University of Reading

# Reanalysis Processes in Native and Non-Native Language Comprehension

A thesis submitted in fulfilment of requirements for the degree of Doctor of Philosophy

**School of Psychology and Clinical Language Sciences**

**University of Reading**

Hiroki Fujita

**October 2019**

Abstract

Temporarily ambiguous sentences (e.g., When Mary dressed the baby laughed happily.) are known to cause comprehension difficulties, as initially assigned interpretations (Mary dressed the baby) need to be revised but are not always fully discarded from memory. The similarities and differences between native (L1) and non-native (L2) sentence processing have been widely debated, and many studies have examined L1 and L2 ambiguity resolution. How L2 speakers deal with misinterpretation is however less known. Further, while studies have looked at ambiguous sentences, how reanalysis occurs in both L1 and L2 speakers in sentences containing filler-dependencies (e.g., It was the book which the boy read the article about.) is not known. This thesis reports three studies investigating these issues in L1 and L2 processing, using offline, eyetracking while reading and structural priming tasks. The results showed that L2 participants performed syntactic reanalysis like L1 participants during the processing of garden-path sentences, with both groups showing evidence of lingering misinterpretation. Lingering misinterpretation was also found in filler-gap sentences, but there were some L1/L2 differences in certain filler-gap constructions such that reanalysis may be less complete for L2 than L1 speakers during online reading, depending on the nature of disambiguating cues and/or reanalysis difficulty. In general, the lingering misinterpretation observed in temporarily ambiguous and filler-gap sentences in both L1and L2 readers results at least partly from failures to discard initially assigned misinterpretations.

Acknowledgements

I thank everyone that I have met (Special thanks have been delivered to each person through an email or in person).

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

Hiroki Fujita

*Table of Contents*

***Declaration of Authorship***

I declare that I am the first author on the following papers, which report original work that I lead in collaboration with my supervisor.

Study 1

Fujita, H., & Cunnings, I. (under review). Reanalysis processes in non-native sentence comprehension. Submitted to *Bilingualism: Language and Cognition*.

Study 2

Fujita, H., & Cunnings, I. (under review). Lingering misinterpretation in native and non-native sentence processing: Evidence from structural priming. Submitted to *Applied Psycholinguistics.*

Study 3

Fujita, H., & Cunnings, I. (under review). Reanalysis and lingering misinterpretation of linguistic dependencies in native and non-native sentence comprehension. Submitted to *Journal of Memory and Language*.

## *Chapter 1. General Introduction*

How native (L1) and non-native (L2) speakers process and comprehend language has been hotly debated over the past 20 years. Some studies report that the mechanisms underlying L1 and L2 sentence processing may differ (e.g., Clahsen & Felser, 2006), while others argue that they may be similar (e.g., McDonald, 2006). One example of this L1 vs. L2 debate is how L1 and L2 speakers comprehend ambiguous sentences.

In language comprehension, even L1 speakers sometimes fail to comprehend the input correctly. One source of this comprehension difficulty comes from ambiguity. Languages are always ambiguous, as many words contain multiple meanings and structures are often compatible with more than one analysis, especially during sentence processing (MacDonald, Pearlmutter, & Seidenberg, 1994). For example, in the following sentence, "When Mary dressed the baby laughed happily", temporary ambiguity arises at "the baby", as at this point, readers can interpret "the baby" either as the direct object of the subordinate clause verb ("dressed") or as the subject of the main clause verb ("laughed"). It is known that readers initially adopt the former analysis (see Chapter 1.1.1) and consequently must conduct reanalysis at the disambiguating region, "laughed".

One recent finding on this syntactic reanalysis is that L1 speakers do not always derive the globally correct interpretation of temporarily ambiguous sentences even after receiving an error signal from the disambiguating word (e.g., Christianson, Hollingworth, Halliwell, & Ferreira, 2001). However, this does not mean that L1 speakers fail to complete reanalysis (e.g., Slattery, Sturt, Christianson, Yoshida, & Ferreira, 2013). Instead, L1 reanalysis difficulty is thought to result from a failure to discard initial misinterpretations[1].

[1] This "lingering misinterpretation" potentially consists of both structural and semantic representations. While teasing these apart may further reveal the nature of lingering misinterpretation, this thesis does not aim to distinguish them and thus treats lingering

One aim of this thesis is to investigate L2 reanalysis processes. While some studies report that L2 speakers initially miscomprehend temporarily ambiguous sentences as L1 speakers do (e.g., Juffs & Harrington, 1996), little is known about how they conduct reanalysis during online reading. This aim is partially motivated by the recent offline finding that L2 speakers may be more likely to persist with misinterpretations than L1 speakers (e.g., Cunnings, Fotiadou, & Tsimpli, 2017; Jacob & Felser, 2016; Pozzan & Trueswell, 2016). This may suggest that unlike L1 speakers, L2 speakers do not complete syntactic reanalysis. This possibility may be compatible with the Shallow Structure Hypothesis (Clahsen & Felser, 2006), which predicts reduced L2 syntactic parsing abilities, under the assumption that L2 speakers do not reconstruct the globally correct parse after reanalysis due to their underspecified parsing mechanisms. Alternatively, the particular L2 reanalysis difficulty may relate to increased susceptibility to initial misinterpretations (Cunnings, 2017). Study 1 addresses this issue during L1 and L2 sentence processing.

This thesis also tests for reanalysis of the subordinate clause verb. As illustrated above, temporarily ambiguous sentences cause reanalysis. This involves reanalysis of not only the temporarily ambiguous noun phrase (“baby”) but also the subordinate clause verb (“dressed”), from a transitive verb to an intransitive verb. To investigate this during language comprehension, Study 2 uses the structural priming paradigm. Structural priming refers to a phenomenon that processing of a certain syntactic structure is facilitated by repeated exposure to it (Bock, 1986). For example, it is known that people who have been exposed to a prepositional phrase (e.g., Tom gave the flower to Lisa.) are more likely to subsequently process similar structures more easily compared with being exposed to a different structure such as a double object phrase (e.g., Tom gave Lisa the flower.). Previous studies have shown that structural priming during language comprehension

---

misinterpretation as a phenomenon consisting of structural and semantic representations following the previous studies.

occurs both in L1 (e.g., Traxler, 2008) and L2 (e.g., Wei, Boland, Cai, Yuan, & Wang, 2019). Importantly, the source of priming depends on a representation being created during language comprehension. That is, if readers successfully adopt the globally correct intransitive interpretation of a temporarily ambiguous sentence after reanalysis, the intransitive representation should be primed. However, if either the memory trace of the initially assigned transitive misinterpretation lingers or reanalysis of this verb is not fully complete, the transitive representation may be primed. Taking advantage of this, van Gompel, Pickering, Pearson and Jacob (2006) examined L1 reanalysis of temporarily ambiguous sentences in a production task. Study 2 similarly adopts the structural priming paradigm but uses eyetracking while reading and end-of-sentence comprehension questions to investigate reanalysis of the subordinate clause verb in L1 and L2 language comprehension.

The third aim is to examine reanalysis of filler-gap dependencies. Filler-gap dependencies are an unbounded grammatical construction such as “It was the book that the boy read very happily last night.”, where the direct object of “read” (gap) is “the book” (filler) but it is dislocated by the relative clause. To assign the thematic interpretation between them, readers need to store the filler in memory and retrieve it at the gap position. Previous studies have shown that the filler and the gap are associated immediately after a gap position appears (e.g., Stowe, 1986). This “active gap filling” may require reanalysis when the predicted gap position is occupied by another entity (the filled-gap effect) such as in “It was the book that the boy read the review about very happily last night.”. According to the active gap filling strategy (Frazier, 1987), in this sentence, readers immediately retrieve the filler, “the book”, at “read”. This interpretation, however, immediately turns out to be wrong, as “the review” occupies the direct object position and the globally correct gap position appears at “about” to mean “the boy read the review about the book”. Reanalysis occurs not only in such filled-gap sentences, but may also occur without filled-

gap effects, in sentences I refer to as non-filled-gap sentences. For example, readers may initially misinterpret the following sentence, “It was the book that the boy read very happily about last night”, as “the boy read the book”, while the globally correct interpretation is “the boy read about the book”. If lingering misinterpretation is a general property of reanalysis, it should underlie reanalysis of filler-gap dependencies as well. However, reanalysis and lingering misinterpretation in filler-gap dependencies have not previously been examined in either L1 or L2 sentence processing. Study 3 explores this issue in L1 and L2 comprehension.

For Study 1–3, this thesis adopted a variety of offline and online experimental tasks. Offline tasks included end-of-sentence comprehension questions and sentence-picture matching, whereas online data were collected using eye-tracking while reading tasks. In the comprehension question task, participants read temporarily ambiguous and unambiguous sentences and later answered comprehension questions probing the temporary ambiguity (e.g., “When Mary dressed the baby laughed happily. Question: Did Mary dress the baby?”). This task is crucial for testing reanalysis, given that it reflects the final representation of the ambiguity. The sentence-picture matching task plays a supplementary role to the comprehension question task. In this task, participants read sentences and then chose pictures that they thought corresponded to what they had read. One motivation for this task comes from criticism of the comprehension question task such that asking an explicit comprehension question may reactivate the initially assigned misinterpretation (e.g., Tabor, Galantucci, & Richardson, 2004; van Gompel et al., 2006), as the question form is orthographically more similar to it (Mary dressed the baby) than to the globally correct interpretation (Mary dressed herself). The sentence-picture matching task may potentially solve this issue, given that pictures do not contain any orthographic information and thus do not bias participants towards the incorrect interpretation. Finally, eye-tracking while reading was used to examine reanalysis processes during online reading.

Having a certain grammatical knowledge does not necessarily mean that one can use it during online reading (Phillips, 1995). Using eye-tracking while reading allows us to examine use of grammar during online reading.

To summarise the results, Study 1 suggested that L2 speakers persist with initial misinterpretations more strongly than L1 speakers. However, Study 1 also found that this increased L2 reanalysis difficulty does not relate to a failure to complete syntactic reanalysis and that part of the L2 reanalysis difficulty results from a failure to erase the memory trace of initially assigned misinterpretations, as has also been argued for L1 reanalysis (Slattery et al., 2013). Study 2 revealed that the initially assigned transitive misinterpretation of temporarily ambiguous garden-path sentences is not fully discarded, and this affects subsequent sentence processing in both L1 and L2. Study 3 similarly showed evidence of lingering misinterpretation in filler-gap dependencies offline for L1 and L2 speakers, though the effect for L1 speakers was elusive under certain conditions, while effects were found for both L1 and L2 speakers during online reading. There was also some evidence from the online experiments that L2 speakers may fail to complete reanalysis of non-filled-gap sentences. These results suggest that lingering misinterpretation generally results from a failure to erase the memory trace of the initially assigned misinterpretation in both L1 and L2 sentence processing (Cunnings, 2017; Slattery et al., 2013), but under some conditions, L2 speakers may fail to complete reanalysis (Clahsen & Felser, 2006).

The remainder of this chapter is outlined as follows. First, I introduce two sentence processing models that account for initial parsing decisions during online reading. Initial parsing processes are crucial in reanalysis, as how readers make initial interpretations determines whether reanalysis occurs. I then discuss the nature of L1 reanalysis that particularly relate to this thesis. I then compare competing accounts of L2 sentence processing, especially as pertaining to ambiguity resolution, before finally discussing

syntactic priming in L1 and L2 speakers, and how it may influence syntactic ambiguity resolution.

### *1.1 Sentence processing in L1*

During sentence processing, readers incrementally analyse each incoming word syntactically and semantically. That is, they create a representation of the sentence and keep updating it without waiting for the end of the sentence. This incrementality means that some sentences may be initially interpreted in a way that is not consistent with the structure of the sentence as a whole. One phenomenon supporting this incremental property of sentence processing is that readers often fall into processing difficulty when reading temporarily ambiguous sentences like (1) (Bever, 1970).[2]

(1) The horse raced past the barn fell.

The temporary ambiguity arises at the first verb region ("raced"), as it can be interpreted as either a past participle or a past tense verb at this point. (1) is known to cause reading difficulty at the second verb region. This indicates that readers initially adopt the past tense verb analysis and realise at the end of the sentence that this analysis is wrong. Reading difficulty in temporarily ambiguous sentences has been confirmed by many studies (e.g., Frazier & Rayner, 1982), but how readers make initial decisions during online reading has been widely debated. Below, I introduce two well-known sentence processing theories that account for readers' initial processing decisions. Providing explanations of how initial decisions are made before discussing reanalysis is crucial, given incremental processing is a prerequisite for reanalysis to occur during online reading.

[2] Due to numbering restarting at (1) in the separate papers, in this thesis, the numbering of examples will start at (1) in each chapter. This also applies to Tables and Figures.

### *1.1.1 Syntax first model*

One influential theory of language comprehension is the syntax-first model (e.g., Frazier & Fodor, 1978), which assumes that reading difficulty caused by temporary ambiguity as in (1) is due completely to syntax. According to the syntax-first model, the goal of sentence processing is to parse sentences with the simplest possible structure. This processing strategy results in reading difficulty, when what is parsed becomes incompatible with the incoming input (e.g., Rayner, Carlson, & Frazier, 1983). Proponents of the syntax-first model call reading difficulty caused by temporarily ambiguous sentences "garden path" effects. The oldest account from the syntax-first model is the Canonical Sentoid Strategy proposed by Bever (1970), which claims that when a sequence of a noun phrase, a verb phrase and a noun phrase appears during sentence processing, it is interpreted as a subject, a verb and an object respectively unless the verb is explicitly marked as subordinate (see also Fodor, Bever, & Garrett, 1974). Following Bever, Frazier (1979) and Frazier and Fodor (1978) proposed two very influential parsing principles based on Kimball (1973): "Minimal Attachment" and "Late Closure", both of which are based on the least effort principle. Minimal Attachment claims that when a temporary ambiguity arises, readers attach it into the syntactic structure with the fewest nodes. For example, at the first verb region in (1), Minimal Attachment assumes that the main clause analysis is initially adopted, because this analysis contains fewer nodes (2a) compared with the alternative reduced relative clause analysis (2b).

(2a) [$_{TP}$ [$_{DP}$ [$_{D}$ The] [$_{NP}$ [$_{N}$ horse]]] [$_{T'}$ [$_{T}$][$_{VP}$ [$_{V}$ raced]]]

(2b) [$_{TP}$ [$_{DP}$ [$_{D}$ The] [$_{NP}$ [$_{N}$ horse] [$_{CP}$ [$_{C}$ (that)][$_{TP}$ [$_{T}$] [$_{VP}$ [$_{V}$ (was) raced]]]]]][$_{VP}$]]

Late Closure, which operates if Minimal Attachment fails to apply due to two possible structures being equally simple, stipulates that incoming input is incorporated into the clause or phrase that is currently being processed rather than starting a new constituent. For example, as illustrated, the sentence (3a) contains temporary ambiguity at "the baby", as both direct object analysis (3b) and subject analysis (3c) are possible at this region.

(3a) When Mary dressed the baby laughed happily.

(3b) [$_{CP}$ [$_{CP}$ [$_{C}$ When] [$_{DP}$ [$_{D}$ Mary] [$_{VP}$ [$_{V}$ dressed] [$_{DP}$ [$_{D}$ the] [$_{NP}$ [$_{N}$ baby]]]]]] [$_{C'}$ [$_{C}$] [$_{TP}$]]

(3c) [$_{CP}$ [$_{CP}$ [$_{C}$ When] [$_{DP}$ [$_{D}$ Mary] [$_{VP}$ [$_{V}$ dressed]]]] [$_{C'}$ [$_{C}$] [$_{TP}$ [$_{DP}$ [$_{D}$ the] [$_{NP}$ [$_{N}$ baby]]]]]]

As shown in (3b/c), both contain the same number of nodes, so Minimal Attachment cannot determine which construction readers prefer. In this case, readers adopt the direct object analysis in accordance with Late Closure, as this analysis is within the currently-being-parsed clause (i.e., subordinate clause) unlike the subject analysis, which requires creating a new clause (i.e., main clause).

The syntax-first two stage model also accounts for how readers process unbounded, filler-gap dependencies like (4).

(4) It was clear which boy John saw last night.

In (4), the wh-constituent ("which boy") is the theme of "saw" but is fronted to the beginning of the clause. Chomsky (1981) assumed that the wh-constituent moves out of the location following the verb to the front of the clause and leaves a gap at the original position. In language comprehension, readers need to associate the filler and the gap to assign the thematic interpretation. The Active Filler account (Frazier, 1987) claims that

readers anticipate the location of a gap as early as possible, and when discovering the first potential gap position, it immediately associates the filer with the gap. Thus, in (4), when "saw" appears, it is immediately assigned the wh-constituent as the direct object. One piece of evidence of such a parsing strategy comes from studies using sentences like (5) (Stowe, 1986).

(5a) My brother wanted to know who Ruth will bring us home to __ at Christmas.

(5b) My brother wanted to know if Ruth will bring us home to Mom at Christmas.

In (5a), although the globally correct gap position appears immediately after the preposition, "to", it is temporarily possible to retrieve the filler ("who") at the embedded clause verb ("bring"), as "who" is a plausible direct object of "bring". Contrarily, no gap appears in (5b) due to the conditional clause. There is evidence that unlike (5b), (5a) causes reading disruption at "us", and this is thought to result from filler-gap dependencies being formed actively, before the actual gap position appears. This is known as the filled-gap effect (e.g., Stowe, 1986). Although both filled-gap and garden-path effects result from temporary ambiguity, in this thesis I will distinguish filled-gap effects from garden-path effects in syntactic ambiguity.

As syntax precedes any other linguistic information in the syntax-first model, non-syntactic cues are argued not to have immediate effects on initial parsing decisions. For example, Pickering and Traxler (2003) reported that L1 speakers actively form filler-gap dependencies even when an optionally transitive verb at the gap position is more frequently used as an intransitive verb (6) (cf. Staub, 2007).

(6a) That's the cat that the dog worried compulsively about after going to the vet because of an injury.

(6b) That's the car that the dog worried compulsively about after going to the vet because of an injury.

(6c) That's the general that the soldier killed enthusiastically for during the war in Korea.

(6d) That's the country that the soldier killed enthusiastically for during the war in Korea.

(6) contains a temporarily available, doubtful gap after the embedded clause verb ("worried/killed") and manipulates semantic plausibility between the filler ("cat/car" and "general/country") and the gap so that the thematic relationship between them is either plausible (6a/c; "the dog worried the cat/the soldier killed the general") or implausible (6b/d; "the dog worried the car/the solder killed the country"). (6) also manipulates transitivity of the embedded clause verb such that the verb in (6a/b) is often used as the intransitive verb ("worried"), whereas the transitive use is more common for the verb of (6c/d; "killed"). If gap filling actively operates, reading times should be longer in (6b/d) than (6a/c) due to implausibility between the filler and the gap. However, if readers immediately make use of verb transitivity information, there should be no such implausibility effects in (6b), as the intransitive use prevents formation of filler-gap dependencies at the temporarily available gap. Contrary to this latter assumption, Pickering and Traxler observed longer reading times in the temporarily implausible conditions, irrespective of transitivity. This shows readers' parsing preferences for gap filling over the use of frequency information.

Similarly, Omaki et al. (2015) reported that readers attempt to associate the filler with intransitive verbs. In their study, L1 speakers read sentences like (7), which compare filler-gap sentences (7a/c) with the baseline control sentences (7b/d) that contain the

critical verb (“wrote/chatted”) embedded inside a relative clause (a so-called “island constraint”; Ross, 1967).

(7a) The city that the author wrote regularly about was named for an explorer.

(7b) The city that the author who wrote regularly saw was named for an explorer.

(7c) The city that the author chatted regularly about was named for an explorer.

(7d) The city that the author who chatted regularly saw was named for an explorer.

In (7), the filler (“the city”) is always an implausible direct object for the verb inside the relative clause (“wrote/chatted”). In the relative clause island (7b/d), forming a filler-gap dependency between the filler and the embedded clause verb is grammatically prohibited, as the filler cannot move out of such a deep construction (Ross, 1967). There is much empirical evidence that readers respect island constraints during sentence processing (e.g., Stowe, 1986; Traxler & Pickering, 1996). Thus, if readers ignore intransitivity during formulation of filler-gap dependencies (what they call “hyper-active gap filling”), reading times should be longer in (7a/c) than (7b/d) due to implausibility. However, if readers respect intransitivity before attempting to associate the filler with the gap (“conservative active gap filling”), implausibility effects should be observed only in (7a). In a sequence of online experiments, Omaki et al. found evidence for the hyper-active gap filling hypothesis, as they constantly observed implausibility effects only in the non-island condition, irrespective of verb intransitivity (cf. Staub, 2007). These studies suggest that readers prioritise constructing syntactic structures and make use of non-syntactic information such as verb subcategorisation at the second stage. They provide strong evidence of active gap filling, which requires reanalysis when a doubtful gap position precedes the ultimately correct gap position.

### *1.1.2 Constraint-based models*

While the syntax-first model predicts many phenomena of human sentence processing, some studies show that readers make use of non-syntactic linguistic information during initial parsing decisions (e.g., Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995). One well-known account against the syntax-first model is interactive accounts, which are also known as constraint-based models. Constraint-based models claim that all possible analyses are activated simultaneously if they receive any support from the available linguistic information, including syntactic, semantic and pragmatic cues, and initial parsing decisions are made after readers take all information into consideration (MacDonald et al., 1994; Taraban & McClelland, 1988; Trueswell, Tanenhaus, & Garnsey, 1994; Trueswell, Tanenhaus, & Kello, 1993; Tyler & Marslen-Wilson, 1977). In constraint-based models, processing difficulty arises when more than two analyses receive the same amount of support (cf. van Gompel, Pickering & Traxler, 2000). For example, Trueswell et al. (1994) tested whether readers rely on plausibility information during initial parsing of sentences like (8).

(8a) The defendant examined by the lawyer turned out to be unreliable.

(8b) The evidence examined by the lawyer turned out to be unreliable.

(8) contains a temporary ambiguity at "examined", as this verb can take either the past form (main clause analysis) or the past participle form modifying the subject noun phrase (reduced relative clause analysis) at this point. (8) also manipulates animacy such that the subject is either animate (8a) or inanimate (8b). The inanimate noun phrase ("The evidence") is an unlikely actor of the verb ("examined"), while it is plausible that "The defendant" takes "examined" as the main clause verb given the naturalness of a direct object appearing after "The defendant examined" (e.g., "The defendant examined the

accident".). The syntax-first two stage models assume that both (8a/b) should cause processing difficulty at the prepositional phrase ("by the lawyer"), as the main clause analysis contains fewer nodes, irrespective of animacy. Trueswell et al. however found increased reading times only in the sentences with the animate head like (8a), and processing difficulty disappeared when the sentences contained the inanimate head like (8b). This result suggests that readers make use of animacy information during initial parsing decisions (cf. Clifton et al., 2003; Ferreira & Clifton, 1986; Hagoort, 2003; McElree & Griffith, 1995; McRae et al., 1998; Rayner et al., 1983). There are also many other studies suggesting that other non-syntactic information influences initial parsing such as frequency (e.g., Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Hale, 2001; Hare, McRae, & Elman, 2003; Levy, 2008; Snedeker & Trueswell, 2004; Trueswell, 1996; Trueswell et al., 1993), prosody (e.g., Blodgett, 2004; Kjelgaard & Speer, 1999; Schepman & Rodway, 2000; Snedeker & Trueswell, 2003; Watson & Gibson, 2005), and discourse context (Altmann, Garnham, & Dennis, 1992; Altmann, Garnham, & Henstra, 1994; Altmann & Steedman 1988; Crain & Steedman, 1985; Spivey & Tanenhaus, 1998; Tanenhaus, Spivey-Knowiton, Eberhard, & Sedivy, 1995; Van Berkum, Brown, & Hagoort, 1999; cf. Ferreira & Clifton, 1986; Mitchell, Corley, & Garnham, 1992; Murray & Liversedge, 1994; Rayner, Garrod, & Perfetti, 1992).

The syntax-first and constraint-based models thus predict different processing mechanisms in initial parsing decisions. However, both models account for the substantial evidence that readers incrementally process language. As stated, this sentence incrementality causes reading difficulty during online reading, when initial parsing turns out to be incompatible with the incoming input (e.g., van Gompel et al., 2000), and this results in reanalysis. In the following, I discuss the nature of L1 reanalysis.

### *1.2 Reanalysis in L1*

Although temporarily ambiguous sentences have been widely examined, few studies explicitly define reanalysis. For example, Sturt and Crocker (1998) assumed that reanalysis occurs when at least one dependency breaks. In this thesis, reanalysis is defined as a phenomenon that occurs when a representation created by the incrementality of sentence processing later turns out to be incompatible with the incoming input by any cue. Early studies on reanalysis examined the conditions under which reanalysis is initiated (e.g., Fodor & Frazier, 1980), while more recent work has focused on how reanalysis may lead to misinterpretation (e.g., Christianson et al., 2001). I discuss each of these issues in turn below.

#### *1.2.1 Reanalysis avoidance*

One early study illustrating the nature of reanalysis is Fodor and Frazier (1980), who proposed the Revision-as-Last-Resort (RALR) principle (cf. Aoshima, Phillips, & Weinberg, 2004). This principle claims that readers do not revise the current analysis unless it turns out to be wrong and reanalysis is required. If this principle works on the basis that reanalysis is not conducted unless the current analysis is "syntactically" inconsistent with the globally correct analysis, it does not account for how readers perform reanalysis, given that plausibility causes reanalysis as well (e.g., Rayner et al., 1983). However, what is important here is that reanalysis causes processing difficulty and it is not performed as a matter of course in the first instance. Several studies support the concept of reanalysis avoidance. For example, Sturt, Pickering, Scheepers, and Crocker (2001) examined if readers disfavour reanalysis using temporarily ambiguous sentences like (9).

(9a) The troops who found the enemy spy had shot himself were later mentioned in the press report.

(9b) The troops who found the enemy spy had shot themselves and were later mentioned in the press report.

Although the first verb in the embedded clause ("found") is temporarily ambiguous, both the syntax-first and constraint-based models assume the noun phrase following this verb ("the enemy spy") to be interpreted as the direct object rather than the subject of the complement clause (e.g., "The troops who found [that] the enemy spy…"). "had shot" is temporarily ambiguous, as it can be attached to either "the enemy" or "The troops" at this point. The reflexive ("himself/themselves") disambiguates towards either analysis. If reanalysis is literally a last resort as predicted by the RALR principle, that is, if readers avoid performing unnecessary revisions, "had shot" should be interpreted as the main clause verb ("The troops…had shot themselves"), as the embedded clause verb interpretation ("The troops who found [that] the enemy had shot himself…") is inconsistent with the initially assigned direct object analysis at "found". However, this main clause verb analysis is presumably disfavoured, given that readers generally prefer low attachment over high attachment in English (e.g., Altmann, van Nice, Garnham, & Henstra, 1998; Phillips & Gibson, 1997). Consistent with the RALR principle, Sturt et al. observed increased reading difficulty at the reflexive in (9a) relative to (9b). This suggests that readers make an effort to attach the temporarily ambiguous phrase ("had shot") to the higher site to avoid reanalysis (see also Schneider & Phillips, 2001 for similar results).

### *1.2.2 Reanalysis cost*

Some studies show or argue that reanalysis difficulty is graded: i.e., some garden-path sentences cause greater reanalysis cost than others (e.g., Sturt et al., 1999). There are several factors that contribute to reanalysis cost. Note that scholars sometimes use conscious and unconscious reanalysis rather than difficult and easy reanalysis to illustrate

degree of reanalysis difficulty. In this thesis, they are treated as the same thing. Traditional studies on reanalysis cost can be divided into two factions. One assumes that computational limitations such as bounds on working memory play a role in the constraints imposed on reanalysis. For example, Gibson (1991) claimed that readers process language in a way such that one analysis is ranked higher than other potential analyses. However, to avoid memory overload, readers prune away analyses requiring more processing load than their available capacity. Reanalysis becomes costly when readers need to reactivate one of the pruned analyses.

Others argue that the degree of deconstruction of the existing structure causes difficult reanalysis. Frazier (1990) for example proposed the minimal revisions principle, which states that reanalysis should keep as much of the already assigned structure as possible. Others have described similar preferences in terms of reanalysis causing particular difficulty when a dominance relation is deconstructed (Gorrell, 1995; Sturt & Crocker, 1996, 1997; Weinberg, 1993; cf. Fodor & Inoue, 1998; see also Pritchett, 1988, 1992). Dominance in these models, which were descended from Description theory (Marcus, Hindle, & Fleck, 1983), is defined such that if node A dominates node B (i.e., A appears above B in the tree structure) and node B dominates node C, then node A necessarily dominates node C. Reanalysis causes particular difficulty when this dominance relation is deconstructed. For example, in (10a), "the baby" is temporarily attached as the theme of "saw", and "saw" dominates "the baby" at this point.

(10a) Mary saw the baby dressed happily.

(10b) When Mary dressed the baby laughed happily.

When reanalysis is required, "the baby" needs to be assigned the actor role from "dressed", but the dominance relation between "saw" and "the baby" continues to hold. In

other words, reanalysis of (10a) simply requires updating dominance relations by adding in a sentence node between "saw" and "the baby" and does not need to delete the existing relations. Thus, (10a) causes little reanalysis cost. Regarding (10b), "the baby" is temporarily attached as the theme of "dressed" and there is a dominance relation between them at the ambiguous region. When "laughed" appears, it assigns the actor role to "the baby" in a similar way to (10a), but unlike (10a), the dominance relation between "dressed" and "the baby" needs to be deconstructed in (10b), as they are no longer in the same local tree. Thus, (10b) causes greater reanalysis cost than (10a). In one of their reported self-paced reading tasks, Sturt et al. (1999) observed larger garden-path effects in (10b) than (10a), which provides evidence of representation-preserving mechanisms (cf. Grodner, Gibson, Argaman, & Babyonyshev, 2003).

Reanalysis difficulty may not be based solely on structural factors. As stated, many non-structural factors (e.g., plausibility, frequency, prosody and context) reduce or remove garden-path effects. Frazier and Clifton (1998) proposed the Semantic Cost Principle, which claims that reanalysis is costly if semantically interpreted analyses need to be revised. For example (1), repeated below in (11a), causes substantial reading difficulty, while difficulty in (11b) is reduced or absent despite the fact that both sentences consist of the same syntactic structure.

