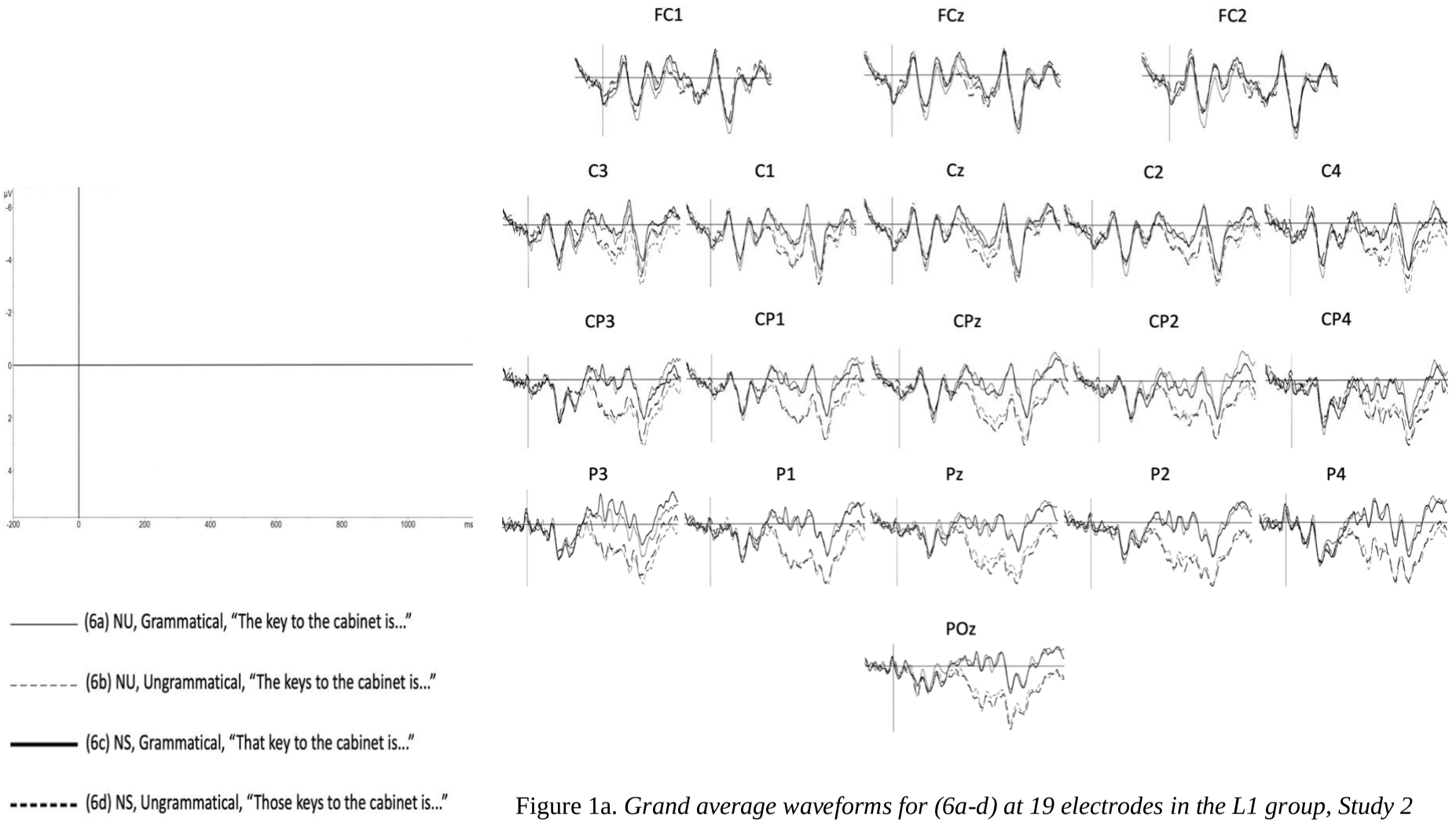

Figure 1a. *Grand average waveforms for (6a-d) at 19 electrodes in the L1 group, Study 2*

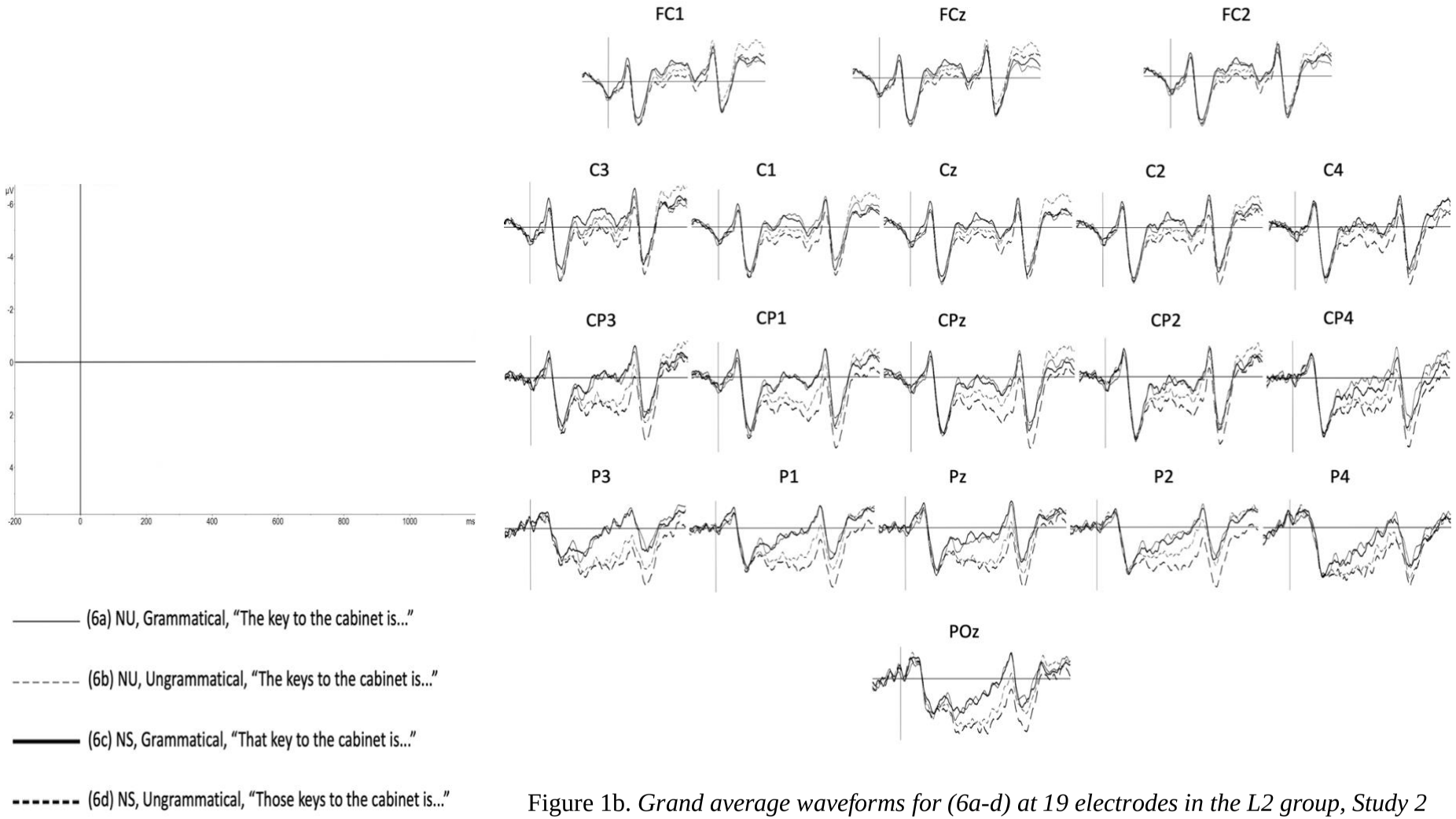

Figure 1b. *Grand average waveforms for (6a-d) at 19 electrodes in the L2 group, Study 2*

Midline analysis results

Regarding the results over the midline electrodes during the 500–1000 ms time window, the main effect of Grammaticality (F (1, 62) = 34.11, $p < .001$) indicated the ungrammatical sentences elicited more positive-going brain responses than the grammatical ones, which corresponds to the P600 effect. Also, the Grammaticality by Number Specification interaction was significant (F (1, 62) = 6.71, $p = .012$). Follow-up t-tests revealed that while both sentences with a number-unspecified determiner and with a number-specified determiner clearly demonstrated the a grammaticality effect (number-unspecified: t = 5, $p < .001$; number-specified: t = 9.99, $p < .001$), the brain responses to the ungrammatical sentences with a number-specified determiner were more positive than those with a number-unspecified one (t = 4.84, $p < .001$) whereas no differences were observed between the grammatical sentences with a number-unspecified determiner and with a number specified one (t = -0.66, $p = .512$). This suggests a larger P600 effect elicited by double number marking in both groups, which is visualised in Figure 1 and Figure 2. Furthermore, the three-way Group by Grammaticality by Caudality interaction was also significant (F (2, 124) = 3.63, $p = .038$). Follow-up analyses suggested the two groups differed in terms of the Grammaticality effect in the posterior region (F (1, 62) = 8.57, $p = .005$). As shown in Figure 3, although both groups exhibited the P600 effect (L1: t = 13.07, $p < .001$; L2: t = 8, $p < .001$), the voltage of the ungrammatical sentences was more positive in the L1 speakers in comparison to in the L2 speakers (t = 3.71, $p < .001$) when there was no between-group difference regarding the grammatical sentences (t = -0.75, $p = .456$). This indicates a larger P600 effect in the posterior area in the L1 group due to its longer duration than in the L2 group, as can be seen in Figure 1. In addition, we found a Group effect in the anterior area (F (1, 62) = 4.9, $p = .03$), showing a significant difference between the two groups in terms of voltage polarity across grammaticality, with positive-going brain responses in the L1 speakers and negative-going responses in the L2 speakers, as displayed in

Figure 3. However, as this effect did not interact with grammaticality, we do not discuss it further.

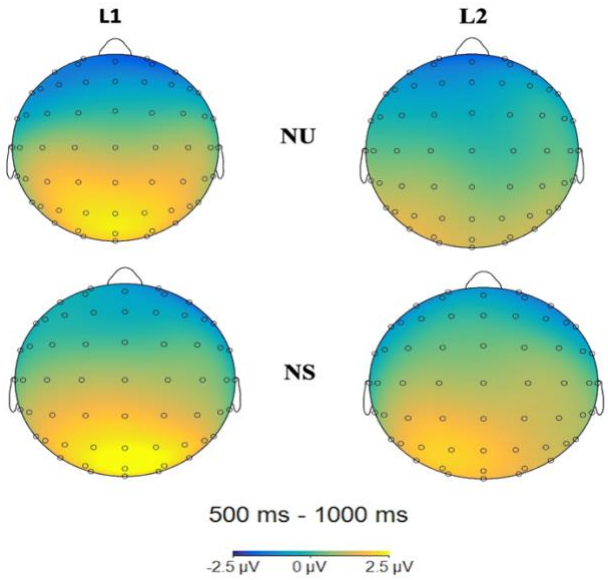

Figure 2. *Topographic distribution of the P600 effects (ungrammatical-minus-grammatical difference) observed in the number unspecified (NU) conditions (6b-6a) and number specified (NS) conditions (6d-6c) during the 500-1000 ms window in the L1 and L2 group, Study 2*

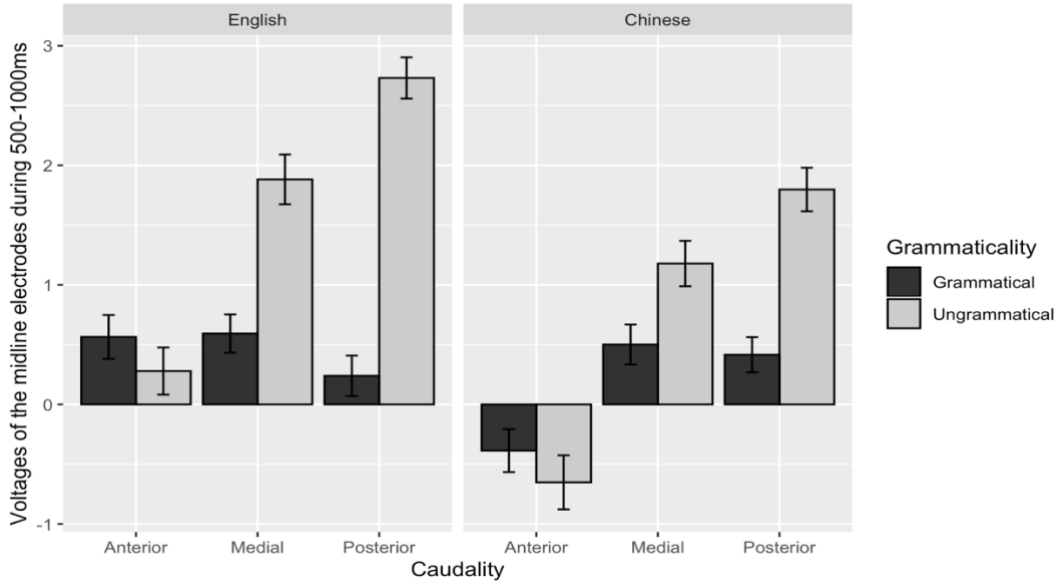

Figure 3. *The Grammaticality effect in each region along the midline electrodes in each group, Study 2*

3.4 General Discussion

This study examined processing of non-local agreement violations in English L1 speakers and Chinese L2 speakers, and tested whether this process was influenced by double marking from determiner-number specification. The results indicated that despite a relatively poorer judgement performance on the sentences containing violations, both L1 and L2 groups showed in general high accuracy to non-local agreement in the whole sentence and EEG GJTs. The EEG data indicated both groups exhibited a P600 during processing of non-local agreement violations during incremental comprehension. Also, the effect of determiner-number specification on detection of violations was attested in both behavioural and neurophysiological measures. We discuss our behavioural and EEG results, along with their implications for theories of L2 sentence processing, in turn below.

Whole sentence and EEG GJTs

Consistent with previous findings (e.g., Armstrong et al., 2018; Chen et al., 2007), the behavioural results indicated both L1 and L2 speakers were equally able to detect syntactic errors in sentences containing non-local agreement violations. However, both groups made more incorrect judgements on ungrammatical sentences than grammatical ones in the whole sentence GJT. The judgement errors here could be either due to response bias that favours grammatical responses (Hammerly, Staub & Dillon, 2019) or attraction from the number match between the intervening NP and verb (e.g., Dillon et al., 2013; Pearlmutter; et al., 1999; Shen et al., 2013). Although attraction is not typically found, or is reduced, in contexts where the intervening NP and verb are singular while the subject is plural (e.g., Bock & Miller, 1991), we do not rule out this possibility. However, it was not our aim to tease apart this issue and our study cannot distinguish between these accounts as we neither manipulated the number of the intervening NP nor neutralised the response bias. Regardless and important for our research

questions, our results showed L2 speakers did not significantly differ from L1 speakers in this regard.