(11a) The horse raced past the barn fell.

(11b) The letter sent on Monday arrived today.

According to the Semantic Cost Principle, interpretation of "raced" in (11a) as the main clause verb is confirmed up until the disambiguating region ("fell"), while in (11b), semantic information such as animacy of the subject ("the letter") and plausibility between the subject and the following verb ("sent") as the main clause verb immediately

disconfirms the incorrect analysis before the disambiguating region (see also Stevenson & Merlo, 1997).

Another potential factor contributing to reanalysis difficulty is length of ambiguity (e.g., Arai & Nakamura, 2016; Frazier & Rayner, 1982; Tabor & Hutchins, 2004). Ferreira and Henderson (1991a, 1991b) examined how constituent length influences reanalysis difficulty using a speeded grammaticality judgement task, where participants judged the grammaticality of temporarily ambiguous/unambiguous sentences like (12) after they were displayed using rapid serial visual presentation.

(12a) After the Martians invaded the town the people were evacuated.

(12b) After the Martians invaded the town was evacuated.

(12c) After the Martians invaded the town that the city bordered was evacuated.

(12b/c) cause reanalysis at the subordinate clause verb (Early Closure), but (12a) does not, as it contains an additional noun phrase ("the people") for the main clause subject (Late Closure). One crucial difference between (12b) and (12c) is that (12c) contains a relative clause modifying the temporarily ambiguous noun phrase ("the town"). Thus, in (12c), readers need to commit to the incorrect analysis longer than (12b), where the ambiguity is disambiguated immediately after the temporary ambiguity. Ferreira and Henderson observed significantly lower acceptability rates in (12c) than (12b) (18% vs. 69%), by which they suggested that strong commitments to the wrong analysis increase the size of garden-path effects. Ferreira and Henderson further showed that what is crucial in ambiguity length is the distance between the head of the ambiguous phrase and the disambiguating region.

(13a) While the boy scratched the dog yawned loudly.

(13b) While the boy scratched the dog that Sally hates yawned loudly.

(13c) While the boy scratched the big and hairy dog yawned loudly.

(13a–c) are temporarily ambiguous sentences, with (13b/c) containing a long ambiguous phrase. (13b/c) are equally long in terms of the number of words, but the location of the head is different. While the head of the ambiguous phrase ("dog") is distant from the disambiguating region ("yawned") in (13b), these words are adjacent in (13c). The grammaticality judgement task showed that (13b) was significantly less acceptable than (13a/c), but there was no statistical difference between (13a/c). Ferreira and Henderson accounted for this result such that thematic roles are assigned when their heads appear and that reanalysis becomes more difficult, the longer readers are committed to the incorrect thematic analysis (see also Ferreira & Henderson, 1998).

Similarly to Ferreira and Henderson, Tabor and Hutchins (2004) claimed that reanalysis becomes difficult when readers commit to the incorrect analysis strongly or when additional supportive evidence of the misanalysis is confirmed ("digging-in" effect). For example, Tabor and Hutchins reported that reading times at "grew" are longer in (14b) than (14a).

(14a) As the author wrote the book grew.

(14b) As the author wrote the book describing Babylon grew.

Tabor and Hutchins claimed that increased reading times in (14b) result from the semantic linkage between "wrote" and "the book" being strengthened by the reduced relative clause. This claim is compatible with the concept of Ferreira and Henderson's reanalysis model.

### *1.2.3. Persistence of misanalysis*

As discussed above once temporarily ambiguous sentences cue disambiguation, readers immediately attempt to conduct reanalysis. Although these previous studies highlight when reanalysis might become particularly difficult, they do not examine the extent to which reanalysis is successful. This is because, until recently, it was thought that initial misinterpretations are completely discarded when reanalysis is completed. However, recent studies suggest that L1 speakers often persist with initial misinterpretations even after reanalysis. As Study 1–3 discuss this issue in detail (Chapter 2–4), here, I briefly introduce some key previous studies.

"Persistence of misinterpretation" was first explored experimentally by Christianson et al. (2001). In their study, L1 speakers read temporarily ambiguous/unambiguous sentences like "When Mary dressed(,) the baby laughed happily" and answered comprehension questions referring to the initial misinterpretation like "Did Mary dress the baby?". Based on the early reanalysis studies (e.g., Frazier & Clifton, 1998; Pritchett, 1992; Sturt & Crocker, 1996; Sturt et al., 1999), reanalysis of "the baby" from the direct object of the subordinate clause to the main clause subject is costly. If reanalysis is successful, participants should correctly respond "no" to this question, as the globally correct structure is that "Mary dressed herself" not "Mary dressed the baby". However, Christianson et al. observed more incorrect "yes" responses when questions followed ambiguous than unambiguous sentences. This suggests that L1 speakers do not always derive the globally correct interpretation after reanalysis.

Sturt (2007) similarly suggested that after reanalysis, readers persist with the initially assigned semantic misinterpretation. In Sturt, L1 speakers read sentences like (15).

(15a) The explorers found (that) the South Pole was actually right at their feet.

(15b) The explorers found (that) the South Pole was actually impossible to reach.

(15a/b) without the complementizer ("that") cause garden paths at the verb ("was") due to parsing preferences for the direct object analysis over the subject analysis at the main clause verb ("found"). The crucial difference between (15a) and (15b) is semantic consistency between the phrase after the disambiguating region and the initially assigned misinterpretation. While in (15a), "right at their feet" is semantically consistent with the initially assigned misinterpretation ("found the South Pole"), "impossible to reach" in (15b) is inconsistent with this misinterpretation. Sturt found increased reading times at the final segment in (15b) relative to (15a), when sentences were temporarily ambiguous. Sturt interpreted this finding as readers persist with initially assigned semantic interpretations after reanalysis (see also Jacob & Felser, 2016 for a similar finding).

The Good Enough Language Processing account, proposed by Ferreira and colleagues, predicts persistence of misinterpretation based on how readers process a language (e.g., Christianson et al., 2001; Christianson, Williams, Zacks, & Ferreira, 2006; Ferreira, Bailey, & Ferraro, 2002; Ferreira et al., 2001; Ferreira & Patson, 2007; Karimi & Ferreira, 2016; Slattery et al., 2013). According to the Good Enough Language Processing account, sentence processing is often "good enough", as readers often dot not create fully specified representations during online reading to minimise processing costs (e.g., Frazier & Fodor, 1978).

Exactly what may be "good enough" during language processing has been debated. One possibility is that readers fail to complete syntactic reanalysis, and instead construct a shallow parse of the input (see discussion in Ferreira et al., 2001; Christianson et al., 2001). Another possibility is that readers complete syntactic reanalysis, such that the ambiguous phrase is fully reanalysed, but that the initially assigned misinterpretation is not fully erased. This second possibility was recently supported by Slattery et al. (2013)

In Slattery et al., L1 speakers read sentences like (16) during an eye-tracking while reading task.

(16a) After the bank manager telephoned David's father/mother grew worried and gave himself approximately five days to reply.

(16b) After the bank manager telephoned, David's father/mother grew worried and gave himself approximately five days to reply.

(16a) is a temporarily ambiguous sentence and thus causes reanalysis at "grew", while (16b) is an unambiguous control. In addition to this ambiguity manipulation, (16) manipulates gender between the temporarily ambiguous noun phrase and a reflexive such that their gender is either matched ("David's father …himself") or mismatched ("David's mother…himself"). For unambiguous sentences, this gender manipulation should cause gender mismatch effects (Sturt, 2003) at the reflexive, with longer reading times for gender mismatching than matching antecedents, as a result of Binding Principle A (Chomsky, 1981). However, if the temporarily ambiguous noun phrase is misinterpreted as the subordinate clause direct object even after reanalysis, gender mismatch effects in the ambiguous conditions (16a) may be absent. This is because the misanalysis places "David's father/mother" and "himself" in different local trees in which they cannot be coreferential. Contrary to this hypothesis, Slattery et al. observed gender mismatch effects at the reflexive in both ambiguous and unambiguous conditions. This result provides strong evidence that L1 speakers complete reanalysis during online reading. Slattery et al. interpreted their results as indicating that L1 speakers conduct syntactic reanalysis, but initially assigned misinterpretations may not be fully discarded. They also described the mechanisms underlying lingering effects based on the lexically guided tree-adjoining grammar parsing model (Lau & Ferreira, 2005) such that before reanalysis is completed,

the structure constructed by the initially assigned misinterpretation is overlaid with the globally correct structure. This newly incorporated structure generally wins over the old one. However, the original misparse remains in the tree via an episodic memory trace (Kaschak & Glenberg, 2004) and influences subsequent sentence processing and language comprehension either structurally (van Gompel et al., 2006) or semantically (Sturt, 2007) (see also Fodor & Inoue, 1998 for an account in terms of co-extant representation of the correct and incorrect interpretations).

So far, I have discussed L1 sentence processing and reanalysis. In what follows, discussions will be focused on L2 sentence processing and reanalysis.

### *1.3. Sentence processing and reanalysis in L2*

Whether L1 and L2 processing are fundamentally similar or different has been widely debated. Over the past few decades, a number of models have been proposed to explain the similarities and differences between L1 and L2 sentence processing (e.g., Cunnings, 2017; Clahsen & Felser, 2006, 2017; Hopp, 2006, 2014; Jiang, 2003, 2007; McDonald, 2006; Sorace, 2011; Ullman, 2001). For example, Ullman (2001) argued that L1/L2 differences reside in different memory systems (declarative/procedural memory) used during learning. Ullman argues that L2 speakers are more likely to rely on declarative storage of grammatical rules, which L1 speakers are more likely to compute using procedural memory. While this model assumes qualitative differences between L1 and L2 processing, other models assume L1 and L2 processing are more similar. For example, McDonald (2006) and Hopp (2006) assume that individual differences in cognitive resources, and the quality of lexical representations, are the main causes of L1/L2 differences. These models predict L1 and L2 processing can be similar once such individual differences are accounted for.

The most relevant models to this thesis are the Shallow Structure Hypothesis proposed by Clahsen and Felser (2006) and the Interference model by Cunnings (2017). Below, I discuss these two L2 processing models first, which are then followed by discussions on L2 reanalysis and how these processing models relate to this thesis.

### *1.3.1 Shallow Structure Hypothesis*

The central concept of the Shallow Structure Hypothesis is that L2 (morpho)syntactic processing is underspecified. This concept is based on the assumption that sentence processing involves dual grammatical processing routes (e.g., Ullman. 2001). One is a deep, fully-specified route associated with grammar. The other is a shallow, underspecified route associated with non-syntactic information. The Shallow Structure Hypothesis specifically claims that L1 speakers engage in both routes during online reading but L2 speakers may be largely limited to the shallow route, potentially due to a reduced syntactic parsing ability. As a result, L2 sentence processing relies more heavily on lexical-semantic and pragmatic information and statistical patterns. There are a number of studies supporting this claim (Felser, & Roberts, 2007; Felser, Roberts, Marinis, & Gross, 2003; Marinis, Roberts, Felser, & Clahsen, 2005; Papadopoulou & Clahsen, 2003). For example, Marinis et al. (2005) examined how L2 speakers process unbounded dependencies containing an intermediate gap in sentences like (17).

(17a) The nurse who$_i$ the doctor argued e'$_i$ that the rude patient had angered e$_i$ is refusing to work late.

(17b) The nurse who$_i$ the doctor's argument about the rude patient had angered e$_i$ is refusing to work late.

(17c) The nurse thought the doctor argued that the rude patient had angered the staff at the hospital.

(17d) The nurse thought the doctor's argument about the rude patient had angered the staff at the hospital.

(17a/b) contain a long dependency between the wh-filler ("The nurse") and the post-verb gap indicated by "$e_i$". In (17a), this dependency crosses two bounding nodes headed by "argued" and "angered". According to Chomsky (1973), a dependency that crosses more than two bounding nodes in a single step violates the constraint of Subjacency. To allow for the formation of unbounded dependency in (17a), Chomsky argued that the wh-constituent needs to move out of the direct object position in the embedded clause via an intermediate gap ("$e'_i$") at the specifier position of the complementizer phrase. Thus, if the intermediate gap is postulated during sentence processing, "that" in (17a), which follows the intermediate gap, should elicit longer reading times than (17c), where the complementizer simply signals the beginning of the embedded clause. Further, at the actual gap position ("angered"), reading times are expected to be shorter in (17a) than (17b) given reduced distance between the filler and the gap in (17a) only if the intermediate gap is reactivated at the complementizer. In a self-paced reading task, Marinis et al. observed reading time patterns consistent with the idea of successful use of the intermediate gap in L1 participants but not in L2 participants. They interpreted this finding to indicate that unlike L1 speakers, L2 speakers may not make use of abstract syntactic information, in this case the intermediate gap, during online reading (but see Pliatsikas & Marinis, 2013, for evidence of more nativelike behaviour with sufficient naturalistic exposure to the L2).

Various other phenomena have also been investigated in relation to shallow L2 processing. For example, Felser, Sato and Bertenshaw (2009) showed L2 speakers' overreliance on non-syntactic information, using sentences like (18).

(18a) John noticed that Richard had cut himself with a very sharp knife.

(18b) Jane noticed that Richard had cut himself with a very sharp knife.

(18c) It was clear to John that Richard had cut himself with a very sharp knife.

(18d) It was clear to Jane that Richard had cut himself with a very sharp knife.

In (18), Binding Principle A syntactically allows the reflexive ("himself") to refer to "Richard" (the "accessible" antecedent) but not "John/Jane" (the "inaccessible" antecedent). The crucial difference between (18a/b) and (18c/d) is that the inaccessible antecedent in (18c/d) does not c-command the reflexive and is placed in a less prominent position. Felser et al. reasoned that if L2 speakers retrieve only the syntactically accessible antecedent during reflexive resolution, properties of the syntactically inaccessible main clause subject should not affect reading times at the reflexive. In an eye-tracking while reading experiment, Felser et al. however observed increased first pass reading times at the reflexive in L2 but not L1 participants when the main clause antecedent matched the reflexive in gender (18a), compared with when it did not (18b). Importantly, such differences were not observed in sentences like (18c/d). Felser et al. interpreted their results to suggest that in (18a/b), L2 participants sometimes initially attempted to resolve the reflexive to the main clause subject ("John/Jane"), as "John/Jane" are discourse-prominent, and thus, L2 participants relied on this non-syntactic information rather than syntactic, binding constraints (see also Felser & Cunnings, 2012).

Syntactic ambiguity resolution and filler-gap dependencies have also been widely examined from the perspective of shallow parsing. Studies specifically on L2 ambiguity resolution are discussed in more detail in section 1.3.3 below.

### *1.3.2 Interference model*

Cunnings's (2017) Interference model predicts L1 and L2 sentence processing differences based on cue-based memory retrieval (e.g., Jager, Engelmann & Vasishth, 2017; Lewis, Vasishth & Van Dyke, 2006; Lewis & Vasishth, 2005). Cue-based parsing models assume that sentence processing involves memory encoding, storage and retrieval. One classic example of this retrieval process is the resolution of linguistic dependencies like (19).

(19) The girl that the car passed in the street walked around the park.

In (19), the dependency between "the girl" and "walked" needs to be completed. According to cue-based parsing models, when "the girl" appears, it is encoded and stored in memory as a memory chunk. This memory representation needs to be retrieved at "walked", which requires "the girl" as the main clause subject based on retrieval cues. In the case of (19), the retrieval cues would be a *grammatical noun* that is *animate* and *walkable*, as only a noun that has these features can be the actor of "walked". Importantly, in cue-based memory retrieval, all the other words of the sentence (e.g., "the car") are also stored as memory chunks and at retrieval, all of them are compared with a set of cues. The item that provides the best match will be retrieved. This memory retrieval causes reading difficulty if more than two items that (partially) match the retrieval cues are stored in memory such as in (20).

(20) The girl that the boy passed in the street walked around the park.

(20) is the same as (19) except that the embedded clause subject, "the car", is replaced with "the boy". As in (19), "the boy" is not a syntactically permissible subject of "walked". However, unlike "the car", "the boy" entails *animate* and *walkable* features; i.e.,

"the boy" partially matches the retrieval cues. This partial match, according to cue-based parsing models, causes interference effects, which should result in longer reading times at "walked" in (20) than (19) (e.g., see Van Dyke, 2007).

Cunnings (2017) assumed that L2 speakers construct parses similarly to L1 speakers but are susceptible to retrieval interference differently to L1 speakers. Specifically, Cunnings argued that L2 speakers are more prone to non-syntactic interference from memory than L1 speakers during memory retrieval. This account provides an alternative explanation to L2 speakers' violation of binding constraints reported by Felser et al. (2009) and Felser and Cunnings (2012) that does not rely on shallow processing. For violation of Binding Principle A in (18), this account claims that L2 speakers may misretrieve the inaccessible antecedent ("John") at the reflexive in (18a/b) due to cue based competition between the accessible antecedent ("Richard") and the discourse-prominent inaccessible antecedent ("John/Jane").

Cunnings (2017) also attempted to account for potential L1/L2 differences in reanalysis based on interference from initially assigned misinterpretations in memory. Below, I first introduce studies on L2 reanalysis and then discuss how the Shallow Structure Hypothesis and the Interference model may relate to this issue.

### *1.3.3 Reanalysis in L2*

L2 speakers face reanalysis difficulty when reading temporarily ambiguous sentences as L1 speakers do. That is, research shows that L2 speakers process temporarily ambiguous sentences incrementally, and when an initially assigned structure becomes incompatible with the input, reanalysis is attempted (e.g., Frenck-Mestre & Pynte, 1997; Hopp, 2015; Jacob & Felser, 2016; Jegerski, 2012; Juffs, 2004; Juffs & Harrington, 1996; Roberts & Felser, 2011). For example, Juffs and Harrington (1996) examined L2 processing of temporarily ambiguous sentences (21a) and unambiguous sentences (21b).

(21a) After Bill drank the water proved to be poisoned.

(21b) After Sam arrived the guests began to eat and drink.

(21b) is treated as unambiguous, as the subordinate clause verb is intransitive and thus may reduce or remove garden-path effects (e.g., Trueswell et al., 1994). In Juffs and Harrington, L2 participants showed increased reading times at the disambiguating region in (21a) as L1 participants did. This suggests that L2 speakers process temporarily ambiguous sentences incrementally and as a result conduct reanalysis.

Juffs (2004) similarly examined L2 reanalysis but targeted Chinese, Japanese, and Spanish speaking learners of English. These languages are known to be null-subject languages: i.e., unlike English, a sentence in these languages can be grammatical without a subject. Juffs reasoned that null-subject language learners of English may show little to no reading difficulty at the disambiguating region in (21a) if their L1 affects ambiguity resolution, given that at reanalysis, the sentence temporarily lacks a main clause subject due to "the water" being misanalysed as the subordinate clause direct object. Juffs however observed similar garden-path effects between L1 and L2 speakers, which suggests that garden-path effects are robust in L2, irrespective of L1 properties.

As has been shown in L1 processing, there are however many studies showing effects of non-syntactic factors such as frequency (e.g., Dussias & Cramer Scaltz, 2008; Dussias, Marful, Gerfen, & Bajo Molina, 2010; Frenck-Mestre & Pynte, 1997; Lee, Lu, & Garnsey, 2013), semantic plausibility (e.g., Hopp, 2015; Roberts & Felser, 2011; Williams, Möbius & Kim, 2001), context (e.g., Pan & Felser, 2011; Pan, Schimke, & Felser, 2015) and prosody (e.g., Dekydtspotter, Donaldson, Edmonds, Fultz, & Petrush, 2008; Hwang & Schafer, 2006; Nickels, Opitz, & Steinhauer, 2013; White, Goad, Goodhue, Hwang, & Lieberman, 2013) on syntactic ambiguity resolution in L2.

L2 speakers also process filler-gap dependencies actively as L1 speakers do (e.g., Dallas, DeDe, & Nicol, 2013; Felser, Cunnings, Batterham, & Clahsen, 2012; Jessen & Felser, 2018; Jessen, Festman, Boxell, & Felser, 2017; Juffs & Harrington, 1995; Omaki & Schulz, 2011; Williams, 2006; Williams, Möbius & Kim, 2001; for a review see Dallas & Kaan, 2008). For example, Williams and colleagues (Williams, 2006; Williams et al., 2001) examined L2 filler-gap dependencies with a stop-making-sense task. In this task, participants read sentences with filler-gap dependencies like (22) in a word-by-word self-paced reading fashion and pressed a button when they thought the sentences became semantically abnormal.

(22a) Which machine did the mechanic fix the motorbike with two weeks ago?

(22b) Which friend did the mechanic fix the motorbike with two weeks ago?

(22) manipulates plausibility between the wh-filler and the verb such that these can be temporarily assigned a thematic interpretation in (22a; "Which machine did the mechanic fix…") but not in (22b; "Which friend did the mechanic fix…"). If L2 speakers actively associate the filler and the gap during incremental processing, temporarily implausible sentences (22b) should elicit a larger number of "stop" responses at or following the verb region than plausible sentences (22a). Williams and colleagues found the same response pattern as this prediction and suggested that L2 speakers actively form filler-gap dependencies.

Previous L2 studies suggest that incremental processing leads L2 speakers to reanalysis in both garden-path and filler-gap sentences. However, little has been explored concerning the nature of L2 reanalysis. Some of the few exceptions are Cunnings, Fotiadou, and Tsimpli (2017), Gerth, Otto, Nam, and Felser (2017), Hopp (2015), Jacob and Felser (2016), Pozzan and Trueswell (2016) and Roberts and Felser (2011). These studies

examined whether L2 speakers derive the correct interpretation after reanalysis by using end-of-sentence comprehension questions probing temporary ambiguity (e.g., Christianson et al., 2001), or act-out tasks, compared to L1 speakers. In Pozzan and Trueswell, L1 and L2 speakers heard sentences like (23) and performed the intended action while seeing picture displays containing a napkin, a box, a frog on a napkin, and a frog on a book while their eye-movements were monitored (using the "visual world paradigm").

(23a) Put the frog on the napkin onto the box.

(23b) Put the frog that's on the napkin onto the box.

(23) causes reanalysis at "onto", as readers/listeners initially misinterpret the first prepositional phrase ("on the napkin") as the destination of the verb ("Put"). The correct action of (23) is put the frog onto the box rather than the napkin. The results showed similar eye-movement patterns between L1 and L2 participants. However, L2 participants performed more incorrect actions than L1 speakers when listening to ambiguous but not unambiguous sentences. Pozzan and Trueswell interpreted their results as indicting that L2 speakers have increased reanalysis difficulty.

Jacob and Felser (2016) tested L2 reanalysis using temporarily ambiguous/unambiguous sentences like (24).

(24) While the gentleman was eating(,) the burgers were still being reheated in the microwave.

Question: Was the gentleman eating the burgers?

The main aim of (24) is to examine whether the initially assigned interpretation, "the gentleman was eating the burgers", which is semantically inconsistent with the subordinate clause verb phrase ("the burgers were still being reheated"), is processed similarly between ambiguous and unambiguous sentences. Offline comprehension questions, which always referred to the initially assigned misinterpretation, followed both ambiguous and unambiguous sentences. Jacob and Felser observed longer reading times at "being reheated" in ambiguous than unambiguous sentences, which shows that initial misinterpretations affected subsequent sentence processing. Further, comprehension accuracy rates for L2 participants were significantly lower than those for L1 participants. Although this reduced comprehension accuracy was observed in both ambiguous and unambiguous conditions, Jacob and Felser claimed that L2 speakers have more difficulty recovering from garden paths than L1 speakers. This increased L2 reanalysis difficulty has also been observed in other studies (e.g., Cunnings et al., 2017; Roberts & Felser, 2011).

How to account for these L1/L2 differences is however debated. Some processing models may predict similar reanalysis mechanisms, with increased L2 proficiency leading to nativelike behaviour (Hopp, 2006; McDonald, 2006). However, other models may make different predictions.

One possibility is that L2 speakers do not complete reanalysis. As reanalysis fails, only the initially assigned incorrect representation is stored in memory after reanalysis. This account may be compatible with the Shallow Structure Hypothesis (Clahsen & Felser, 2006), which claims that L2 syntactic parsing may be underspecified compared with L1 speakers. That is, it is possible that L2 speakers fail to complete reanalysis because they are unable to construct the syntactic structure of the globally correct interpretation.

Another possibility, which was recently proposed by Cunnings (2017), is that L2 speakers complete reanalysis but have particular difficulty discarding initial misinterpretations. Cunnings argued that this may result from L2 speakers being more

prone to interference from initial misinterpretations in memory than L1 speakers. That is, in "When Mary dressed the baby laughed", "the baby" is successfully reanalysed as the main clause subject. However, the initial misinterpretation, "Mary dressed the baby", is stored in memory after reanalysis, and L2 speakers are more likely than L1 speaker to misretrieve this incorrect interpretation, possibly as a result that the initial misinterpretation is more prominent than the globally correct interpretation. This predicts L2 speakers should attempt syntactic reanalysis during online reading like L1 speakers but be more susceptible to initially assigned lingering misinterpretations. That is, in "When Mary dressed the baby laughed", "the baby" is successfully reanalysed as the main clause subject. However, the initial misinterpretation, "Mary dressed the baby", is stored in memory after reanalysis, and L2 speakers are more likely than L1 speaker to misretrieval this incorrect interpretation, possibly as a result that the initial misinterpretation is more prominent than the globally correct interpretation. The three studies reported in this thesis attempted to tease these issues apart. Before reporting these studies however, I discuss structural priming briefly, as this phenomenon is important in Study 2.

### *1.4. Effects of structural priming on ambiguity resolution*

Structural priming refers to a facilitatory phenomenon that occurs when readers or listeners are exposed to a certain structure (Bock, 1986). For example, readers, who have been exposed to a prepositional phrase, tend to process a subsequent prepositional phrase more easily than before exposure. This structural priming has been observed widely in both production and comprehension in L1 (e.g., production: Bock, 1986; Bock & Griffin, 2000; Cleland & Pickering, 2003, 2006; Pickering & Branigan, 1998; comprehension: Arai, van Gompel, & Scheepers, 2007; Branigan, Pickering, & McLean, 2005; Ledoux, Traxler, & Swaab, 2007; Tooley & Bock, 2014; Tooley, Swaab, Boudewyn, Zirnstein, & Traxler, 2014; Tooley, Traxler, & Swaab, 2009; Traxler, 2008; Traxler & Tooley, 2008; Traxler,

Tooley, & Pickering, 2014), and recently in L2 (e.g., production: McDonough, 2006; McDonough & Chaikitmongkol, 2010; McDonough & Kim, 2009; Shin & Christianson, 2012; comprehension: Nitschke, Kidd, & Serratrice, 2010; Nitschke, Serratrice, & Kidd, 2014; Weber & Indefrey, 2009; Wei, Boland, & Brennan, 2018; Wei, Boland, Cai, Yuan, & Wang, 2019). As an example of priming effects, Branigan et al. (2005) tested priming effects on the interpretation of prepositional phrases like (25).

(25) The policeman prodded the doctor with the gun.

(25) is globally ambiguous, as it can mean that the policeman used the gun to prod the doctor (high attachment) or that the policeman prodded the doctor who had the gun (low attachment). In Branigan et al., L1 participants initially read globally ambiguous prime sentences like (25) and then saw two pictures, one that matched either the low or high attachment interpretation and the other matched neither. Participants were then provided globally ambiguous target sentences that were structurally similar to (25) like "The waitress prodded the clown with the umbrella" and saw two pictures, which corresponded to either the low or high attachment interpretation. The participants' task was to choose which picture matched the presented sentence. The results showed that participants tended to choose the picture for the target sentence that matched the interpretation assigned to the prime sentence. This suggests that previous linguistic experience affects subsequent language comprehension.

The most relevant previous priming studies to this thesis are Traxler (2015) and van Gompel et al. (2006). In Traxler, L1 participants read temporarily ambiguous early closure sentences (26a) and late closure sentences (26b).

(26a) As Jason watched the birds came closer and closer.

(26b) As Jason watched the birds the fox came closer and closer.

(26a) causes reanalysis at "came" as a result of "the birds" being misanalysed as the direct object of "watched", whereas (26b) does not, because another noun phrase, "the fox", appears as the main clause subject so that "the birds" can remain as the subordinate clause theme. In a self-paced reading task, L1 participants, who read temporarily ambiguous prime sentences like (26a), showed reduced garden-path effects in temporarily ambiguous target sentences, compared with when reading late closure sentences like (26b) as a prime. This result suggests that L1 speakers reinterpret the temporarily ambiguous verb intransitively too some degree.

In van Gompel et al., L1 participants read temporarily ambiguous/unambiguous sentences (27a) as primes and then completed sentence fragments (27b) as targets.

(27a) While the man was visiting(,) the children who were surprisingly pleasant and funny played outside.
(27b) When the doctor was visiti…

As discussed previously, readers initially misinterpret the subordinate clause verb of ambiguous sentences like (27a) as being transitive. van Gompel et al. reasoned that if L1 speakers persist with this initial misinterpretation after reanalysis, they may be more likely to complete sentence fragments (27b) with a verb being used transitively after reading temporarily ambiguous than unambiguous sentences. Consistent with this prediction, van Gompel et al. observed more transitive interpretations after ambiguous sentences. Traxler and van Gompel et al. thus demonstrated that structural priming effectively informs us with regards to the representation that is derived after reanalysis. However, whether the

initially assigned transitive misinterpretation lingers at the comprehension level is unexplored.

Recent L2 studies similarly showed that L2 speakers benefit from structural priming in language comprehension. For example, Wei et al. (2018) tested structural priming effects on L2 sentence processing of reduced relative clauses like (28).

(28a) The speaker selected by the group would be perfect for the program.

(28b) The director observed by the cop was in a bad part of the town

(28c) The architect selected by the manager was educated at Yale.

(28a/b) are prime sentences while (28c) is a target sentence. (28a) but not (28b) shares the same verb ("selected") with (28c). All sentences contain reduced relative clauses, which cause substantial reading difficulty at the preposition ("by") following the past participle verb. In a self-paced reading task, Wei et al. observed reduced reading times at the preposition in target sentences only when the verb was repeated between prime and target. This suggests that structural priming in L2 comprehension may be dependent on lexical overlap. However, in a follow-up study that used similar experimental items to (28), Wei et al. (2019) observed structural priming effects, irrespective of whether the verb was shared or not. Thus, while whether L2 structural priming is lexically independent is not fully elucidated, an issue that also has been raised in relation to structural priming in L1 comprehension (e.g., Arai et al., 2007; Traxler, 2008), previous studies have shown that L2 speakers are subject to structural priming during language comprehension.

### *1.5. This thesis*

I have hitherto discussed L1 and L2 sentence processing, reanalysis and priming effects on reanalysis. In summary, previous studies have shown that L1 and L2 speakers process

temporarily ambiguous sentences incrementally and consequently conduct reanalysis. Recent L1 studies suggest that L1 speakers successfully complete syntactic reanalysis but persist with initially assigned misinterpretations during online reading (e.g., Slattery et al., 2013). L2 speakers similarly persist with initial misinterpretations, with greater persistence than L1 speakers. Previous priming studies have shown that L1 speakers persist with the initially assigned transitive misinterpretation at the production level but attempt to reanalyse the subordinate clause verb as intransitive to some degree. What drives L2 lingering misinterpretations and the L1/L2 differences is, however, little known. L2 speakers may fail to complete reanalysis. Alternatively, they may have particular difficulty erasing the memory trace of initially assigned misinterpretations (Cunnings, 2017). Further, it is currently unknown how initially assigned misinterpretations may influence filler-gap dependencies in L1 and L2 processing. If lingering effects are a general property of reanalysis, L1 and L2 speakers should persist with initial misinterpretations in filler-gap dependencies as has previously been shown in garden-path sentences.

The following three chapters report three studies examining these issues. Study 1 aims to explore L2 reanalysis in garden-path sentences. Specifically, Study 1 tests the incomplete reanalysis hypothesis and Cunnings's Interference model. Study 2 similarly examines reanalysis of garden-path sentences, using the structural priming paradigm with a focus on the temporarily ambiguous subordinate clause verb. Finally, Study 3 tests reanalysis of filler-gap dependencies in L1 and L2 processing.

*Chapter 2. Study 1: Reanalysis processes in non-native sentence comprehension*


## Abstract

We report two offline and two eye-movement experiments examining non-native (L2) sentence processing during and after reanalysis of temporarily ambiguous sentences like "While Mary dressed the baby laughed happily". Such sentences cause reanalysis at the main clause verb ("laughed"), as the temporarily ambiguous noun phrase ("the baby") may initially be misanalysed as the direct object of the subordinate clause verb ("dressed"). The offline experiments revealed that L2ers have difficulty reanalysing temporarily ambiguous sentences with greater persistence of the initially assigned misinterpretation than native (L1) speakers. In the eye-movement experiments, we found that L2ers complete reanalysis similarly to L1ers but fail to fully erase the memory trace of the initially assigned interpretation. Our results suggested that the source of L2 reanalysis difficulty is a failure to erase the initially assigned misinterpretation from memory rather than a failure to conduct syntactic reanalysis.



**Keywords**: non-native sentence processing; syntactic ambiguity, eye-movements; reading.

## 2.1 Introduction

Syntactic ambiguity resolution has played an important role in motivating research in native (L1) and non-native (L2) sentence processing. Previous studies have shown that L1 and L2 speakers encounter difficulty when reading temporarily ambiguous sentences (e.g., Frazier & Rayner, 1982; Juffs & Harrington, 1996). For example, in (1), a temporary ambiguity emerges at "the baby", which can be interpreted either as the subordinate clause object or the main clause subject. Although the globally correct interpretation turns out to be the latter, the first interpretation ("Mary dressed the baby") may be initially considered during incremental processing. This initially assigned misinterpretation requires reanalysis later in the sentence however, once it is ruled out at the disambiguating verb ("laughed").

(1) While Mary dressed the baby laughed happily.

Importantly, research has shown that the initially assigned misinterpretation may persist even after reanalysis (Christianson, Holingworth, Halliwell & Ferreira, 2001), and L2ers have particular difficulty revising misinterpretations (Jacob & Felser, 2016). However, why L2ers have particular reanalysis difficulty is debated.

To explore this issue, we report four experiments investigating reanalysis in L1 and L2 processing. Two experiments adopted offline methods to tap the final interpretation assigned to garden-path sentences, while the others used eye-tracking during reading to investigate the time-course of reanalysis during L1 and L2 sentence processing. Our results suggest that L1 and L2 reanalysis difficulty resides mostly in a difficulty in discarding the initial misinterpretation from memory, rather than an inability to complete syntactic reanalysis. Below, we begin by discussing approaches to reanalysis and misinterpretation in L1ers, before discussing potential differences between L1 and L2 processing.

### 2.1.1 Reanalysis in L1 sentence processing

Many studies have shown that in sentences like (1), reading difficulty arises at "laughed" (e.g., Ferreira & Henderson, 1991; Pickering & Traxler, 1998; Sturt, Pickering, & Crocker, 1999). This suggests that "the baby" is initially interpreted as the theme of the subordinate clause verb "dressed", and that reanalysis of this phrase as the main clause subject occurs at the main clause verb. Recent studies have focused on the nature of reanalysis. Christianson et al. (2001) used end-of-sentence questions (e.g., "Did Mary dress the baby?") that probed interpretation of the temporarily ambiguous phrase. Although the correct answer to such questions following reanalysis should be "no", they observed more incorrect "yes" responses when questions followed temporarily ambiguous sentences like (1) than unambiguous sentences disambiguated with a comma (e.g., While Mary dressed, the baby laughed happily.). They interpreted this as evidence that the initially assigned misinterpretation ("Mary dressed the baby") often persists following reanalysis.

One potential counterargument to this claim is that lower accuracy rates for ambiguous sentences may be an artefact of the design, due to reactivation of the misinterpretation as a result of the comprehension question being more similar to the ambiguous than unambiguous sentences (e.g., Nakamura & Arai, 2016; Tabor, Galantucci, & Richardson, 2004; van Gompel, Pickering, Pearson, & Jacob, 2006). However, corroborating results have been found with various designs that avoid such repetition, including sentence-picture matching (Malyutina & den Ouden, 2016; Nakamura & Arai, 2016; Patson, Darowski, Moon, & Ferreira, 2009; van Gompel et al., 2006).

While many studies thus show that readers persist with misinterpretation, different accounts have been proposed as to why it occurs (e.g., Christianson et al., 2001; Kaschak & Glenberg, 2004; Sturt, 2007). The Good Enough Language Processing account predicts the lingering effect based on how readers process language (e.g., Christianson et al., 2001; Ferreira, Christianson, & Hollingworth, 2001; Slattery, Sturt, Christianson, Yoshida, &

Ferreira, 2013). However, the precise nature of "good enough" processing has been debated. In earlier instantiations of the theory (e.g., Christianson et al., 2001), it was hypothesised that readers may not complete syntactic reanalysis, and instead adopt underspecified or "shallow" syntactic parses of garden-path sentences. This account claimed that lingering misinterpretation arises due to a failure to complete syntactic reanalysis. More recently however, research has suggested that L1ers do successfully conduct syntactic reanalysis, but that lingering effects arise because the initially assigned misinterpretation is not completely erased (Qian, Garnsey, & Christianson, 2018; Slattery et al., 2013). Consider Slattery et al. (2013), who monitored L1ers' eye-movements in two experiments.

In one experiment, participants read temporarily ambiguous and unambiguous sentences like (2a) and (2b), respectively, which additionally manipulated gender (mis)match between a reflexive ("himself") and its antecedent ("father/mother").

(2a) After the bank manager telephoned David's father/mother grew worried and gave himself approximately five days to reply.

(2b) After the bank manager telephoned, David's father/mother grew worried and gave himself approximately five days to reply.

Slattery et al. expected longer reading times at the disambiguating verb "grew" in (2a) than (2b) but were particularly interested in subsequent processing at the reflexive. Following Sturt (2003), in unambiguous (2b), longer reading times were predicted when the reflexive mismatched in gender with its antecedent, the main clause subject ("David's mother grew worried and gave himself…"), compared to when it matched ("David's father… himself"). Slattery et al. reasoned that if syntactic reanalysis following garden-paths is incomplete, such that "David's father/mother" remains as the subordinate clause

object rather than the main clause subject, gender mismatch effects should be reduced or absent in (2a). This is because according to Binding Principle A (Chomsky, 1981), the reflexive's antecedent in (2) must be the main clause subject. If reanalysis is incomplete however, the reflexive may fail to find an antecedent. Contrary to this incomplete reanalysis hypothesis, Slattery et al. observed gender mismatch effects in both ambiguous and unambiguous sentences, suggesting L1ers perform syntactic reanalysis of the temporarily ambiguous noun phrase.

In a second experiment, Slattery et al. tested texts like (3).

(3a) While Frank dried off the truck/grass that was dark green was peed on by a stray dog.

Frank quickly finished drying himself off then yelled out the window at the dog.

(3b) While Frank dried off, the truck/grass that was dark green was peed on by a stray dog.

Frank quickly finished drying himself off then yelled out the window at the dog.

The first sentence is either temporarily ambiguous (3a) or unambiguous (3b). It also manipulates whether the main clause subject is a plausible or implausible theme of the subordinate clause verb (plausible "dried off the truck" vs. implausible "dried off the grass"). Slattery et al. reasoned that for ambiguous sentences, plausible initially assigned misinterpretations may linger, compared with implausible ones.

The second sentence always referred to the globally correct interpretation of the first sentence ("Frank quickly finished drying himself off"). It is however inconsistent with the initially assigned misinterpretation (e.g., "Frank dried off the truck"). Thus, if the initial misinterpretation is completely erased, there should be no reading time differences between conditions at the reflexive in the second sentence. However, if misinterpretation

lingers in the plausible conditions, reading times at "himself" may become longer in ambiguous than unambiguous conditions, as evidence that the initially assigned misinterpretation ("Franck dried off the truck") lingered after reanalysis. Consistent with this latter prediction, Slattery et al. observed longer reading times for ambiguous than unambiguous sentences at the reflexive of the second sentence in plausible conditions. Taking the results of both experiments together, Slattery et al. argued L1ers conduct syntactic reanalysis of the temporarily ambiguous phrase, but that the initially assigned misinterpretation is not fully erased from memory.

### 2.1.2 Reanalysis in L2 sentence processing

Previous research has shown that L2ers are susceptible to garden-path effects like L1ers (e.g., Juffs & Harrington, 1996; for review see Cunnings, 2017). One noteworthy finding related to L2 reanalysis is that L2ers may have more difficulty reanalysing temporary ambiguities than L1ers. For example, several studies report that L2ers answer comprehension questions probing the interpretation of temporary ambiguities less accurately than L1ers (e.g., Gerth et al., 2017; Jacob & Felser, 2016; Pozzan & Trueswell, 2016).

While studies suggest L2ers may persist with misinterpretation more strongly than L1ers, the mechanisms underlying L2 reanalysis have not been fully examined. One potential account is that L2ers cannot perform syntactic reanalysis completely. This may be compatible with the Shallow Structure Hypothesis (Clahsen & Felser, 2006; 2017), which claims L2ers have difficulty constructing abstract syntactic structures during sentence processing (Felser, & Roberts, 2007; Felser, Roberts, Marinis, & Gross, 2003; Marinis, Roberts, Felser, & Clahsen, 2005; Papadopoulou & Clahsen, 2003). Although the Shallow Structure Hypothesis was not originally formulated to account for L2 reanalysis processes, it might be that L2ers have difficulty constructing the correct syntactic structure

after reanalysis as a result of shallow parsing. Alternatively, L2ers may perform syntactic reanalysis like L1ers but have increased difficulty in erasing the initially assigned interpretation from memory (Cunnings, 2017). This would predict that L2ers should complete syntactic reanalysis like L1ers, but that misinterpretations should be more likely to linger during L2 processing.

One study that bears on this issue is Jacob and Felser (2016). They conducted an eye-movement study on lingering misinterpretation following garden-paths and interpreted their results as evidence that L2ers do not complete syntactic reanalysis. They tested sentences like (4).

(4) While the gentleman was eating(,) the burgers were still being reheated in the microwave.

The primary motivation of their experiment was to examine whether the main clause ("the burgers were still being reheated"), which is semantically inconsistent with the initially assigned interpretation of the preceding temporary ambiguity ("the gentleman was eating the burgers"), is processed similarly between ambiguous and unambiguous sentences. In addition to the finding of longer reading times at "being reheated" for ambiguous sentences, indicating lingering effects, Jacob and Felser observed smaller garden-path effects for L2 than L1 participants, which they interpreted as indicating that L2ers may not initiate reanalysis as consistently or quit reanalysis earlier than L1ers.

### 2.1.3 The Present Study

Against this background, we report two studies investigating L2 reanalysis. We aimed to tease apart whether reanalysis difficulty during L2 processing relates to incomplete syntactic reanalysis, or difficulty in erasing initially assigned misinterpretations. In the first

study, participants completed an offline task that used comprehension questions to investigate the final interpretation assigned to garden-path sentences (Experiment 1a) and an online eye-tracking during reading experiment to investigate reanalysis during processing (Experiment 1b). In the second study, participants completed a sentence-picture matching task to further investigate how garden-path sentences are interpreted (Experiment 2a), and an online eye-tracking experiment that investigated persistence of misinterpretation (Experiment 2b). For both studies, although we report the offline tasks first, participants completed the eye-tracking task first in a separate experimental session.

### 2.2 Experiment 1a

Experiment 1a manipulates ambiguity such that (5a) is temporarily ambiguous while (5b) is unambiguous due to the presence of the comma. Thus, (5a), but not (5b), causes reanalysis at the disambiguating region ("laughed"). Each experimental sentence was followed by one of two comprehension questions, (6a) or (6b), that tested two aspects of reanalysis. Specifically, reanalysis in (5a) involves reinterpreting the subordinate clause verb as intransitive, and the temporarily ambiguous noun phrase as the main clause subject. (6a) tested the former while (6b) examined the latter case. We avoided yes/no questions that involve repetition of the temporary ambiguity to minimise potential reactivation of the misinterpretation (van Gompel et al., 2006). The two answers always denoted a misinterpretation and the globally correct interpretation.

(5a) *Ambiguous*

After the mother dressed the baby in the living room laughed very happily.

(5b) *Unambiguous*

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

(6a) *Subordinate clause question*

What happened? 1. The mother dressed herself 2. The mother dressed the baby

(6b) *Main clause question*

Who laughed very happily? 1. The mother 2. The baby

We expected lower question accuracy rates for ambiguous than unambiguous sentences (Christianson et al., 2001). If the temporarily ambiguous noun phrase is reanalysed as the main clause subject but the initially assigned misinterpretation persists (Slattery et al., 2013), accuracy rates should be higher for ambiguous main clause questions than ambiguous subordinate clause questions, as main clause questions specifically refer to reanalysis of the temporarily ambiguous noun phrase, while subordinate clause questions test the interpretation of the subordinate clause which is thought to linger after reanalysis. It was also expected that accuracy rates would be lower for L2 than L1 participants when questions followed ambiguous sentences (e.g., Jacob & Felser, 2016).

### 2.2.1 Method

*Participants*

40 L1 English speakers (5 males, mean age = 19; range = 18–23) and 40 L2 English speakers (10 males, mean age = 25; range = 18–43), of various L1 backgrounds[3], from the University of Reading community participated in Experiment 1a. Participants received either course credit or a small monetary incentive, immediately after completing the experiment.

[2] First languages were Chinese (6), Spanish (5), French (4), Malay (4), Turkish (4), Cantonese (2), Kazakh (2), Polish (2), Swedish (2), Greek (1), Indonesian (1), Italian (1), Norwegian (1), Russian (1), Serbian (1), Thai (1).

L2 participants started learning English in a school environment after age five (mean age of onset = 8.6; SD = 3.1; range = 5–18). They also completed the Oxford Placement Test (OPT) (Allan, 2004). Their mean scores were 72/100 (SD = 10.3; range = 52–89), showing they were upper intermediate-advanced English language learners.

*Materials*

Experimental materials consisted of 24 sets of sentences as in (5) paired with comprehension questions as in (6). The materials manipulated ambiguity such that a comma was either inserted (5a) or removed (5b). The subordinate clause always utilised reflexive absolute transitive (RAT) or reciprocal verbs (Ferreira & McClure, 1997). RAT verbs, when used intransitively, must be interpreted reflexively. For example, "the mother dressed" in (5b) can only mean "the mother dressed herself" (compare to "the mother called", where even when used intransitively the "calling" action is made toward another implied theme). Reciprocal verbs share a similar property when the subject is plural. These verbal properties are crucial in Experiment 1a, as they create a situation where there is always only one absolute correct answer to the comprehension questions.

The experiment also contained 72 filler sentences with a variety of syntactic structures, all of which were followed by a comprehension question. Experimental stimuli were constructed with four counterbalanced presentation lists in a Latin Square design. The full set of experimental items used for each experiment reported here is included in Appendix A–D.

*Procedure*

The experiment was administered as a whole-sentence reading comprehension task using Linger (Rohde, 2010). Each trial began with a cross onscreen. Upon pressing the space bar, an entire sentence was shown. After reading the sentence, participants pressed the space

bar again, at which point the sentence disappeared and was replaced with a question containing two options. Participants answered each question by pressing either the "1" key for the first option or the "2" key for the second option. The experimental and filler sentences were pseudo-randomised so that at least two filler sentences always appeared between each experimental sentence. The order of the answers was also randomised to assign the correct and incorrect answers to the two options equally. Participants were instructed to simply read each sentence and answer the comprehension questions accurately. The experiment began with some practice trials.

After the experiment, L2 participants completed the OPT. The main experiment took approximately 20–25 minutes with an additional 25–30 minutes for the OPT.

*Data analysis*

We analysed accuracy rates to examine lingering misinterpretations. Data analysis was conducted in R (R Core Team, 2018) using generalised linear mixed-effects models. Models included sum-coded fixed effects of ambiguity (ambiguous/unambiguous), question type (subordinate clause question/main clause question) and group (L1/L2).

All models were fit using the maximal random effects structure that converged, including by-subject and by-item random intercepts, and random slopes for each within-item and within-subject fixed effect (Barr, Levy, Scheepers, & Tily, 2013).[4] For each fixed effect, *p* values were estimated using the Laplace Approximation.

When a three-way interaction was observed, follow-up analyses were performed on question type to examine potentially different accuracy rate patterns within each question type. In the case of interactions between ambiguity and group, planned comparisons tested

[4] If the maximal model failed to converge, we first removed the random correlation parameters. If this model still did not converge, the random effect accounting for the least variance (generally zero) was iteratively removed until convergence was achieved.

effects of group at the two levels of ambiguity to examine L1/L2 differences. Data and analysis code for all experiments reported here is available at the first author's Open Science Framework webpage (https://osf.io/bt637/).

### 2.2.2 Results

Average accuracy rates for filler sentences were 93% for both L1 (range = 83–100) and L2 (range = 79–100) participants, suggesting participants read the sentences carefully. Accuracy rates and a summary of the statistical analysis are provided in Tables 1 and 2.

There was a significant main effect of ambiguity due to lower accuracy rates for the ambiguous than unambiguous conditions, and a significant main effect of question type, with more correct responses to main clause than subordinate clause questions. There was also a reliable three-way interaction between ambiguity, question type and group. Planned 2×2 analyses on each question type showed significantly lower accuracy for ambiguous than unambiguous sentences for subordinate clause questions (*estimate* = 1.792, $z$ = 5.71, *SE* = 0.31, $p < .001$). Although numerically there was a tendency towards lower accuracy rates for L2 than L1 participants, neither the main effect of group (*estimate* = 0.537, $z$ = 1.53, *SE* = 0.35, $p = .127$) or interaction (*estimate* = 0.110, $z$ = 0.20, *SE* = 0.55, $p = .840$) was significant. Regarding main clause questions, the main effect of ambiguity but not group was significant (ambiguity: *estimate* = 2.036, $z$ = 3.36, *SE* = 0.61, $p < .001$; group: *estimate* = 0.224, $z$ = 0.59, *SE* = 0.38, $p = .556$). There was also a marginal two-way interaction (*estimate* = 1.333, $z$ = 1.71, *SE* = 0.78, $p = .088$). Planned comparisons by ambiguity revealed lower accuracy rates for L2 than L1 participants in the ambiguous condition (*estimate* = 0.881, $z$ = 2.11, *SE* = 0.42, $p = .035$) but not in the unambiguous condition (*estimate* = 0.261, $z$ = 0.42, *SE* = 0.62, $p = .673$).

| | L1 | | | | L2 | | | |
|---|---|---|---|---|---|---|---|---|
| | SC Question | | MC Question | | SC Question | | MC Question | |
| | Amb | Unamb | Amb | Unamb | Amb | Unamb | Amb | Unamb |
| Experiment 1a | | | | | | | | |
| | 58 (49) | 88 (32) | 75 (44) | 93 (25) | 50 (50) | 82 (38) | 60 (49) | 95 (22) |
| Experiment 2a | | | | | | | | |
| | 69 (46) | 83 (37) | 81 (39) | 92 (27) | 57 (50) | 81 (39) | 65 (48) | 93 (26) |

Table 1. *Accuracy rates for comprehension questions following ambiguous and unambiguous sentences in Experiment 1a and 2a (SDs in parentheses). SC = subordinate clause, MC = main clause, Amb = Ambiguous, Unamb = Unambiguous.*

| | Experiment 1a | | | Experiment 2a | | |
|---|---|---|---|---|---|---|
| | *Estimate (SE)* | *t(z) value* | *p* value | *Estimate (SE)* | *t(z) value* | *p* value |
| Ambiguity | 2.30 (0.24) | 9.71 | < .001 | 1.76 (0.22) | 7.94 | < .001 |
| Group | 0.46 (0.30) | 1.55 | .120 | 0.62 (0.33) | 1.88 | .060 |
| QT | 0.93 (0.22) | 4.24 | < .001 | 0.92 (0.31) | 2.95 | .003 |
| Ambiguity:Group | 0.45 (0.46) | 0.97 | .331 | 0.72 (0.39) | 1.84 | .066 |
| Ambiguity:QT | 0.31 (0.32) | 0.98 | .328 | 0.48 (0.40) | 1.21 | .228 |
| Group:QT | 0.25 (0.32) | 0.77 | .444 | 0.09 (0.35) | 0.26 | .797 |
| Ambiguity:Group:QT | 1.34 (0.64) | 2.10 | .036 | 0.56 (0.61) | 0.92 | .359 |

Table 2. *Summary of statistical analyses for Experiment 1a and 2a. QT = Question Type.*

### 2.2.3 Discussion

Experiment 1a showed that L1 and L2 participants persisted with misinterpretation. Although accuracy rates were generally higher for main clause than subordinate clause questions, the low accuracy rates for ambiguous main clause questions may contrast with Slattery et al. (2013) and Christianson et al. (2001), who claimed that L1ers successfully reanalyse the temporarily ambiguous noun phrase as the main clause subject. This issue is discussed in detail in General Discussion.

Importantly, there was evidence that reanalysis was more costly for L2 than L1 participants in main clause questions. However, for subordinate clause questions, L2ers did not have significantly lower accuracy than L1ers. This result is inconsistent with some previous studies, especially Jacob and Felser (2016) who used similar materials and observed more incorrect responses for L2 participants. One potential reason for the conflicting results is a difference in the questions used, as Jacob and Felser used yes/no

questions while the present study asked "What happened?". We return to the replicability of offline L2 reanalysis effects in Experiment 2a, which used a sentence-picture matching task. We first however report Experiment 1b, which tested online reanalysis processes.

### 2.3 Experiment 1b

Experiment 1b investigates L2 reanalysis during sentence processing. We adapted Slattery et al.'s (2013) research design. Participants read texts like (7) while their eye-movements were monitored.

(7) Some people had a party at a friend's house at the weekend.

(a) *Ambiguous, Gender Match*

After the neighbour visited Ken's dad decided to prepare himself a cold drink.

(b) *Ambiguous, Gender Mismatch*

After the neighbour visited Ken's mum decided to prepare himself a cold drink.

(c) *Unambiguous, Gender Match*

After the neighbour visited, Ken's dad decided to prepare himself a cold drink.

(d) *Unambiguous, Gender Mismatch*

After the neighbour visited, Ken's mum decided to prepare himself a cold drink.

It was very tasty.

(7a/b) are temporarily ambiguous while (7c/d) are unambiguous. We further manipulated gender (mis)match between the reflexive ("himself") and its antecedent ("Ken's dad/mum"). In (7a/c), the antecedent matches the reflexive's gender, whereas in (7b/d), it does not. In line with previous studies, garden-path effects are expected with

increased reading times for ambiguous sentences at the disambiguating region ("decided"). In the unambiguous condition, the gender mismatch condition should elicit longer reading times at the reflexive than the gender match condition (Slattery et al., 2013).

For the ambiguous conditions, if reanalysis is complete, similar gender mismatch effects are expected. However, if reanalysis is incomplete, there should be no or reduced gender mismatch effects for the ambiguous conditions. This would suggest that the temporary ambiguous noun phrase remained as the direct object of the subordinate clause verb. For L1ers, we expected to replicate Slattery et al. (2013), and observe gender mismatch effects in both ambiguous and unambiguous conditions. If L2ers fail in syntactic reanalysis, gender mismatch effects should emerge only in the unambiguous conditions. Alternatively, if L2 syntactic reanalysis is successful, they should behave like L1ers.[5]

### 2.3.1 Method

*Participants*

The participants from Experiment 1a took part in Experiment 1b. Experiment 1b was conducted in a separate session, at least one week before Experiment 1a.

*Materials*

24 sets of experimental sentences as in (7) were created. Each set began with a lead-in sentence which always appeared on the first line. The critical target sentence appeared across two lines, with the line-break always appearing immediately after the time

[5] Note that some studies indicate L2ers are more prone to violating Binding Principle A during processing than L1ers (e.g., Felser & Cunnings, 2012; Felser, Sato & Bertenshaw, 2009). Irrespective of this possibility, for present purposes, it is the *relative* difference in gender mismatch effects between ambiguous and unambiguous conditions that is of primary importance.

conjunction. Half of the experimental sentences employed the masculine reflexive and half the feminine. A third wrap-up sentence, which took up one line onscreen, was inserted on the third line.

72 fillers were also constructed with a variety of syntactic structures, which took up either two or three lines on the screen. All experimental and two-third of filler texts were followed by a yes/no comprehension question. No question asked about the temporary ambiguity or the antecedent of the reflexive.

*Procedure*

Although viewing was binocular, eye-movements were recorded from the right eye using an SR Research Eyelink 1000. Before each experimental session, calibration of the eye-tracker was conducted on a nine-point grid, and recalibration was conducted when needed. Before each text was shown, participants had to fixate on a black square above the first word of the sentence. Upon fixation of the square, the text appeared. When participants thought they comprehended the text, they pressed a button on a game pad to make the text disappear. Either the next trial then began, or a comprehension question was shown, which participants answered by pressing an appropriate button on a gamepad. Experimental and filler texts were presented in a pseudo-randomised Latin-square design.

After the experimental session, L2 participants looked through a vocabulary list containing critical vocabulary used for the subordinate clause verb and the main clause subject. Participants ticked a box next to each word that they were unsure of. The entire experiment lasted 40–60 minutes.

*Data Analysis*

Three reading time measures are reported for four regions of text. To test for garden-path effects, we analysed the disambiguating region ("decided"), along with a first spillover

region containing the words up to but not including the reflexive ("to prepare"). To test for gender mismatch effects, we analysed the reflexive region ("himself"), along with a second spillover region that contained the rest of the critical sentence ("a cold drink"). Reading time measures consisted of first pass reading times, the summed duration of all fixations within a region until an eye-movement away from the region, regression path duration, the summed duration of fixations from the first fixation entering a region from the left up until but not including the first fixation in a region to the right, and total viewing times, the summed duration of all fixations within a region[6]. Fixations shorter than 80ms that were within one degree of visual arc of another fixation were merged, and any other fixations shorter than 80ms or longer than 800ms were removed. Further, any region that a participant skipped was removed from data analysis, which affected less than 7% of the L1 data and 3% of L2 data. Trials including vocabulary that the L2ers did not know were also removed, which affected less than 0.1% of the L2 data.

Data analysis used linear mixed-effect models (Baayen, Davidson, & Bates, 2008). Each model included log-transformed reading times as the dependent variable. Sum-coded fixed effects included ambiguity (ambiguous/unambiguous), gender (gender match/mismatch) and group (L1/L2). To reduce the number of independent statistical tests

[6] First pass reading times indicate early processing stages, while regression path duration and total viewing times reflect later processing stages. The nature of each measure is useful particularly when examining the timing of a certain effect during online reading (e.g., the timing of garden-path effects). However, this thesis examines how L1 and L2 speakers conduct reanalysis, where the presence of an effect, rather than the timing of an effect, is important (e.g., the presence of gender mismatch effects). Thus, while this thesis reports the three eye-movement measures across Study 1–3 following the traditional eye-movement studies, it does not interpret the results based on the timing of an effect.

conducted (von der Malsburg & Angele, 2017), we additionally included region (disambiguating region/first spillover region or reflexive region/second spillover region) as a fixed effect (see Cunnings & Sturt, 2018). Each model was fit with the maximal random effects structure that converged. As treating region as a fixed effect involves two non-independent datapoints from the same trial, we also included random intercepts for trial. For each fixed effect, *p* values were estimated using the Satterthwaite approximation implemented by the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017).

In the case of interactions between region and ambiguity or gender, follow-up comparisons were performed to examine garden-path or gender mismatch effects at each region. In the case of an interaction between group and ambiguity or gender, follow-up tests examined effects of ambiguity or gender at the two levels of group. In the case of an interaction between ambiguity and gender, gender mismatch effects were examined at the two levels of ambiguity.