The findings also showed that double number marking led to greater accuracy for sentences containing non-local agreement violations in both groups, which was attested in both whole sentence and EEG GJTs. Even though ungrammatical sentences were generally more poorly judged than the grammatical ones, this difference disappeared in sentences with a number-specified determiner, suggesting determiner-number specification facilitates detection of non-local violations. It could be that the number representation of the subject NP becomes more salient because of double number marking and hence number violations more noticeable. Therefore, these findings suggested double number marking from determiner-number specification increases sensitivity to non-local number violations. This effect is not limited to quantification, the domain tested in Tanner & Bulkes (2015), but indeed to demonstratives as well. Additionally, in the EEG GJT, judgement accuracy for grammatical sentences with a demonstrative determiner was slightly lower than those with a number-unspecified determiner. This might be because grammatical sentences were judged to be more felicitous when there was a bare determiner compared to a demonstrative. However, it is also possible that this difference is spurious as it was not found in the whole sentence GJT or EEG data themselves.

In summary, L1 and L2 speakers were sensitive to number violations in non-local agreement in the two judgement tasks. Number cues from determiner-number specification were similarly processed by L1 and L2 speakers as double number marking facilitated detection of non-local agreement violations in both groups.

ERP effects in L1 and L2

The ERP results during the 500-1000 ms time window from both lateral and midline electrodes showed a typical P600 effect elicited by sentences containing non-local agreement violations

irrespective of number specification in the L1 and L2 groups, suggesting both L1 and L2 speakers detected the non-local violations during incremental processing. Also, the P600 effect was mainly distributed in the medial and posterior areas of the scalp across the board, which confirms that the P600 effect is largely displayed in the centro-parietal region. However, some between-group differences were also observed, as the posterior P600 effect was larger in the L1 speakers than the L2 speakers.

Our findings are consistent with previous L1 literature (e.g., Osterhout & Mobley, 1995; Osterhout et al., 1996; Tanner et al., 2012; Tanner & Bulkes, 2015) and some existing L2 studies (e.g., Armstrong et al., 2018; Lim & Christianson, 2015) that suggest L2 processing of non-local dependencies is not fundamentally different from L1 processing in an immersion setting, even when it comes to processing of a linguistic feature absent in the L1. In comparison to L2 studies that indicated agreement computation in non-immersed L2 speakers whose L1 does not have subject-verb agreement is qualitatively different from that in L1 speakers (e.g., Chen et al., 2007; Ojima et al., 2005), our results provided further neurocognitive evidence demonstrating that neural responses to non-local agreement computation in Chinese speakers of English are not destined to remain distinct from L1 speakers', at least when the L2 speakers have ample experience in a native immersion setting. Therefore, our findings suggest it is likely that immersive input is at least partially deterministic in explaining differences between studies conducted in an immersion setting and those with L2 speakers who lack this relevant experience. In other words, the boost in quality input and opportunity to use the L2 that immersion provides could be responsible for the neurocognitive substrates underlying native-like grammatical processing in our L2 learners in juxtaposition to other similar studies reviewed herein. As such, as opposed to claims that L2 speakers cannot acquire features, in this case number, that are not instantiated in their L1 (Hawkins & Chan, 1997), our findings suggest that it is possible for Chinese speakers of English to process non-local linguistic

dependencies similarly to L1 speakers, even when the relevant feature is not realised in their L1.

Despite the L1/L2 similarities discussed above, it was not the case that our study provides evidence that L1 and L2 processing are exactly the same. Within the 500-1000 ms time window, the posterior P600 effect was longer in the L1 group than in the L2 group, as shown in Figure 1. The P600 effect extended beyond 1000 ms in the L1 group but ended around 800 ms in the L2 group.[1] We argue that the fact that both groups reliably showed the P600 effect in the same time window with no significant distributional differences indicates quantitative rather than qualitative differences in neural responses to non-local agreement violations between L1 and (immersed) L2 speakers. This quantitative difference might indicate that agreement violations were detected online more consistently by the L1 group than the L2ers. Given the nature of this L1/L2 difference, it is compatible with theories which predict quantitative differences between L1 and L2 processing (e.g., Hopp, 2014). Although our findings are not compatible with a strong view of 'shallow' L2 processing that would predict L2ers cannot construct well-specified syntactic representations, the possibility that L2 speakers may not compute agreement as consistently as L1 speakers might be compatible with a weaker version of the SSH (Clahsen & Felser, 2006, 2018). Note also, that given that in all our experimental sentences the noun that intervened between the verb and sentence subject was singular and thus matched the number properties of the verb, the smaller P600 for L2ers might

[1] Following comments from two reviewers, we also conducted an analysis using a 500-800ms time window, which is the same time window as in Armstrong et al. (2018). Although other aspects of our findings stayed the same in this time window, the between-groups difference in the size of the P600 effect was no longer significant. This is inline with our claims that the P600 extended longer in the L1 than the L2 group.

also be suggestive of L2ers being more sensitive to interference from intervening constituents (Cunnings, 2017). While some existing research has investigated interference/attraction in L2 processing (Lago & Felser, 2018; Lim & Christianson, 2015; Tanner et al., 2012), further ERP research that manipulates the number properties of the intervening noun is required here to tease these accounts apart.

Double number marking and ERP effects in L1 and L2 processing

The processing of agreement violations was also found to be modulated by double number marking over the midline electrodes across the board as the P600 effect was larger when the sentences had a number-specified determiner compared to when they had a number-unspecified determiner for both groups, indicating double marking enhanced the neural signal to non-local agreement violations in L1 and L2 processing. Tanner & Bulkes (2015) argued that readers start predicting the number of an upcoming verb based on the number features of the subject NP. They argued that double number marking has a higher degree of predictability and allows readers to make earlier anticipations as a quantifier, or demonstrative determiner in our case, clearly indicates the number features of the verb before encountering the subject NP. Previous studies have suggested stronger brain responses are associated with increased predictability and stronger predictions in lexical and syntactic processing (e.g., Brothers, Swaab, & Traxler, 2015; DeLong, Urbach, & Kutas, 2005; Wlotko & Federmeier, 2012). Therefore, the larger P600 effect we observed for double marking is compatible with the hypothesis that double number marking leads to a stronger prediction being made in both groups.

Our results are consistent with Tanner & Bulkes (2015) for the L1 speakers but contrast in ways with Armstrong et al. (2018) for the L2 speakers. Similar to Armstrong et al., our L2ers demonstrated P600 effects to number violations, but our results differ to Armstrong et al. in

relation to the effect of double number marking. Recall that Armstrong et al. observed smaller, rather than larger, P600 effects in their Chinese L2 speakers in sentences with double marking. They hypothesized that although Chinese speakers of English could acquire the underlying syntactic features of plurality marking in English there was an L1-influence effect for double marking. In other words, Armstrong et al. claimed that because Chinese exclusively marks plurality via quantifiers/demonstratives alone (there is no double marking), when Chinese speakers encounter plurality marked in the way that seemingly overlaps with Chinese, i.e. in a prenominal position via a determiner/quantifier in English, they allocate processing resources to this shared cross-linguistic cue, and consequently, less cognitive resources are allocated to the processing of morphosyntactic agreement cues. As noted previously, one difference between our study and Armstrong et al. is that while we used number-marked demonstratives (that/those), they used quantifiers (many/some). From the perspective of Armstrong et al.'s transfer-based account however, it is not clear that the type of double-number marking should matter, as Chinese has both quantifiers and demonstratives. As such, it does not immediately follow from their account that the type of prenominal (double) marking element should matter. However, as mentioned previously, Armstrong et al. (2018) used quantifiers, some of which (e.g. *some*) can occur with both plural and singular nouns with appropriate verbal agreement. This fact alone could possibly lead to a reduced P600 effect in the L2 speakers. As a result, in our study we avoided this issue by employing demonstratives that are strictly confined to either singular ("this/that") or plural ("these/those") nouns and thus obligatorily either singular (e.g., "is") or plural verbs (e.g., "are"). In doing so, we found the effect of determiner-number specification modulated L1 and L2 processing in the same direction without the need for further consideration. Another methodological difference between the two studies is the structures tested. While Armstrong et al. (2018) used local agreement, we adopted non-local agreement. Thus, future research is required to determine whether different findings between the two

studies are related to structural complexity. Regardless, our data suggest that Chinese speakers do not merely rely on or prioritise the lexical cue from the determiner (when available) for number encoding, but also utilise morphological cues for number agreement computation.

3.5 Conclusion

We observed sensitivity to non-local agreement violations and its interaction with double number marking in both L1 and L2 groups across three tasks. Therefore, we suggest that, despite some observed quantitative differences, the Chinese speakers of English we tested in an immersion setting were, like English L1 speakers, able to compute agreement violations in non-local dependencies, and that double marking from determiner-number specification facilitates detection of number violations in both L1 and L2 processing. The P600 effects we observed suggest that Chinese speakers of English, at least in an immersion setting, have similar neural responses to L1 speakers when processing a novel agreement feature absent in the L1.

Chapter 4. Study 3: Determiner-number specification and non-local agreement computation in L1 and L2 processing: A self-paced reading study

Abstract

The present study employed a self-paced reading task in conjunction with concurrent acceptability judgements to examine how similar or different English natives and Chinese learners of English are when processing non-local agreement. We also tested how determiner-number specification modulates number agreement computation in both native and non-native processing by manipulating number marking with demonstrative determiners (*the* versus *that/these*). Results suggest both groups were sensitive to non-local agreement violations, indexed by longer reading times for sentences containing number violations. Furthermore, we found determiner-number specification facilitated processing of number violations in both native and non-native groups in an acceptability judgement task only, with stronger sensitivity to violations with demonstrative determiners than those with bare determiners. Contrary to some theories that predict qualitative differences between native and non-native processing, we did not find any significant differences between native and non-native speakers, despite the fact that the Chinese speakers of English had to process a novel linguistic feature absent in their native language.

Keywords: non-native sentence processing; non-local agreement; determiner-number specification; self-paced reading

4.1 Introduction

What underlies the similarities and differences with respect to grammatical acquisition and processing between native (L1) speakers and second language (L2) learners has been strongly debated. In research in L2 acquisition, representational deficit approaches to adult L2 acquisition, such as the Failed Functional Features hypothesis (FFFH; Hawkins & Chan,1997), claim that adult L2 learners have limited recourse to acquire native-like mental representations for L2 morphosyntactic features absent in their L1. Updated versions of such accounts, such as the Interpretability Hypothesis (e.g., Hawkins & Casillas, 2008; Tsimpli & Dimitrakopoulou, 2007), propose that it is only uninterpretable (grammatical) features that are not instantiated in the L1 that become inaccessible in adult L2 acquisition, such as number agreement on verbs. Alternatively, other L2 acquisition theories, such as the Full transfer/Full access (FT/FA; Schwartz & Sprouse, 1996), maintain that L2 learners can eventually acquire novel L2 features by modifying the interlanguage until it becomes L1-like.

Similar debates are found in research examining real-time L2 processing. Processing accounts, such as the Shallow Structure Hypothesis (SSH; Clahsen & Felser, 2006, 2018) predict that L2ers, irrespective of L1 background, rely mainly on non-syntactic information and, as a result, construct less detailed syntactic representations of complex structures, such as non-local linguistic dependencies, compared to L1ers. Alternatively, others argue that proficient L2ers can compute syntactic structures in the same way as L1ers, though quantitative rather than qualitative differences between L1ers and L2ers may arise from issues related to memory retrieval interference, cognitive efficiency and/or participant-level individual differences (e.g., Cunnings, 2017; Hopp, 2014; Kaan, 2014; Lempert, 2016; McDonald, 2006; Sagarra & Herchensohn, 2010; Tanner, Inoue, & Osterhout, 2014).

Subject-verb agreement, as in (1) and (2), forms an important test case to inform the abovementioned theoretical debates, as it is not a universal feature across languages and can

occur in non-local dependencies. In (1/2), the verb ("are") must agree with the sentential subject in number. As a result, both (1) and (2) contain an agreement violation due to the number mismatch between the subject and the verb. (1) contains local agreement, as the sentence subject and verb are directly adjacent. However, (2) is a more complex structure, termed a "non-local dependency", as the verb ("are") is not adjacent to its subject ("The park") because of the intervening noun phrase (intervening NP; "the flats") in the prepositional phrase ("near the flats).

(1) *The window are clean (local agreement violation)

(2) *The park near the flats are huge (non-local agreement violation)

The FFFH and Interpretability Hypothesis argue that difficulty should arise when an L2er attempts to acquire a novel grammatical feature. As a result, agreement should not be fully acquirable by an L2er whose L1 does not instantiate the relevant features. Meanwhile, the SSH emphasises that L2ers may under-utilise structural information during parsing, which may give rise to processing difficulty with complex syntactic structures such as non-local agreement.

Note that the FFFH and Interpretability Hypothesis predict difficulty only in L2ers where the L1 does not have agreement, while the SSH predicts non-local agreement processing may be generally difficult in all L2ers. Regardless, both theories assume that subject-verb agreement should cause difficulty to L2ers from non-agreement backgrounds (e.g., speakers of Chinese, Korean and Japanese). Conversely, other accounts would predict that all L2ers can acquire and process agreement similarly to L1ers given enough exposure to the L2 (e.g., Hopp, 2014; McDonald, 2006; Schwartz & Sprouse, 1996).

Subsequently, a substantial number of studies have investigated subject-verb agreement during processing, especially with L2 speakers of L1s that do not license agreement features, to adjudicate between these competing accounts (e.g., Alemán Bañón, Miller & Rothman, 2017; Armstrong, Bulkes, & Tanner, 2018; Chen, Shu, Liu, Zhao, & Li, 2007; Jiang, 2004, 2007; Lim & Christianson, 2015; Ojima, Nakata, & Kakigi, 2005; Tanner, Nicol, Herschensohn, & Osterhout, 2012). While L2ers whose L1 has agreement have been reported to show native-like parsing when processing either local or non-local violations (e.g., Frenck-Mestre, Osterhout, McLaughlin, Foucart, 2008; Tanner, McLaughlin, Herschensohn, & Osterhout, 2013; Tanner, Inoue et al., 2014), findings from L2ers whose L1 does not have agreement have been mixed. For example, some studies suggest similar patterns between L1 and L2ers (e.g., Lim & Christianson, 2015), while others indicate problematic agreement computation in L2ers, indexed by insensitivity to agreement violations (e.g, Jiang, 2004). Therefore, whether agreement computation is particularly problematic to L2ers without agreement in the L1 still remains unclear and needs to be further investigated with a view at unpacking what underlies the inconsistency in the available findings.

Recently, some findings have shown that how number is marked on determiners modulates sensitivity to agreement violations in L1 and L2 processing, though in different directions. For example, while Tanner & Bulkes (2015) found a stronger sensitivity to violations when the number feature of the subject was double marked on both the determiner and the NP (“many cookies”), compared to when a bare determiner was used (“the cookies”), Armstrong et al. (2018) observed a reverse pattern in Chinese L2ers of English, with reduced sensitivity to violations following double number marking. However, these studies tested local agreement only, and this phenomenon has not been examined elsewhere in contexts of non-local agreement thus far.

(3) *The/Many cookies tastes the best when dipped in milk.

To extend on prior findings, we used a self-paced reading paradigm with concurrent acceptability judgements to measure: (i) sensitivity to non-local agreement violations like (2) in English L1ers and Chinese L2ers of English and (ii) how double number marking from determiner-number specification influences sensitivity to violations in L1 and L2.