### 2.3.2 Results

Mean accuracy rates across experimental and filler trials were 89% for L1 participants (range = 75–100%) and 88% for L2 participants (range = 71–97%), indicating participants read carefully. A summary of the reading time data and inferential statistics are provided in Table 3 and 4. For brevity, all statistical models showed a main effect of group due to longer reading times for L2 than L1 participants. Also, we do not discuss main effects of region, as these are difficult to interpret due to the different lexical material across regions. Effects of region thus have little meaning unless they interact with another fixed effect.[7]

*Disambiguating Region*

[7] We also do not discuss interactions between region and group alone, as these are not informative about effects of ambiguity or gender (mis)match.

For both first pass and regression path times at the disambiguating and first spillover regions, there were significant main effects of ambiguity, with longer reading times for ambiguous than unambiguous sentences. In first pass times there was also a significant ambiguity by region interaction, with larger garden-path effects at the disambiguating region. In first-pass times the effect of gender and the three-way interaction between ambiguity, gender and region were marginal. Follow-up analyses conducted on each region suggested a marginal interaction between ambiguity and gender only in the disambiguating region (*estimate* = 0.083, $t$ = 1.82, *SE* = 0.05, $p$ = .073), with apparent gender mismatch effects in the ambiguous conditions only (for ambiguous conditions, *estimate* = 0.071, $t$ = 2.16, *SE* = 0.03, $p$ = .034; for unambiguous conditions, *estimate* = 0.012, $t$ = 0.41, *SE* = 0.03, $p$ = .682). The main effect of gender and two-way interaction between ambiguity and gender were also significant in regression path times. Follow-up analyses by ambiguity showed apparent gender mismatch effects only in ambiguous sentences (ambiguous: *estimate* = 0.090, $t$ = 2.62, *SE* = 0.03, $p$ = .009; unambiguous: *estimate* = 0.001, $t$ = 0.03, *SE* = 0.03, $p$ = .976). We did not expect gender mismatch effects in these measures, and as they appear before the reflexive, we assume these effects are spurious. In total viewing times, there were significant main effects of ambiguity, with longer reading times in ambiguous sentences, and gender, with longer reading times in gender mismatch conditions. Gender mismatch effects are expected in this measure, as total times can include reading after the reflexive was encountered. There was a marginal three-way interaction between ambiguity, gender and region. Follow-up analyses by region showed main effects of ambiguity (*estimate* = 0.283, $t$ = 8.08, *SE* = 0.04, $p < .001$) and gender (*estimate* = 0.125, $t$ = 2.88, *SE* = 0.04, $p$ = .009) with a marginal interaction between them (*estimate* = 0.109, $t$ = 1.77, *SE* = 0.06, $p$ = .092) in the disambiguating region. Pairwise comparisons conducted at the two levels of ambiguity indicated gender mismatch effects only in the ambiguous condition (ambiguous: *estimate* = 0.178, $t$ = 3.03, *SE* = 0.06, $p$

= .006; unambiguous: *estimate* = 0.072, *t* = 1.54, *SE* = 0.05, *p* = .138). Analysis of the first spillover region indicated main effects of ambiguity (*estimate* = 0.201, *t* = 5.84, *SE* = 0.03, *p* < .001) and gender (*estimate* = 0.148, *t* = 3.88, *SE* = 0.04, *p* < .001) only, due to garden-path and gender mismatch effects respectively.

*Reflexive Region*

At the reflexive and second spillover regions, a significant main effect of gender was observed in all eye-movement measures, with longer reading times in gender mismatch than gender match conditions. This effect in total viewing time is illustrated in Figure 1. Additionally, in first pass reading times, the model showed a significant interaction between ambiguity and group, but pairwise comparisons revealed no main effect of ambiguity in either group (L1: *estimate* = 0.041, *t* = 1.65, *SE* = 0.02, *p* = .114; L2: *estimate* = 0.022, *t* = 1.09, *SE* = 0.02, *p* = .276).

In regression path duration, there was a significant interaction between gender and region, due to larger gender mismatch effects at the spillover region (*estimate* = 0.236, *t* = 6.50, *SE* = 0.04, *p* < .001) than the reflexive region (*estimate* = 0.129, *t* = 4.50, *SE* = 0.03, *p* < .001). There was also a marginal interaction between ambiguity and gender. However, pairwise comparisons showed significant gender mismatch effects in both ambiguous (*estimate* = 0.134, *t* = 3.99, *SE* = 0.03, *p* < .001) and unambiguous (*estimate* = 0.231, *t* = 5.87, *SE* = 0.04, *p* < .001) conditions.

Total viewing times showed a similar interaction between gender and region due to different sizes of gender mismatch effects between the reflexive (*estimate* = 0.247, *t* = 7.88, *SE* = 0.03, *p* < .001, reading time differences between gender match and mismatch conditions: 119ms) and spillover (*estimate* = 0.173, *t* = 6.96, *SE* = 0.02, *p* < .001, 143ms) regions.

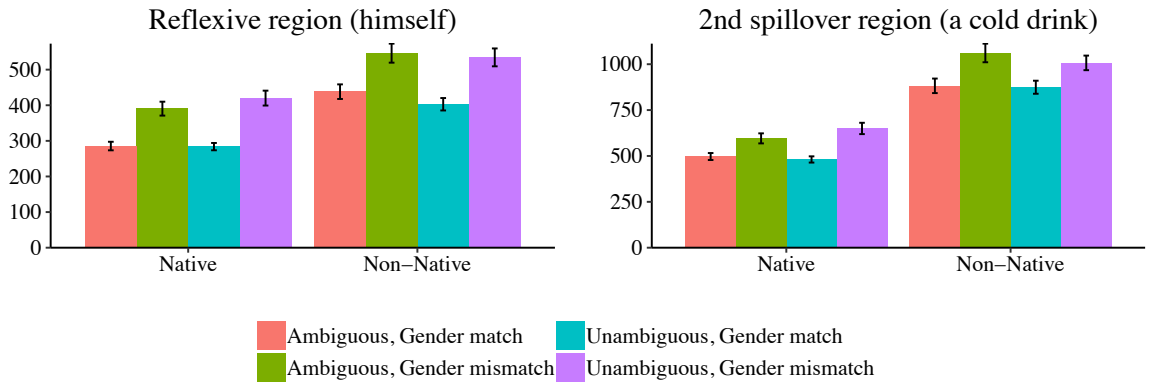


Figure 1. *Total viewing times in milliseconds at the reflexive and 2nd spillover regions in Experiment 1b.*

| | Disambiguating region (decided) | | Spillover Region (to prepare) | | Reflexive Region (himself) | | 2nd Spillover Region (a cold drink) | |
|---|---|---|---|---|---|---|---|---|
| | L1 | L2 | L1 | L2 | L1 | L2 | L1 | L2 |
| **First pass time** | | | | | | | | |
| Ambiguous, Gender Match | 295 (157) | 352 (179) | 283 (141) | 369 (183) | 207 (73) | 268 (133) | 378 (215) | 594 (333) |
| Ambiguous, Gender Mismatch | 323 (164) | 372 (220) | 276 (131) | 377 (180) | 241 (141) | 271 (115) | 387 (216) | 610 (339) |
| Unambiguous, Gender Match | 270 (133) | 336 (164) | 268 (113) | 362 (192) | 210 (71) | 259 (110) | 393 (217) | 578 (334) |
| Unambiguous, Gender Mismatch | 277 (133) | 325 (157) | 283 (123) | 365 (177) | 234 (104) | 272 (131) | 428 (274) | 600 (351) |
| **Regression path time** | | | | | | | | |
| Ambiguous, Gender Match | 522 (726) | 605 (654) | 568 (689) | 874 (1262) | 332 (524) | 350 (332) | 834 (1236) | 1115 (1396) |
| Ambiguous, Gender Mismatch | 625 (793) | 747 (1107) | 700 (927) | 900 (1256) | 387 (879) | 440 (691) | 1007 (1674) | 1396 (1797) |
| Unambiguous, Gender Match | 400 (440) | 430 (399) | 404 (460) | 409 (294) | 252 (215) | 280 (168) | 566 (555) | 1116 (1324) |
| Unambiguous, Gender Mismatch | 410 (546) | 447 (467) | 382 (413) | 435 (309) | 387 (671) | 357 (373) | 965 (1112) | 1382 (1351) |
| **Total viewing time** | | | | | | | | |
| Ambiguous, Gender Match | 517 (379) | 746 (524) | 447 (266) | 695 (497) | 285 (186) | 438 (317) | 497 (294) | 882 (612) |
| Ambiguous, Gender Mismatch | 685 (545) | 871 (651) | 558 (406) | 821 (636) | 390 (302) | 546 (410) | 596 (424) | 1061 (779) |
| Unambiguous, Gender Match | 422 (271) | 539 (422) | 371 (221) | 529 (339) | 284 (159) | 403 (269) | 481 (264) | 875 (549) |
| Unambiguous, Gender Mismatch | 477 (369) | 574 (366) | 454 (333) | 617 (372) | 420 (324) | 534 (386) | 650 (477) | 1007 (611) |

Table 3. *Reading times for three eye-movement measures at four regions of texts in Experiment 1b (SDs in parentheses).*

| | Disambiguation / Spillover Region (decided / to prepare) | | | Reflexive / 2nd Spillover Region (himself / a cold drink) | | |
|---|---|---|---|---|---|---|
| | *Estimate (SE)* | *t* | *p* | *Estimate (SE)* | *t* | *p* |
| **First pass time** | | | | | | |
| Ambiguity | 0.05 (0.02) | 3.08 | .005 | 0.01 (0.01) | 0.67 | .50 |
| Group | 0.23 (0.04) | 5.55 | < .001 | 0.29 (0.05) | 6.26 | < .001 |
| Gender | 0.02 (0.01) | 1.76 | .078 | 0.05 (0.01) | 3.67 | < .001 |
| Region | 0.02 (0.03) | 0.83 | .413 | 0.62 (0.04) | 14.40 | < .001 |
| Ambiguity:Group | 0.006 (0.03) | 0.20 | .840 | 0.06 (0.03) | 2.10 | .05 |
| Ambiguity:Gender | 0.04 (0.03) | 1.19 | .239 | 0.01 (0.03) | 0.29 | .77 |
| Group:Gender | 0.02 (0.03) | 0.85 | .404 | 0.03 (0.03) | 1.00 | .33 |
| Ambiguity:Region | 0.09 (0.03) | 2.92 | .005 | 0.009 (0.03) | 0.33 | .75 |
| Group:Region | 0.10 (0.03) | 3.73 | < .001 | 0.23 (0.05) | 4.36 | < .001 |
| Gender:Region | 0.01 (0.03) | 0.42 | .678 | 0.02 (0.03) | 0.84 | .41 |
| Ambiguity:Group:Gender | 0.05 (0.07) | 0.72 | .478 | 0.03 (0.06) | 0.43 | .67 |
| Ambiguity:Group:Region | 0.07 (0.06) | 1.05 | .305 | 0.07 (0.06) | 1.21 | .24 |
| Ambiguity:Gender:Region | 0.10 (0.05) | 1.93 | .054 | 0.03 (0.05) | 0.58 | .56 |
| Group:Gender:Region | 0.05 (0.05) | 0.89 | .374 | 0.04 (0.06) | 0.70 | .49 |
| Ambiguity:Group:Gender:Region | 0.06 (0.10) | 0.59 | .563 | 0.04 (0.14) | 0.31 | .76 |
| **Regression path time** | | | | | | |
| Ambiguity | 0.28 (0.03) | 9.03 | < .001 | 0.03 (0.03) | 1.12 | .274 |
| Group | 0.21 (0.06) | 3.57 | < .001 | 0.29 (0.06) | 4.86 | < .001 |
| Gender | 0.04 (0.02) | 2.10 | .036 | 0.18 (0.03) | 6.81 | < .001 |
| Region | 0.06 (0.04) | 1.54 | .134 | 0.95 (0.05) | 19.23 | < .001 |
| Ambiguity:Group | 0.08 (0.06) | 1.30 | .201 | 0.009 (0.06) | 0.15 | .881 |
| Ambiguity:Gender | 0.09 (0.04) | 2.08 | .037 | 0.09 (0.05) | 1.87 | .076 |
| Group:Gender | 0.03 (0.04) | 0.61 | .541 | 0.03 (0.05) | 0.74 | .464 |
| Ambiguity:Region | 0.10 (0.06) | 1.75 | .093 | 0.08 (0.06) | 1.50 | .147 |
| Group:Region | 0.08 (0.05) | 1.43 | .157 | 0.31 (0.08) | 3.92 | < .001 |
| Gender:Region | 0.01 (0.04) | 0.26 | .794 | 0.11 (0.04) | 2.46 | .014 |
| Ambiguity:Group:Gender | 0.11 (0.08) | 1.28 | .199 | 0.14 (0.09) | 1.55 | .122 |
| Ambiguity:Group:Region | 0.13 (0.10) | 1.32 | .191 | 0.13 (0.09) | 1.40 | .167 |
| Ambiguity:Render:Region | 0.10 (0.09) | 1.15 | .250 | 0.06 (0.09) | 0.63 | .533 |
| Group:gender:Region | 0.03 (0.09) | 0.32 | .749 | 0.02 (0.09) | 0.28 | .780 |
| Ambiguity:Group:Gender:Region | 0.07 (0.18) | 0.40 | .693 | 0.16 (0.17) | 0.93 | .354 |
| **Total viewing time** | | | | | | |
| Ambiguity | 0.24 (0.02) | 10.2 | < .001 | 0.01 (0.02) | 0.52 | .606 |
| Group | 0.32 (0.07) | 4.74 | < .001 | 0.43 (0.07) | 6.40 | < .001 |
| Gender | 0.14 (0.04) | 3.58 | .002 | 0.21 (0.02) | 8.96 | < .001 |
| Region | 0.04 (0.04) | 1.05 | .303 | 0.60 (0.04) | 13.82 | < .001 |
| Ambiguity:Group | 0.06 (0.04) | 1.45 | .146 | 0.07 (0.05) | 1.39 | .177 |
| Ambiguity:Gender | 0.05 (0.05) | 1.14 | .268 | 0.04 (0.04) | 1.08 | .279 |
| Group:Gender | 0.02 (0.05) | 0.41 | .685 | 0.02 (0.05) | 0.41 | .687 |
| Ambiguity:Region | 0.08 (0.03) | 2.90 | .004 | 0.003 (0.03) | 0.09 | .933 |
| Group:Region | 0.11 (0.04) | 3.02 | .003 | 0.23 (0.05) | 4.54 | < .001 |
| Gender:Region | 0.03 (0.03) | 0.87 | .396 | 0.07 (0.03) | 2.49 | .021 |
| Ambiguity:Group:Gender | 0.08 (0.09) | 0.89 | .377 | 0.06 (0.08) | 0.75 | .452 |
| Ambiguity:Group:Region | 0.005 (0.06) | 0.10 | .922 | 0.06 (0.06) | 1.00 | .319 |
| Ambiguity:Gender:Region | 0.11 (0.06) | 1.95 | .051 | 0.01 (0.06) | 0.26 | .797 |
| Group:gender:Region | 0.02 (0.06) | 0.31 | .759 | 0.02 (0.05) | 0.36 | .723 |
| Ambiguity:Group:Gender:Region | 0.04 (0.11) | 0.37 | .711 | 0.10 (0.11) | 0.94 | .360 |

Table 4. *Summary of statistical analyses for Experiment 1b*.

### 2.3.3 Discussion

Consistent with previous studies, Experiment 1b showed that both L1 and L2 participants encountered reading difficulty upon disambiguation of temporarily ambiguous sentences (Frazier & Rayner, 1982; Juffs & Harrington, 1996). Regarding reanalysis, there was evidence of gender mismatch effects, irrespective of ambiguity, at the reflexive and second spillover regions that were not significantly modulated by group. This suggests that both L1 and L2 participants constructed the correct syntactic structure after reanalysis, at least to the extent that Binding Principle A was at play, which replicates the L1 findings from Slattery et al. (2013).

Experiment 1b suggests that L2ers conduct syntactic reanalysis during online reading. This indicates that L2 reanalysis difficulty cannot be accounted for entirely by the incomplete reanalysis hypothesis. This interpretation raises the possibility that L2 reanalysis difficulty observed in Experiment 1a results not completely from failures to construct the correct syntactic structure, but at least partly from the persistence of misinterpretation (Cunnings, 2017; Slattery et al., 2013).

Experiments 2a/b further explored L2 reanalysis. Experiment 2a replicated Experiment 1a using a different task, while Experiment 2b tested whether and to what extent misinterpretation lingers during online processing.

## 2.4 Experiment 2a

Experiment 2a aimed to replicate Experiment 1a, using sentence-picture matching. Participants read temporarily ambiguous (8a) and unambiguous (8b) sentences, and were then shown one of two pictures pairs, (9a) or (9b). The subordinate clause picture pair in (9a) denotes either the correct or incorrect action of the subordinate clause (“the lady woke up”/“the lady woke up her husband”), while the main clause picture pair (9b) depicts either the correct or incorrect action of the main clause (“her husband drank some coffee”/“the

lady drank some coffee"). Participants chose which picture they thought best corresponded to the sentence.

(8a) *Ambiguous*

After the lady woke up her husband in the apartment drank some coffee.

(8b) *Unambiguous*

After the lady woke up, her husband in the apartment drank some coffee.

(9a) *Subordinate clause pictures*

Figure 2. *Example picture pairs used for the subordinate clause questions in Experiment 2a*.

(9b) *Main clause pictures*

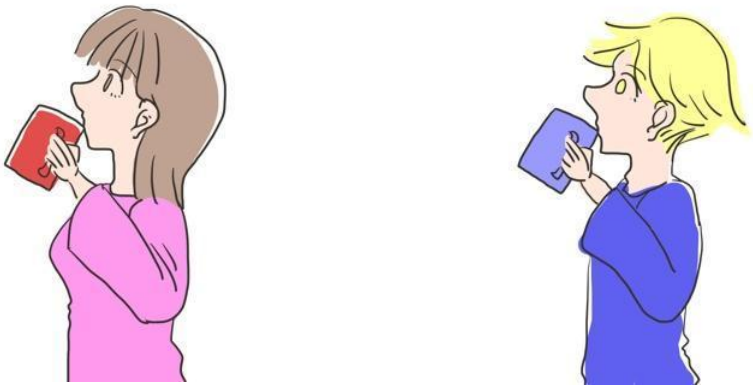

Figure 3. *Example picture pairs used for the main clause questions in Experiment 2a*.

We expected higher error rates for ambiguous than unambiguous sentences with lower accuracy rates for subordinate clause than main clause questions in the ambiguous condition (Christianson et al., 2001). If L2ers persist with misinterpretation more often than L1ers, L2ers' accuracy rates should be lower than L1ers' in the ambiguous condition.

### 2.4.1 Method

*Participants*

40 L1 English speakers (7 males, mean age = 19; range = 18–23) and 40 L2 English speakers of various L1 backgrounds[8] (14 males, mean age = 23; range = 18–47), none of whom took part in Experiment 1a or 1b, from the University of Reading community participated in Experiment 2a. Participants received course credit or payment. L2 participants started learning English from age eight onwards (mean age of onset = 8.9; SD = 1.1; range = 8–11) and also completed the OPT, which showed that they were upper intermediate-advanced English language learners (mean = 76; SD = 10.6; range = 51–94).

*Materials*

Experiment 2a used 24 temporarily ambiguous and unambiguous sentences as in (8), using only RAT or reciprocal verbs, and four coloured pictures for each experimental set. Two of the pictures tapped the interpretation of the subordinate clause (9a) while the other two examined the interpretation of the main clause (9b). The experiment additionally included 84 fillers of a variety of different constructions, accompanied by two pictures. The experimental and filler sentences were presented in a counterbalanced Latin square design.

[8] First languages were Greek (8), Bulgarian (5), Italian (5), French (3), Indonesian (3), Polish (3), Lithuanian (2), Spanish (2), Turkish (2), Arabic (1), Bangladeshi (1), Chinese (1), Dutch (1), Finnish (1), German (1), and Thai (1).

*Procedure and Data Analysis*

The procedure was identical to Experiment 1a, except that after the critical sentences participants saw one picture pair, either (9a) or (9b), and chose which picture that they felt best matched the content of the sentence. The experiment was administered using the IBEX Farm web-based experiment presentation platform (Drummond, 2013) but was completed by participants in a traditional lab setting. All other aspects of the procedure and data analysis were identical to Experiment 1a.

### 2.4.2 Results

Descriptive and inferential statistics are encapsulated in Table 1 and 2. Accuracy rates for fillers were 94% for L1 participants (range = 69–100) and 95% for L2 participants (range = 83–100).

Analysis showed that there was a significant main effect of question type, with lower accuracy rates for subordinate clause than main clause questions. There was also a significant main effect of ambiguity, with lower accuracy for ambiguous sentences, qualified by a marginal two-way interaction between ambiguity and group. This suggested significantly lower accuracy rates for L2 than L1 participants in the ambiguous conditions only (ambiguous: *estimate* = 0.960, $z = 2.52$, *SE* = 0.38, $p = .012$; unambiguous: *estimate* = 0.155, $z = 0.30$, *SE* = 0.52, $p = .765$).

### 2.4.3 Discussion

As in Experiment 1a, Experiment 2a indicated lingering misinterpretation in L1ers and L2ers. In line with previous studies (Jacob & Felser, 2016), there was also evidence that L2ers had particular difficulty with reanalysis. Importantly, L2ers were less accurate for both subordinate and main clause ambiguous conditions, unlike Experiment 1a that

showed an interaction between ambiguity and group only in main clause questions. Additionally, as mentioned for Experiment 1a, the finding that both groups chose pictures incorrectly some of the time in the ambiguous, main clause condition, may not be expected if readers reanalyse the temporarily ambiguous noun phrase as the main clause subject (Slattery et al., 2013). We return to these issues in the General Discussion, but first report Experiment 2b, which examined lingering misinterpretation during online processing.

### 2.5 Experiment 2b

Experiment 2b examines whether misinterpretation influences subsequent sentence processing. Slattery et al. (2013) used an eye-tracking paradigm whereby a continuation sentence following their critical sentences was always consistent with the correct interpretation of the temporarily ambiguous sentences. We extended their design by including not only consistent but also inconsistent sentences as in (10).

(10a) *Ambiguous, Consistent Continuation*

When the mother dressed her son at home called the dog.

It was clear that the mother was dressing herself formally for an important ceremony.

(10b) *Unambiguous, Consistent Continuation*

When the mother dressed, her son at home called the dog.

It was clear that the mother was dressing herself formally for an important ceremony.

(10c) *Ambiguous, Inconsistent Continuation*

When the mother dressed her son at home called the dog.

It was clear that the mother was dressing her son formally for an important ceremony.

(10d) *Unambiguous, Inconsistent Continuation*

When the mother dressed, her son at home called the dog.

It was clear that the mother was dressing her son formally for an important ceremony.

It was tiring.

The first sentence is either temporarily ambiguous (10a/10c) or unambiguous (10b/10d). The subordinate clause consisted of either a RAT or reciprocal verb so that the interpretation of the initially adopted analysis ("the mother dressed her son") was always incorrect after reanalysis. In (10a/b), the continuation sentence ("the mother was dressing herself") is consistent with the correct analysis of the first sentence, while the continuation sentence in (10c/d) ("the mother was dressing her son") is inconsistent with this analysis. It is however consistent with the initial misinterpretation.

Importantly, this in(consistency) may be reversed or attenuated depending on how the first sentence is interpreted and how strongly misinterpretation lingers. If the correct analysis of the first sentence is constructed, a main effect of consistency is expected in the second sentence, with longer reading times for inconsistent (10c/d) than consistent (10a/b) continuations, irrespective of ambiguity. However, if the initially assigned misinterpretation ("the mother dressed her son") lingers, inconsistency effects should be modulated by ambiguity, such that in the inconsistent conditions, reading times of the continuation sentence should be shorter for ambiguous (10c) than unambiguous sentences (10d), as a result of the continuation sentence in (10c) being misinterpreted as being consistent with the ambiguous, first sentence. For the consistent conditions, we may observe the opposite effect: i.e., if the misinterpretation ("the mother dressed her son") lingers, the ambiguous, consistent continuation condition (10a – "the mother was dressing herself") may be misperceived as being inconsistent. This would lead to longer reading times during the second sentence in (10a) than (10b). Thus, the crucial prediction is whether a main effect of consistency is observed, or an interaction between consistency and ambiguity.

Regarding L1/L2 differences, the initial misinterpretation may persist more strongly in L2ers than L1ers (Cunnings, 2017). In this case, lingering effects in consistent or inconsistent sentences may be larger for L2ers than L1ers.

### 2.5.1 Methods

*Participants*

The participants in Experiment 2b were identical to those in Experiment 2a. Experiment 2b was conducted at least one week before Experiment 2a.

*Materials*

Experimental materials comprised 24 sets of texts like (10) in a Latin square design with two levels of ambiguity (ambiguous/unambiguous) and consistency (consistent/inconsistent). Each text contained a temporarily ambiguous or unambiguous sentence on the first line, a critical continuation sentence across two lines and a wrap-up sentence at the end of the second line. For continuation sentences, the line-break appeared after the complementizer ("that"). The experiment also contained 72 filler texts, which always took up two lines on the screen. All experimental texts and two-thirds of the fillers were followed by a yes/no comprehension question that did not tap any of the critical manipulations.

*Procedure and Data analysis*

The procedure was the same as Experiment 1b. L2 participants completed a vocabulary test that tested their knowledge of the words used for the subordinate clause subject and verb, and the main clause subject.

The calculation of reading time measures, data exclusion criteria and data analysis procedure were identical to Experiment 1b, except that a fixed effect of consistency

(consistent/inconsistent) was included instead of gender. To test for garden-path effects, we specified the main clause verb of the first sentence as the disambiguating region ("called") and the rest of the sentence as the first spillover region ("the dog"). To test for consistency effects, the critical region was defined as the text that denoted the (in)consistency effect ("dressing her son/herself") and the second spillover region ("formally for an") was defined as the rest of the sentence except the last two words to avoid wrap-up effects. Skipping rates were 8% for the L1 data and 3% for the L2 data across all regions. Trials including words that L2 participants did not know were removed from analysis, which affected less than 0.1% of the L2 data.

### 2.5.2 Results

Overall accuracy rates of comprehension questions were 89% for L1 participants (range = 75–97%) and 88% for L2 participants (range = 76–97%). The reading time data and inferential analyses are summarised in Table 5 and 6.

*Disambiguating Region*

At the disambiguating and first spillover regions, there was a significant main effect of ambiguity that was qualified by a significant ambiguity by region interaction in first pass times. At the disambiguating region, first pass times were marginally longer for ambiguous than unambiguous sentences (*estimate* = 0.044, $t$ = 1.97, *SE* = 0.02, $p$ = .058), but this pattern was reversed at the spillover region (*estimate* = 0.128, $t$ = 4.36, *SE* = 0.03, $p$ < .001).

While this effect at the spillover region may appear counterintuitive, shorter first pass times for ambiguous conditions may occur if readers quickly regressed out of this region (e.g., Sturt, 2007). Indeed, consistent with this interpretation of first pass times, regression path times indicated a significant main effect of ambiguity, with longer reading

times for ambiguous sentences. This was qualified by a significant ambiguity by region interaction, with longer reading times for ambiguous than unambiguous sentences at both disambiguating (*estimate* = 0.202, *t* = 6.22, *SE* = 0.03, *p* < .001) and spillover (*estimate* = 0.471, *t* = 8.88, *SE* = 0.05, *p* < .001) regions, though the effect was larger at the spillover region. There was also a two-way interaction between ambiguity and group. However, pairwise comparisons on group showed garden paths for both groups (L1: *estimate* = 0.272, *t* = 5.04, *SE* = 0.05, *p* < .001; L2: *estimate* = 1.32, *t* = 3.67, *SE* = 0.04, *p* < .001), and the interaction seemed to be driven by larger garden-path effects for L1 participants (L1: 579ms vs. L2: 375ms)

In total viewing times, there was a significant main effect of ambiguity, an ambiguity by region interaction and a significant three-way interaction between ambiguity, group and region. However, 2×2 analysis by region showed only reliable and marginal main effects of ambiguity due to garden-path effects at the disambiguating (*estimate* = 0.315, *t* = 8.10, *SE* = 0.04, *p* < .001) and spillover (*estimate* = 0.068, *t* = 1.97, *SE* = 0.03, *p* = .058) regions.

*Continuation Sentence*

At the critical and second spillover regions in the continuation sentence, there was a significant main effect of consistency in all measures, with longer reading times in inconsistent conditions. Although we are cautious in interpreting this effect due to differences in lexical material between conditions, longer reading times for inconsistent sentences are expected here.

In first pass reading times, there was a marginal interaction between consistency, group and region. Planned comparisons by region revealed a main effect of consistency only in the critical region due to longer reading times for inconsistent than consistent sentences (critical region: *estimate* = 0.212, *t* = 6.23, *SE* = 0.03, *p* < .001; second spillover

region: *estimate* = 0.011, *t* = 0.38, *SE* = 0.03, *p* = .705). There was also an interaction between ambiguity and region. Pairwise comparisons at the two levels of region showed a main effect of ambiguity only in the critical region due to decreased reading times for ambiguous than unambiguous sentences (*estimate* = 0.041, *t* = 2.27, *SE* = 0.02, *p* = .023). Regression path duration showed a significant interaction between ambiguity and consistency. Pairwise comparisons indicated that for inconsistent sentences, reading times were significantly shorter for ambiguous than unambiguous sentences (*estimate* = 0.087, *t* = 3.40, *SE* = 0.03, *p* < .001), showing lingering misinterpretation. Reading times did not differ in consistent sentences (*estimate* = 0.002, *t* = 0.06, *SE* = 0.03, *p* = .955). This effect of lingering misinterpretation is illustrated in Figure 4. There was also a consistency by region interaction, as reading times were longer for inconsistent than consistent sentences only in the critical region (critical region: *estimate* = 0.211, *t* = 6.25, *SE* = 0.03, *p* < .001; second spillover region: *estimate* = 0.074, *t* = 1.63, *SE* = 0.05, *p* = .118).

Total viewing times showed a significant interaction between consistency and region, but planned comparisons by region revealed a main effect of consistency at critical (*estimate* = 0.317, *t* = 8.30, *SE* = 0.04, *p* < .001) and spillover (*estimate* = 0.106, *t* = 2.39, *SE* = 0.04, *p* = .024) regions due to inconsistency effects. There was also a significant interaction between ambiguity and region. Planned comparisons showed a main effect of ambiguity only at the critical region due to reduced reading times for ambiguous sentences (consistency region: *estimate* = 0.118, *t* = 440, *SE* = 0.03, *p* < .001; second spillover region: *estimate* = 0.045, *t* = 1.49, *SE* = 0.03, *p* = .144).

| | Disambiguating region (called) | | Spillover Region (the dog) | | Critical Region (dressing her son/herself) | | 2nd Spillover Region (formally for an) | |
|---|---|---|---|---|---|---|---|---|
| | L1 | L2 | L1 | L2 | L1 | L2 | L1 | L2 |
| **First pass time** | | | | | | | | |
| Ambiguous, Consistent | 296 (150) | 286 (130) | 362 (226) | 475 (256) | 237 (106) | 293 (132) | 401 (232) | 540 (359) |
| Unambiguous, Consistent | 271 (119) | 282 (130) | 424 (269) | 530 (285) | 268 (159) | 287 (138) | 409 (269) | 506 (305) |
| Ambiguous, Inconsistent | 293 (142) | 316 (162) | 363 (213) | 500 (294) | 288 (127) | 373 (220) | 420 (247) | 498 (298) |
| Unambiguous, Inconsistent | 284 (142) | 281 (132) | 421 (268) | 550 (288) | 310 (147) | 384 (214) | 405 (260) | 496 (325) |
| **Regression path time** | | | | | | | | |
| Ambiguous, Consistent | 602 (753) | 393 (360) | 1763 (1778) | 1753 (1748) | 317 (612) | 408 (687) | 552 (512) | 704 (617) |
| Unambiguous, Consistent | 363 (354) | 377 (398) | 887 (859) | 1093 (1155) | 308 (222) | 381 (288) | 540 (530) | 617 (439) |
| Ambiguous, Inconsistent | 624 (754) | 475 (466) | 1682 (1669) | 1696 (1708) | 344 (196) | 439 (436) | 625 (688) | 691 (826) |
| Unambiguous, Inconsistent | 389 (401) | 339 (288) | 835 (858) | 1065 (1023) | 429 (615) | 484 (548) | 701 (838) | 699 (479) |
| **Total viewing time** | | | | | | | | |
| Ambiguous, Consistent | 657 (399) | 771 (615) | 745 (465) | 1160 (773) | 336 (216) | 457 (291) | 639 (453) | 923 (646) |
| Unambiguous, Consistent | 457 (308) | 563 (342) | 695 (451) | 1021 (654) | 376 (253) | 477 (305) | 643 (444) | 885 (580) |
| Ambiguous, Inconsistent | 721 (466) | 835 (611) | 748 (455) | 1178 (800) | 453 (332) | 622 (458) | 705 (461) | 902 (614) |
| Unambiguous, Inconsistent | 513 (402) | 579 (390) | 761 (569) | 1068 (622) | 539 (376) | 722 (492) | 732 (441) | 1001 (661) |

Table 5. *Reading times for three eye-movement measures at three regions of texts in Experiment 2b (SDs in parentheses)*.

| | Disambiguating/Spillover Region (called / the dog) | | | Critical/2nd Spillover Region (dressing her son/herself / formally for an) | | |
|---|---|---|---|---|---|---|
| | *Estimate (SE)* | *t* | *p* | *Estimate (SE)* | *t* | *p* |
| **First pass time** | | | | | | |
| Ambiguity | 0.04 (0.02) | 2.51 | .022 | 0.01 (0.02) | 0.54 | .598 |
| Group | 0.16 (0.04) | 4.06 | < .001 | 0.21 (0.05) | 4.27 | < .001 |
| C | 0.02 (0.02) | 1.24 | .221 | 0.10 (0.02) | 4.86 | < .001 |
| Region | 0.38 (0.05) | 7.79 | < .001 | 0.36 (0.06) | 6.23 | < .001 |
| Ambiguity:Group | 0.02 (0.03) | 0.59 | .554 | 0.04 (0.03) | 1.12 | .268 |
| Ambiguity:C | 0.02 (0.03) | 0.61 | .549 | 0.01 (0.03) | 0.20 | .842 |
| Group:C | 0.03 (0.03) | 0.81 | .420 | 0.00 (0.04) | 0.11 | .910 |
| Ambiguity:Region | 0.17 (0.04) | 4.35 | < .001 | 0.06 (0.03) | 1.96 | .063 |
| Group:Region | 0.28 (0.06) | 4.75 | < .001 | 0.05 (0.04) | 1.07 | .292 |
| C:Region | 0.00 (0.03) | 0.05 | .957 | 0.22 (0.05) | 4.87 | < .001 |
| Ambiguity:Group:C | 0.04 (0.06) | 0.62 | .534 | 0.07 (0.07) | 0.97 | .343 |
| Ambiguity:Group:Region | 0.03 (0.06) | 0.44 | .666 | 0.07 (0.07) | 1.01 | .324 |
| Ambiguity:C:Region | 0.01 (0.06) | 0.11 | .910 | 0.03 (0.05) | 0.64 | .521 |
| Group:C:Region | 0.003 (0.07) | 0.05 | .964 | 0.12 (0.06) | 1.87 | .074 |
| Ambiguity:Group:C:Region | 0.12 (0.12) | 0.96 | .336 | 0.04 (0.11) | 0.32 | .750 |
| **Regression path times** | | | | | | |
| Ambiguity | 0.34 (0.03) | 9.83 | < .001 | 0.04 (0.02) | 2.18 | .042 |
| Group | 0.02 (0.06) | 0.31 | .758 | 0.18 (0.05) | 3.37 | .0012 |
| C | 0.01 (0.02) | 0.30 | .767 | 0.14 (0.03) | 4.93 | < .001 |
| Region | 0.98 (0.06) | 15.78 | < .001 | 0.46 (0.05) | 9.06 | < .001 |
| Ambiguity:Group | 0.15 (0.05) | 2.79 | .010 | 0.03 (0.04) | 0.63 | .535 |
| Ambiguity:C | 0.03 (0.05) | 0.56 | .573 | 0.09 (0.04) | 2.49 | .015 |
| Group:C | 0.01 (0.05) | 0.19 | .849 | 0.06 (0.04) | 1.45 | .150 |
| Ambiguity:Region | 0.27 (0.05) | 5.07 | < .001 | 0.06 (0.04) | 1.28 | .204 |
| Group:Region | 0.24 (0.09) | 2.55 | .013 | 0.004 (0.05) | 0.07 | .945 |
| C:Region | 0.06 (0.06) | 1.08 | .293 | 0.14 (0.06) | 2.49 | .021 |
| Ambiguity:Group:C | 0.05 (0.09) | 0.54 | .593 | 0.01 (0.07) | 0.08 | .934 |
| Ambiguity:Group:Region | 0.01 (0.13) | 0.05 | .964 | 0.005 (0.09) | 0.05 | .958 |
| Ambiguity:C:Region | 0.08 (0.10) | 0.79 | .427 | 0.10 (0.07) | 1.35 | .177 |
| Group:C:Region | 0.01 (0.12) | 0.11 | .916 | 0.04 (0.07) | 0.49 | .624 |
| Ambiguity:Group:C:Region | 0.29 (0.21) | 1.40 | .178 | 0.07 (0.14) | 0.48 | .637 |
| **Total viewing time** | | | | | | |
| Ambiguity | 0.19 (0.03) | 6.18 | < .001 | 0.08 (0.02) | 3.76 | .0012 |
| Group | 0.29 (0.07) | 4.29 | < .001 | 0.31 (0.06) | 4.94 | < .001 |
| C | 0.05 (0.03) | 1.68 | .108 | 0.21 (0.04) | 6.02 | < .001 |
| Region | 0.39 (0.06) | 7.02 | < .001 | 0.49 (0.07) | 7.48 | < .001 |
| Ambiguity:Group | 0.01 (0.06) | 0.16 | .874 | 0.02 (0.05) | 0.36 | .722 |
| Ambiguity:C | 0.003 (0.05) | 0.05 | .958 | 0.08 (0.05) | 1.53 | .141 |
| Group:C | 0.002 (0.04) | 0.05 | .958 | 0.01 (0.06) | 0.20 | .844 |
| Ambiguity:Region | 0.25 (0.03) | 7.45 | < .001 | 0.08 (0.03) | 2.46 | .022 |
| Group:Region | 0.27 (0.06) | 4.77 | < .001 | 0.02 (0.04) | 0.52 | .607 |
| C:Region | 0.02 (0.04) | 0.42 | .680 | 0.22 (0.04) | 5.36 | < .001 |
| Ambiguity:Group:C | 0.05 (0.08) | 0.54 | .592 | 0.06 (0.08) | 0.79 | .430 |
| Ambiguity:Group:Region | 0.11 (0.06) | 2.04 | .042 | 0.05 (0.06) | 0.76 | .453 |
| Ambiguity:C:Region | 0.10 (0.06) | 1.56 | .132 | 0.01 (0.06) | 0.23 | .817 |
| Group:C:Region | 0.03 (0.07) | 0.38 | .705 | 0.09 (0.06) | 1.55 | .136 |
| Ambiguity:Group:C:Region | 0.07 (0.13) | 0.60 | .559 | 0.09 (0.11) | 0.80 | .425 |

Table 6. *Summary of statistical analyses for Experiment 2b. C = Consistency.*

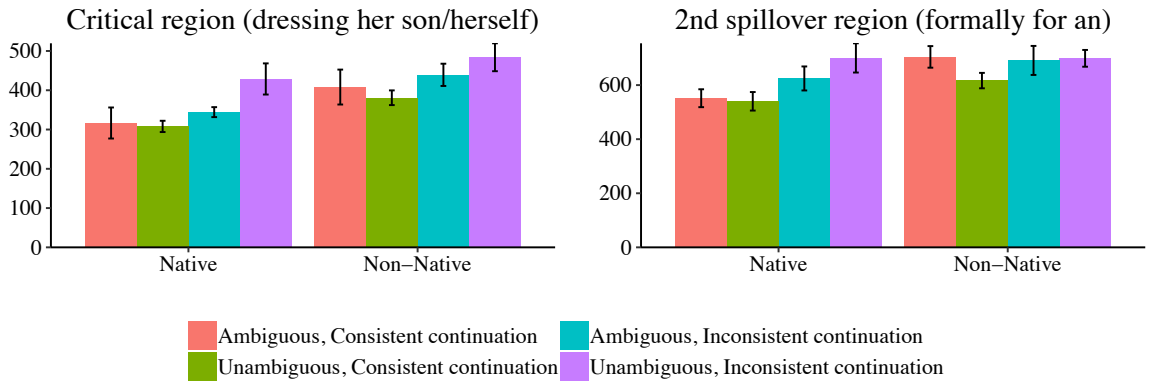


Figure 4. *Regression path duration in milliseconds at the critical and 2nd spillover regions in Experiment 2b*.

### 2.5.3 Discussion

As in Experiment 1b, Experiment 2b showed that L1 and L2 participants had more difficulty reading ambiguous than unambiguous sentences due to reanalysis cost. The significant interaction between ambiguity and group may indicate smaller garden-path effects for L2ers (Jacob & Felser, 2016).

In the second sentence, longer reading times for inconsistent conditions suggests both groups generally adopted the globally correct interpretation after reanalysis. Importantly, there was also evidence of lingering misinterpretation. This was most evident in regression path times, which were significantly shorter following ambiguous than unambiguous sentences in the inconsistent conditions. This suggests the initially assigned misinterpretation persisted to the extent that it influenced reading of the continuation sentence. Although this is compatible with misinterpretation lingering in memory, we did not fully replicate the findings of Experiment 2 reported in Slattery et al., (2013), who did not test inconsistent sentences but did report lingering effects in consistent sentences. We however, observed differences between inconsistent but not consistent conditions. Additionally, we did not find significant differences in terms of lingering misinterpretation between L1ers and L2ers. We discuss these results, along with our other findings, in more detail below.

## 2.6 General Discussion

The aims of the present study were to investigate whether L2ers have more difficulty in reanalysis than L1ers and why L2 reanalysis difficulty occurs. The two offline experiments suggested that L2ers have more difficulty reanalysing temporary ambiguities than L1ers. Experiment 1b provided evidence that both L1ers and L2ers complete syntactic reanalysis, and Experiment 2b suggested that misinterpretation persists past syntactic disambiguation and influences subsequent sentence processing. Below, the implications of these results are discussed.

### *L2 reanalysis processes*

The offline experiments reported in the present study provide some support for the claim that L2ers show greater reanalysis difficulty than L1ers (e.g., Jacob & Felser, 2016; Pozzan & Trueswell, 2016). In Experiment 1a, we found significantly lower accuracy for L2ers than L1ers in ambiguous sentences, though this was restricted to main clause questions. In Experiment 2a, although the interaction was marginal, L2ers tended to have lower comprehension accuracy than L1ers for ambiguous sentences, for both questions types. As such, the direction of effects is compatible with previous studies indicating initially assigned misinterpretations persist more often for L2ers than L1ers.

Both online experiments showed clear garden path effects. Some studies have reported different garden-path effects for L1ers and L2ers. Jacob and Felser (2016) reported larger garden-path effects for L1ers than L2ers, which may suggest L2ers do not conduct reanalysis as consistently as L1ers. Although we found a similar pattern in regression path times at the disambiguating region in Experiment 2b, this effect was not found in other measures in this experiment, nor in any measure in Experiment 1b. As such, we did not consistently find evidence that L2ers hesitate to conduct reanalysis or quit reanalysis earlier than L1ers.

We considered two accounts of L1/L2 differences. Firstly, we considered that L2ers may not complete syntactic reanalysis, and secondly, that L2ers may be more prone to interference from the initially assigned misinterpretation (Cunnings, 2017). In terms of gender mismatch effects in Experiment 1b, we did not find significant evidence to suggest L2ers reanalysed the temporarily ambiguous noun phrase less successfully than L1ers. The reading time data for the continuation sentence in Experiment 2b suggested lingering misinterpretation in both L1ers and L2ers. Although we did not find evidence of increased lingering misinterpretation in L2ers than L1ers in this experiment, as predicted by Cunnings (2017), we believe that our overall pattern of results is consistent with the idea that L1ers and L2ers attempt syntactic reanalysis to a similar degree, with lingering misinterpretation being the primary source of comprehension difficulty in both groups.

The size of the observed L1/L2 differences in comprehension accuracy, particularly for ambiguous conditions in the offline tasks, is smaller than in some previous studies (e.g. Pozzan & Trueswell, 2016). As illustrated by the standard deviations in Table 1, comprehension accuracy rates following ambiguous sentences are widely distributed for L2 participants, with some achieving comprehension rates comparable to L1 participants. This suggests that participant-specific factors may account for at least some of the L1/L2 differences. Given some studies show that individual differences influence the size of garden-path effects (Dussias & Piñar, 2012; Jegerski, 2012; Havik et al., 2009; Hopp, 2006; Roberts & Felser, 2011), how individual differences may influence both garden-path effects and lingering misinterpretation may be one key to clarifying L1/L2 differences.

*Good enough language processing in L1 and L2 comprehension*

Experiment 1a and 2a/b showed that initial misinterpretations linger in both L1 and L2 speakers at the offline and online levels. This lingering effect is compatible with the "good enough" account of language processing, which predicts that comprehenders may process

new incoming input without fully erasing representations that they previously created but which turn out to be incompatible with the globally correct interpretation (e.g., Ferreira et al., 2001). Slattery et al. (2013) claimed L1ers construct the correct syntactic structure after reanalysis and argued that lingering effects result from co-existing representations of the globally correct and initially constructed but globally incorrect interpretations. Experiment 1b replicated this finding and extended it to L2ers.

One finding from Experiments 1a and 2a, which the good enough account might not predict, is that L1ers and L2ers sometimes answered ambiguous main clause questions incorrectly. This is because if the temporarily ambiguous noun phrase is successfully reanalysed as the main clause subject (Slattery et al., 2013), we would expect high accuracy for questions about the main clause agent. In their Experiment 2, Christianson et al. (2001) tested interpretation of both the subordinate clause and the main clause in a similar way to the present offline experiments. They reported a significant interaction between question type and ambiguity. Although errors occurred following ambiguous sentences for questions tapping both clauses, errors were much more frequent for subordinate clauses (62%) than main clauses (12%), compatible with the claim that the temporarily ambiguous noun phrase is reanalysed but the initially misinterpretation lingers. However, across both our experiments, the numerical differences between error rates for subordinate clause questions (L1ers 37%, L2ers 47%) and main clause questions (L1ers 22%, L2ers 38%) were smaller. One difference between our study and Christianson et al. relates to how questions were asked. While we asked wh-questions (e.g. “Who laughed very loudly?”; Experiment 1a) or used sentence-picture matching (Experiment 2a), Christianson et al.’s questions always asked yes/no questions (equivalent to “Did the baby laugh?” for our example (5)) that always referred to the correct interpretation. This may have biased participants in Christianson et al. towards the correct interpretation (Tabor et al., 2004) more often than in our study.

While our offline results for main clause questions may thus suggest reanalysis is not always complete for L1ers and L2ers, the results from Experiment 1b, like Slattery et al. (2013), found no evidence to suggest syntactic reanalysis is incomplete during L1 or L2 processing. One possibility is that the reflexive in Experiment 1b (and Slattery et al.), in referring to the main clause subject, may have reinforced the correct interpretation of the temporarily ambiguous noun, compared to our offline experiments that did not include a reflexive. How such effects may influence lingering misinterpretation in both L1ers and L2ers may be a fruitful avenue of future research.

Finally, another issue relates to how L1ers and L2ers are prone to misinterpretation during sentence processing. This derives from different results between Experiment 2 of Slattery et al. (2013) and Experiment 2b of the present study. Specifically, while Slattery et al. showed lingering misinterpretation when reading subsequent text that was *consistent* with the globally correct analysis of the temporary ambiguity, the present study showed such effects only when the subsequent text contained *inconsistent* information. Although both effects are consistent with lingering misinterpretation, further research examining the relative size of these two effects is required. More generally, although our results are broadly consistent with "good enough" processing, these inconsistencies with previous L1 studies highlight the need for increased replication in psycholinguistics (Vasishth, Mertzen, Jäger & Gelman, 2018), in both L1 and L2 processing.

## 2.7 Conclusion

The present study reported four experiments investigating L1 and L2 reanalysis. Two offline experiments were consistent with L2ers having more difficulty reanalysing garden-path sentences than L1ers. We argued that our two eye-movement experiments showed that L1ers and L2ers complete syntactic reanalysis during online reading, and that reanalysis difficulty in both groups results at least partly from failures to erase initial

misinterpretations from memory. Taken together, we argue that our results suggest that as in L1 processing, L2ers attempt to conduct syntactic reanalysis and L2 reanalysis difficulty at least partly results from failures to erase initially assigned misinterpretations.

## *Chapter 3. Study 2: Lingering misinterpretation in native and non-native sentence processing: Evidence from structural priming*


### Abstract

Syntactic ambiguity resolution has been widely examined in native (L1) and non-native (L2) sentence comprehension. Temporarily ambiguous sentences like "When Mary dressed the baby laughed happily" cause reading difficulty at the disambiguating region, "laughed", at least in part because the subordinate clause verb ("dressed") needs to be reanalysed from a transitive to an intransitive verb. The present study investigates whether L1 and L2 speakers persist with the initially assigned transitive misinterpretation after reanalysis, using the structural priming paradigm and the eye-tracking while reading task. Reading times revealed that unambiguous prime sentences facilitated the intransitive interpretation in ambiguous target sentences, especially when prime and target sentences shared the subordinate clause verb (lexical boost). However, no such facilitation was observed when prime sentences were ambiguous, suggesting the initially assigned transitive misinterpretation was not fully discarded. Comprehension accuracy rates were not significantly influenced by priming effects. These results provide experimental evidence that the initially assigned transitive misinterpretation lingers in memory and that one source of reanalysis difficulty observed in subject-object ambiguities during L1 and L2 sentence processing is misanalysis of the temporarily ambiguous subordinate clause verb.

### 3.1 Introduction

A central aim of psycholinguistics is to reveal how readers parse sentences during online reading. One well-attested nature of sentence parsing is that readers incrementally select (e.g., Frazier & Rayner, 1982) or rank (e.g., Gibson, 1991) a certain analysis among several grammatically permissible alternatives. As a result, readers sometimes encounter processing difficulty during online reading. For example, it is known that readers have difficulty reading temporarily ambiguous garden-path sentences like (1).

(1) When Mary dressed the baby laughed happily.

The temporary ambiguity occurs at "the baby", as at this point, readers can interpret it either as the direct object of the subordinate clause verb ("dressed") or as the subject of the main clause verb ("laughed"). The globally correct interpretation turns out to be the latter at the main clause verb. However, many previous studies have shown that readers initially adopt the former interpretation and subsequently are required to abandon or rerank the selected interpretation (e.g., Frazier & Rayner, 1982). Thus, reanalysis of (1) involves reinterpretation of the temporally ambiguous noun phrase as the main clause subject rather than the direct object of the subordinate clause verb, and of the subordinate clause verb as an intransitive rather than transitive verb.

Traditionally, it was considered that the initial misinterpretation was discarded once it turned out to be incompatible with the globally correct interpretation. However, recent studies provide considerable evidence that both native (L1) and non-native (L2) speakers often persist with the initially assigned misinterpretation (e.g., Christianson, Holingworth, Halliwell & Ferreira, 2001; Jacob & Felser, 2016). Whether such lingering misinterpretation results from the same mechanisms during L1 and L2 processing has

however been debated (e.g., Cunnings, 2017; Jacob & Felser, 2016; Pozzan & Trueswell, 2016).

The present study reports an eye-tracking while reading experiment using the structural priming paradigm in language comprehension to examine reanalysis during L1 and L2 comprehension. Specifically, while previous studies have focused on reanalysis of the temporarily ambiguous noun phrase, we aimed to examine the representation of the subordinate clause verb in sentences like (1) after reanalysis. Our results showed that the initially assigned transitive misinterpretation lingers even after reanalysis in both L1 and L2 speakers. This finding provides further evidence, along with previous reanalysis studies, that both L1 and L2 speakers persist with misinterpretations even after reanalysis of garden-path sentences, with its novelty being that it provides direct evidence that reanalysis of the transitivity of the subordinate clause verb is one source of reanalysis difficulty during processing, not just reanalysis of the temporarily ambiguous noun phrase[9].

Below, we begin by discussing reanalysis and lingering misinterpretation in L1 processing, before discussing recent work on L2 processing in this domain. We then discuss how structural priming can inform our understanding of reanalysis in L1 and L2 comprehension.

[9] Reanalysis of the temporarily ambiguous noun phrase and subordinate clause verb are assumed to be different processes. For example, during reanalysis, readers, after being garden-pathed, may initially access the erroneous region (i.e., the temporarily ambiguous noun phrase). Following Slattery et al. (2013) and Study 1 of this thesis, readers successfully reanalyse this noun phrase as the main clause subject. Subsequently, readers may attempt to reanalyse the subordinate clause verb as the intransitive verb, but this reanalysis of the subordinate clause verb may not be complete (e.g., Fodor & Inoue, 1998).

### 3.1.1 Reanalysis in L1 sentence processing

Readers encounter processing difficulty when reading garden-path sentences like (1) at the main clause verb, "laughed", as they initially misinterpret the temporarily ambiguous noun phrase, "the baby", as the direct object, while the globally correct interpretation is as the main clause subject. As a result, reanalysis is required (e.g., Ferreira & Henderson, 1991; Pickering & Traxler, 1998; Sturt, Pickering, & Crocker, 1999).

One property of reanalysis, which has been recently elucidated, is that readers often persist with initially assigned misinterpretations. The first study showing this experimentally was Christianson et al. (2001). In their study, L1 participants read temporarily ambiguous sentences like (1) and unambiguous sentences like "When Mary dressed, the baby laughed happily", where the comma prevents formation of the thematic relation between the subordinate clause verb and the temporarily ambiguous noun phrase. After reading these ambiguous/unambiguous sentences, participants answered comprehension questions that referred to the initially assigned misinterpretation (e.g., Did Mary dress the baby?). The correct response to this question is "no", as "Mary dressed herself", not "the baby". However, Christianson et al. observed more "yes" responses when comprehension questions followed ambiguous than unambiguous sentences. This result suggests that the initial misinterpretation remains activated even after reanalysis and influences subsequent language comprehension. Importantly, a number of subsequent works have corroborated this "persistence of misinterpretation", using a variety of research designs (e.g., Christianson, Williams, Zacks, & Ferreira, 2006; Malyutina & den Ouden, 2016; Kaschak & Glenberg, 2004; Nakamura & Arai, 2016; Patson, Darowski, Moon, & Ferreira, 2009; Staub, 2007; Sturt, 2007; van Gompel, Pickering, Pearson, & Jacob, 2006). These studies provide evidence that readers do not always derive the correct interpretation of garden-path sentences during reanalysis.

The Good Enough Language Processing model proposed and developed by Ferreira and colleagues attempts to account for this persistence of misinterpretation (Christianson et al., 2001; Ferreira, Bailey, & Ferraro, 2002; Ferreira, Christianson, & Hollingworth, 2001; Ferreira & Patson, 2007; Karimi & Ferreira, 2016; Slattery, Sturt, Christianson, Yoshida, & Ferreira, 2013). According to this model, language comprehension is not always accurate, as readers tend to avoid creating a fully detailed representation during sentence processing to minimise processing cost (e.g., Frazier, 1979). Previously, this model attributed persistence of misinterpretation to a failure to complete syntactic reanalysis; i.e., in sentences like (1), readers misinterpret the temporarily ambiguous noun phrase as the direct object even after reanalysis (Christianson et al., 2001). However, Slattery et al. (2013) recently provided evidence against this account. In their Experiment 1, L1 participants read sentences like (2), which manipulate sentence ambiguity by inclusion or removal of the comma, along with gender (mis)match between a reflexive ("himself") and its antecedent, the temporarily ambiguous noun phrase ("David's father/mother").

(2a) After the bank manager telephoned(,) David's father grew worried and gave himself approximately five days to reply.

(2b) After the bank manager telephoned(,) David's mother grew worried and gave himself approximately five days to reply.

When the comma is absent, (2) requires reanalysis at "grew" due to misanalysis of the temporarily ambiguous noun phrase. The reflexive also requires an antecedent which, according to Binding Principle A (Chomsky,1981), must be in the local syntactic domain. Crucially, this antecedent is the temporarily ambiguous noun phrase, which either matches or mismatches the reflexive in gender ("David's father/mother"). For unambiguous sentences, reading times at the reflexive were expected to be longer in (2b) compared to

(2a), as a result of gender mismatch effects (e.g., Sturt, 2003). For ambiguous sentences, if reanalysis is syntactically incomplete, the temporarily ambiguous noun phrase would remain as the direct object after reanalysis. In this syntactic configuration, it cannot act as an antecedent for the reflexive as it would not be in the reflexive's local domain. Indeed, it can only act as an antecedent for the reflexive, if reanalysis is complete. Therefore, Slattery et al. predicted that if syntactic reanalysis is incomplete, this should lead to absent or reduced gender mismatch effects in the ambiguous conditions. Contrary to this prediction however, Slattery et al. observed gender mismatch effects in both ambiguous and unambiguous conditions. They took this as evidence that readers conduct syntactic reanalysis of the temporarily ambiguous noun phrase during online reading.

In a second experiment, Slattery et al. further investigated if persistence of misinterpretation results from a failure to erase the memory trace of the initially assigned misinterpretation, using texts like (3) that contain a continuation sentence after the temporarily ambiguous/unambiguous sentence.

(3a) While Frank dried off(,) the truck that was dark green was peed on by a stray dog. Frank quickly finished drying himself off then yelled out the window at the dog.

(3b) While Frank dried off(,) the grass that was dark green was peed on by a stray dog. Frank quickly finished drying himself off then yelled out the window at the dog.

The first sentence is either ambiguous or unambiguous. It also manipulates the temporarily ambiguous noun phrase to be either a plausible ("truck") or implausible ("grass") theme for the subordinate clause verb ("dried off"). The critical region is "himself" in the second continuation sentence, which is consistent with the globally correct interpretation of the first sentence ("Frank dried himself off") but importantly inconsistent with the initially assigned misinterpretation in the plausible, ambiguous condition ("Frank

dried off the truck"). Slattery et al. reasoned that if misinterpretation lingers in memory after reanalysis, readers may misinterpret the continuation sentence as being inconsistent with the first sentence in the plausible, ambiguous condition. Indeed, Slattery et al. reported longer reading times at the reflexive in this condition, suggesting a failure to discard the initial misinterpretation from memory. Taken together, Slattery et al. claimed that their experiments suggest syntactic reanalysis of the temporarily ambiguous noun phrase is complete, but that initially assigned misinterpretations linger in memory. Slattery et al. accounted for lingering misinterpretations based on a lexically guided tree-adjoining grammar parsing model (Ferreira, Lau, & Bailey, 2004) such that after reanalysis, the temporarily ambiguous noun phrase is syntactically reinterpreted as the main clause subject but the initially constructed direct object misparse may remain in the syntactic tree (see also Ferreira et al., 2001; Fodor & Inoue, 1998).

### 3.1.2 Reanalysis in L2 sentence processing

L2 speakers have reanalysis difficulty when reading temporarily ambiguous sentences like (1) as L1 speakers do: i.e., they initially misinterpret (1) as "Mary dressed the baby" and as a consequence are required to reanalyse it as "Mary dressed herself" at the main clause verb (e.g., Juffs & Harrington, 1996; Hopp, 2015; Roberts & Felser, 2011). Recent L2 studies have shown that L2 speakers persist with misinterpretations as L1 speakers do, with greater L2 reanalysis cost for L2 than L1 speakers (Cunnings, Fotiadou, & Tsimpli, 2017; Jacob & Felser, 2016, Pozzan & Trueswell, 2016; Roberts & Felser, 2011). Pozzan and Trueswell for example examined L2 reanalysis of prepositional phrases like (4), using the visual world paradigm.

(4a) Put the frog on the napkin onto the box.

(4b) Put the frog that's on the napkin onto the box.