4.1.1 Agreement processing in L1

A large amount of L1 literature has studied processing of local and non-local number agreement violations during real-time comprehension, using various methods (e.g., Alemán-Bañón & Rothman, 2019; Dillon, Mishler, Sloggett, & Phillips, 2013; Hammerly, Staub, & Dillon, 2019; Osterhout & Mobley, 1995; Osterhout, McKinnon, Bersick, & Corey, 1996; Pearlmutter, Garnsey, & Bock, 1999; Shen, Staub, & Sanders, 2013; Tanner, 2011; Tanner, Nicol, & Brehm, 2014; Wagers, Lau, & Phillips, 2009). L1ers are consistently sensitive to agreement violations, and as a result, typically show slower reading times for sentences containing violations like (1) and (2) in comparison to sentences without violations (e.g., Dillon et al., 2013; Pearlmutter et al., 1999; Wager et al., 2009). However, in contexts where a non-local agreement violation occurs as in (2), L1ers may have difficulty with detecting such violations when the intervening NP matches the verb in number. One influential account describes this finding in terms of similarity-based retrieval interference, which predicts that the intervening NP may be retrieved as the grammatical subject as it matches the verb on number properties, despite the fact that the subject ("the park") is the grammatical agreement controller (e.g., Dillon et al.,2013; Jäger, Engelmann, & Vasishth, 2017; Pearlmutter et al., 1999; Shen et al., 2013; Tanner et al., 2012; Tanner, Grey, & van Hell, 2017).

How number is specified can influence L1 agreement processing. English allows double number marking of an NP by specifying the number feature of the determiner that modifies the noun (e.g., “many cookies”). Using such number-specified determiners (e.g., quantifiers, demonstratives), number properties are more explicit compared to when using bare determiners that do not mark number (e.g., “the cookies”). In an event-related potential (ERP) study with concurrent sentence judgements, Tanner & Bulkes (2015) manipulated this factor by using stimuli like (4) to investigate whether determiner-number specification with quantifiers, as in (4c/d), would render a stronger sensitivity to local number violations compared to cases like (4a/b), without determiner-number specification.

(4a) The cookies taste the best when dipped in milk. (Grammatical, Number Unspecified)
(4b) *The cookies tastes the best when dipped in milk. (Ungrammatical, Number Unspecified)
(4c) Many cookies taste the best when dipped in milk. (Grammatical, Number Specified)
(4d) *Many cookies tastes the best when dipped in milk. (Ungrammatical, Number Specified)

Ungrammatical sentences like (4c/d) yielded a P600 effect compared to grammatical sentences like (4a/b). As predicted, they found a higher judgement accuracy and a larger amplitude of the P600 effect in sentences like (4d), where the plural subject NP was preceded by a number-specified determiner, compared to in (4b), where it was headed by a number-unspecified determiner. This finding indicates that determiner-number specification using the quantifier facilitates detection of local agreement violations in L1 processing. However, whether this influences non-local agreement is as yet unknown.

4.1.2 Agreement processing in L2

A large literature has examined the acquisition and processing of agreement in L2 learners (e.g., Chen et al., 2007; Jiang, 2004; Lardiere, 2007; Lim & Christianson, 2015; Tanner et al., 2012; White, Valenzuela, Kozlowska-Macgregor, & Leung, 2004). Findings from these studies have failed to provide converging evidence. While some studies suggested that L2ers may struggle with subject-verb agreement in acquisition or processing if their L1 does not instantiate number agreement (e.g., Chen et al., 2007; Jiang, 2004; Lardiere, 1998a, 1998b, 2007), other research indicates that successful acquisition and native-like processing of agreement is achievable in this population (e.g., Lempert, 2016; Lim & Christianson, 2015).

Lim & Christianson (2015) investigated processing of non-local violations in Korean L2ers of English in an eye-tracking during reading experiment, using stimuli as in (5) which contained an intervening NP that is inside a relative clause rather than a prepositional phrase like (2). Besides grammaticality, they also manipulated the number properties of the intervening NP so that it either matched or mismatched the verb on number. Their results found that, similar to L1ers, L2ers showed longer reading times for (5c) and (5d) compared to (5a) and (5b), suggesting that they detected agreement errors in both ungrammatical conditions irrespective of number match. In addition, like the L1 counterparts, L2ers showed shorter reading times for (5c), where the intervening NP matched the verb in number, than (5d) when it did not. This was taken to indicate that both groups were affected by similarity-based interference. In summary, Lim & Christianson (2015) argued that Korean L2ers process subject-verb agreement similarly to L1ers.

(5a) The teachers who instructed the students were very strict. (Grammatical, Match)

(5b) The teachers who instructed the student were very strict. (Grammatical, Mismatch)

(5c) *The teacher who instructed the students were very strict. (Ungrammatical, Match

(5d) *The teacher who instructed the student were very strict. (Ungrammatical, Mismatch)

Jiang (2004) examined processing of non-local agreement in Chinese L2ers of English in a self-paced reading experiment. The results, however, as opposed to Lim & Christianson (2015), suggested that L2ers lacked sensitivity to morphologically marked number information in non-local agreement as they, unlike the L1ers, failed to show significantly longer reading times for sentences containing violations than grammatical sentences. However, differences in the statistical analyses conducted may in part explain these discrepancies. Specifically, Jiang (2004) did not test for a statistical interaction between the L1 and L2 groups, and instead drew conclusions based on a significant difference found between grammatical and ungrammatical conditions in the L1 group that was not significant in the L2 group. While the relevant L2 comparison did not reach statistical significance, the Chinese group did show numerical trends that were in the same direction as the L1 group. As the crucial group by grammaticality interaction was not tested, it is difficult to be certain that the L1 and L2 groups did indeed behave differently here, and any conclusions about group differences need to be considered with caution. Furthermore, the intervening NP in the stimuli used in Lim and Christianson (2015) was in a relative clause whereas in Jiang (2004) it was inside a prepositional phrase. This could have contributed to the cross-study difference as previous findings have indicated that interference from the intervening NP may become stronger when having a prepositional phrase rather than a relative clause (e.g., Tanner, 2011; Tanner et al., 2012).

Another L2 study on non-local agreement is Chen et al. (2007), who also tested Chinese L2ers of English using a similar design to Lim & Christianson (2015), but with ERPs. Unlike Jiang (2004) and Lim & Christianson (2015) who only used a reading comprehension task to check if participants paid attention to the experiment (e.g., “Did the teacher instruct the

student?"), Chen et al. (2007) employed a grammaticality judgement task during online reading to tap into L2ers' integrated linguistic knowledge (e.g., "Is this sentence grammatical or ungrammatical?"). Even though the judgement data showed the Chinese L2ers detected agreement violations during online reading, reflected by high judgment accuracy, the ERP responses suggested that they employed a distinct neural process, a late negative shift, from L1ers who exhibited a P600 effect. Therefore, the results were taken to indicate that Chinese L2ers cannot process agreement in a native-like way.

Only one published study has examined how determiner-number specification influences agreement processing in L2 processing. Armstrong et al. (2018) tested Chinese L2ers using the same materials and design as in Tanner & Bulkes (2015) (see 4), in an ERP with concurrent sentence judgement study. Unlike English, Chinese does not license double number marking. Instead, number properties of a noun are mainly only marked using number-specified determiners (e.g., quantifiers, demonstratives), without any morphological marker (e.g., "many cookie"). Armstrong et al.'s results showed no differences between the ungrammatical conditions (4b), where a bare determiner "the" was used, and (4d), where a number-specified determiner "many" was used, in grammaticality judgement accuracy. In the ERP data, and similar to Tanner and Bulkes (2015), the Chinese learners of English showed a P600 effect for ungrammatical (4c/d) compared to grammatical (4a/b). However, unlike Tanner and Bulkes, the Chinese learners in Armstrong et al.'s study showed a smaller P600 effect for (4d) compared to (4b), suggesting a decreased sensitivity to local violations following determiner-number specification using quantifiers.

Armstrong et al. attributed the above difference to an L1 processing strategy transfer. Specifically, they argued that since number can be marked by determiners in both Chinese and English (e.g., both languages use quantifiers and demonstratives to mark number), this overlap prompted the Chinese L2ers to focus on the determiner that carries number properties of the

noun (e.g., “many”), a strategy they would use when processing their L1. As a result of this strategy, Armstrong et al. argued, less attention was given to the plural morpheme “-s” on the noun, which contributed to the attenuated P600 effect. Nevertheless, Armstrong et al. added that this could also be a general L2 processing strategy given that they only tested one group of L2ers. However, note that some of the quantifiers used in their materials (e.g., “some”) can also occur with singular nouns (e.g., “Some bread is on the table”), which could have contributed to the null effect in the grammaticality judgement task and the reduced P600 effect found in their L2ers. Therefore, this effect needs to be further examined using other types of determiners that mark number unambiguously, such as demonstratives (e.g., these, those). Also, by examining the case of demonstratives, we can ascertain whether the reported effects in previous studies are attributed to quantification itself or, a more inclusive category, determiner-number specification. Meanwhile, no published studies have systematically examined the effect of determiner-number specification in contexts of non-local agreement in L2 processing.

4.1.3 The present study

Against the aforementioned issues, we conducted a self-paced reading experiment with concurrent acceptability judgements to examine how Chinese L2ers of English process non-local agreement containing an intervening constituent in a prepositional phrase, as in Jiang (2004), directly compared to English L1ers. Findings will inform debates surrounding whether native-like acquisition of a novel L2 feature is possible as well as the extent to which native-like L2 processing is attainable in a domain of non-local agreement violations (e.g., The keys to the cabinet is rusty). We also delved a bit deeper to see the extent to which specific properties of the target grammar might facilitate or otherwise hinder native-like processing. To this end, we probed whether double marking, determiner-number specification via demonstratives,

modulates sensitivity to non-local agreement violations in L1 and L2 processing. Herein, the following research questions are addressed:

(i)Will Chinese L2ers of English be able to detect non-local agreement violations that contain a prepositional phrase?

(ii)Will determiner-number specification via demonstratives increase or decrease sensitivity to non-local agreement violations in English L1ers and Chinese L2ers of English?

According to the FFFH/Interpretability Hypothesis and SSH, Chinese L2ers may behave differently from L1ers, such that they may be insensitive to non-local agreement violations. Finding that L2 learners are not sensitive to subject verb agreement violations during processing would provide strong support for such theories. Conversely, accounts such as FT/FA would predict that Chinese L2ers may behave similarly to L1ers in terms of sensitivity to agreement violations. Furthermore, L1 sensitivity to non-local agreement violations should be enhanced via determiner-number specification (Tanner & Bulkes, 2015), leading to higher judgment accuracy and longer reading times for violations following determiner-number specification. If our L2ers process double number marking differently to L1ers (Armstrong et al., 2018), they should exhibit a decreased sensitivity to non-local violations following determiner-number specification and hence lower judgement accuracy and shorter reading times for those sentences. Otherwise, if our L2ers can acquire and process number marking in a native-like way, higher judgement accuracy and longer reading times should be observed in sentences containing violations following determiner-number specification.

4.2 Method

Participants

The experiment was conducted with 40 English L1ers (mean age = 20.7) and 40 immersed Chinese L2ers of English (mean age = 25.7; mean age of acquisition = 8.2 years, range = 4-14 years). All participants were right-handed and had normal or corrected to normal vision. The L1 participants were undergraduates at the University of Reading. They received a small payment or course credit upon completion of the study. The L2ers learned English in a school setting in China where they were born and raised. They were all studying a higher education degree in the UK at the time of testing and reported their lengths of immersion experience (mean = 31.6 months, range = 5-120 months, SD =30.1). Their English proficiency scores, as measured by a quick Oxford Placement Test (Oxford University Press, 2004), ranged from 26-56 out of 60 (mean = 43, SD = 6.33).

Materials

We employed a self-paced reading task in conjunction with a concurrent acceptability judgement task to test participants' online processing and comprehension of agreement violations in non-local linguistic dependencies. The motivation for using acceptability judgments is that they may be a better way to prompt participants to use grammatical information implicitly during online reading, compared to grammatical judgements which instruct participants to focus on grammatical information and may involve more use of explicit knowledge (e.g., Guo, Guo, Yan, Jiang, & Peng, 2009). The reading task consisted of 32 critical items like (6) with 4 experimental conditions, and 64 fillers, pseudo-randomised in a latin-square design so that each participant read a different list and only one condition of each item, and, therefore, 8 sentences per condition. The critical items manipulated grammaticality (grammatical vs. ungrammatical). Half the sentence subjects were singular and half were plural, while the critical verb was always singular ("is"), such that half the sentences were grammatical like (6a&6c) and half were ungrammatical like (6b&6d). Also, half the

experimental sentences had “is” as the critical verb and half had “has”. The intervening noun was always singular so that it matched the critical verb on number properties. Number specification on the determiner (number-specified vs. number-unspecified) was also manipulated, using demonstratives. Half the items had a number-unspecified determiner (“The”) as in (6a&6b) whereas half had a number-specified determiner, demonstrative (“This/These”), as in (6c&6d). Within the items, the numbers of demonstratives “this/these” and “that/those” were equally distributed. All the critical sentences were followed by acceptability judgements where participants were asked to indicate whether the sentence they read was acceptable or not.

(6a) The picture of the lake is so beautiful. (Grammatical, Number-Unspecified (NU))

(6b) The pictures of the lake is so beautiful. (Ungrammatical, Number-Unspecified (NU))

(6c) This picture of the lake is so beautiful. (Grammatical, Number-Specified (NS))

(6d) These pictures of the lake is so beautiful. (Ungrammatical, Number-Specified (NS))

Of the filler sentences, 32 of the 64 were followed by acceptability judgements, half of which were grammatical and half ungrammatical. Some of these fillers had a similar structure to the experimental sentences but contained a plural verb (e.g., The motorbikes in the street are really cool.) to stop participants being strategic given that all the verbs were singular in the experimental items. The remaining 32 fillers were all grammatical and did not require participants to make a judgement. We adopted this procedure to minimise any task-related effects on reading behaviour based on participants seeing multiple ungrammatical sentences. The reading task was carried out in a web-based self-paced reading paradigm where sentences were presented word by word.

Procedure

The study was conducted online, with participants completing the experiment in their own time and setting. All participants were first asked to complete a participant form and give informed consent. Following this, they were instructed to complete a self-paced reading task where they first saw a row of dashes that covered up the sentence to-be-read and then progressed with reading at their own pace by pressing the space bar to uncover one word at a time. Participants were told to read as naturally as they could and to make sure they understood the sentences. After all experimental sentences and half of the fillers, a question "Acceptable or unacceptable?" appeared on a different page and participants had to judge whether the sentence was acceptable or not by pressing 1 (acceptable) and 2 (unacceptable) on the keyboard. Before the actual experiment, participants familiarised themselves with the procedure by completing several practice trials. Finally, the Chinese speakers of English completed the proficiency test.

Data analysis

Reading times were calculated at two regions of text as underlined in (6). The critical region consisted of the critical verb ("is"), while the spillover region contained the word after the critical verb ("so"). Datapoints containing reading times less than 100ms or over 10,000ms were removed, accounting for less than 1% of the data. Such data likely index lapses in attention.

For the reading time task, the data were log-transformed to minimise skew (see Vasishth & Nicenboim, 2016) and analysed using mixed-effects models (Baayen, Davidson, & Bates, 2008) that included sum coded (-1/1) fixed effects of Group (L1/L2), Grammaticality (grammatical and ungrammatical), Number Specification (number-specified and number-unspecified) and their interactions. Post hoc analyses were conducted for any further interactions. For the judgement data, correct answers were coded as 1 and incorrect answers

were coded as 0. As such, a value closer to 1 indicates higher accuracy. For analysis, we used a binomial generalised mixed-effects model (Jaeger, 2008) containing the same sum coded fixed effects as the reading time data.

The models were fit using the maximal random effects model that converged (Barr, 2013; Barr, Levy, Scheepers, & Tily, 2013). Random intercepts and slopes were included. By-subject random slopes included grammaticality*number specification, and by-item random slopes included group*grammaticality*number specification. When the maximal model failed to converge, we first removed the random correlations. If it still failed to converge, we then iteratively removed the random effect that accounted for the least variance until the model converged. The experimental materials, data and analysis code for our experiments is available at the Open Science Framework (OSF) website (https://osf.io/pnux2/).

4.3 Results

The summaries of descriptive and inferential statistics for both the acceptability judgement and reading time tasks are presented in Tables 1 and 2.

Acceptability judgment results

The judgment accuracy of the fillers was 0.93 for the L1 group and 0.8 for the L2 group (all participants scored above 0.66), indicating that participants paid attention to the experiment. The overall accuracy across all four conditions was 0.89 in the L1 group (SD = 0.31) and 0.84 in the L2 group (SD = 0.37).

The statistical analysis revealed a significant Group main effect, showing that the overall accuracy was higher in the L1ers than the L2ers. This main effect further interacted with Grammaticality. Post-hoc analyses indicated that the two groups were significantly different in terms of judgement accuracy of the ungrammatical sentences only (ungrammatical

estimate = -0.66, SE = 0.17, z = -3.97, $p < .001$; grammatical estimate = 0.11, SE = 0.21, z = 0.54, $p = .588$), with the L2 group having a lower accuracy in the conditions containing agreement violations compared to the L1 group. The two-way Grammaticality by Number Specification interaction was also significant, in the absence of any further significant interactions with group. Follow-up analyses revealed that Number Specification effect was significant in both grammatical sentences (estimate = 0.40, SE = 0.13, z = 3.12, $p = .002$) and ungrammatical sentences (estimate = -0.31, SE = 0.09, z = -3.38, $p < .001$), but in different directions. While accuracy for sentences containing number-specified determiners was lower than those containing number-unspecified determiners in grammatical conditions, it was higher for ungrammatical conditions.

	L1 Speakers	L2 Speakers
Acceptability Judgements		
Grammatical, Number Unspecified	0.92 (0.02)	0.93 (0.01)
Ungrammatical, Number Unspecified	0.88 (0.02)	0.73 (0.02)
Grammatical, Number Specified	0.84 (0.02)	0.88 (0.02)
Ungrammatical, Number Specified	0.93 (0.01)	0.81 (0.02)
Reading Times (Verb)		
Grammatical, Number Unspecified	433 (12)	621 (25)
Ungrammatical, Number Unspecified	441 (11)	680 (33)
Grammatical, Number Specified	429 (16)	599 (22)
Ungrammatical, Number Specified	449 (14)	704 (31)
Reading Times (Spillover)		
Grammatical, Number Unspecified	407 (9)	542 (16)
Ungrammatical, Number Unspecified	417 (10)	579 (24)
Grammatical, Number Specified	395 (8)	573 (22)
Ungrammatical, Number Specified	415 (10)	581 (23)

Table 1. *Descriptive statistics (standard errors in parentheses) for acceptability judgements and reading times, Study 3*

	Estimate (SE)	t / z	*p*
Acceptability Judgements			
Group	**-0.302 (0.151)**	**-2.00**	**.046**
Grammaticality	-0.229 (0.127)	-1.81	.071
Specification	0.031 (0.072)	0.43	.669
Group*Grammaticality	**-0.422 (0.108)**	**-3.91**	**< .001**
Group*Specification	0.012 (0.071)	0.16	.870
Grammaticality*Specification	**-0.376 (0.086)**	**-4.37**	**< .001**
Group*Grammaticality*Specification	0.067 (0.071)	0.95	.345
Reading Times (Verb)			
Group	**0.162 (0.030)**	**5.37**	**< .001**
Grammaticality	**0.026 (0.010)**	**2.60**	**.014**
Specification	0.001 (0.008)	0.12	.903
Group*Grammaticality	0.012 (0.008)	1.51	.136
Group*Specification	-0.010 (0.007)	-1.33	.182
Grammaticality*Specification	-0.009 (0.008)	-1.17	.247
Group*Grammaticality*Specification	-0.002 (0.008)	-0.28	.783
Reading Times (Spillover)			
Group	**0.133 (0.025)**	**5.37**	**< .001**
Grammaticality	0.008 (0.008)	0.96	.338
Specification	0.000 (0.007)	0.05	.961
Group*Grammaticality	-0.008 (0.009)	-0.93	.357
Group*Specification	-0.010 (0.007)	-1.53	.127
Grammaticality*Specification	-0.001 (0.007)	-0.07	.947
Group*Grammaticality*Specification	0.004 (0.007)	0.49	.624

Table 2. *Inferential statistics for acceptability judgements and reading times, Study 3*

Reading time results

At the critical region, the results indicated a significant main effect of Group without any further interactions, suggesting that the L2ers had longer reading times than the L1ers. The main effect of Grammaticality was also significant, showing longer reading times spent on sentences containing agreement violations than the grammatical conditions (see Table 1).

At the spillover region, only the main effect of Group was significant, indicating longer reading times across conditions for the L2ers. No other significant main effects or interactions of theoretical interest were observed (all $t < 1.53$, all $p > .127$).

4.4 General Discussion

Our results showed that although the L2ers had a lower judgement accuracy relative to the English L1ers in sentences containing non-local agreement violations, both groups were capable of detecting agreement violations in non-local linguistic dependencies, as indicated by overall accuracy in all conditions being high. Both groups also exhibited longer reading times for ungrammatical than grammatical sentences at the critical verb. Furthermore, we found a facilitation effect of determiner-number specification in the judgement data but not reading time data. We discuss the implications of our judgement and reading times data in turn below.

Acceptability judgements

The judgement data suggested that the L2 group exhibited good knowledge of the target linguistic feature and judged non-local agreement accurately most of the time, which is in line with previously reported L2 findings in judgement tasks (e.g., Armstrong et al., 2018; Chen et al., 2007; Ojima et al., 2005). Also, consistent with some existing L2 studies (e.g., Tanner et al., 2012), we found judgement accuracy for the ungrammatical sentences was comparatively lower in the L2ers than L1ers. This judgement error might result from interference from the singular intervening NP that has matching number features to the verb (e.g., Dillon et al., 2013; Shen et al., 2013; Tanner et al., 2012). Therefore, the lower accuracy for the ungrammatical sentences in the L2 group may be compatible with the claim that L2ers are more susceptible to interference than L1ers when processing non-local linguistic dependencies (Cunnings, 2017). The error could also result from a response bias towards grammatical/acceptable responses (e.g., Hammerly et al, 2019) as our L2ers did show a preference for more acceptable responses than the L1ers in the fillers (60% versus 53%). Nevertheless, we do not draw any strong conclusions about the interference effect or response bias here as we neither included an

ungrammatical baseline condition with a mismatching intervening NP (e.g., “The pictures of the lakes is so beautiful”), nor manipulated acceptable/unacceptable response proportions.

We also found that judgement accuracy for sentences containing non-local agreement violations increased by determiner-number specification in both groups. The higher judgement accuracy in the ungrammatical sentences with a plural demonstrative determiner reflected that the L1 and L2ers found the sentences with violations more unacceptable when the determiner is number-specified compared to when it is number-unspecified. According to Tanner & Bulkes (2015), as opposed to the number-unspecified determiner “the”, the number-specified determiner (e.g., those) allows readers to predict the number properties of the upcoming subject NP and verb as it clearly marks plurality, leading to an early and stronger prediction, and therefore a more pronounced agreement error when the prediction is violated. Alternatively, it could also be the case that the number representation of the subject becomes stronger as its number features are expressed twice with determiner-number specification, and hence violations at the verb are more pronounced. Regardless, our results suggest determiner-number specification enhances sensitivity to non-local violations in acceptability judgements, and the L1ers and L2ers were not significantly different in this regard.

However, we found a higher accuracy for grammatical sentences with a number-unspecified determiner compared to those with a number-specified determiner in both groups, indicating that our participants found grammatical sentences with a bare determiner more acceptable than those with a demonstrative. Since this difference was not present in the reading time data, we can only speculate that it is most likely spurious and might have something to do with the judgement task we employed. As we used acceptability rather than grammaticality judgements, people may have judged the sentences using non-syntactic information such as pragmatics and hence found the sentences with ‘the’ more acceptable than those with demonstratives (e.g., “this”, “those”). This could also explain why our participants found

ungrammatical sentences with a demonstrative more unacceptable than those with a bare determiner.

Our results concerning the effect of determiner-number specification are aligned with previous L1 findings (Tanner & Bulkes, 2015), but inconsistent with Armstrong et al.'s (2018) L2 findings, who reported no such effect in their Chinese learners of English in a grammaticality judgement task. This could be explained by differences in the materials between studies. As mentioned previously, the use of quantifiers could have contributed to the null effect in the Chinese L2ers in Armstrong et al. (2018) as some of the quantifiers (e.g., *some*) can occur with both plural and singular nouns and corresponding verbal agreement. Conversely, our study avoided this potential confound by using demonstratives that are strictly tied to either singular ("this/that") or plural ("these/those") nouns and their according verbal inflections. Furthermore, the pattern in our judgement data is not consistent with the account in Armstrong et al. (2018), who reported reduced sensitivity to morphosyntactic violations following determiner-number specification in Chinese L2ers and accounted for this result using an L1 transfer strategy. Specifically, they argued that since Chinese uses lexical cues alone, such as quantifiers, to mark number, when Chinese speakers encounter double number marking using quantification and morphological cues in English, they attend more to the feature that exists in their L1 (quantification) and consequently become less sensitive to the obligatorily concurrent morphological cue ("-s"). According to their transfer-based account, we should have observed similar results (reduced sensitivity to violations) in our L2ers as Chinese also uses demonstratives to mark number. However, our results suggested a different direction of the effect. As explained earlier, the apparent reduced sensitivity in their L2ers could have arisen from their choice of quantifiers used. Given that some of the quantifiers used by Armstrong et al. could occur with singular nouns and corresponding verbal agreement, some of the violated singular verbs in their stimuli could have been processed as grammatical, which possibly

reduced the overall strength of the effect in the Chinese L2ers. Therefore, we believe that Chinese L2ers do not overly rely on lexical cues from the determiner to encode number when both lexical and morphological cues are available. Instead, they utilised cues from both levels for non-local agreement computation. Our data also suggested that the effect of determiner-number specification is not limited to quantification and also applies to demonstratives.

In conclusion, our L1 and L2 participants detected number agreement violations in non-local dependencies in the judgement task. The use of demonstratives led to enhanced sensitivity to non-local agreement violations in both groups, suggesting that double number marking was similarly processed by the L1ers and L2ers.

Reading times

The Grammaticality effect observed at the critical verb showed longer reading times elicited by the sentences containing non-local agreement violations, indicating that both L1 and L2ers noticed syntactic violations during incremental comprehension, irrespective of double number marking from determiner-number specification.

Our online findings corroborate prior L1 agreement literature (e.g., Dillon et al., 2013; Pearlmutter et al., 1999; Shen et al., 2013; Wagers et al., 2009) and some existing L2 studies, indicating detection of agreement violations in L2ers with an L1 that has no agreement (e.g., Armstrong et al., 2018; Lim & Christianson, 2015). In contrast to some L2 studies that suggested qualitative differences in non-local agreement computation between English L1ers and Chinese L2ers of English (e.g., Chen et al., 2007; Jiang, 2004), our reading time data did not show any significant differences in non-local agreement processing in our Chinese L2ers as compared to English L1ers during real-time sentence comprehension. Previous research has suggested that immersive naturalistic L2 input may be associated with native-like syntactic processing (e.g., Armstrong et al., 2018; Dussias, 2003; Morgan-Short, Sanz, Steinhauer, &

Ullman, 2010; Morgan-Short, Steinhauer, Sanz, & Ullman, 2012; Pliatsikas & Marinis, 2013), which could account for the different findings between Chen et al. (2007) and our study as their participants were tested in China whereas our participants were tested in the UK.

Recall that Jiang (2004) also tested immersed Chinese L2ers but argued that the Chinese L2ers in his study were not sensitive to non-local agreement violations during processing. However, as mentioned previously, Jiang (2004) only conducted statistical comparisons between conditions within each group and did not test for the crucial Grammaticality by Group interaction. Our study is similar to Jiang (2004) with respects to the target feature (i.e., non-local agreement with a prepositional phrase), testing paradigm (i.e., self-paced reading) and testing setting (i.e., immersion setting), though we directly compared the L1 and L2 groups. Therefore, we believe that the conflicting results between studies may result from the fact that Jiang (2004) did not conduct direct group comparisons rather than other factors, especially given that the numeral trends in Jiang's L2 group were similar to the patterns in his L1 group. Most importantly, in contrast to what might be expected under the SSH and FFFH/Interpretability Hypothesis, in our study we did not find any significant differences between L1 and L2 speakers in terms of sensitivity to non-local agreement violations during incremental processing.

Our reading time data did not replicate the effect of determiner-number specification observed in our judgement data, which would most clearly predict a significant grammaticality by number specification interaction. There was, however, a numerical trend of this effect, which can be observed at the critical verb region in both groups, showing a larger reading time difference between grammatical and ungrammatical sentences with a number-specified determiner compared to a bare determiner (L1: 20ms vs 8ms respectively; L2: 105ms vs 59ms respectively). We are cautious in overinterpreting this effect here, but emphasise that we did not observe significant L1/L2 differences in relation to either grammaticality or number

specification effects during online reading. Although our results clearly show effects of number specification in our judgement data, future research is required to further examine the effects of number specification during online processing in L1ers and L2ers. Most importantly for present purposes, despite the lack of significant effects of number specification during processing, both L1ers and L2ers demonstrated sensitivity to non-local agreement violations.

4.5 Conclusion

We examined L1 and L2 processing of non-local agreement violations using a self-paced reading task with concurrent acceptability judgements. Both L1 and L2 groups detected non-local agreement violations in both acceptability judgements and reading times. Despite some quantitative differences in judgement accuracy for ungrammatical sentences, we did not find any signification differences between Chinese L2ers and English L1ers when processing agreement violations in non-local dependencies during online reading. Indeed, cues for number marking were processed similarly by both groups in the judgement task, with double marking from determiner-number specification facilitating detection of non-local agreement violations in L1 and L2 processing. In summary, we did not find significant differences between L1 and L2 readers in detecting number violations during processing, and as such our results do not provide support for theories that predict qualitative differences between L1 and L2 acquisition and processing.

Chapter 5. General discussion and conclusion

This thesis investigated L1 and L2 processing of ambiguous RCs and non-local number agreement. Recall that the L2 group consisted of L2ers of English with an L1 that, in contrast to English, is reported to have a high attachment preference in Study 1. Chinese L2ers of English were tested in Study 2 and Study 3, as Chinese lacks subject-verb agreement.