(4a) causes reanalysis at the second prepositional phrase ("onto the box"), as listeners initially misinterpret the first prepositional phrase ("on the napkin") as the destination of the verb ("put") while it is in fact a modifier. (4b) is unambiguous due to the overt relativer "that". In Pozzan and Trueswell, L1 and L2 participants heard sentences like (4) while viewing a display containing the referents mentioned in the sentence, and then had to act out the instruction. Eye-movements during listening showed similar processing patterns between L1 and L2 participants, with "on the napkin" being temporarily misinterpreted as the destination of the verb in both groups. However, L2 participants performed more incorrect actions than L1 speakers following ambiguous sentences. This result suggests that L2 participants persisted with the initial misinterpretation in ambiguous sentences more often than L1 speakers.

Jacob and Felser (2016) examined L2 reanalysis of subject-object ambiguities like (5) in a reading experiment that manipulated ambiguity via the presence or absence of a comma.

(5) While the gentleman was eating(,) the burgers were still being reheated in the microwave.

Question: Was the gentleman eating the burgers?

(5) manipulates the main clause verb phrase ("were still being reheated") such that it is inconsistent with the initial misinterpretation of the subordinate clause verb phrase in the ambiguous condition ("eating the burgers"), as it is not possible to "eat the burgers" that "are being reheated in the microwave". Comprehension questions always referred to the initially assigned misinterpretation of ambiguous sentences (e.g., Christianson et al.,

2001). If L2 speakers persist with the initial misinterpretation, reading times at the main clause verb phrase may be longer for ambiguous than unambiguous sentences due to inconsistency effects. Reading times showed this inconsistency effect in both L1 and L2 participants, suggesting that L2 speakers persist with misinterpretations as L1 speakers do. Comprehension accuracy showed more incorrect responses for L2 than L1 participants. Although this group effect did not interact with ambiguity, Jacob and Felser interpreted it as indicating that L2 speakers have more difficulty recovering from garden paths than L1 speakers.

There are potentially different accounts for why L2 speakers have difficulty in reanalysis. One is that unlike L1 speakers, L2 speakers do not complete syntactic reanalysis. This possibility may be compatible with the Shallow Structure Hypothesis, which claims that L2 speakers do not construct fully specified syntactic parses during online reading (Clahsen & Felser, 2006, 2017). Indeed, some studies report weaker garden-path effects (e.g., Jacob & Felser, 2016) and greater reanalysis difficulty (e.g., Cunnings, Fotiadou, & Tsimpli, 2017; Gerth, Otto, Nam, & Felser, 2017; Pozzan & Trueswell, 2016) for L2 than L1 speakers, which may suggest that L2 speakers are reluctant to initiate or complete reanalysis (see Jacob & Felser, 2016). Another possibility, recently proposed by Cunnings (2017), is that L2 speakers conduct syntactic reanalysis successfully but are more prone to initially assigned misinterpretations lingering in memory than L1 speakers. The present study does not intend to tease apart these accounts. Rather, while previous studies have focused on reanalysis of the temporarily ambiguous noun phrase, this study aims to test specifically for whether the initially assigned transitive misinterpretation of the verb lingers after reanalysis in L2 sentence processing, which has not previously been systematically examined.

### 3.1.3 Structural priming

Structural priming refers to the phenomenon that processing of a certain grammatical structure is facilitated by (repeated) exposure to it (Bock, 1986; see Branigan & Pickering, 2017; Pickering & Ferreira, 2008 and Tooley & Traxler, 2010 for a review). For example, Bock (1986) showed that during a picture description task, participants produced more prepositional phrases after being exposed to prepositional phrases (e.g., A rock star sold some cocaine to an undercover agent.), while more double object phrases were produced after double object phrases (e.g., A rock star sold an undercover agent some cocaine.). This structural priming has been observed widely in both production and comprehension in L1 speakers (e.g., production: Bock, 1986; Bock & Griffin, 2000; Cleland & Pickering, 2003, 2006; Pickering & Branigan, 1998; comprehension: Branigan, Pickering, & McLean, 2005; Ledoux, Traxler, & Swaab, 2007; Tooley & Bock, 2014; Tooley, Swaab, Boudewyn, Zirnstein, & Traxler, 2014; Tooley, Traxler, & Swaab, 2009; Traxler, 2008; Traxler & Tooley, 2008; Traxler, Tooley, & Pickering, 2014), and more recently in L2 processing (e.g., production: McDonough, 2006; McDonough & Chaikitmongkol, 2010; McDonough & Kim, 2009; Shin & Christianson, 2012; comprehension: Nitschke, Kidd, & Serratrice, 2010; Nitschke, Serratrice, & Kidd, 2014; Weber & Indefrey, 2009; Wei, Boland, & Brennan, 2018; Wei, Boland, Cai, Yuan, & Wang, 2019). Structural priming in language comprehension is often indicated by decreased reading times, showing facilitated sentence processing and end-of-sentence language comprehension indicative of regularisation to a certain interpretation. For example, Traxler (2015) tested whether L1 structural priming occurs in early closure sentences like (6a), compared with late closure sentences like (6b) during self-paced reading,

(6a) As Jason watched the birds came closer and closer.

(6b) As Jason watched the birds the fox came closer and closer.

(6a) is similar to (1) in that reanalysis occurs at the main clause verb ("came"). No such reading difficulty is expected in (6b) because another noun phrase ("the fox"), which is inserted following the theme of the subordinate clause verb, acts as the subject of the main clause predicate. Traxler observed reduced reading times when L1 participants read early closure sentences consecutively, compared with when early closure sentences followed late closure sentences.

One factor that influences priming is lexical overlap between prime and target sentences. Priming effects are larger when prime and target sentences share a certain word (e.g., Pickering & Branigan, 1998). This lexical boost may be particularly important in structural priming during language comprehension, as some studies report absence of structural priming in language comprehension when prime and target sentences do not share a critical word (e.g., Arai et al., 2007; Branigan et al., 2005; Ledoux et al., 2007; Traxler, 2015). However, recent studies tend to show both lexically-mediated and abstract structural priming in language comprehension (e.g., Pickering, McLean, & Branigan, 2013; Tooley & Bock, 2014; Traxler, 2008).

There are broadly two accounts for the cause of structural priming. One is Pickering and Branigan (1998)'s residual activation account. This account assumes that the recently represented word and its related structure remains activated after being processed for a short term. As a result, these are processed more easily than before exposure. Their account directly explains the lexical boost effect such that when prime and target sentences share the same content word, both lexical and structural information will be primed. However, when no content word is shared between prime and target sentences, only abstract structural information will be primed. The other account is Bock and Griffin (2000) and Chang, Dell and Bock (2006)'s implicit learning account. Contrary to Pickering and Branigan, this account assumes that priming results from long-term implicit learning

effects: i.e., when a certain structure is processed repeatedly, readers regularise to the structure, which results in (cumulative) priming. As priming effects in this account are thought to be error-driven, priming becomes larger when what is processed differs from what is predicted due to a greater adjustment to it. As the implicit learning account predicts long-term priming, it does not account for the increased priming effects due to the lexical overlap based on the lexical boost, which is short-lived. Instead, it claims that the cause of the lexical boost is explicit memory being activated by the repeated word.

Importantly, both accounts predict what is represented in the comprehension system is the source of priming. van Gompel et al. (2006) made use of this to investigate how garden-path sentences like (1) are represented after reanalysis at the production level. In their study, L1 participants read temporarily ambiguous/unambiguous sentences like (7a) and then completed target fragments like (7b).

(7a) While the man was visiting(,) the children who were surprisingly pleasant and funny played outside.

(7b) When the doctor was visiti…

van Gompel et al. reasoned that if L1 speakers fully discard the initial transitive misinterpretation after reanalysis, temporarily ambiguous prime sentences should facilitate the intransitive interpretation to the same degree as unambiguous sentences do. However, van Gompel et al. observed a higher proportion of transitive sentence productions after ambiguous than unambiguous sentences, irrespective of whether or not the subordinate clause verb was shared between prime sentences and target fragments. They took this result as evidence that the initially assigned transitive misinterpretation remains activated after reanalysis.

Fewer studies have examined structural priming in L2 speakers. However, most relevant to current purposes, existing studies have shown that L2 speakers are similarly subject to immediate and long-lasting structural priming in language comprehension (Nitschke et al, 2010; Weber & Indefrey, 2009; Wei et al., 2019). Some studies have shown lexically-independent structural priming during comprehension (e.g., Nitschke et al., 2014; Wei et al., 2019), while others suggest lexically-mediated structural priming (e.g., Wei et al., 2018). Thus, as also found in L1 studies of priming during comprehension discussed above, priming during comprehension in L2 speakers appears stronger in the presence of lexical overlap. These existing studies examined structural priming effects on interpretation and processing of (reduced) relative clauses. However, no study, to our knowledge, has directly examined how structural priming affects garden-path sentences like (1) in L1 and L2 speakers.

### 3.1.4 The present study

Many previous studies have shown that L1 and L2 readers encounter reading difficulty when reading temporarily ambiguous sentences and often persist with misinterpretations after reanalysis. However, currently little research has examined the nature of reanalysis during L1 and L2 language comprehension. As illustrated above, temporarily ambiguous sentences, such as “When May dressed the baby laughed happily”, involve reanalysis of “the baby” as the main clause subject and “dressed” as an intransitive verb. However, most previous studies do not tease apart these two aspects of reanalysis processes. One of the few exceptions is Slattery et al. (2013), which specifically examined reanalysis of the temporarily ambiguous noun phrase, making use of Binding Principle A and gender mismatch effects.

The present study aimed to examine the nature of reanalysis during language comprehension in L1 and L2 processing. Specifically, we examine representation of the

temporarily ambiguous subordinate clause verb after reanalysis of garden-path sentences like (1). Slattery et al. revealed that the temporarily ambiguous noun phrase is successfully reinterpreted as the main clause subject. However, whether readers reinterpret the initially assigned transitive verb as intransitive is unexplored. Slattery et al. assumed that the subordinate clause verb may remain as transitive even after reanalysis, because the temporarily ambiguous noun phrase is interpreted both as the subject in the main clause and as the direct object in the subordinate clause. We explored this possibility in L1 and L2 language comprehension, using the structural priming (van Gompel et al., 2006) and eye-tracking while reading tasks. We reasoned that if representation of the initial transitive misinterpretation lingers in memory even after reanalysis, ambiguous prime sentences should not prime the intransitive interpretation of a subsequent garden-path sentence to the same degree as unambiguous prime sentences do during online reading. Note, we do not aim to distinguish the residual activation and implicit learning accounts for structural priming. Also, the present study differs from Traxler (2015) in that in Traxler, the subordinate clause verb of unambiguous sentences is always interpreted as transitive, while our present study uses a comma after the subordinate clause verb of unambiguous sentences, which should be interpreted intransitively. Traxler's manipulation indicates that ambiguous sentences may prime the intransitive interpretation to some degree but does not show whether the initial transitive misinterpretation lingers. Our research design can inform us as to whether readers persist with the initial transitive misinterpretation during sentence comprehension.

Another aim of the present study is to examine the nature of L2 reanalysis and compare L1 and L2 speakers. As stated above, previous studies have shown that L2 speakers often persist with misinterpretations as L1 speakers do. However, the cause of L2 lingering misinterpretation is debated. If L2 lingering misinterpretation results at least partially from a failure to discard the initial transitive misinterpretation of the subordinate

clause verb, ambiguous sentences should not prime the intransitive interpretation to the same degree as unambiguous sentences, as also predicted for L1 processing.

### 3.2 Experiment

#### 3.2.1 Method

*Participants*

48 native English speakers (9 males, mean age = 20; range = 18–48) and 48 L2 English speakers (12 males, mean age = 21; range = 17–36) of various L1 backgrounds[10], from the University of Reading community, participated for either course credit or payment. The L2 participants started learning English in a school environment after age five onwards. After the main experiment, the L2 participants completed the Quick Placement test (Quick placement test: Version 1, 2004) to measure their L2 proficiency. This indicated an average score 48 out of 60 (range = 31–59), showing that L2 participants were upper intermediate to advanced English language learners.

*Materials*

We created 36 sets of experimental texts as in (8/9). Each set contains a prime sentence and a target sentence manipulating ambiguity and lexical overlap, resulting in 6 conditions.

*(8a) Lexical Overlap, Prime-Unambiguous, Target-Unambiguous*

While James washed, his child waited very quietly in the bathroom.

After the lady washed, the dog started eating some food quickly.

[10] First languages of the L2 participants were Greek (11), Italian (6), Bulgarian (3), German (3), Romance (3), Cantonese (2), Danish (2), French (2), Polish (2), Russian (2), Slovak (2), Spanish (2), Bahasa (1), Chinese (1), Croatian (1), Dutch (1), Lithuanian (1), Malay (1), Portuguese (1), Sinhala (1).

*(8b) Lexical Overlap, Prime-Unambiguous, Target-Ambiguous*

While James washed, his child waited very quietly in the bathroom.

After the lady washed the dog started eating some food quickly.

*(8c) Lexical Overlap, Prime-Ambiguous, Target-Ambiguous*

While James washed his child waited very quietly in the bathroom.

After the lady washed the dog started eating some food quickly.

*(9a) Non-Lexical Overlap, Prime-Unambiguous, Target-Unambiguous*

While James called, his child waited very quietly in the bathroom.

After the lady washed, the dog started eating some food quickly.

*(9b) Non-Lexical Overlap, Prime-Unambiguous, Target-Ambiguous*

While James called, his child waited very quietly in the bathroom.

After the lady washed the dog started eating some food quickly.

*(9c) Non-Lexical Overlap, Prime-Ambiguous, Target-Ambiguous*

While James called his child waited very quietly in the bathroom.

After the lady washed the dog started eating some food quickly.

Question: Did the lady wash the dog?

The first sentence is a prime sentence followed by the second, target sentence. Each sentence was presented as a separate trial. Participants first read the prime sentence in full, pressing a button once complete. They then separately read the target sentence, again

pressing a button once complete. Target (but not prime) sentences were followed by a comprehension question referring to the initial misinterpretation of the sentence (Christianson et al., 2001) that required a yes/no push button response.

The subordinate clause verb ("washed/called") is identical across prime and target sentences in (8a/b/c) but different in (9a/b/c). Prime sentences in (8c) and (9c) are ambiguous while those in (8a/b) and (9a/b) are unambiguous. Target sentences also manipulated ambiguity such that (8b/c/) and (9b/c) are ambiguous and (8a) and (9a) are unambiguous. The subordinate clause verb of target sentences always consisted of either a reflexive absolute transitive (RAT) verb or a reciprocal verb such as "wash" and "hug" (Ferreira & McClure, 1997). When being used without any direct object, RAT verbs must be interpreted reflexively. For example, "the lady washed" in (8) must be interpreted as "the lady washed herself", unlike "the lady called", which can take another implied noun phrase as the direct object. Similarly, reciprocal verbs need to be interpreted reciprocally when the subject is plural. For example, "the lady and the girl hugged" can only mean "the lady and the girl hugged each other". These verb properties allow each comprehension question to have only one absolute correct answer, which is crucial in examining the final interpretation of garden-path sentences (Christianson et al., 2001). The full set of experimental items used for the present study is included in Appendix.

Both prime and target sentences were presented on one line of text. The experiment also contained 86 filler sentences of which two sentences were structurally similar to target ambiguous sentences and two sentences to target unambiguous sentences. Two to four filler sentences appeared between each set of experimental sentences. 56 filler sentences were followed by a yes/no comprehension question that asked about different parts of the sentences equally.

We expected longer reading times at the disambiguating verb in both prime and target sentences following ambiguous than unambiguous sentences, as evidence of garden-

path effects (e.g., Frazier & Rayner, 1982). We also expected lower accuracy rates for comprehension questions following ambiguous than unambiguous target sentences, as evidence of lingering misinterpretation (Christianson et al., 2001; Jacob & Felser, 2016). If the initial transitive misinterpretation lingers in language comprehension (van Gompel et al., 2006), ambiguous target sentences should have longer reading times and higher error rates when they follow ambiguous (8c/9c) than unambiguous (8b/9b) prime sentences. Similarly, if this lingering transitive misinterpretation affects offline comprehension (e.g., Christianson et al., 2001), participants may answer target sentence comprehension questions less accurately after ambiguous (8c/9c) than unambiguous (8b/9b) prime sentences. Also, if structural priming in language comprehension is fully lexically dependent (e.g., Arai et al., 2007), any priming effect should be observed only in (8a/b/c). L1/L2 differences were expected to emerge such that L2 participants would persist with misinterpretations more strongly than L1 participants (e.g., Pozzan & Trueswell, 2016).

*Procedure*

Eye-movements were recorded from the participant's right eye though viewing was binocular, using an SR Research Eyelink 1000. Before the experiment began, calibration of the eye-tracker was conducted on a nine-point grid. Recalibration was conducted where appropriate between trials. Care was taken not to conduct recalibration between prime and target sentences to avoid priming effects being reduced due to lapse in concentration. Before each sentence appeared, a gaze trigger was presented above the first word of the sentence. Once participants fixated on it, the sentence appeared. Participants pressed a button on a game pad once they completed reading each sentence. Either the next sentence or a yes/no comprehension question then appeared onscreen as appropriate. Participants answered the question by pressing a button on a game pad. After the main experiment, L2 participants looked through a vocabulary list containing words used for the subordinate

clause verb (e.g., “washed/called”) to check if there was any word that they did not know, and then completed the Quick Placement test. Experimental and filler sentences were presented in a pseudo-randomised order with a Latin-square design. The entire experiment lasted approximately 40 minutes with an additional 20 minutes for the Quick Placement test.

*Data analysis*

Each prime and target sentence was divided into two regions for analysis. The disambiguating region consisted of the main clause verb (“waited/started”) while a spillover region (“very quietly in/eating some”) was defined as the lexical material after the disambiguating region except the last two words of each sentence, to minimise end-of-trial effects influencing reading times. From the recorded eye-movements, we calculated three reading times measures: first pass reading time, the sum of fixations within a region entered from the left up until an eye-movement away from the region; regression path duration, the summed duration of all fixations measured from when a region is first fixated from the left, up until but not including the first fixation in a region to the right; and total viewing times, the summed duration of all fixations in a region. Prior to the calculation of reading time measures, fixations shorter than 80ms that were within one degree of visual arc of another fixation were merged. Any other fixations below 80ms or above 800ms were then removed. Any region that a participant skipped was also removed from data analysis, which affected 9% of the L1 data and 6% of the L2 data. Trials including a subordinate clause verb that L2 participants did not know were also removed, which resulted in 0.1% data loss of the L2 data.

Dependent variables were reading times and comprehension accuracy rates. Data analysis was conducted in R (R Core Team, 2019). We fit linear mixed effect models (Baayen, Davidson, & Bates, 2008) to reading times after log-transforming them.

Accuracy rates were fit to binomial distributions using generalised linear mixed effect models. For prime sentences, the models included fixed effects of ambiguity to test for garden-path effects (ambiguous/unambiguous), group (L1/L2), and lexical overlap (same/different). For target sentences, in addition to these fixed effects, the models also included another fixed effect of prime sentence ambiguity (ambiguous/unambiguous). For reading times, to minimise the number of independent statistical tests run on eye-movement data (see von der Malsburg & Angele, 2017), we also included region (disambiguating region/spillover region) as an additional fixed effect (see Cunnings & Sturt, 2018; Paape, Nicenboim, & Vasishth, 2017).

The fixed effects for prime sentences were all sum coded (-1/1). For target sentences, fixed effects of group, lexical overlap and region were sum coded likewise. For ambiguity, we adopted helmert coding with two contrasts, one (ambiguity) that compared unambiguous and ambiguous conditions, and a second that tested for priming effects within the ambiguous conditions (i.e. effect of unambiguous vs. ambiguous prime sentences on ambiguous target sentences).

All models were fit using the maximal random effects structure that converged, including by-subject and by-item random intercepts and random slopes for each within-subject and within-item fixed effect (Barr, Levy, Scheepers, & Tily, 2013).[11] In addition, as analysing region as a fixed effect includes two non-independent data points from the same trial, a by-trial intercept and a random slope of region under subject, item and trial

[11] If the maximal model failed to converge, we first removed the random correlation parameters. If this model still did not converge, the by-trial random slope for region was initially removed under the assumption that estimation of this random slope is the main cause of convergence failure, given that it consists of only two data points from each trial. If convergence failure still occurred, we iteratively removed the random effect accounting for the least variance until convergence was achieved.

were also included. For each fixed effect, *p* values were estimated using the Satterthwaite approximation implemented by the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017).

When ambiguity interacted with region or lexical overlap, follow-up analysis was performed at the two levels of region or lexical overlap to examine effects of ambiguity at each region and lexical overlap condition, respectively. In the case of an interaction between ambiguity and group, pairwise comparisons were planned at the two levels of ambiguity to test for L1/L2 differences. Data and analysis code for all experiments reported here is available at the first author's Open Science Framework webpage (https://osf.io/dqnsj/).

### 3.2.2 Results

Mean accuracy rates for filler sentences were 91% for L1 participants (range = 74–98%) and 92% for L2 participants (range = 75–100%), indicating that participants paid attention during the experiment. The reading time and comprehension accuracy data and inferential statistics are provided in Tables 1–6. Below, we do not discuss main effects of region or group by region interactions, as these are not related to our research questions and have little meaning unless they interact with another fixed effect of theoretical interest. For brevity, there were main effects of group in most of the measures and regions, as reading times were longer for L2 than L1 participants.

*Prime sentences*

First pass reading times showed a main effect of ambiguity due to longer reading times for ambiguous than unambiguous sentences.

In regression path duration, there was a significant interaction between ambiguity and region. Pairwise comparisons by region showed a significant main effect of ambiguity

only in the spillover region (disambiguating region: *estimate* = -0.044, *SE* = 0.02, *t* = -1.54, *p* = .133, 38ms; spillover region: *estimate* = 0.101, *SE* = 0.03, *t* = 2.92, *p* = .005, 71ms). There was a marginal ambiguity by group interaction, suggesting larger ambiguity effects for L1 (reading time differences between ambiguous and unambiguous sentences: 39ms) than L2 participants (5ms) when averaged across both regions.

Total viewing times showed a main effect of ambiguity, as ambiguous sentences induced longer reading times. This main effect was modulated by a significant interaction with region due to a larger garden-path effect for the spillover region (*estimate* = 0.351, *SE* = 0.04, *t* = 9.16, *p* < .001; 380ms) than the disambiguating region (*estimate* = 0.236, *SE* = 0.04, *t* = 6.05, *p* < .001; 156ms).

In sum, reading times for the prime sentences indicated garden-path effects for both L1 and L2 participants.

| | Disambiguating region (waited) | | | | Spillover region (very quietly in) | | | |
|---|---|---|---|---|---|---|---|---|
| | Native | | Non-Native | | Native | | Non-Native | |
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| *First Pass Time* | | | | | | | | |
| Same Verb, Unambiguous | 265 | 146 | 300 | 175 | 412 | 254 | 485 | 285 |
| Same Verb, Ambiguous | 275 | 154 | 287 | 150 | 450 | 267 | 540 | 334 |
| Different Verb, Unambiguous | 264 | 144 | 308 | 206 | 416 | 282 | 505 | 298 |
| Different Verb, Ambiguous | 267 | 144 | 307 | 180 | 447 | 255 | 494 | 283 |
| *Regression Path Duration* | | | | | | | | |
| Same Verb, Unambiguous | 393 | 314 | 431 | 368 | 585 | 607 | 696 | 547 |
| Same Verb, Ambiguous | 389 | 301 | 385 | 269 | 642 | 597 | 746 | 628 |
| Different Verb, Unambiguous | 419 | 380 | 451 | 376 | 558 | 460 | 647 | 467 |
| Different Verb, Ambiguous | 374 | 283 | 395 | 295 | 680 | 680 | 704 | 639 |
| *Total Viewing Time* | | | | | | | | |
| Same Verb, Unambiguous | 449 | 356 | 591 | 481 | 689 | 477 | 911 | 600 |
| Same Verb, Ambiguous | 589 | 578 | 739 | 579 | 1031 | 915 | 1331 | 1011 |
| Different Verb, Unambiguous | 441 | 322 | 607 | 476 | 703 | 515 | 916 | 608 |
| Different Verb, Ambiguous | 640 | 609 | 743 | 629 | 1078 | 877 | 1298 | 933 |

Table 1. *Reading times of prime sentences for three eye-movement measures at two regions of text*.

| | *First pass time* | | | | *Regression path time* | | | | *Total viewing time* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *E* | *SE* | *t* | *p* | *E* | *SE* | *t* | *p* | *E* | *SE* | *t* | *p* |
| Amb | 0.045 | 0.014 | 3.270 | **0.003** | 0.030 | 0.021 | 1.445 | 0.157 | 0.292 | 0.029 | 10.118 | **<0.001** |
| LO | 0.003 | 0.009 | 0.302 | 0.765 | -0.007 | 0.011 | -0.617 | 0.542 | 0.002 | 0.013 | 0.189 | 0.852 |
| Group | 0.070 | 0.018 | 3.862 | **<0.001** | 0.060 | 0.023 | 2.680 | **0.009** | 0.133 | 0.032 | 4.115 | **<0.001** |
| R | 0.235 | 0.027 | 8.607 | **<0.001** | 0.231 | 0.033 | 7.114 | **<0.001** | 0.271 | 0.038 | 7.160 | **<0.001** |
| Amb: LO | -0.015 | 0.015 | -0.973 | 0.339 | -0.008 | 0.019 | -0.407 | 0.687 | 0.006 | 0.016 | 0.378 | 0.708 |
| Amb:Group | -0.021 | 0.013 | -1.583 | 0.123 | -0.032 | 0.018 | -1.799 | 0.075 | -0.025 | 0.026 | -0.978 | 0.332 |
| LO:Group | 0.003 | 0.006 | 0.489 | 0.625 | -0.004 | 0.008 | -0.581 | 0.562 | -0.002 | 0.008 | -0.260 | 0.796 |
| Amb:R | 0.030 | 0.019 | 1.582 | 0.122 | 0.071 | 0.024 | 2.929 | **0.006** | 0.059 | 0.026 | 2.283 | **0.028** |
| LO:R | 0.002 | 0.009 | 0.183 | 0.856 | -0.013 | 0.011 | -1.164 | 0.252 | 0.001 | 0.010 | 0.134 | 0.894 |
| Group:R | 0.018 | 0.008 | 2.223 | **0.029** | 0.026 | 0.011 | 2.266 | **0.027** | 0.011 | 0.009 | 1.218 | 0.228 |
| Amb:LO:Group | -0.006 | 0.013 | -0.472 | 0.638 | -0.002 | 0.018 | -0.132 | 0.896 | -0.013 | 0.017 | -0.766 | 0.449 |
| Amb:LO:R | -0.018 | 0.014 | -1.287 | 0.206 | 0.013 | 0.020 | 0.623 | 0.537 | -0.002 | 0.015 | -0.104 | 0.918 |
| Amb:Group:R | -0.015 | 0.017 | -0.875 | 0.387 | -0.011 | 0.017 | -0.640 | 0.527 | -0.003 | 0.015 | -0.195 | 0.847 |
| LO:Group:R | -0.004 | 0.007 | -0.516 | 0.607 | -0.010 | 0.008 | -1.221 | 0.225 | -0.002 | 0.006 | -0.265 | 0.793 |
| Amb:LO:Group:R | -0.020 | 0.013 | -1.530 | 0.126 | -0.014 | 0.017 | -0.776 | 0.440 | 0.004 | 0.012 | 0.333 | 0.739 |

Table 2. *Summary of statistical analyses for prime sentences*. *LO = Lexical Overlap, Amb = Ambiguity and R = Region*.

*Target sentences*

First pass reading time revealed a main effect of ambiguity, as ambiguous sentences induced longer reading times. This was qualified by a significant four-way interaction between ambiguity, group, lexical overlap and region. As planned, we conducted additional analyses for each region. For the disambiguating region, this analysis showed a significant main effect of ambiguity due to garden-path effects (*estimate* = -0.095, *SE* = 0.02, *t* = -6.00, *p* < .001) and a significant interaction between ambiguity and lexical overlap (*estimate* = 0.030, *SE* = 0.02, *t* = 1.97, *p* = .049). There was also a marginal three-way interaction between ambiguity, lexical overlap and group (*estimate* = 0.033, *SE* = 0.02, *t* = 2.01, *p* = .053). Separate analyses at the two levels of lexical overlap showed that for the lexical-overlap condition, there was a further significant ambiguity by group interaction (*estimate* = -0.053, *SE* = 0.02, *t* = -2.43, *p* = .015), as garden-path effects were larger for L2 (*estimate* = -0.177, *SE* = 0.03, *t* = -5.57, *p* < .001; 63ms) than L1 participants (*estimate* = -0.071, *SE* = 0.03, *t* = -2.20, *p* = .035; 21ms). The non-overlap condition showed a significant main effect of ambiguity due to reanalysis cost (*estimate* = -0.065, *SE* = 0.02, *t* = -3.02, *p* = .003), in the absence of any further significant effects (all *t* < 0.66, all *p* > .507). No effects were significant for the spillover region (all *t* < 1.57, all *p* > .127).

In regression path duration, the only fully significant effect of theoretical interest was the effect of ambiguity, with longer reading times in ambiguous than unambiguous sentences. There was a marginal effect of priming, suggesting longer reading times for ambiguous target sentences that followed ambiguous than unambiguous primes. There was also a marginal four-way interaction for the ambiguity effect. Additional analysis by region showed that for the disambiguating region, there was a significant three-way interaction between ambiguity, lexical overlap and group (*estimate* = 0.052, *SE* = 0.02, *t* = 2.17, *p* = .037). Further follow-up analyses by lexical overlap showed main effects of ambiguity in both overlap conditions (lexical overlap: *estimate* = -0.291, *SE* = 0.04, *t* = -8.11, *p* < .001;

non-lexical overlap: *estimate* = -0.290, *SE* = 0.03, *t* = -8.66, *p* < .001), with a further marginal ambiguity by group interaction in non-overlap conditions (*estimate* = 0.055, *SE* = 0.03, *t* = 1.65, *p* = .100). Planned comparisons at the two levels of group showed a significant main effect of ambiguity for each group (L1: *estimate* = -0.345, *SE* = 0.05, *t* = -6.82, *p* < .001; L2: *estimate* = -0.234, *SE* = 0.05, *t* = -5.05, *p* < .001), with effects being larger for L1 (225ms) than L2 speakers (188ms). The spillover region showed a significant main effect of ambiguity due to increased reading times for ambiguous sentences (*estimate* = -0.241, *SE* = 0.02, *t* = -9.96, *p* < .001).