Study 1 suggested that both L1 and L2ers demonstrated a low attachment preference in both offline and online tasks, with the L1 group showing a stronger low attachment preference during reading, indexed by a larger reading time difference between globally ambiguous RCs and high attachment RCs. Also, attachment resolution was influenced by the RC's syntactic position, reading span and lexical automaticity similarly in both L1 and L2 processing.

Grammaticality judgement data and ERP responses in Study 2 showed that Chinese L2ers of English detected non-local agreement violations during both offline and online reading. The ERP data also indicated that L2ers exhibited a qualitatively similar albeit smaller ERP effect as L1ers to agreement violations. Furthermore, double number marking influenced processing of non-local agreement similarly in both L1 and L2 group, with double number marking increasing judgement accuracy and neural sensitivity for and to non-local agreement violations.

In Study 3, both judgement data and reading time results demonstrated that both L1ers and Chinese L2ers of English were sensitive to non-local agreement violations during online reading, even though only Chinese L2ers had a lower judgement accuracy for ungrammatical sentences than grammatical ones. In addition, as in Study 2, double number marking increased judgement accuracy for sentences containing agreement violations in both L1ers and L2ers. However, despite numerical trends, this effect was not significant in the reading time data.

Implications of these results will be discussed in relation to similarities and differences between L1 and L2 sentence processing in more detail below.

5.1 Similarities and differences between L1 and L2 attachment resolution

As introduced in Chapter 1, this thesis considered three models predicting how L1ers resolve attachment ambiguities. The garden path model predicts a low attachment preference, guided by the parsing strategy Late Closure. The competition models predict that globally ambiguous RCs are more difficult to process than temporarily ambiguous ones as readers compute multiple analyses in parallel. Conversely, the unrestricted race model claims that globally ambiguous RCs should be the easiest to process as readers variably attach the RC because they consider all sources of information that may influence attachment preferences. Offline and reading time data clearly showed that L1ers of English preferred low attachment, which is consistent with previous findings (e.g., Cuetos & Mitchell, 1988; Hopp, 2014) and compatible with the garden path model. Since we did not find any evidence of competition or variable attachment, our results in this respect do not support the competition and the unrestricted models.

Regarding L2 attachment resolution, recall that the Shallow Structures Hypothesis (SSH) predicts null attachment preferences due to L2ers' reduced ability to use syntactic information during parsing. Other theories such as the Lexical Bottleneck Hypothesis predict native-like attachment preferences when individual differences are considered. The findings from both offline and online measures suggested a low attachment preference in L2ers, even though this preference emerged at a later point in time and was not as strong as that in the L1 group, suggesting that L2 processing is also guided by Late Closure. While our L2 findings stand in contrast with some previous findings suggesting no clear attachment preferences in L2ers in either offline or online tasks (e.g., Felser, Roberts et al., 2003; Kim & Christianson, 2017; Omaki, 2005), they are in line with other L2 studies showing native-like parsing preferences (e.g., Bidaoui et al., 2016; Hopp, 2014). Thus, Study 1 showed no evidence of shallow L2 parsing and is seemingly most consistent with the models that only predict quantitative L1/L2 differences (e.g., Cunnings, 2017; Hopp, 2014).

Study 1 also presented some evidence that attachment resolution is modulated by various kinds of non-syntactic factors to a similar extent in L1ers and L2ers. Both offline and online results found that where the RC is embedded influenced the strength of the low attachment preference, with both L1 and L2 groups showing a stronger low attachment preference when the RC modified a complex NP that served as the object of the sentence. This indicates that both L1 and L2ers took into consideration such discourse-based cues during parsing. Our offline findings differ from previous studies that showed no such position effect in offline measures (e.g., Hemforth et al., 2015; Kim & Christianson, 2017). As discussed previously, it could be that Study 1 had a much larger sample size compared to these studies.

Since previous studies only examined individual differences effects in either L1 or L2ers but not both, Study 1 investigated whether individual differences modulate L1/L2 attachment resolution similarly. Firstly, we found that attachment preferences interacted with reading span in the offline task, with higher reading span being associated with a stronger low attachment preference in both groups, which replicated previous findings (e.g., Omaki, 2005; Swets et al., 2007). Secondly, the eye-tracking data showed that lexical automaticity influenced attachment resolution in L1 and L2ers similarly. Although both L1/L2 individuals with high and low automaticity preferred low over high attachment, individuals with high automaticity resolved the processing difficulty from object-modifying RCs that forced high attachment more efficiently.

Our L2 findings are not entirely consistent with Hopp (2014) in relation to the effect of lexical automaticity. While Hopp (2014) showed that only L2ers with faster lexical access demonstrated a native-like preference, we found that both our L2ers with slow and fast lexical access showed a low attachment preference, but those with slower access showed larger effects. As discussed previously, this could be due to our L2ers being more efficient in terms of lexical processing than those tested in Hopp (2014) given that our L2ers were recruited in an

immersion context whereas those in Hopp (2014) were not. Finally, we found that the low attachment preference increases as a function of L2 proficiency in the offline experiment, which is compatible with models that predict native-like parsing with increased L2 proficiency (e.g., Hopp, 2006). These suggest that although L1 and L2 attachment resolution are guided by syntactic information, non-syntactic factors (e.g., discourse, working memory) also play a role during this process. Thus, our findings could be viewed as partially compatible with the unrestricted race model in terms of parsing being serial and interactive.

In summary, both offline and online findings from Study 1 did not provide evidence of qualitative differences between our L1 and L2ers given that both L1 and L2 processing were guided by the same parsing principle and modulated by a range of factors to a similar degree.

5.2 Similarities and differences between L1 and L2 processing of agreement and number marking

Recall that there are several models that make testable claims regarding the extent to which L1 and L2 acquisition and processing are similar or divergent. In terms of acquisition, the Failed Function Feature Hypothesis (FFFH) and the Interpretability Hypothesis maintain that adult L2ers cannot acquire agreement in a native-like way when, as in the case of Chinese learners of English, doing so requires the acquisition of a novel function feature. Conversely, the Full Transfer/Full Access Hypothesis (FT/FA) claims that L2ers can acquire such features on the basis of exposure to relevant cues in the input, though difficulties may arise from transfer from the L1 grammar rather than default syntactic deficits in the language faculty. In terms of processing, while the Shallow Structure Hypothesis (SSH) holds that adult L2ers may have difficulty processing agreement due to reduced ability to parse structures, the Interference model argues that they can compute such structures, though they may have processing difficulties due to memory retrieval interference rather than shallow parsing. The judgement

results from Study 2 and Study 3 showed that, like the English L1ers, our Chinese L2ers of English could detect non-local agreement violations. In Study 2, both L1and L2ers made more incorrect judgements on sentences containing non-local agreement violations than grammatical ones. In Study 3, only L2ers made more incorrect judgements on ungrammatical sentences while L1ers did not show this difference. The online findings from Study 2 and Study 3 provided further evidence that the Chinese L2ers were sensitive to agreement violations during online reading. The ERP evidence from Study 2 demonstrated that, despite a smaller P600 effect in the Chinese L2ers, ERP responses to agreement violations were qualitatively similar across the groups. The reading time data from Study 3 also showed that both L1 and L2 groups showed longer reading times for sentences containing agreement violations and that there were no significant differences between the two groups. In summary, Study 2 and Study 3 do not support theories predicting shallow parsing or inability to acquire novel L2 features, as our Chinese L2ers detected agreement errors and could compute non-local agreement similarly to L1ers. Thus, the results lent support for full access to UG in adult L2ers as suggested by the FT/FA model. The quantitative differences indicated by the reduced P600 effect in Study 2 and lower judgement accuracy for ungrammatical sentences in Study 3 in the Chinese L2ers of English may be indicative of some influence from L1 transfer due to the fact that Chinese does not have agreement. These findings also seem compatible with the Interference model in relation to L2ers' increased sensitivity to interference from the intervening constituent. Therefore, the implications of these quantitative differences should be further investigated in future research.

Regarding processing of double number marking, there was no significant L1/L2 differences. Recall that Tanner & Bulkes (2015) predicted stronger responses/sensitivity to number agreement violations when number is marked more explicitly. Consistent with this account, our behavioural results in both Study 2 and Study 3 showed that judgement accuracy

for sentences containing non-local agreement violations increased in the condition of double number marking in both L1 and L2 groups. This effect was also attested in the ERP data of Study 2, with a larger P600 effect to agreement violations following double number marking in L1ers and Chinese L2ers of English. However, despite numerical trends, the online reading time data in Study 3 did not replicate this double number marking effect. Our L2 results are similar to Armstrong et al. (2018) in terms of L2ers' sensitivity to agreement violations, but different in terms of the double marking effect. Unlike Study 2 and Study 3 where this effect was found in the judgement tasks, Armstrong et al. did not show this effect in their behavioural data. While Study 2 showed a larger P600 effect for violations following double number marking, Armstrong et al. showed a smaller P600 effect under the same condition. As discussed previously, the different findings between our study and Armstrong et al may have been due to different stimuli (demonstratives vs. quantifiers) and structures (non-local agreement vs. local agreement) being tested.

Also, Study 2 and Study 3 differ in terms of the online findings in relation to the double number marking effect, which was significant in the ERP data in Study 2 but not the self-paced reading data in Study 3, even though they adopted the same stimuli. One possibility is that Study 2 used ERPs which typically require more trials compared to a self-paced reading task as used in Study 3. As a result, Study 2 has much more observations per participant than Study 3 (160 vs. 32), despite a slightly smaller sample size (64 participants vs. 80 participants). This may suggest that studies that adopt reading time measures could either enlarge their sample sizes or include more trials per condition to obtain more total observations for capturing effects like double marking. Also, it could be that ERPs may be a more sensitive method to address such questions compared to a self-paced reading paradigm given its advantage of high temporal resolution in the measure of milliseconds and its direct and immediate measure of automatic brain responses to a given stimulus from its onset. However, although online effects of double

number marking were significant in the ERP data but not the self-paced reading data, the pattern of results, with increased sensitivity to violations with double number marking, was numerically in the same direction in Study 2 and Study 3. Further research is required here to examine potential differences between self-paced reading and ERPs.

Furthermore, both Study 2 and Study 3 showed that felicity may have influenced judgement accuracy differences between grammatical sentences with and without demonstratives, with those containing a demonstrative being judged as more ungrammatical or unacceptable. One may argue that felicity could also have led to the differences between ungrammatical sentences with and without demonstratives in the judgement tasks in Study 2 and Study 3 and the online double number marking effect in Study 2. However, if that was the case, we would have also observed some differences in the ERP responses between grammatical sentences with and without demonstratives, which is absent in the findings. Thus, we argue that the observed differences between ungrammatical sentences, at least in the ERP data, do not index any influence from felicity but double number marking. Regardless of this issue, the L1 and L2ers were not significantly different in this regard, suggesting double number marking was processed similarly across the groups.

In summary, the judgement data and online findings from Study 2 and Study 3 indicate that, Chinese L2ers of English can acquire and process non-local agreement, despite the fact that agreement is not a realised feature in their L1 Chinese.

5.3 Limitations and future directions

While this thesis investigated similarities and differences between L1 and L2 processing on various linguistic aspects, there are some limitations that could be addressed in future research. Firstly, the fact that the object-modifying RCs in Study 1 contained either perception (e.g., "We liked the brother of the man...") or non-perception matrix verbs (e.g., "We met the brother

of the man...") could have affected our results as previous research has shown that the use of perception verbs may lead to a high attachment bias in attachment resolution (e.g., Grillo, Costa, Fernandes, & Santi, 2015; Rohde, Levy, & Kehler, 2011). Thus, future research should either manipulate perception and non-perception verbs as a variable or avoid this issue by only using verbs that do not elicit attachment biases.

Also, while Study 1 examined English RC attachment preferences in L2ers of English from various L1 backgrounds that are reported to prefer high attachment, the linguistic subgroups differed in terms the number of participants and perhaps the strength of attachment biases in the L1. This heterogeneity within our L2 group may have differentially influenced our findings. Therefore, future research will need to investigate whether our results can be replicated with a more homogenous L2 group.

In addition, future research should include a group of L2ers with a high attachment preference in their L1 and directly compare them to L2ers with an L1 with a low attachment preference, in addition to L1ers of English within the same study using the design and materials of Study 1. With this design, one can more stringently test whether L1 transfer plays a role in L2 attachment resolution.

While Study 2 and Study 3 showed some quantitative differences between English L1ers and Chinese L2ers of English in terms of agreement processing, whether these differences relate to L1 transfer or a general L2 processing issue (e.g., retrieval interference) is unknown. Future research should address this question by testing an L2 group with an L1 that has agreement and double number marking (e.g., Spanish L2ers of English) with the design and materials of Study 2 and Study 3. If such an L2 group also shows similar processing profiles to those of our Chinese L2ers of English, it may suggest that those L1/L2 differences arise from retrieval interference, a general issue in L2 processing. Alternatively, if this L2

group does not behave any differently to L1 group, it may indicate that L1 transfer plays a role in L1/L2 differences.

Study 2 and Study 3 filled a gap in the literature by examining the effect of double number marking on processing of non-local agreement. The L2 findings in this respect were inconsistent to the account in Armstrong et al. (2018), who had addressed the same research question in the contexts of local agreement. Nevertheless, whether these different findings are attributed to different structures tested in these studies (local versus non-local agreement) is not clear and thus, should be examined in future research, using a within-subjects design to directly compare double number marking effects in local and non-local agreement. Also, as discussed previously, different stimuli between our studies and Armstrong et al. could have led to the contradicting findings. While Armstrong et al. adopted quantifiers, Study 2 and Study 3 used demonstratives. Thus, their account on how Chinese L2ers of English process double marking in the case of local agreement should be further tested using demonstratives that also exist in Chinese.

While Study 2 and Study 3 investigated how double number marking influences non-local agreement processing, the effect has been only examined in cases where double number marking occurs on the subject NP (e.g., Those windows of the house...). Future research should also examine the effect in contexts where double number marking occurs on the intervening NP, such as “The park behind those houses were very busy”. Addressing such a question could inform us about whether different positions of double number marking would lead to different effects.

In addition, despite the numerical trends in the reading time data, Study 3 did not replicate the online ERP findings in Study 2 in relation to the double number marking effect, which may have something to do with issues related to sample size or the number of

observations per participant. Thus, testing with a larger sample size or more experimental items in future research may help to replicate such effects in reading time experiments.

5.4 Conclusion

This thesis aimed to adjudicate between competing acquisition and processing theories by examining the potential for similarities and differences between L1 and L2 sentence processing in terms of attachment resolution and agreement computation. The three studies reported in this thesis provided evidence from various linguistic phenomena for native-like L2 sentence processing, despite some quantitative differences between L1 and L2ers. Thus, the findings do not suggest shallow parsing in L2ers but lend support to the L2 processing theories arguing for only quantitative L1 and L2 differences, such as the Interference model and the Lexical Bottleneck Hypothesis. Study 2 and Study 3 further informed the theoretical debate on L2 acquisition by showing that L2ers can acquire grammatical features not instantiated in their L1. Therefore, both studies provided evidence against the FFFH and the Interpretability Hypothesis but supported the FT/FA. Taken together, the findings from this thesis favour theories that suggest no fundamental L1/L2 differences between L1 and L2 acquisition and processing.

References

Ågren, M., Michot, M. E., Granget, C., Gerolimich, S., Hadermann, P., & Stabarin, I. (2021). "Les copains* dit au revoir": On Subject–Verb Agreement in L2 French and Cross-Linguistic Influence. *Languages*, *6*(1), 7.

Alemán Bañón, J., Fiorentino, R., & Gabriele, A. (2014). Morphosyntactic processing in advanced second language (L2) learners: An event-related potential investigation of the effects of L1–L2 similarity and structural distance. *Second Language Research*, *30*(3), 275-306.

Alemán Bañón, J., Miller, D. & Rothman, J. (2017). Morphological Variability in Second Language Learners: An Examination of Electrophysiological and Production Data. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *43*(10), 1509-1536, doi: 10.1037/xlm0000394

Alemán Bañón, J., & Rothman, J. (2019). Being a participant matters: Event-related potentials show that markedness modulates person agreement in Spanish. *Frontiers in psychology*, *10*, 746.

Armstrong, A., Bulkes, N., & Tanner, D. (2018). Quantificational cues modulate the processing of English subject-verb agreement by native Chinese speakers: An ERP study. *Studies in Second Language Acquisition*, *40*(4), 731-754.

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.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*, 328.

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.

Bidaoui, A., Foote, R., & Abunasser, M. (2016). Relative Clause Attachment in Native and L2 Arabic. *International Journal of Arabic Linguistics*, *2*(2), 75-95.

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning. *Linguistic perspectives on second language acquisition*, *4*, 1-68.

Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive psychology*, *23*(1), 45-93.

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition*, *136*, 135-149.

Brysbaert, M., & Mitchell, D. C. (1996). Modifier attachment in sentence parsing: Evidence from Dutch. *The Quarterly Journal of Experimental Psychology Section A*, *49*(3), 664-695.

Carreiras, M., & Clifton, C. (1999). Another word on parsing relative clauses: Eyetracking evidence from Spanish and English. *Memory & Cognition*, *27*(5), 826-833.

Carreiras, M., Salillas, E., & Barber, H. (2004). Event-related potentials elicited during parsing of ambiguous relative clauses in Spanish. *Cognitive Brain Research*, *20*(1), 98-105.

Carstens, V. (2000). Concord in minimalist theory. *Linguistic inquiry*, *31*(2), 319-355.

Chen, L., Shu, H. U. A., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject–verb agreement in L2 learning. *Bilingualism: Language and Cognition*, *10*(2), 161-174.

Chernova, D., & Chernigovskaya, T. V. (2015). Syntactic Ambiguity Resolution in Sentence Processing: New Evidence from a Morphologically Rich Language. In *EAPCogSci*.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*(1), 3-42.

Clahsen, H., & Felser, C. (2018). Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition*, *40*(3), 693-706.

Clifton, C., Frazier, L., & Deevy, P. (1999). Feature manipulation in sentence comprehension: 2703. *Italian journal of linguistics*, *11*(1), 11-40.

Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition*, *30*(1), 73-105.

Cunnings, I. (2017). Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition*, *20*(4), 659-678.

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., Donaldson, B., Edmonds, A. C., Fultz, A. L., & Petrush, R. A. (2008). Syntactic and prosodic computations in the resolution of relative clause attachment ambiguity by English-French learners. *Studies in Second Language Acquisition*, *30*(4), 453-480.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature neuroscience*, *8*(8), 1117-1121.

Desmet, T., Brysbaert, M., De Baecke, C. (2002). The correspondence between sentence production and corpus frequencies in modifier attachment. *The Quarterly Journal of Experimental Psychology: Section A*, *55*(3), 879-896.

Desmet, T., De Baecke, C., Drieghe, D., Brysbaert, M., & Vonk, W. (2006). Relative clause attachment in Dutch: On-line comprehension corresponds to corpus frequencies when lexical variables are taken into account. *Language and Cognitive Processes*, *21*(4), 453-485.

De Vincenzi, M., & Job, R. (1993). Some observations on the universality of the late-closure strategy. *Journal of Psycholinguistic Research*, *22*(2), 189-206.

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.

Dinçtopal-Deniz, N. (2010). Relative clause attachment preferences of Turkish L2 speakers of English. *Research in Second Language Processing and Parsing*, *53*, 27.

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

Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making syntax of sense: number agreement in sentence production. *Psychological review*, *112*(3), 531.

Felser, C., Marinis, T., & Clahsen, H. (2003a). Children's processing of ambiguous sentences: A study of relative clause attachment. *Language Acquisition*, *11*(3), 127-163.

Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003b). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, *24*(3), 453-489.

Fernández, E. M. (2003). *Bilingual Sentence Processing: Relative clause attachment in English and Spanish*. Amsterdam: Benjamins.

Fodor, J. D. (2002). Prosodic disambiguation in silent reading. *North East Linguistic Society*, *32*, 113–132.

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies.* Unpublished doctoral dissertation, University of Connecticut.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 559–586). Hillsdale, NJ: Erlbaum.

Frenck-Mestre, C., Osterhout, L., McLaughlin, J., & Foucart, A. (2008). The effect of phonological realization of inflectional morphology on verbal agreement in French: Evidence from ERPs. *Acta Psychologica*, *128*(3), 528-536.

Gibson, E., Pearlmutter, N., Canseco-González, E., & Hickok, G. (1996). Cross-linguistic attachment preferences: Evidence from English and Spanish. *Cognition*, *59*, 23–59.

Grillo, N., Costa, J., Fernandes, B., & Santi, A. (2015). Highs and lows in english attachment. *Cognition*, *144*, 116-122.

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

Grüter, T., Rohde, H., & Schafer, A. J. (2014). The role of discourse-level expectations in non-native speakers' referential choices. In Orman, W. & Valleau, M. J. (Eds.), *Proceedings of the 38th Boston University Conference on Language Development* (pp.179-191). Somerville, MA: Cascadilla Press.

Guo, J., Guo, T., Yan, Y., Jiang, N., & Peng, D. (2009). ERP evidence for different strategies employed by native speakers and L2 learners in sentence processing. *Journal of Neurolinguistics*, *22*(2), 123-134.

Hammerly, C., Staub, A., & Dillon, B. (2019). The grammaticality asymmetry in agreement attraction reflects response bias: Experimental and modeling evidence. *Cognitive psychology*, *110*, 70-104.

Hawkins, R., & Casillas, G. (2008). Explaining frequency of verb morphology in early L2 speech. *Lingua*, *118*, 595–612.

Hawkins, R., & Chan, C. Y. H. (1997). The partial availability of Universal Grammar in second language acquisition: The 'failed functional features hypothesis'. *Second language research*, *13*(3), 187-226.

Hawkins, R., & Liszka, S. (2003). Locating the source of defective past tense marking in advanced L2 English speakers. *Language Acquisition and Language Disorders*, *30*, 21-44.

Hemforth, B., Fernandez, S., Clifton Jr, C., Frazier, L., Konieczny, L., & Walter, M. (2015). Relative clause attachment in German, English, Spanish and French: Effects of position and length. *Lingua*, *166*, 43-64.

Hemforth, B., Konieczny, L., Seelig, H., & Walter, M. (2000). Case matching and relative clause attachment. *Journal of Psycholinguistic Research*, *29*(1), 81-88.

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, *22*(3), 369-397.

Hopp, H. (2014). Working memory effects in the L2 processing of ambiguous relative clauses. *Language Acquisition*, *21*(3), 250-278.

Hopp, H. (2016). The timing of lexical and syntactic processes in second language sentence comprehension. *Applied Psycholinguistics*, *37*(5), 1253-1280.

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.

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.

Jäger, L. A., Mertzen, D., Van Dyke, J. A., & Vasishth, S. (2020). Interference patterns in subject-verb agreement and reflexives revisited: A large-sample study. *Journal of Memory and Language*, *111*, 104063.

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

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

Jiang, N., Novokshanova, E., Masuda, K., & Wang, X. (2011). Morphological congruency and the acquisition of L2 morphemes. *Language Learning*, *61*(3), 940-967.

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different?. *Linguistic Approaches to Bilingualism*, *4*(2), 257-282.

Keating, G. D. (2010). The effects of linear distance and working memory on the processing of gender agreement in Spanish. *Research in second language processing and parsing*, *53*, 113-134.

Kim, J. H., & Christianson, K. (2013). Sentence complexity and working memory effects in ambiguity resolution. *Journal of Psycholinguistic Research*, *42*(5), 393-411.

Kim, J. H., & Christianson, K. (2017). 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.

Lago, S., & Felser, C. (2018). Agreement attraction in native and non-native speakers of German. *Applied Psycholinguistics*, *39* (1), 619-647.

Lardiere, D. (1998a). Case and tense in the "fossilized" steady state. Second Language Research, 14, 1–26.

Lardiere, D. (1998b). Dissociating syntax from morphology in a divergent L2 end-state grammar. Second Language Research, 14, 359–375.

Lardiere, D. (2007). Ultimate attainment in second language acquisition: A case study. Mahwah, NJ: Erlbaum.

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second language research*, *25*(2), 173-227.

Lempert, H. (2016). Late L2ers can acquire grammatical features that do not occur in their L1: Evidence from the effect of animacy on verb agreement in L1 Chinese. *Memory & cognition*, *44*(4), 538-553.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive science*, *29*(3), 375-419.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in cognitive sciences*, *10*(10), 447-454.

Lim, J. H., & Christianson, K. (2015). Second language sensitivity to agreement errors: Evidence from eye movements during comprehension and translation. *Applied Psycholinguistics*, *36*(6), 1283-1315.

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, *9*(2), 157-201.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive psychology*, *24*(1), 56-98.

Maia, M., Costa, A., Fernández, E., & Lourenço-Gomes, M. D. C. (2004). A compreensão de orações relativas ambíguas em Português Brasileiro e Europeu: Um estudo comparativo. *Revista ABRALIN*, *3*(1), 2.

Maia, M., Fernández, E. M., Costa, A., & Lourenço-Gomes, M. D. C. (2007). Early and late preferences in relative clause attachment in Portuguese and Spanish. *Journal of Portuguese Linguistics*, *5*(6), 227-250.

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.

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language learning*, *60*(1), 154-193.

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of cognitive neuroscience*, *24*(4), 933-947.

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.

Omaki, A. (2005). *Working memory and relative clause attachment in first and second language processing* (Doctoral dissertation).

Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and language*, *59*(3), 494-522.

Osterhout, L., McKinnon, R., Bersick, M., & Corey, V. (1996). On the language specificity of the brain response to syntactic anomalies: Is the syntactic positive shift a member of the P300 family?. *Journal of Cognitive Neuroscience*, *8*(6), 507-526.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and language*, *34*(6), 739-773.

Oxford University Press (2004). *Quick Placement Test: Version 1*. Oxford: OUP.

Pan, H. Y., & Felser, C. (2011). Referential context effects in L2 ambiguity resolution: Evidence from self-paced reading. *Lingua*, *121*(2), 221-236.

Pan, H. Y., Schimke, S., & Felser, C. (2015). Referential context effects in non-native relative clause ambiguity resolution. *International Journal of Bilingualism*, *19*(3), 298-313.

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition*, *25*(4), 501-528.

Payne, B. R., Grison, S., Gao, X., Christianson, K., Morrow, D. G., & Stine-Morrow, E. A. (2014). Aging and individual differences in binding during sentence understanding: Evidence from temporary and global syntactic attachment ambiguities. *Cognition*, *130*(2), 157-173.

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and language*, *41*(3), 427-456.

Phillips, C., Wagers, M. W., & Lau, E. F. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. *Experiments at the Interfaces*, *37*, 147-180.

Pliatsikas, C., & Marinis, T. (2013). Processing empty categories in a second language: When naturalistic exposure fills the (intermediate) gap. *Bilingualism: Language and Cognition*, *16*(1), 167-182.

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.

Rohde, H., Levy, R., & Kehler, A. (2011). Anticipating explanations in relative clause processing. *Cognition*, *118*(3), 339-358.

Rothman, J. (2010). On the typological economy of syntactic transfer: Word order and relative clause high/low attachment preference in L3 Brazilian Portuguese. *International Review of Applied Linguistics in Language Teaching*, *48*(2-3), 245-273.

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.

Scheepers, C., Sturt, P., Martin, C. J., Myachykov, A., Teevan, K., & Viskupova, I. (2011). Structural priming across cognitive domains: From simple arithmetic to relative-clause attachment. *Psychological Science*, *22*(10), 1319-1326.

Shen, E. Y., Staub, A., & Sanders, L. D. (2013). Event-related brain potential evidence that local nouns affect subject–verb agreement processing. *Language and Cognitive Processes*, *28*(4), 498-524.

Spychalska, M., Kontinen, J., & Werning, M. (2016). Investigating scalar implicatures in a truth-value judgement task: Evidence from event-related brain potentials. *Language, Cognition and Neuroscience*, *31*(6), 817-840.

Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*, *36*(1), 201-216.

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.

Tabor, W., & Tanenhaus, M. K. (1999). Dynamical models of sentence processing. *Cognitive Science*, *23*(4), 491-515.

Tanner, D. (2011). *Agreement mechanisms in native and nonnative language processing: Electrophysiological correlates of complexity and interference*. University of Washington.

Tanner, D. (2019). Robust neurocognitive individual differences in grammatical agreement processing: A latent variable approach. *Cortex*, *111*, 210-237.

Tanner, D., & Bulkes, N. Z. (2015). Cues, quantification, and agreement in language comprehension. *Psychonomic bulletin & review*, *22*(6), 1753-1763.

Tanner, D., Goldshtein, M., & Weissman, B. (2018). Individual differences in the real-time neural dynamics of language comprehension. In *Psychology of learning and motivation* (Vol. 68, pp. 299-335). Academic Press.

Tanner, D., Grey, S., & van Hell, J. G. (2017). Dissociating retrieval interference and reanalysis in the P600 during sentence comprehension. *Psychophysiology*, *54*(2), 248-259.

Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition*, *17*(2), 277-293.

Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition*, *16*(2), 367-382.

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.

Tanner, D., Nicol, J., Herschensohn, J., & Osterhout, L. (2012). Electrophysiological markers of interference and structural facilitation in native and nonnative agreement processing. In *Proceedings of the 36th Boston university conference on language development* (pp. 594-606). Cascadilla Somerville, MA.

Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia*, *56*, 289-301.

Taraban, R., & McClelland, J. L. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of memory and language*, *27*(6), 597-632.

Tily, H., Fedorenko, E., & Gibson, E. (2010). The time-course of lexical and structural processes in sentence comprehension. *The Quarterly Journal of Experimental Psychology*, *63*(5), 910-927.

Traxler, M. J. (2007). Working memory contributions to relative clause attachment processing: A hierarchical linear modeling analysis. *Memory & Cognition*, *35*(5), 1107-1121.

Traxler, M. J., Pickering, M. J., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, *39*(4), 558-592.

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.

Tsimpli, I. M., & Dimitrakopoulou, M. (2007). The interpretability hypothesis: Evidence from wh-interrogatives in second language acquisition. *Second Language Research*, *23*, 215–242.

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 407.

Van Dyke, J. A., & Lewis, R. L. (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.

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of memory and language*, *65*(3), 247-263.