Total viewing times showed a main effect of ambiguity due to reanalysis cost, which was modified by a significant four-way interaction. There was also a significant three-way interaction between priming, lexical overlap and region.

For the four-way interaction with ambiguity, analysis by region showed that there was a three-way interaction between ambiguity, lexical overlap and group in the disambiguating region (*estimate* = 0.059, *SE* = 0.02, *t* = 2.90, *p* = .007). Further comparisons by lexical overlap showed that for the lexical-overlap condition, there was a two-way interaction between ambiguity and group (*estimate* = -0.103, *SE* = 0.03, *t* = -3.31, *p* = .0013) due to larger garden-path effects for L2 (*estimate* = -0.464, *SE* = 0.05, *t* = -10.10, *p* < .001; 285ms) than L1 participants (L1: *estimate* = -0.256, *SE* = 0.04, *t* = -6.18, *p* < .001; 147ms). For the non-lexical-overlap condition, there was a main effect of ambiguity due to garden paths (*estimate* = -0.328, *SE* = 0.03, *t* = -9.90, *p* < .001). The spillover region similarly showed garden-path effects irrespective of lexical overlap (*estimate* = -0.099, *SE* = 0.02, *t* = -4.62, *p* < .001), with a marginal interaction by group (*estimate* = -0.037, *SE* = 0.02, *t* = -1.71, *p* = .096), as garden-path effects were larger for L2 (*estimate* = -0.135, *SE* = 0.03, *t* = -4.29, *p* < .001; 109ms) than L1 participants (*estimate* = -0.069, *SE* = 0.03, *t* = -1.99, *p* = .058; 48ms).

For the three-way interaction (priming), analysis by region showed a significant ambiguity by lexical overlap interaction in the disambiguating region (*estimate* = 0.029, *SE* = 0.013, *t* = 2.20, *p* = .036). Further analysis by lexical overlap showed a significant effect of priming due to longer reading times for ambiguous target sentences following ambiguous than unambiguous prime sentences in lexical-overlap but not non-lexical overlap conditions. (lexical overlap: *estimate* = -0.048, *SE* = 0.02, *t* = -2.65, *p* = .012; non-lexical overlap: *estimate* = 0.011, *SE* = 0.02, *t* = 0.58, *p* = .568). This priming effect found in both groups for lexical overlap conditions only is shown in Figure 1. The spillover region did not show any significant nor marginal effects related to priming (all $t < 0.81$, all $p > .418$).

| | Disambiguating region (started) | | | | Spillover region (eating some) | | | |
|---|---|---|---|---|---|---|---|---|
| | Native | | Non-Native | | Native | | Non-Native | |
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| *First Pass Time* | | | | | | | | |
| SV, Unambiguous Prime, Unambiguous Target | 287 | 147 | 277 | 133 | 392 | 246 | 426 | 214 |
| SV, Unambiguous Prime, Ambiguous Target | 298 | 143 | 329 | 164 | 389 | 245 | 430 | 278 |
| SV, Ambiguous Prime, Ambiguous Target | 319 | 177 | 351 | 218 | 406 | 281 | 418 | 218 |
| DV, Unambiguous Prime, Unambiguous Target | 271 | 120 | 300 | 146 | 403 | 261 | 433 | 215 |
| DV, Unambiguous Prime, Ambiguous Target | 300 | 158 | 322 | 165 | 366 | 220 | 419 | 225 |
| DV, Ambiguous Prime, Ambiguous Target | 294 | 132 | 322 | 163 | 377 | 230 | 428 | 231 |
| *Regression Path Duration* | | | | | | | | |
| SV, Unambiguous Prime, Unambiguous Target | 340 | 226 | 332 | 247 | 576 | 652 | 530 | 369 |
| SV, Unambiguous Prime, Ambiguous Target | 504 | 459 | 553 | 644 | 704 | 761 | 805 | 857 |
| SV, Ambiguous Prime, Ambiguous Target | 506 | 599 | 579 | 675 | 790 | 778 | 796 | 881 |
| DV, Unambiguous Prime, Unambiguous Target | 310 | 199 | 363 | 263 | 487 | 321 | 493 | 309 |
| DV, Unambiguous Prime, Ambiguous Target | 532 | 517 | 539 | 696 | 719 | 716 | 833 | 977 |
| DV, Ambiguous Prime, Ambiguous Target | 538 | 577 | 564 | 615 | 793 | 829 | 822 | 869 |
| *Total Viewing Time* | | | | | | | | |
| SV, Unambiguous Prime, Unambiguous Target | 387 | 243 | 393 | 249 | 621 | 439 | 642 | 368 |
| SV, Unambiguous Prime, Ambiguous Target | 492 | 395 | 654 | 508 | 647 | 469 | 786 | 503 |
| SV, Ambiguous Prime, Ambiguous Target | 576 | 511 | 702 | 583 | 679 | 536 | 749 | 499 |
| DV, Unambiguous Prime, Unambiguous Target | 369 | 221 | 451 | 299 | 609 | 431 | 662 | 398 |
| DV, Unambiguous Prime, Ambiguous Target | 566 | 496 | 660 | 517 | 665 | 527 | 751 | 461 |
| DV, Ambiguous Prime, Ambiguous Target | 552 | 406 | 630 | 419 | 664 | 455 | 758 | 451 |

Table 3. *Reading times of target sentences for three eye-movement measures at two regions of text. SV = Same Verb and DV = Different Verb.*

| | *First Pass Time* | | | | *Regression Path Duration* | | | | *Total Viewing Time* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *E* | *SE* | *t* | *p* | *E* | *SE* | *t* | *p* | *E* | *SE* | *t* | *p* |
| Amb | -0.031 | 0.012 | -2.517 | **0.014** | -0.264 | 0.021 | -12.679 | **<0.001** | -0.221 | 0.021 | -10.759 | **<0.001** |
| P | -0.011 | 0.007 | -1.698 | 0.101 | -0.017 | 0.010 | -1.670 | 0.098 | -0.012 | 0.010 | -1.154 | 0.258 |
| LO | -0.007 | 0.006 | -1.266 | 0.214 | -0.001 | 0.008 | -0.132 | 0.895 | 0.007 | 0.008 | 0.795 | 0.432 |
| Group | 0.049 | 0.018 | 2.743 | **0.007** | 0.035 | 0.025 | 1.412 | 0.161 | 0.083 | 0.028 | 2.989 | **0.004** |
| R | 0.126 | 0.018 | 6.813 | **<0.001** | 0.181 | 0.022 | 8.330 | **<0.001** | 0.143 | 0.021 | 6.752 | **<0.001** |
| Amb:LO | 0.025 | 0.011 | 2.203 | **0.028** | -0.018 | 0.017 | -1.087 | 0.277 | 0.006 | 0.017 | 0.374 | 0.710 |
| P:LO | 0.006 | 0.007 | 0.933 | 0.353 | 0.007 | 0.013 | 0.517 | 0.609 | 0.010 | 0.009 | 1.084 | 0.278 |
| Amb:Group | -0.009 | 0.012 | -0.704 | 0.484 | -0.005 | 0.018 | -0.304 | 0.762 | -0.040 | 0.019 | -2.096 | **0.039** |
| P:Group | 0.003 | 0.008 | 0.454 | 0.653 | 0.000 | 0.010 | 0.007 | 0.995 | 0.009 | 0.010 | 0.859 | 0.397 |
| LO:Group: | 0.006 | 0.005 | 1.162 | 0.253 | 0.004 | 0.008 | 0.483 | 0.629 | 0.000 | 0.008 | 0.026 | 0.980 |
| Amb:R | 0.064 | 0.014 | 4.545 | **<0.001** | 0.024 | 0.017 | 1.462 | 0.144 | 0.123 | 0.013 | 9.677 | **<0.001** |
| P:R | 0.001 | 0.007 | 0.087 | 0.931 | -0.011 | 0.011 | -0.974 | 0.337 | 0.007 | 0.007 | 0.944 | 0.348 |
| LO:R | 0.002 | 0.006 | 0.407 | 0.687 | -0.004 | 0.008 | -0.447 | 0.658 | -0.005 | 0.005 | -0.870 | 0.384 |
| Group:R | 0.012 | 0.008 | 1.495 | 0.140 | 0.003 | 0.012 | 0.280 | 0.780 | -0.004 | 0.009 | -0.470 | 0.640 |
| Amb:LO:Group | 0.008 | 0.011 | 0.747 | 0.455 | 0.022 | 0.017 | 1.270 | 0.212 | 0.036 | 0.017 | 2.171 | **0.033** |
| P:LO:Group | -0.001 | 0.008 | -0.140 | 0.889 | -0.001 | 0.012 | -0.050 | 0.961 | -0.010 | 0.011 | -0.886 | 0.382 |
| Amb:LO:R | -0.006 | 0.011 | -0.486 | 0.627 | -0.019 | 0.018 | -1.100 | 0.279 | -0.010 | 0.012 | -0.774 | 0.444 |
| P:LO:R | -0.008 | 0.010 | -0.808 | 0.424 | 0.003 | 0.011 | 0.292 | 0.772 | -0.019 | 0.008 | -2.325 | **0.026** |
| Amb:Group:R | 0.010 | 0.012 | 0.846 | 0.400 | -0.009 | 0.017 | -0.509 | 0.611 | 0.003 | 0.012 | 0.240 | 0.811 |
| P:Group:R | 0.001 | 0.007 | 0.092 | 0.927 | 0.014 | 0.011 | 1.268 | 0.208 | 0.000 | 0.007 | -0.001 | 0.999 |
| LO:Group:Region | 0.001 | 0.006 | 0.202 | 0.841 | -0.002 | 0.010 | -0.201 | 0.842 | -0.002 | 0.006 | -0.270 | 0.789 |
| Amb:LO:Group:R | -0.024 | 0.011 | -2.154 | **0.031** | -0.030 | 0.017 | -1.768 | 0.077 | -0.025 | 0.011 | -2.134 | **0.033** |
| P:LO:Group:R | 0.002 | 0.007 | 0.263 | 0.793 | -0.005 | 0.012 | -0.462 | 0.647 | 0.001 | 0.007 | 0.196 | 0.845 |

Table 4. *Summary of statistical analyses for target sentences. LO = Lexical Overlap, Amb = Ambiguity, R = Region and P = Priming.*

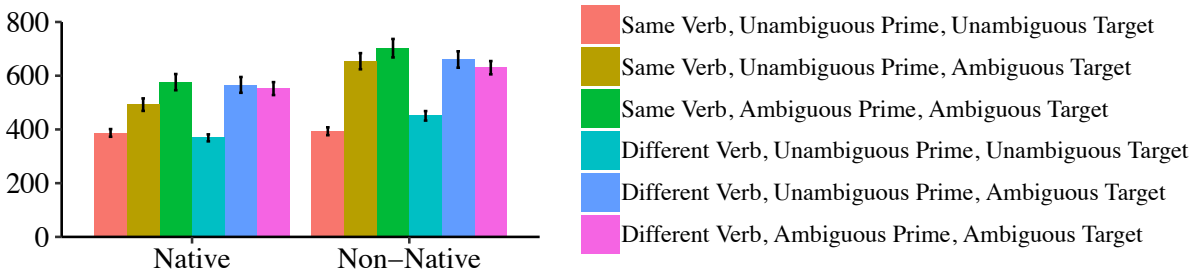


Figure 1. *Total viewing times in milliseconds at the disambiguating region in target sentences.*

*Comprehension question accuracy*

Comprehension accuracy rates revealed a marginal main effect of group, as overall accuracy rates were higher for L1 than L2 participants. There was also a main effect of ambiguity, with lower accuracy for ambiguous sentences, showing persistence of the incorrect transitive interpretation. This ambiguity effect was qualified by an ambiguity by lexical overlap interaction, as accuracy rates were lower for the non-lexical-overlap conditions (*estimate* = 2.099, *SE* = 0.28, *z* = 7.58, *p* < .001; differences of accuracy rates between the ambiguous and unambiguous conditions; 17%) than the lexical-overlap conditions (*estimate* = 1.295, *SE* = 0.23, *z* = 5.68, *p* < .001; 11%). However, there was no indication that ambiguous prime sentences facilitated either transitive or intransitive verb interpretations.

| | Native | | Non-Native | |
|---|---|---|---|---|
| | Mean | SD | Mean | SD |
| Same Verb, Unambiguous Prime, Unambiguous Target | 0.92 | 0.27 | 0.90 | 0.30 |
| Same Verb, Unambiguous Prime, Ambiguous Target | 0.87 | 0.34 | 0.75 | 0.44 |
| Same Verb, Ambiguous Prime, Ambiguous Target | 0.86 | 0.35 | 0.72 | 0.45 |
| Different Verb, Unambiguous Prime, Unambiguous Target | 0.95 | 0.22 | 0.92 | 0.27 |
| Different Verb, Unambiguous Prime, Ambiguous Target | 0.80 | 0.40 | 0.72 | 0.45 |
| Different Verb, Ambiguous Prime, Ambiguous Target | 0.82 | 0.39 | 0.73 | 0.45 |

Table 5. *Accuracy rates for comprehension questions following target sentences*.

| | *Comprehension Accuracy Rate* | | | |
|---|---|---|---|---|
| | *Log-Odds* | *SE* | *z value* | *p value* |
| Ambiguity | 1.681 | 0.197 | 8.515 | **<0.001** |
| Priming | 0.013 | 0.064 | 0.206 | 0.837 |
| Lexical Overlap | -0.029 | 0.074 | -0.385 | 0.700 |
| Group | -0.341 | 0.176 | -1.934 | 0.053 |
| Ambiguity:Lexical Overlap | 0.387 | 0.143 | 2.707 | **0.007** |
| Priming:Lexical Overlap | -0.068 | 0.063 | -1.072 | 0.284 |
| Ambiguity:Group | 0.179 | 0.191 | 0.934 | 0.350 |
| Priming:Group | 0.036 | 0.064 | 0.563 | 0.574 |
| Lexical Overlap:Group: | 0.040 | 0.062 | 0.641 | 0.522 |
| Ambiguity:Lexical Overlap:Group | -0.201 | 0.143 | -1.403 | 0.161 |
| Priming:Lexical Overlap:Group | 0.000 | 0.064 | 0.007 | 0.994 |

Table 6. *Summary of statistical analyses for comprehension accuracy rates*.

### 3.2.3 Discussion

The results showed reading difficulty when participants read temporarily ambiguous compared to unambiguous sentences in both prime and target sentences. There was some evidence of group differences in the size of these ambiguity effects, though these were not consistent across prime and target sentences. For prime sentences, regression path duration showed a marginal trend indicating that the size of garden-path effects was larger for L1 than L2 participants. In target sentences, the opposite pattern, with larger garden-path effects for L2 participants, was observed in a number of measures.

Importantly, total reading times of the disambiguating region of ambiguous target sentences were increased when prime sentences were ambiguous, compared with when they were unambiguous. This indicates that ambiguous prime sentences did not facilitate the globally correct intransitive interpretation to the same degree as unambiguous prime sentences. This suggests the initial transitive interpretation in ambiguous prime sentences lingered. Total viewing times further provided evidence that this lingering misinterpretation is dependent on lexical overlap: i.e., longer reading times induced by ambiguous prime sentences were observed only in the lexical-overlap condition. There was also evidence from comprehension accuracy rates that participants persisted with the

initially assigned misinterpretation with marginally lower accuracy for L2 participants in both ambiguous and unambiguous conditions. However, there were no significant effects to indicate that accuracy rates differed between ambiguous and unambiguous prime sentences.

### 3.3 General Discussion

The present study investigated the structural representation that L1 and L2 speakers derive during reanalysis of the subject-object ambiguity. The novelties of our results are that firstly, the initially assigned transitive misinterpretation of a temporarily ambiguous verb is not fully discarded after reanalysis in language comprehension and this affects subsequent online reading and secondly, this effect was observed for both L1 and L2 participants. Below, we discuss the implications of these findings.

*Persistence of misinterpretation in L1*

The present study observed that the initially assigned transitive misinterpretation of a temporarily ambiguous verb lingers after reanalysis. This finding is consistent especially with van Gompel et al. (2006), which reported that L1 speakers produce more transitive interpretations after reading temporarily ambiguous than unambiguous sentences. Our results thus extend van Gompel et al.'s findings from priming during production to priming during comprehension. One finding from the present study, which is different to van Gompel et al., is that while they showed priming irrespective of lexical overlap between primes and targets, the clearest evidence of priming in the present study in total viewing times indicated that unambiguous prime sentences facilitated the intransitive interpretation only in the lexical-overlap condition. This discrepancy may be accounted for by the asymmetrical finding between production and comprehension such that comprehension priming is heavily lexically mediated (e.g., Arai et al., 2007; Branigan et al., 2005). Also,

the present study cannot clarify whether the observed priming effects resulted from residual activation of the original transitive interpretation (Pickering & Branigan, 1998) or the unerased explicit memory of the initial misanalysis (Bock & Griffin, 2000; Chang et al., 2006). Investigating the nature of structural priming was not our aim. Rather, the present study was designed to examine the nature of reanalysis with a particular focus on the representation of the subordinate clause verb after reanalysis of the subject-object ambiguity, and our results revealed that the initially assigned transitive misinterpretation is not fully discarded after reanalysis during language comprehension.

Traxler (2015) recently showed that early closure prime sentences facilitate the transitive interpretation of subsequent target sentences, compared with late closure prime sentences, which indicates that L1 speakers reanalyse the subordinate clause verb transitively at least to some degree. Our results additionally show that the initially assigned transitive interpretation is not fully discarded in comparison to unambiguously intransitive verbs. We believe that the lingering misinterpretation observed in the present study can be accounted for by the Good Enough Language Processing model, which assumes that L1 readers often create linguistic representations unfaithful to the input so as to avoid costly language processing (e.g., Ferreira et al., 2001). Recent L1 studies suggest that lingering misinterpretations result from a failure to erase the memory trace of the initially assigned misinterpretation rather than a failure to complete syntactic reanalysis (e.g., Slattery et al., 2013). Given the results of Slattery et al., which provided experimental evidence that reanalysis of the temporarily ambiguous noun phrase is complete, we interpret our results as indicating that the source of the lingering transitive misinterpretation is interference from the memory trace of the initially assigned transitive misinterpretation that is not fully erased from memory, rather than incomplete reanalysis.

Although priming influenced reading times of target sentences, it did not influence comprehension accuracy rates. This may suggest priming has an influence on reading

times but not overall comprehension accuracy. However, we note here that comprehension accuracy for ambiguous sentences was generally quite high, which may have made it difficult to observe small differences related to priming. Further research is required to further examine how priming may influence comprehension accuracy of garden-path sentences.

*L2 reanalysis*

As shown in previous studies (e.g. Jacon & Felser, 2016; Juffs & Harrington, 1996; Hopp, 2015; Roberts & Felser, 2011), L2 participants showed garden-path effects with longer reading times for temporarily ambiguous than unambiguous sentences. Some previous studies reported different garden path effects for L1 and L2 speakers. Jacob and Felser (2016), for example, reported reduced garden-path effects for L2 than L1 speakers, which they interpreted as indicating that L2 speakers may hesitate to initiate reanalysis or quit reanalysis earlier than L1 speakers. Although we observed a similar marginal trend in regression path duration, several measures for target sentences showed the opposite pattern, with larger garden path effects for L2 speakers. As such, while we are cautious in interpreting potential differences in the size of garden path effects between L1 and L2 speakers in the present study, we did not consistently find reading time evidence suggesting that L2 speakers are more hesitant to conduct reanalysis than L1 speakers.

Also, our L2 participants had marginally more difficulty answering comprehension questions following target sentences than L1 speakers. This is partially compatible with recent L2 reanalysis studies showing that L2 speakers persist with misinterpretation more strongly than L1 speakers (e.g., Pozzan & Trueswell, 2016). The strongest evidence of this would have been lower comprehension accuracy for L2 speakers in ambiguous conditions only. However, although there was a numerical trend for larger differences between unambiguous and ambiguous sentences in L2 than L1 participants, this group effect needs

to be taken with caution, given that it did not significantly interact with ambiguity (see also Jacob & Felser, 2016): i.e., overall accuracy to both ambiguous and unambiguous sentences was lower for L2 than L1 participants rather than ambiguous sentences alone. Further research is required to elucidate the extent to which L2 speakers have increased difficulty in understanding garden-path sentences compared to L1 speakers.

Most importantly for present purposes, L2 participants in the present study showed that unambiguous prime sentences facilitated the transitive interpretation as in L1 participants, which suggests that the structural representation that L2 participants created during online reading is stable enough to cause priming effects during language comprehension. This finding is consistent with recent studies, showing that L2 priming effects occur in language comprehension (e.g., Nitschke et al., 2010, 2014; Wei et al., 2019). However, unlike these recent studies, the present study showed clear structural priming effects only in lexical overlap conditions (see also Wei et al., 2018). Like L1 speakers, this may suggest that priming during comprehension in L2 processing is strongly lexically mediated. The present study also showed that unlike unambiguous prime sentences, ambiguous prime sentences did not facilitate the intransitive interpretation in L2 sentence processing. This suggests that the initially assigned transitive misinterpretation is not fully discarded after reanalysis and that reanalysis of the subordinate clause verb is one source of the reding difficulty caused by garden-path sentences in L2. As discussed, there are two potential sentence processing accounts for the L2 reanalysis difficulty. One is that unlike L1 speakers, L2 speakers do not complete reanalysis. The other is that L2 speakers complete reanalysis but are prone to interference from initially assigned misinterpretations (Cunnings, 2017). Recently, Fujita and Cunnings (under review), which extended the research design of Slattery et al. (2013) to L2 speakers, showed that L2 speakers conduct syntactic reanalysis but fail to erase the memory trace of the initial misinterpretation as L1 speakers do. Thus, what we discussed for L1 lingering misinterpretation may apply here, in

that reanalysis difficulty in both L1 and L2 speakers relates to difficulty in erasing initially assigned misinterpretations from memory, rather than an inability to conduct syntactic reanalysis. Irrespective of this issue, our results nonetheless suggest similar priming effects during comprehension of garden-path sentences for L1 and L2 speakers.

### 3.4 Conclusion

The present study investigated the nature of language comprehension during reanalysis of garden-path sentences with the subject-object ambiguity (e.g., When Mary dressed the baby laughed happily), using the structural priming paradigm. While a number of previous studies have shown that reading difficulty occurs during reanalysis processes, little has been experimentally examined about the nature of reanalysis, especially in terms of potential similarities and differences between L1 and L2 processing. In the present study, comprehension accuracy rates confirmed that reanalysis difficulty occurs during L1 and L2 language comprehension, with initially assigned misinterpretations lingering following reanalysis (Christianson et al., 2001; Slattery et al., 2013). Reading time measures showed that unambiguous prime sentences facilitate the intransitive interpretation, while such priming effects did not appear to the same degree in ambiguous prime sentences in either L1 or L2 speakers. These priming effects however were observed only when the subordinate clause verb was shared between prime and target sentences. These results suggest that lexically-mediated structural priming caused by unambiguous sentences reduces garden-path effects, but ambiguous sentences do not do so to the same degree as a result of the initially assigned transitive misinterpretation not being fully discarded in both L1 and L2 speakers. These findings suggest that initially assigned transitive misinterpretations lingers, and that reanalysis of the transitivity properties of a temporarily ambiguous verb is one of the sources of reanalysis difficulty found in L1 and L2 language comprehension.

## *Chapter 4. Study 3: Reanalysis and lingering misinterpretation of linguistic dependencies in native and non-native sentence comprehension*


### Abstract

Research on temporarily ambiguous "garden path" sentences (e.g., After Mary dressed the baby laughed very happily) has shown that initially assigned misinterpretations linger after reanalysis of the temporarily ambiguous phrase in both native (L1) and non-native (L2) readers. How lingering misinterpretation may influence other aspects of language processing has however not been systematically examined. We report three offline and two online experiments investigating reanalysis and misinterpretation of filler-gap dependences (e.g., Elisa noticed the truck which the policeman watched the car from) in L1 and L2 sentence processing and comprehension. Our results showed that L1 and L2 speakers are prone to interference from initially assigned misinterpretations during filler-gap dependency resolution. L1/L2 differences were observed such that L2 speakers had increased difficulty reanalysing some filler-gap dependencies, however this was restricted to certain disambiguating cues. Taken together, the present study indicates that lingering misinterpretation after reanalysis is a general property of language comprehension.



**Keywords**: syntactic ambiguity; eye-movements; filler-gap dependencies; good enough processing; non-native sentence processing.

## 4.1 Introduction

Syntactic ambiguity resolution has played an important role in informing our understanding of sentence comprehension in both native (L1) and non-native (L2) language speakers. Temporarily ambiguous, "garden-path" sentences such as (1) have been particularly influential in informing theoretical accounts of both the parsing strategies that individuals may use during sentence processing (e.g., Ferreira & Henderson, 1991; Frazier & Rayner, 1982), and which factors may lead to misinterpretation during language comprehension (e.g., Christian, Hollingworth, Halliwell & Ferreira, 2001; Slattery, Sturt, Christianson, Yoshida, & Ferreira, 2013).

(1) After Mary dressed the baby laughed very happily.

In (1), the noun phrase "the baby" is temporarily ambiguous as it can be interpreted either as the direct object of the subordinate clause verb ("dressed") or as the subject of the main clause verb ("laughed"). Though the temporary ambiguity disambiguates towards the latter at "laughed", many previous studies have shown that readers initially adopt the former analysis (e.g., Ferreira & Henderson, 1991; Frazier & Rayner, 1982; Pickering & Traxler, 1998; Sturt, Pickering, & Crocker, 1999). (1) thus requires reanalysis to derive the correct sentence interpretation.

Recent studies report that readers do not necessarily derive the correct interpretation of sentences like (1) after reanalysis. For example, Christianson et al. (2001) showed that readers persist with the initial misinterpretation, "Mary dressed the baby", after reanalysis even though this interpretation is not licensed by the globally correct sentence structure. Such lingering misinterpretations have also been shown to persist more strongly for L2 speakers (Jacob & Felser, 2016; Pozzan & Trueswell, 2016), although the

cause of this L1/L2 difference is debated (Cunnings, 2017; Jacob & Felser, 2016; Pozzan & Trueswell, 2016).

While misinterpretation of garden-path sentences has now been widely documented (e.g., Christianson et al., 2001; Christianson, Williams, Zacks, & Ferreira, 2006; Patson, Darowski, Moon, & Ferreira, 2009; Slattery et al., 2013; Sturt, 2007; van Gompel, Pickering, Pearson, & Jacob, 2006), existing research in this vein has examined only a very narrow set of sentence structures. This leaves open the question of the extent to which such findings indicate a general property of sentence processing or are restricted to these specific sentence types. Temporary ambiguities that may lead to misinterpretation also arise in other types of sentences however, such as those containing "filler-gap" dependencies as in (2). Here, successful comprehension requires that the displaced "filler" ("the car") must be associated with a corresponding "gap", adjacent to "near", such that the sentence is interpreted as "the policeman stopped near the car". However, although the gap in (2) is ultimately at the preposition "near", there is a temporary potential gap at the verb "stopped". Thus, "the car" may initially be interpreted as the direct object of "stopped" (i.e., "the policeman stopped the car"), even though this interpretation turns out to be subsequently incorrect.

(2) John saw the car that the policeman stopped quite abruptly near yesterday morning.

Although the temporary ambiguities that occur in (1) and (2) are both a type of garden-path, we refer here to sentences like (1) as garden-path sentences and sentences like (2) as filler-gap sentences, to distinguish between them. While misinterpretation of garden-path sentences has been widely examined in L1 and L2 processing, to date, little is known about whether initially assigned misinterpretations of filler-gap sentences linger. Potential differences between L1 and L2 speakers for such sentences have also not previously been

systematically examined. To address these issues, we report five experiments that examined lingering misinterpretation in filler-gap dependencies, to both test the generalisability of lingering effects of misinterpretation during language comprehension beyond previously tested garden-path sentences, and to tease apart different theoretical accounts of L1 and L2 sentence processing. We begin by discussing theoretical accounts of lingering misinterpretation of garden-path sentences in L1 and L2 processing, before discussing the processing of filler-gap dependencies in more detail.

### 4.1.1 Lingering misinterpretation in L1 and L2 processing

Many studies have shown that garden-path sentences like (1) cause reanalysis difficulty, with longer reading times at the disambiguating region ("laughed"), compared to an unambiguous control that contains a comma after "dressed" (e.g., Sturt et al., 1999). As mentioned above, Christianson et al. (2001) showed that L1 English speakers also sometimes misinterpret such sentences even after reanalysis. They asked participants comprehension questions probing the temporary ambiguity (e.g., "Did Mary dress the baby?"). The correct response to this question is "no", as in (1), the globally correct interpretation is "Mary dressed herself", not "Mary dressed the baby". However, in a sequence of offline experiments, they observed that participants more often answered "yes" to such questions following ambiguous sentences like (1) compared with the unambiguous condition. This suggests that the initially assigned misinterpretation lingers even after the globally correct interpretation is confirmed. Since the publication of Christianson et al.'s results, several studies have replicated this lingering misinterpretation using different methods (Christianson et al., 2006; Jacob & Felser, 2016; Malyutina & den Ouden, 2016; Nakamura & Arai, 2016; Patson et al., 2009; Slattery et al., 2013; Sturt, 2007; van Gompel et al., 2006).