Van Gompel, R. P., Pickering, M. J., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*, *52*(2), 284-307.

Van Gompel, R. P., Pickering, M. J., & Traxler, M. J. (2000). Unrestricted race: A new model of syntactic ambiguity resolution. In A. Kennedy (Ed.), *Reading as a Perceptual Process* (pp. 621-648). Amsterdam, The Netherlands: Elsevier.

Vasishth, S. & Nicenboim, B. (2016). Statistical methods for linguistic research: Foundational Ideas – Part I. *Language and Linguistics Compass*, *10*, 349-369.

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.

White, L., Valenzuela, E., Kozlowska-Macgregor, M., & Leung, Y. -. I. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics*, *25*(1), 105-133.

Witzel, J., Witzel, N., & Nicol, J. (2012). Deeper than shallow: Evidence for structure-based parsing biases in second-language sentence processing. *Applied Psycholinguistics*, *33*(2), 419-456.

Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *NeuroImage*, *62*(1), 356-366.

Zagar, D., Pynte, J., & Rativeau, S. (1997). Evidence for early closure attachment on first pass reading times in French. *The Quarterly Journal of Experimental Psychology Section A*, *50*(2), 421-438.

Appendices for Study 1

Appendix A

Below are the experimental items from the offline task. The subject and object modifying relative clause conditions are shown respectively, delimited by a /.

1. The brother of the man who bought himself some books got married yesterday. / We knew the brother of the man who bought himself some books.

2. The nephew of the man who embarrassed himself at the party studied psychology. / Someone greeted the nephew of the man who embarrassed himself at the party.

3. The brother of the boy who amused himself by watching TV was popular at the school. / Everyone liked the brother of the boy who amused himself by watching TV.

4. The father of the young man who injured himself by accident was a professor. / Somebody met the father of the young man who injured himself by accident.

5. The older brother of the man who made himself happy by travelling was very friendly. / Someone disliked the older brother of the man who made himself happy by travelling.

6. The older brother of the man who cooked himself some tasty food was a famous writer. / Someone met the older brother of the man who cooked himself some tasty food.

7. The brother of the man who educated himself by visiting the museum divorced yesterday. / We knew the brother of the man who educated himself by visiting the museum.

8. The stepfather of the man who treated himself to some dessert was very polite. / Someone greeted the stepfather of the man who treated himself to some dessert.

9. The uncle of the man who hurt himself by accident was very rich. / Everybody liked the uncle of the man who hurt himself by accident.

10. The stepbrother of the man who kept himself happy by cooking owned a company. / Someone met the stepbrother of the man who kept himself happy by cooking.

11. The sister of the lady who taught herself a foreign language bought a big house. / We met the sister of the lady who taught herself a foreign language.

12. The mother of the woman who enjoyed herself at dinner was a nurse. / Everybody knew the mother of the woman who enjoyed herself at dinner.

13. The sister of the lady who bought herself some nice shoes worked at a local school. / We visited the sister of the lady who bought herself some nice shoes.

14. The stepsister of the girl who stopped herself eating too much used to live here. / Someone liked the stepsister of the girl who stopped herself eating too much.

15. The sister of the woman who entertained herself by reading lived in New York. / Someone knew the sister of the woman who entertained herself by reading.

16. The grandmother of the girl who helped herself to some pie was very chatty. / We greeted the grandmother of the girl who helped herself to some pie.

17. The aunt of the lady who amused herself by watching comedies had a big family. / We visited the aunt of the lady who amused herself by watching comedies.

18. The sister of the lady who prepared herself for the conference was allergic to nuts. / We knew the sister of the lady who prepared herself for the conference.

19. The stepsister of the girl who enjoyed herself at the library had an accident yesterday. / Somebody knew the stepsister of the girl who enjoyed herself at the library.

20. The aunt of the woman who served herself some cake was teaching at university. / Everybody greeted the aunt of the woman who served herself some cake.

Appendix B

Below are the experimental items from the eye-movement experiment. The subject and object modifying relative clause conditions are shown respectively, delimited by a /. Gender manipulations are shown in parenthesis.

1. The brother (sister) of the man (woman) who accidentally hurt himself yesterday afternoon lived in town. / We saw the brother (sister) of the man (woman) who accidentally hurt himself yesterday afternoon. Luckily it wasn't serious in the end.

2. The brother (sister) of the boy (girl) who badly injured himself in the garden was popular in school. / Everyone knew the brother (sister) of the boy (girl) who badly injured himself in the garden. An ambulance was on its way.

3. The son (daughter) of the old man (woman) who happily bought himself a nice gift was a singer. / Everybody liked the son (daughter) of the old man (woman) who happily bought himself a nice gift. It was a lovely day.

4. The grandfather (grandmother) of the boy (girl) who gladly helped himself to some food retired last year. / We greeted the grandfather (grandmother) of the boy (girl) who gladly helped himself to some food. Everyone was in a good mood.

5. The brother (sister) of the man (woman) who politely introduced himself at the party had some drink. / We saw the brother (sister) of the man (woman) who politely introduced himself at the party. It was very boring.

6. The stepbrother (stepsister) of the man (woman) who happily entertained himself in the pub worked in finance. / We knew the stepbrother (stepsister) of the man (woman) who happily entertained himself in the pub. There were always lots of people at the bar.

7. The younger brother (sister) of the boy (girl) who often taught himself some English words looked very happy. / We saw the younger brother (sister) of the boy (girl) who often taught himself some English words. Their parents educated them very well.

8. The brother (sister) of the man (woman) who greatly enjoyed himself at the concert was a nice person. / Nobody knew the brother (sister) of the man (woman) who greatly enjoyed himself at the concert. It was a fun evening.

9. The older brother (sister) of the little boy (girl) who always hated himself for lying lived in London. / We phoned the older brother (sister) of the little boy (girl) who always hated himself for lying. It was very awkward.

10. The brother (sister) of the man (woman) who busily prepared himself for a meeting was very kind. / We knew the brother (sister) of the man (woman) who busily prepared himself for a meeting. The meeting was in the afternoon.

11. The son (daughter) of the king (queen) who usually dressed himself very nicely walked in the forest. / Someone saw the son (daughter) of the king (queen) who usually dressed himself very nicely. It was a nice day.

12. The stepbrother (stepsister) of the man (woman) who finally stopped himself from eating the cake had a nice house. / Everyone recognised the stepbrother (stepsister) of the man (woman) who finally stopped himself from eating the cake. The party was awesome.

13. The brother (sister) of the man (woman) who always cooked himself tasty food watched a movie. / We greeted the brother (sister) of the man (woman) who always cooked himself tasty food. It was a lovely evening.

14. The brother (sister) of the boy (girl) who often entertained himself by doing sports saw the teacher. / Everyone knew the brother (sister) of the boy (girl) who often entertained himself by doing sports. The school year had just started.

15. The uncle (aunt) of the policeman (policewoman) who gladly served himself some delicious cake loved classical music. / Someone liked the uncle (aunt) of the policeman (policewoman) who gladly served himself some delicious cake. It was a lovely day.

16. The brother (sister) of the man (woman) who always kept himself in a good mood could speak five languages. / We knew the brother (sister) of the man (woman) who always kept himself in a good mood. They were nice people.

17. The stepbrother (stepsister) of the man (woman) who just found himself in big trouble lived by the river. / Someone met the stepbrother (stepsister) of the man (woman) who just found himself in big trouble. Luckily the problem was fixed.

18. The brother (sister) of the man (woman) who luckily saved himself from the car accident had a decent job. / We liked the brother (sister) of the man (woman) who luckily saved himself from the car accident. Everybody knew them.

19. The niece (nephew) of the woman (man) who accidentally embarrassed herself at dinner studied chemistry. / We met the niece (nephew) of the woman (man) who accidentally embarrassed herself at dinner. Everybody loved the food.

20. The sister (brother) of the girl (boy) who repeatedly blamed herself for the mistake liked reading. / We phoned the sister (brother) of the girl (boy) who repeatedly blamed herself for the mistake. It was just a small accident.

21. The aunt (uncle) of the saleswoman (salesman) who unexpectedly hurt herself in the street waited for help. / Nobody knew the aunt (uncle) of the saleswoman (salesman) who unexpectedly hurt herself in the street. There had been a few accidents recently.

22. The mother (father) of the lady (man) who severely criticized herself for being rude sat on the sofa. / Everyone knew the mother (father) of the lady (man) who severely criticized herself for being rude. They looked so upset.

23. The older sister (brother) of the girl (boy) who always told herself to be happy knew history very well. / We saw the older sister (brother) of the girl (boy) who always told herself to be happy. They lived a wonderful life.

24. The sister (brother) of the woman (man) who always educated herself by reading books lived in a flat. / Someone greeted the sister (brother) of the woman (man) who always educated herself by reading books. Reading is a good way to pass the time.

25. The aunt (uncle) of the lady (man) who quietly amused herself in the garden had three sons. / We knew the aunt (uncle) of the lady (man) who quietly amused herself in the garden. It was a lovely afternoon.

26. The sister (brother) of the young girl (boy) who badly cut herself in the kitchen had a good education. / We recognised the sister (brother) of the young girl (boy) who badly cut herself in the kitchen. Luckily the cut soon healed.

27. The daughter (son) of the woman (man) who loudly explained herself to the court wore a black hat. / We met the daughter (son) of the woman (man) who loudly explained herself to the court. Nobody knew the truth.

28. The sister (brother) of the woman (man) who recently bought herself a new car was good at cooking. / We liked the sister (brother) of the woman (man) who recently bought herself a new car. The car was quite expensive.

29. The stepsister (stepbrother) of the lady (man) who stupidly shot herself in the forest knew everyone. / Someone knew the stepsister (stepbrother) of the lady (man) who stupidly shot herself in the forest. We were worried and nervous.

30. The sister (brother) of the girl (boy) who badly burnt herself by accident had a flat in London. / Somebody encountered the sister (brother) of the girl (boy) who badly burnt herself by accident. The injury was quite serious.

31. The stepsister (stepbrother) of the girl (boy) who always forced herself to keep fit moved to New York. / We liked the stepsister (stepbrother) of the girl (boy) who always forced herself to keep fit. Everyone knew it.

32. The sister (brother) of the woman (man) who recently devoted herself to a new job was really tall. / Everybody knew the sister (brother) of the woman (man) who recently devoted herself to a new job. They were very nice people.

33. The sister (brother) of the girl (boy) who always kept herself in good shape graduated yesterday. / We knew the sister (brother) of the girl (boy) who always kept herself in good shape. They were moving to New York soon.

34. The aunt (uncle) of the lady (man) who constantly worried herself about the weather was very rich. / Everyone liked the aunt (uncle) of the lady (man) who constantly worried herself about the weather. It was freezing outside.

35. The sister (brother) of the lady (man) who really enjoyed herself at lunch got married last year. / Nobody disliked the sister (brother) of the lady (man) who really enjoyed herself at lunch. Their lives were getting better.

36. The stepsister (stepbrother) of the girl (boy) who quickly calmed herself after the accident was born here. / Someone saw the stepsister (stepbrother) of the girl (boy) who quickly calmed herself after the accident. It was pretty surprising.

Appendices for Study 2

Appendix A

Below are the experimental items from the EEG task. The grammatical and ungrammatical conditions are shown respectively, delimited by a /. Determiner-number specification manipulations are shown in parenthesis.

1. The (This) window of the house is very clean. / The (These) windows of the house is very clean.

2. The (This) picture of the lake is so beautiful. / The (These) pictures of the lake is so beautiful.

3. The (This) email from the company is really boring. / The (These) emails from the company is really boring.

4. The (This) cup on the table is very dirty. / The (These) cups on the table is very dirty.

5. The (This) gift from the party is so nice. / The (These) gifts from the party is so nice.

6. The (This) book on the desk is very heavy. / The (These) books on the desk is very heavy.

7. The (This) film about the scientist is quite long. / The (These) films about the scientist is quite long.

8. The (This) colleague of the lady is so lazy. / The (These) colleagues of the lady is so lazy.

9. The (This) patient of the doctor is very angry. / The (These) patients of the doctor is very angry.

10. The (This) apple on the table is very sweet. / The (These) apples on the table is very sweet.

11. The (This) train to the city is quite fast. / The (These) trains to the city is quite fast.

12. The (This) name on the postcard is not clear. / The (These) names on the postcard is not clear.

13. The (This) mistake in the article is so obvious. / The (These) mistakes in the article is so obvious.

14. The (This) picture on the wall is very funny. / The (These) pictures on the wall is very funny.

15. The (This) problem in the school is extremely serious. / The (These) problems in the school is extremely serious.

16. The (This) photo from the trip is very nice. / The (These) photos from the trip is very nice.

17. The (This) building in the street is quite old. / The (These) buildings in the street is quite old.

18. The (This) road in the mountain is not safe. / The (These) roads in the mountain is not safe.

19. The (This) door of the building is always open. / The (These) doors of the building is always open.

20. The (This) notebook on the desk is quite new. / The (These) notebooks on the desk is quite new.

21. The (This) key to the room is very big. / The (These) keys to the room is very big.

22. The (This) guitar for the concert is quite old. / The (These) guitars for the concert is quite old.

23. The (This) entrance to the building is not obvious. / The (These) entrances to the building is not obvious.

24. The (This) student of the teacher is really smart. / The (These) students of the teacher is really smart.

25. The (This) bridge to the island is not safe. / The (These) bridges to the island is not safe.

26. The (This) guy with the actor is very rich. / The (These) guys with the actor is very rich.

27. The (This) lawyer from the company is very professional. / The (These) lawyers from the company is very professional.

28. The (This) desk in the office is really small. / The (These) desks in the office is really small.

29. The (This) bridge over the river is so old. / The (These) bridges over the river is so old.

30. The (This) map of the city is very detailed. / The (These) maps of the city is very detailed.

31. The (This) window in the kitchen is always closed. / The (These) windows in the kitchen is always closed.

32. The (This) jacket on the chair is very dirty. / The (These) jackets on the chair is very dirty.

33. The (This) exam for the course is really difficult. / The (These) exams for the course is really difficult.

34. The (This) knife on the plate is not sharp. / The (These) knives on the plate is not sharp.

35. The (This) store in the street is so busy. / The (These) stores in the street is so busy.

36. The (This) computer in the office is very new. / The (These) computers in the office is very new.

37. The (This) waiter with the manager is very nice. / The (These) waiters with the manager is very nice.

38. The (This) story about the city is available online. / The (These) stories about the city is available online.

39. The (This) student in the class is very quiet. / The (These) students in the class is very quiet.

40. The (This) movie at the cinema is really interesting. / The (These) movies at the cinema is really interesting.

41. The (That) handbag in the shop is not cheap. / The (Those) handbags in the shop is not cheap.

42. The (That) hotel near the station is really busy. / The (Those) hotels near the station is really busy.

43. The (That) friend of the girl is very helpful. / The (Those) friends of the girl is very helpful.

44. The (That) answer to the question is really funny. / The (Those) answers to the question is really funny.

45. The (That) song by the singer is so beautiful. / The (Those) songs by the singer is so beautiful.

46. The (That) magazine on the sofa is very boring. / The (Those) magazines on the sofa is very boring.

47. The (That) customer of the designer is really rich. / The (Those) customers of the designer is really rich.

48. The (That) train to the airport is so busy. / The (Those) trains to the airport is so busy.

49. The (That) river near the village is quite clear. / The (Those) rivers near the village is quite clear.

50. The (That) document in the folder is extremely important. / The (Those) documents in the folder is extremely important.

51. The (That) café outside the mall is very popular. / The (Those) cafés outside the mall is very popular.

52. The (That) report about the conference is really good. / The (Those) reports about the conference is really good.

53. The (That) comment about the policy is quite stupid. / The (Those) comments about the policy is quite stupid.

54. The (That) problem with the plan is really obvious. / The (Those) problems with the plan is really obvious.

55. The (That) email about the meeting is not clear. / The (Those) emails about the meeting is not clear.

56. The (That) poster for the concert is very creative. / The (Those) posters for the concert is very creative.

57. The (That) wall of the flat is not clean. / The (Those) walls of the flat is not clean.

58. The (That) market at the festival is always busy. / The (Those) markets at the festival is always busy.

59. The (That) story in the book is very interesting. / The (Those) stories in the book is very interesting.

60. The (That) airport outside the city is quite big. / The (Those) airports outside the city is quite big.

61. The (That) office in the school is usually busy. / The (Those) offices in the school is usually busy.

62. The (That) house near the park is really modern. / The (Those) houses near the park is really modern.

63. The (That) model with the designer is very famous. / The (Those) models with the designer is very famous.

64. The (That) dress for the party is so colourful. / The (Those) dresses for the party is so colourful.

65. The (That) orange on the tree is really small. / The (Those) oranges on the tree is really small.

66. The (That) book for the course is so good. / The (Those) books for the course is so good.

67. The (That) toy in the box is very dirty. / The (Those) toys in the box is very dirty.

68. The (That) monkey behind the tourist is very cute. / The (Those) monkeys behind the tourist is very cute.

69. The (That) secretary of the manager is not polite. / The (Those) secretaries of the manager is not polite.

70. The (That) book on the shelf is quite old. / The (Those) books on the shelf is quite old.

71. The (That) library in the city is always busy. / The (Those) libraries in the city is always busy.

72. The (That) doctor of the patient is very young. / The (Those) doctors of the patient is very young.

73. The (That) bus to the school is really convenient. / The (Those) buses to the school is really convenient.

74. The (That) farm near the forest is so big. / The (Those) farms near the forest is so big.

75. The (That) task in the game is quite difficult. / The (Those) tasks in the game is quite difficult.

76. The (That) product of the company is still popular. / The (Those) products of the company is still popular.

77. The (That) lady behind the guy is talking loudly. / The (Those) ladies behind the guy is talking loudly.

78. The (That) dog behind the girl is very small. / The (Those) dogs behind the girl is very small.

79. The (That) kid with the volunteer is very happy. / The (Those) kids with the volunteer is very happy.

80. The (That) assistant of the scientist is really excellent. / The (Those) assistants of the scientist is really excellent.

81. The (This) boy with the teacher has been educated. / The (These) boys with the teacher has been educated.

82. The (This) girl with the boy has been invited. / The (These) girls with the boy has been invited.

83. The (This) plane in the airport has been upgraded. / The (These) planes in the airport has been upgraded.

84. The (This) thief with the policeman has been punished. / The (These) thieves with the policeman has been punished.

85. The (This) spoon on the table has been washed. / The (These) spoons on the table has been washed.

86. The (This) rule in the school has been changed. / The (These) rules in the school has been changed.

87. The (This) cinema in the town has been closed. / The (These) cinemas in the town has been closed.

88. The (This) restaurant by the sea has been reopened. / The (These) restaurants by the sea has been reopened.

89. The (This) seminar of the course has been cancelled. / The (These) seminars of the course has been cancelled.

90. The (This) nurse with the doctor has been helpful. / The (These) nurses with the doctor has been helpful.

91. The (This) requirement of the machine has been described. / The (These) requirements of the machine has been described.

92. The (This) dancer with the trainer has been invited. / The (These) dancers with the trainer has been invited.

93. The (This) bedroom of the flat has been decorated. / The (These) bedrooms of the flat has been decorated.

94. The (This) bike in the garden has been cleaned. / The (These) bikes in the garden has been cleaned.

95. The (This) boy with the teacher has been naughty. / The (These) boys with the teacher has been naughty.

96. The (This) umbrella in the office has been lost. / The (These) umbrellas in the office has been lost.

97. The (This) toilet for the office has been cleaned. / The (These) toilets for the office has been cleaned.

98. The (This) employee with the manager has been friendly. / The (These) employees with the manager has been friendly.

99. The (This) girl with the policeman has been quiet. / The (These) girls with the policeman has been quiet.

100. The (This) door of the house has been locked. / The (These) doors of the house has been locked.

101. The (This) hospital of the city has been excellent. / The (These) hospitals of the city has been excellent.

102. The (This) email about the conference has been read. / The (These) emails about the conference has been read.

103. The (This) baby of the lady has big eyes. / The (These) babies of the lady has big eyes.

104. The (This) wall in the house has been painted. / The (These) walls in the house has been painted.

105. The (This) museum in the town has many visitors. / The (These) museums in the town has many visitors.

106. The (This) report about the accident has some mistakes. / The (These) reports about the accident has some mistakes.

107. The (This) singer near the journalist has many fans. / The (These) singers near the journalist has many fans.

108. The (This) lesson in the textbook has been taught. / The (These) lessons in the textbook has been taught.

109. The (This) website about the singer has been updated. / The (These) websites about the singer has been updated.

110. The (This) reader of the writer has some questions. / The (These) readers of the writer has some questions.

111. The (This) article about the school has many mistakes. / The (These) articles about the school has many mistakes.

112. The (This) trip to the seaside has excited us. / The (These) trips to the seaside has excited us.

113. The (This) factory near the town has some trouble. / The (These) factories near the town has some trouble.

114. The (This) footballer in the team has no confidence. / The (These) footballers in the team has no confidence.

115. The (This) classroom for the exam has been cleaned. / The (These) classrooms for the exam has been cleaned.

116. The (This) room of the house has no windows. / The (These) rooms of the house has no windows.

117. The (This) presentation for the meeting has been prepared. / The (These) presentations for the meeting has been prepared.

118. The (This) book on the shelf has been read. / The (These) books on the shelf has been read.

119. The (This) class with the teacher has been rescheduled. / The (These) classes with the teacher has been rescheduled.

120. The (This) train to the seaside has many passengers. / The (These) trains to the seaside has many passengers.

121. The (That) flower behind the tree has been watered. / The (Those) flowers behind the tree has been watered.

122. The (That) mirror in the house has been wiped. / The (Those) mirrors in the house has been wiped.

123. The (That) interview with the officer has been arranged. / The (Those) interviews with the officer has been arranged.

124. The (That) gym near the mall has been popular. / The (Those) gyms near the mall has been popular.

125. The (That) cat of the neighbour has been fed. / The (Those) cats of the neighbour has been fed.

126. The (That) patient of the therapist has no complaints. / The (Those) patients of the therapist has no complaints.

127. The (That) hunter in the forest has saved us. / The (Those) hunters in the forest has saved us.

128. The (That) project about the research has already started. / The (Those) projects about the research has already started.

129. The (That) museum near the school has many visitors. / The (Those) museums near the school has many visitors.

130. The (That) painting on the wall has been sold. / The (Those) paintings on the wall has been sold.

131. The (That) book by the professor has been published. / The (Those) books by the professor has been published.

132. The (That) superstar with the policeman has been arrested. / The (Those) superstars with the policeman has been arrested.

133. The (That) flat behind the park has been sold. / The (Those) flats behind the park has been sold.

134. The (That) report about the research has some mistakes. / The (Those) reports about the research has some mistakes.

135. The (That) letter in the box has become dusty. / The (Those) letters in the box has become dusty.

136. The (That) mail in the box has been read. / The (Those) mails in the box has been read.

137. The (That) mountain near the river has no tourists. / The (Those) mountains near the river has no tourists.

138. The (That) message from the company has been deleted. / The (Those) messages from the company has been deleted.

139. The (That) oven in the kitchen has some problems. / The (Those) ovens in the kitchen has some problems.

140. The (That) letter for the company has been sent. / The (Those) letters for the company has been sent.

141. The (That) tree near the farm has been cut. / The (Those) trees near the farm has been cut.

142. The (That) hall near the university has been repainted. / The (Those) halls near the university has been repainted.

143. The (That) secretary of the manager has been fired. / The (Those) secretaries of the manager has been fired.

144. The (That) lecture on the course has been cancelled. / The (Those) lectures on the course has been cancelled.

145. The (That) shop in the town has no customers. / The (Those) shops in the town has no customers.

146. The (That) club near the church has been closed. / The (Those) clubs near the church has been closed.

147. The (That) bank in the town has good service. / . The (Those) banks in the town has good service.

148. The (That) piano in the shop has been sold. / The (Those) pianos in the shop has been sold.

149. The (That) file on the computer has been saved. / The (Those) files on the computer has been saved.

150. The (That) cow of the farmer has been fed. / The (Those) cows of the farmer has been fed.

151. The (That) sandwich in the fridge has been eaten. / The (Those) sandwiches in the fridge has been eaten.

152. The (That) soldier with the nurse has some injuries. / The (Those) soldiers with the nurse has some injuries.

153. The (That) movie about the war has been downloaded. / The (Those) movies about the war has been downloaded.

154. The (That) article on the website has many readers. / The (Those) articles on the website has many readers.

155. The (That) island near the beach has some animals. / The (Those) islands near the beach has some animals.

156. The (That) university near the city has good facilities. / The (Those) universities near the city has good facilities.

157. The (That) patient of the dentist has already left. / The (Those) patients of the dentist has already left.

158. The (That) essay for the course has been written. / The (Those) essays for the course has been written.

159. The (That) painting near the mirror has many colours. / The (Those) paintings near the mirror has many colours.

160. The (That) video about the sport has been uploaded. / The (Those) videos about the sport has been uploaded.

Appendix B

Below are the experimental items from the whole sentence grammaticality judgement task. The grammatical and ungrammatical conditions are shown respectively, delimited by a /. Determiner-number specification manipulations are shown in parenthesis.

1. The (That) officer with the man is very kind. / The (Those) officers with the man is very kind.

2. The (That) bag on the floor is really heavy. / The (Those) bags on the floor is really heavy.

3. The (That) printer in the office is convenient. / The (Those) printers in the office is convenient.

4. The (That) ticket for the show is expensive. / The (Those) tickets for the show is expensive.

5. The (That) boy in the park is happy. / The (Those) boys in the park is happy.

6. The (That) book on the shelf is very interesting. / The (Those) books on the shelf is very interesting.

7. The (That) experiment in the laboratory is really long. / The (Those) experiments in the laboratory is really long.

8. The (That) cup in the cupboard is new. / The (Those) cups in the cupboard is new.

9. The (That) monkey by the river is really fast. / The (Those) monkeys by the river is really fast.

10. The (That) baker in the kitchen is extremely lazy. / The (Those) bakers in the kitchen is extremely lazy.

11. The (That) girl on the sofa is beautiful. / The (Those) girls on the sofa is beautiful.

12. The (That) assistant of the manager is always helpful. / The (Those) assistants of the manager is always helpful.

13. The (That) project by the scientist has been finished. / The (Those) projects by the scientist has been finished.

14. The (That) room in the hotel has been cleaned. / The (Those) rooms in the hotel has been cleaned.

15. The (That) door in the shop has been damaged. / The (Those) doors in the shop has been damaged.

16. The (That) restaurant near the hotel has many customers. / The (Those) restaurants near the hotel has many customers.

17. The (That) bottle from the kitchen has been thrown away. / The (Those) bottles from the kitchen has been thrown away.

18. The (That) report about the accident has been submitted. / The (Those) reports about the accident has been submitted.

19. The (That) bike from the shop has been stolen. / The (Those) bikes from the shop has been stolen.

20. The (That) poster of the meeting has been made. The (Those) posters of the meeting has been made.

21. The (That) problem from the meeting has been solved. / The (Those) problems from the meeting has been solved.

22. The (That) towel in the bathroom has been washed. / The (Those) towels in the bathroom has been washed.

23. The (That) phone from the office has been stolen. / The (Those) phones from the office has been stolen.

24. The (That) letter to the school has been posted. / The (Those) letters to the school has been posted.

Appendices for Study 3

Appendix A

Below are the experimental items from the self-paced reading task. The grammatical and ungrammatical conditions are shown respectively, delimited by a /. Determiner-number specification manipulations are shown in parenthesis.

1. The (This) window of the house is very clean. / The (These) windows of the house is very clean.

2. The (This) picture of the lake is so beautiful. / The (These) pictures of the lake is so beautiful.

3. The (This) email from the company is really boring. / The (These) emails from the company is really boring.

4. The (This) cup on the table is very dirty. / The (These) cups on the table is very dirty.

5. The (This) gift from the party is so nice. / The (These) gifts from the party is so nice.

6. The (This) book on the desk is very heavy. / The (These) books on the desk is very heavy.

7. The (This) film about the scientist is quite long. / The (These) films about the scientist is quite long.

8. The (This) colleague of the lady is so lazy. / The (These) colleagues of the lady is so lazy.

9. The (That) handbag in the shop is not cheap. / The (Those) handbags in the shop is not cheap.

10. The (That) hotel near the station is really busy. / The (Those) hotels near the station is really busy.

11. The (That) friend of the girl is very helpful. / The (Those) friends of the girl is very helpful.

12. The (That) answer to the question is really funny. / The (Those) answers to the question is really funny.

13. The (That) song by the singer is so beautiful. / The (Those) songs by the singer is so beautiful.

14. The (That) magazine on the sofa is very boring. / The (Those) magazines on the sofa is very boring.

15. The (That) customer of the designer is really rich. / The (Those) customers of the designer is really rich.

16. The (That) train to the airport is so busy. / The (Those) trains to the airport is so busy.

17. The (This) bedroom of the flat has been decorated. / The (These) bedrooms of the flat has been decorated.

18. The (This) door of the house has been locked. / The (These) doors of the house has been locked.

19. The (This) plane in the airport has been upgraded. / The (These) planes in the airport has been upgraded.

20. The (This) spoon on the table has been washed. / The (These) spoons on the table has been washed.

21. The (This) rule in the school has been changed. / The (These) rules in the school has been changed.

22. The (This) cinema in the town has been closed. / The (These) cinemas in the town has been closed.

23. The (This) restaurant by the sea has been reopened. / The (These) restaurants by the sea has been reopened.

24. The (This) seminar of the course has been cancelled. / The (These) seminars of the course has been cancelled.

25. The (That) flower behind the tree has been watered. / The (Those) flowers behind the tree has been watered.

26. The (That) mirror in the house has been wiped. / The (Those) mirrors in the house has been wiped.

27. The (That) interview with the officer has been arranged. / The (Those) interviews with the officer has been arranged.

28. The (That) gym near the mall has been popular. / The (Those) gyms near the mall has been popular.

29. The (That) cat of the neighbour has been fed. / The (Those) cats of the neighbour has been fed.

30. The (That) hunter in the forest has saved us. / The (Those) hunters in the forest has saved us.

31. The (That) project about the research has already started. / The (Those) projects about the research has already started.

32. The (That) museum near the school has many visitors. / The (Those) museums near the school has many visitors.