While these studies generally indicate that lingering misinterpretation is a robust effect, different accounts about why the initially assigned misinterpretation persists have been proposed. The Good-Enough Sentence Processing Account, proposed by Ferreira and colleagues accounts for the lingering effect from the perspective of how the comprehender processes language (Ferreira, Bailey, & Ferraro, 2002; Ferreira, Christianson, & Hollingworth, 2001; Ferreira & Patson, 2007; Karimi & Ferreira, 2016; Slattery et al., 2013). Slattery et al. (2013) considered different ways in which sentence processing may be "good enough" (see also discussion in Christianson et al., 2001; Ferreira et al., 2001). Firstly, they considered that readers may not complete syntactic reanalysis of the temporary ambiguity, such that readers maintain the initially assigned misinterpretation and do not construct a fully specified structure for the sentence. Alternatively, syntactic reanalysis may be complete, but the initially assigned misinterpretation may linger in memory. In two experiments, Slattery et al. (2013) argued for the latter possibility. In their Experiment 1, participants read sentences like (3), which was either temporarily ambiguous or unambiguous due to the comma. The sentences also manipulated gender match between a reflexive ("himself") and its syntactically "accessible" antecedent, which is also the temporarily ambiguous noun phrase ("David's father/mother").

(3a) After the bank manager telephoned(,) David's <u>father</u> grew worried and gave himself approximately five days to reply.

(3b) After the bank manager telephoned(,) David's <u>mother</u> grew worried and gave himself approximately five days to reply.

For unambiguous sentences, Slattery et al. expected longer reading times at the reflexive in gender mismatch (3b) than gender match (3a) conditions (Sturt, 2003). In ambiguous sentences, the temporarily ambiguous noun phrase may initially be interpreted

as the direct object of the subordinate clause verb ("telephoned"), but during reanalysis, it needs to be reassigned the subject role. Slattery et al. hypothesised that if reanalysis is syntactically incomplete, the gender mismatch effect may be absent or reduced in the ambiguous condition. Indeed, if reanalysis is incomplete, the temporarily ambiguous phrase would remain as the direct object in the subordinate clause, where it cannot be a syntactically accessible antecedent for the reflexive, as a result of Binding Principle A (Chomsky, 1981). It can be an antecedent for the reflexive only if syntactic reanalysis, as the main clause subject, is complete. In an eye-tracking while reading task, Slattery et al. observed gender mismatch effects, irrespective of ambiguity. They took this as evidence that L1 readers complete syntactic reanalysis.

In their Experiment 2, Slattery et al. tested texts like (4).

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

The first sentence was either ambiguous or unambiguous. The second continuation sentence referred back to the globally correct interpretation of the first sentence ("Frank quickly finished drying himself off"). It is however inconsistent with the initially assigned misinterpretation in the ambiguous condition ("Frank dried off the truck"). If the initial misinterpretation lingers, reading times at the reflexive in the second sentence may be longer in ambiguous than unambiguous conditions. Alternatively, if the initial misinterpretation is completely erased, there should be no reading time differences between conditions here. Slattery et al. observed longer reading times for ambiguous than unambiguous sentences at the critical reflexive, indicating a failure to erase the initial misinterpretation. Together, the results of both experiments suggest that L1 speakers complete syntactic reanalysis, but initially assigned misinterpretations linger in memory.

L2 speakers encounter reading difficulty when reading garden-path sentences as L1 speakers do (e.g., Hopp, 2015; Jacob & Felser, 2016; Jegerski, 2012; Juffs, 2004; Juffs & Harrington, 1996; Roberts & Felser, 2011). Some studies also show that L2 speakers may have more difficulty recovering from garden paths than L1 speakers (e.g., Gerth, Otto, Nam, & Felser, 2017; Jacob & Felser, 2016; Pozzan & Trueswell, 2016; Roberts & Felser, 2011), although successful reanalysis is more likely as L2 proficiency improves (Gerth et al., 2017; Hopp, 2006). For example, Jacob and Felser (2016) investigated L1/L2 differences in the final interpretation of garden-path sentences like (1), using end-of-sentence comprehension questions referring to the initially assigned misinterpretation as in Christianson et al. (2001). Their results showed that L2 speakers had lower accuracy in answering the comprehension questions than L1 speakers.

How to account for this L1/L2 difference is debated. One possibility is that unlike L1 speakers, L2 speakers cannot construct the globally correct syntactic structure licensed by reanalysis. As a result, they persist with the initial misinterpretation. This account may be compatible with the Shallow Structure Hypothesis, which claims that L2 syntactic parsing is shallow or underspecified (Clahsen & Felser, 2006, 2017). Alternatively, as Slattery et al. (2013) showed for L1 speakers, L2 speakers may construct the correct parse after reanalysis, but have increased difficulty in erasing the memory trace of the initial misinterpretation (Cunnings, 2017).

### 4.1.2 Filler-gap dependency resolution in language comprehension

Filler-gap dependencies are unbounded long-distance syntactic dependencies like (5), where the direct object (“the car”), called “the filler”, is dislocated from the post-verb region in the relative clause called “the gap” (indicated by an underline). In order to assign the correct thematic interpretation to (5), readers need to associate the filler and the gap.

(5) Mary saw the car which her son liked __ last weekend.

One well-known aspect of filler-gap processing is that readers assign an identified filler as an argument of a gap at the first possible opportunity (e.g., Aoshima, Phillips, & Weinberg, 2004; Chacón et al., 2016; Felser, Clahsen, & Münte, 2003; Frazier, 1987; Frazier & Clifton, 1989; Garnsey, Tanenhaus, & Chapman, 1989; Nakano, Felser, & Clahsen, 2002; Nicol, & Swinney, 1989; Omaki et al., 2015; Parker, 2017; Pickering & Traxler, 2001, 2003; Sussman & Sedivy, 2003; Wagers & Phillips, 2009, 2014). One piece of evidence of this so-called "active gap filling" comes from the observation of reading disruption in wh-fronting sentences like (6a), compared to control sentences like (6b).

(6a) My brother wanted to know who Ruth will bring us home to __ at Christmas.

(6b) My brother wanted to know if Ruth will bring us home to Mom at Christmas.

In (6a), although the globally correct gap position appears immediately after the preposition "to", the filler ("who") can be temporarily associated with the embedded clause verb ("bring"), as "who" is a plausible direct object of "bring". In (6b), there is no such dependency due to the conditional clause. Sentences like (6a) causes reading disruption at "us" compared to (6b). This suggests readers initially interpret "who" to be the direct object of "bring", but this initial misinterpretation is reanalysed when it becomes clear that this verb has an overt argument ("us"). This is known as the filled-gap effect (Stowe, 1986).

Evidence that readers engage in active gap filling is also found in the absence of filled-gap effects. For example, Traxler and Pickering (1996) examined sentences like (7).

(7a) We like the city that the author wrote unceasingly and with great dedication about __ while waiting for a contract.

(7b) We like the book that the author wrote unceasingly and with great dedication about __ while waiting for a contract.

In (7), the gap is ultimately after the preposition "about" but there is an earlier potential gap at the verb "wrote". We refer to these types of constructions as non-filled-gap sentences. While the first temporary gap in (6a) is filled by the overt direct object of the verb "bring", in (7) the temporary gap at "wrote" does not have an overt direct object. As such, while the initial gap is disambiguated immediately in filled-gap sentences like (6a), it is not disambiguated until the preposition is reached in non-filled gap sentences like (7). Thus, the type, and potentially timing, of disambiguating cues in filled-gap and non-filled-gap sentences differ.

Traxler and Pickering manipulated plausibility between the filler and the embedded clause verb such that in the implausible condition (7a), the filler ("the city") is not a semantically appropriate direct object for the verb, while in the plausible condition (7b) the filler ("the book") is. Traxler and Pickering found longer reading times at "wrote" for implausible (7a) than plausible (7b). This suggests that readers postulated a gap at the first available position, before reaching the actual gap position at the preposition.

Several studies have also explored how L2 speakers process filler-gap dependencies (e.g., Dallas, DeDe, & Nicol, 2013; Felser, Cunnings, Batterham, & Clahsen, 2012; Jessen & Felser, 2018; Jessen, Festman, Boxell, & Felser, 2017; Johnson, Fiorentino, & Gabriele, 2016; Juffs, 2005; Juffs & Harrington 1995; Omaki & Schulz, 2011; Williams, 2006; Williams, Möbius & Kim, 2001; for review see Dallas & Kaan, 2008). These studies generally report that L2 speakers actively associate a filler with the first available gap like L1 speakers.

Jessen and Felser (2018), who used event related potentials to examine L1/L2 differences in the processing of filler-gap dependencies like (8), recently showed that L2 speakers have more difficulty reanalysing filler-gap dependencies than L1 speakers.

(8a) Bill liked the house that Bob built some ornaments for __ at his workplace.
(8b) Bill liked the women that Bob built some ornaments for __ at his workplace.

Sentences like (8) cause reanalysis difficulty at "some ornaments" due to filled-gap effects. (8) also manipulates whether the filler ("the house/women") is a plausible direct object for the embedded clause verb ("built"). While both L1 and L2 participants showed an N400 effect at "built" in (8b), only L2 participants showed a P600 effect at the disambiguating region ("ornaments") and at the post-preposition region marked by an underline in (8a). They interpreted this as evidence that L2 speakers have more difficulty reanalysing filled-gap sentences than L1 speakers.

While many studies show evidence of active gap filling, little is known about how misanalysed filler-gap dependencies affect language comprehension. We are aware of only two studies that have examined this issue, which both examined L1 but not L2 comprehension, and utilised offline tasks only. In Lassotta, Omaki and Franck (2016), L1 French speaking adults were provided with cartoons depicting a story and then answered globally ambiguous (9a) and temporarily ambiguous questions (9b).

(9a) Où est-ce qu'Aline a expliqué qu'elle allait attraper des papillons?
where Q Aline has explained that she went catch some butterflies
"Where did Aline explain that she was going to catch butterflies?"
(9b) Où est-ce qu'Aline a expliqué dans le salon qu'elle allai attraper des papillons?

where Q Aline has explained in the room that she was going to catch some butterflies

"Where did Aline explain in the living room that she was going to catch butterflies?"

In (9a), the wh-constituent can be associated with either the main clause verb ("expliqué" *explain*) or the embedded clause verb ("attraper" *catch*). In (9b), the wh-constituent can only be associated with the embedded clause verb, but may temporarily be associated with the main clause verb during incremental processing. The main clause interpretation is however ruled out at the following filled-gap prepositional phrase ("dans le salon" *in the room*). Lassotta et al. found a strong preference for the main clause interpretation in sentences like (9a). Although the subordinate clause interpretation should be preferred in (9b), participants still picked the main clause interpretation approximately 50% of the time. This suggests the main clause misinterpretation was initially assigned during processing, and lingered even after disambiguation.

Omaki, Davidson White, Goro, Lidz and Phillips (2014) previously conducted an experiment with similar sentences in Japanese. Importantly, in the Japanese equivalent of (9b) ("Doko-de Yukiko-chan-wa kouen-de choucho-o tsukamaeru-to itteta-no?" *Where was Yukiko telling someone that she would catch a butterfly at the park?*) the disambiguating prepositional phrase not only rules out the embedded clause interpretation, but it also appears before either the main clause or subordinate clause verb, unlike in French where it occurs after the main clause verb. Unlike Lossotta et al., in Omaki et al., adult L1 Japanese speakers exhibited a strong preference for the main clause interpretation of the Japanese equivalents of (9b). This suggest that the timing of disambiguating information plays an important role in filler-gap interpretation.

### 4.1.3 The present study

While the processing of filler-gap dependencies during L1 and L2 comprehension has been widely studied, extant studies have not systematically examined the reanalysis processes involved in filler-gap dependency resolution, nor whether initially assigned misinterpretations linger for filler-gap dependencies. The small literature that has examined this issue has utilised offline tasks only. Using online methods is crucial especially in testing lingering misinterpretation, as although offline methods are of advantage to examine the final interpretation assigned to temporarily ambiguous sentences, their explicitness may not always reflect how the comprehender actually interpreted the sentences in real-time during reading (e.g., Tabor, Galantucci, & Richardson, 2004). Furthermore, no study has examined potential L1/L2 differences in lingering misinterpretations of filler-gap dependencies. Given previous studies, it is possible that L2 speakers may persist with initially assigned filler-gap interpretations more strongly than L1 speakers.

To explore these issues, the present study aimed to investigate whether and to what extent initially assigned misinterpretation lingers for filler-gap dependencies in L1 and L2 speakers, in a series of experiments. We utilised both offline comprehension tasks and online eye-movement measures, to investigate the final interpretation assigned to sentences containing filler-gap dependencies, and to investigate how such sentences are processed in real-time. Across experiments, we examined lingering misinterpretations in sentences containing filled-gaps and non-filled-gaps, to examine how different disambiguating cues influence lingering misinterpretation in filler-gap dependencies, and to test the generalisability of our results across these two constructions.

### 4.2 Experiment 1a

The aim of Experiment 1a was to test for lingering misinterpretation in the resolution of filler-gap dependencies using an offline task. We tested both filled-gap and non-filled-gap sentences, as in (10/11).

(10) *Filled-gap*

(a) *Ambiguous*

John saw the car which the officer stopped the bicycle near earlier today.

(b) *Unambiguous*

John saw the car near which the officer stopped the bicycle earlier today.

(c) What did the officer stop?

1. The car 2. The bicycle

(11) *Non-filled-gap*

(a) *Ambiguous*

The teacher saw the desk which the student moved very quickly around during the lunch break.

(b) *Unambiguous*

The teacher saw the desk around which the student moved very quickly during the lunch break.

(c) Did the student move the desk?

1. Yes 2. No

In (10a/11a), the gap is ultimately at the prepositions “near” and “around”, respectively, but there is an earlier possible gap following the verbs “stopped” and “moved”. In filled-gap (10a), this initial misinterpretation is ruled out at the following

noun phrase ("the bicycle"), while in (11a) it is not ruled out until the preposition. (10b/11b) are unambiguous controls in which the preposition is fronted. Each sentence was followed by a question, as in (10c/11c), to test whether initially assigned misinterpretations linger.

If initially assigned misinterpretations linger, participants should choose the incorrect answer ("the car" in (10c) and "Yes" in (11c)) more frequently in ambiguous than unambiguous sentences. If L2 participants are more persistent with initial misinterpretation (Jacob & Felser, 2016; Pozzan & Trueswell, 2016), accuracy rates should be lower for L2 than L1 participants in the ambiguous conditions.

### 4.2.1 Method

*Participants*

40 L1 English speakers (5 males, mean age = 19; range = 18–23) and 40 L2 English speakers (10 males, mean age = 25; range = 18–43), of various L1 backgrounds[12], from the University of Reading community, took part in Experiment 1a. Participants received course credit or a small monetary incentive.

The L2 participants started learning English in a school environment after age five. They completed the grammar part of the Oxford Placement Test (OPT) (Allan, 2004) after the main experiment, which indicated an average score of 72 out of 100 (SD = 10.3; range = 52–89). This places them as upper intermediate to advanced English language learners.

*Materials*

[12] First languages of the L2 participants were Chinese (6), Spanish (5), French (4), Malay (4), Turkish (4), Cantonese (2), Kazakh (2), Polish (2), Swedish (2), Greek (1), Indonesian (1), Italian (1), Norwegian (1), Russian (1), Serbian (1), Thai (1).

Experimental materials consisted of 12 sets of filled-gap sentences (10a/b) and 12 sets of non-filled-gap sentences (11a/b), which manipulated ambiguity, and were followed by comprehension questions (10c/11c). The full set of experimental items is provided in Appendix A.

72 filler sentences were also constructed, all of which were accompanied by a binary comprehension question. Experimental items were presented with four counterbalanced lists in a Latin Square design.

*Procedure*

The experimental and filler sentences were presented using Linger (Rohde, 2010). At the start of each trial, a cross appeared onscreen. After pressing the space bar, the cross disappeared, and the sentence appeared in full. When participants pressed the space bar again, the sentence was replaced with a comprehension question containing two options, which participants answered by pressing an appropriate key on the keyboard. The two options were pseudo-randomised so that half of the correct answers were presented on the left side and half on the right side. The experimental order was also pseudo-randomised for each participant such that at least two filler sentences appeared between each experimental sentence. Participants were instructed to simply read each sentence for comprehension and answer the questions accurately. The experiment began with some practices. L2 participants completed the OPT after the experiment. The experiment took 20–25 minutes on average with an additional 25–30 minutes for the OPT.

*Data analysis*

Filled-gap and non-filled-gap sentences were analysed separately, as lexical material and the question form differed between the sentence types. Comprehension accuracy rates were treated as a dependent variable to assess lingering misinterpretation. The data were

analysed in R (R Core Team, 2018) by fitting generalised linear mixed-effects models using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015). Each model included sum-coded fixed effects of ambiguity (ambiguous/unambiguous) and group (L1/L2). In the case of an ambiguity by group interaction, planned comparisons tested the effect of group at the two levels of ambiguity to examine L1/L2 differences within each ambiguous and unambiguous condition. Models were fit with the maximal random effects structure that converged (Barr, Levy, Scheepers, & Tily, 2013).[13] For fixed effects, *p*-values were estimated using the Laplace Approximation implemented by the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017). Data and analysis code for all experiments reported in the present study is available at the first author's Open Science Framework webpage (https://osf.io/6uz95/)

### 4.2.2 Results

Mean accuracy rates of filler sentences were 93% for both groups (L1 range = 83–100; L2 range = 79–100), suggesting participants paid attention during the experiment. Comprehension accuracy rates and inferential statistics are summarised in Tables 1 and 2.

For the filled-gap condition, there was a main effect of ambiguity due to lower accuracy rates for the ambiguous than unambiguous conditions. For the non-filled-gap condition, models similarly revealed a marginal main effect of ambiguity in the same direction. However, these effects did not interact with group.

[13] When this maximal model did not converge, we first removed the random correlation parameters. If this model still did not converge, the random effect accounting for the least variance (generally zero) was iteratively removed until converged was achieved.

| | L1 | | | | | L2 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Filled-Gap | | Non-Filled-Gap | | | Filled-Gap | | Non-Filled-Gap | | |
| | Amb | Unamb | Amb | Unamb | Control | Amb | Unamb | Amb | Unamb | Control |
| Experiment 1a | | | | | | | | | | |
| | 71 (45) | 76 (43) | 34 (48) | 39 (49) | - | 65 (48) | 78 (41) | 23 (42) | 35 (48) | - |
| Experiment 2a | | | | | | | | | | |
| | 71 (45) | 82 (38) | 70 (46) | 68 (47) | 86 (35) | 70 (46) | 77 (42) | 50 (50) | 69 (46) | 86 (35) |
| Experiment 3 | | | | | | | | | | |
| | 68 (47) | 73 (45) | 47 (50) | 50 (50) | - | 69 (46) | 78 (42) | 36 (48) | 59 (49) | - |

Table 1. *Accuracy rates (%) for comprehension questions following ambiguous and unambiguous sentences in Experiments 1a, 2a and 3 (SDs in parentheses). Amb = Ambiguous and Unamb = Unambiguous.*

| | Experiment 1a | | | Experiment 2a | | |
|---|---|---|---|---|---|---|
| | *Estimate (SE)* | *t(z) value* | *p* value | *Estimate (SE)* | *t(z) value* | *p* value |
| **Filled-gap sentences** | | | | | | |
| Ambiguity | 0.62 (0.21) | 2.91 | .004 | 0.48 (0.21) | 2.25 | .025 |
| Group | 0.11 (0.30) | 0.39 | .700 | 0.25 (0.29) | 0.87 | .384 |
| Ambiguity:Group | 0.51 (0.41) | 1.26 | .209 | 0.40 (0.39) | 1.01 | .311 |
| **Non-Filled-gap sentences** | | | | | | |
| Ambiguity | 0.43 (0.25) | 1.70 | .090 | 0.49 (0.20) | 2.39 | .017 |
| Group | 0.51 (0.33) | 1.53 | .125 | 0.53 (0.30) | 1.79 | .073 |
| Ambiguity:Group | 0.22 (0.51) | 0.43 | .668 | 1.10 (0.37) | 3.01 | .003 |

Table 2. *Summary of statistical analyses for Experiments 1a and 2a*

### 4.2.3 Discussion

Though the results were clearest in the filled-gap conditions, these results provide some preliminary evidence that both L1 and L2 speakers have more difficulty answering comprehension questions following ambiguous than unambiguous filler-gap sentences, suggesting lingering misinterpretation. However, contrary to some previous studies (e.g., Gerth et al., 2017; Jacob & Felser, 2016; Pozzan & Trueswell, 2016), there was no statistically significant difference in accuracy rates between L1 and L2 participants.

One unexpected result was that comprehension accuracy rates were generally low for the non-filled-gap sentences. Given the high accuracy rates for filler sentences, we do not think this results from a lack of participant attention. One potential account of this is that participants may have inferred a plausible theme for the intransitive verb, even in the unambiguous condition. For example, experimental items included trials such as "The boy bought the novel about which the girl read very happily last night. Did the girl read the novel?". Although the expected answer was "no", it is conceivable that participants inferred that "the girl" both "read about" and "read" "the novel", even in the unambiguous conditions given the semantic similarity between the two interpretations ("read the novel" and "read about the novel"). Indeed, similar effects have been observed in previous studies investigating garden-path sentences containing the subject-object ambiguities when optionally transitive verbs such as "visit" are used rather than reflexive absolute transitive verbs such as "dress" (e.g., Christianson et al., 2001). Another possibility is that the nature

of the question ("Did the girl read the novel?") may have biased toward misinterpretation. Note that despite this potential inference, we found suggestive evidence that the initially assigned misinterpretation lingers in non-filled-gap sentences.

Experiment 2a reports an offline sentence-picture matching task that attempts to address these issues and replicate our findings using a different design. We first report Experiment 1b however, which tested misinterpretation of filled-gap sentences during online reading.

### 4.3 Experiment 1b

The aim of Experiment 1b was to test whether L1 and L2 speakers persist with initial misinterpretations during sentence processing following filled-gaps, and tested texts in four conditions like (12a-d). Inspired by the design of Slattery et al. (2013, Experiment 2), this experiment used continuation sentences to test for lingering effects of misinterpretation.

(12) Some chores needed to be done.

(a) *Ambiguous, Consistent Continuation*

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

It seemed that the maid was cleaning the floor while thinking about dinner.

(b) *Unambiguous, Consistent Continuation*

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

It seemed that the maid was cleaning the floor while thinking about dinner.

(c) *Ambiguous, Inconsistent Continuation*

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

It seemed that the maid was cleaning the brush while thinking about dinner.

(d) *Unambiguous, Inconsistent Continuation*

The child noticed the brush with which the maid was cleaning the floor very carefully.
It seemed that the maid was cleaning the brush while thinking about dinner.

Each item in Experiment 1b contained three sentences. The first sentence is a lead-in sentence, which is followed by a second, filler-gap sentence. This is either temporarily ambiguous (12a/12c) or unambiguous (12b/12d). The temporary gap at "cleaning" in ambiguous (12a/12c) is disambiguated by a filled-gap ("the floor"). The third sentence is a continuation sentence whose meaning is either consistent or inconsistent with the globally correct interpretation of the second sentence. Specifically, in (12a/b), the continuation is consistent with the globally correct interpretation of the second sentence ("the maid was cleaning the floor") but inconsistent with the initially assigned misinterpretation ("the maid was cleaning the brush"). On the other hand, the continuation sentence in (12c/d) is inconsistent with the globally correct interpretation but consistent with the initial misinterpretation.

For the second, filler-gap sentences, we expected to elicit longer reading times at "the floor" for ambiguous than unambiguous sentences due to filled-gap effects. Regarding the continuation sentence, if reanalysis is complete, reading times should be longer for inconsistent (12c/12d) than consistent (12a/12b). This inconsistency effect may be reversed or attenuated if the initial misinterpretation is not fully erased. That is, if initial misinterpretations linger, readers may interpret the inconsistent region of the continuation sentence in (12c) as being consistent with the second sentence, as this region is consistent with the initial misinterpretation. This would lead to an attenuation of the inconsistency effect, with shorter reading times in the continuation sentence for (12c) than (12d). For consistent conditions, lingering misinterpretation may make readers interpret the consistent region of the continuation sentence in (12a) as being inconsistent with the second sentence, as it is inconsistent with the initial misinterpretation. This would lead to longer reading

times in ambiguous (12a) than unambiguous (12b). Thus, the crucial prediction for lingering misinterpretation is an interaction between ambiguity and consistency in the third sentence.

### 4.3.1 Method

*Participants*

The participants from Experiment 1a took part in Experiment 1b. We tested the same participants because, as is typical in L2 research, we aimed to test both offline comprehension and online processing in the same learners. Although we reported Experiment 1a first, Experiment 1b was completed by participants before Experiment 1a, so as to avoid the offline task influencing the more sensitive online experiment. Participants completed the two tasks in separate experimental sessions, at least one week apart.

*Materials*

We created 24 sets of experimental sentences as in (12) (see Appendix B for full list). Each set began with a lead-in sentence that always appeared on the first line. The second filler-gap sentence appeared across the first and second lines, with a line break after the relative pronoun ("which"). The third continuation sentence appeared across lines two and three, with a line break before the complementiser ("that"). Words used for the consistency manipulation in the continuation sentence (e.g., "floor/brush") were matched for frequency, length and lexical decision speed according to the norms provided by the English Lexicon Project (Balota et al., 2007). The experiment also contained 72 filler texts with a variety of syntactic structures, which always took up either two or three lines onscreen. All experimental and two-third of filler texts were followed by a binary yes/no comprehension question. Comprehension questions of experimental texts asked about different parts of the

text equally but never probed the interpretation of the filler-gap dependency in the second target sentence.

*Procedure*

Although viewing was binocular, eye-movements were recorded from the participant's right eye using an SR Research Eyelink 1000. Each session began with calibration of the eye-tracker on a nine-point grid. Recalibration was performed between trials if any drift in calibration was observed. Before each trial appeared, a gaze trigger was presented above the first word of the text to be displayed. Upon fixation on this gaze trigger, the text appeared. Participants were instructed to press a button on a game pad after reading each text. A yes/no question then appeared onscreen if appropriate, which participants answered by pressing a button on a game pad. Experimental and filler texts were presented in a pseudo-randomised order with a Latin-square design. The entire experiment lasted 40–60 minutes.

*Data analysis*

The experimental texts were divided into three regions for analysis. To test for filled-gap effects, we analysed the disambiguating region ("the floor") in the filler-gap sentence. To test for effects of consistency, the critical region was specified as the noun phrase that manipulated consistency in the continuation sentence ("the floor/brush") while a spillover region ("while thinking") contained the rest of the sentence except the last two words, which were not included to avoid end-of-trial effects influencing reading times. We calculated three reading times measures. These included first pass reading time, the sum of fixations within a region entered from the left up until an eye-movement away from the region, and regression path duration, the summed duration of all fixations measured from when a region is first fixated from the left, up until but not including the first fixation in a

region to the right. We also calculated total viewing times, the summed duration of all fixations in a region. Prior to the calculation of reading time measures, fixations shorter than 80ms that were within one degree of visual arc of another fixation were merged. Any other fixations shorter than 80ms or over 800ms were removed. Any region that a participant skipped was removed from data analysis, which affected less than 9% of the L1 data and 4% of L2 data.

For data analysis, we fitted linear mixed effect models (Baayen, Davidson, & Bates, 2008) to each reading time measure in R using the lme4 package. Reading times were log-transformed prior to analysis. For the disambiguating region, the mixed effect models included sum-coded fixed effects of ambiguity (ambiguous/unambiguous), group (L1/L2), and consistency (consistent/inconsistent). For the critical/spillover regions, in addition to these fixed effects, the models included region (critical/spillover regions) as a fixed effect. To minimise the number of independent tests conducted at each region (von der Malsburg & Angele, 2018), we conducted a single analysis of each measure with region as a fixed effect (see Cunnings & Sturt, 2018). By including region as a fixed-effect, we can explicitly test any potential time-course effects across regions, as it is possible that one effect may appear at one region (e.g., critical region) but not another (e.g., spillover region). Each model was initially specified with the maximal random effects structure. If this model failed to converge, correlation parameters and random effects were removed as described in Experiment 1a until it successfully converged. As including region as a fixed effect involves including two non-independent datapoints from a single trial, a random intercept for trial was also included.

When an interaction appeared between region and ambiguity or consistency, follow-up comparisons were performed at the two levels of region to examine the effect of ambiguity or consistency at each region. In the case of an interaction between group and ambiguity or consistency, follow-up tests examined effects of ambiguity or consistency at

the two levels of group. In the case of an interaction between ambiguity and consistency, ambiguity effects were examined at the two levels of consistency.

### 4.3.2 Results

Mean accuracy rates of experimental and filler texts were 89% for L1 participants (range = 75–100%) and 88% for L2 participants (range = 71–97%), which ensures that all participants paid attention to the texts. A summary of the reading time data and inferential statistics are provided in Table 3 and 4.

For brevity, a main effect of group was observed in all measures and at all regions due to longer reading times for L2 than L1 participants. Also, we do not discuss main effects of region nor group by region interactions below, as these are unrelated to the research questions that the present study address, and thus have little meaning unless they interact with another fixed effect.

*Disambiguating region*

There were marginal or significant main effects of ambiguity in all measures, with increased reading times for ambiguous than unambiguous sentences. This is evidence of the classic filled-gap effect in both L1 and L2 participants. There was also a significant main effect of consistency in total viewing times, indicating longer reading times for inconsistent than consistent conditions.

*Critical and spillover regions*

There was a significant main effect of consistency in all measures due to longer reading times for inconsistent than consistent sentences. This main effect was modulated by marginal and significant two-way interactions between ambiguity and consistency in regression path and total viewing times.

For regression path duration, planned comparisons at the two levels of consistency indicated shorter reading times for the ambiguous than unambiguous condition in inconsistent sentences (*estimate* = 0.057, *t* = 2.21, *SE* = 0.03, *p* = .028), but no significant differences between consistent sentences (*estimate* = 0.026, *t* = 0.80, *SE* = 0.03, *p* = .431).

Total viewing times revealed significant differences in both consistent and inconsistent conditions. For inconsistent sentences, the ambiguous condition again had significantly shorter reading times than the unambiguous conditions (*estimate* = 0.078, *t* = 2.44, *SE* = 0.03, *p* = .024), while the effect was reversed in consistent sentences, with longer reading times in the ambiguous condition (*estimate* = 0.057, *t* = 2.18, *SE* = 0.03, *p* = .030). This pattern of results is illustrated in Figure 1. It indicates lingering misinterpretation, given that reading times for consistent and inconsistent sentences were influenced by the ambiguity of the prior filler-gap sentences.

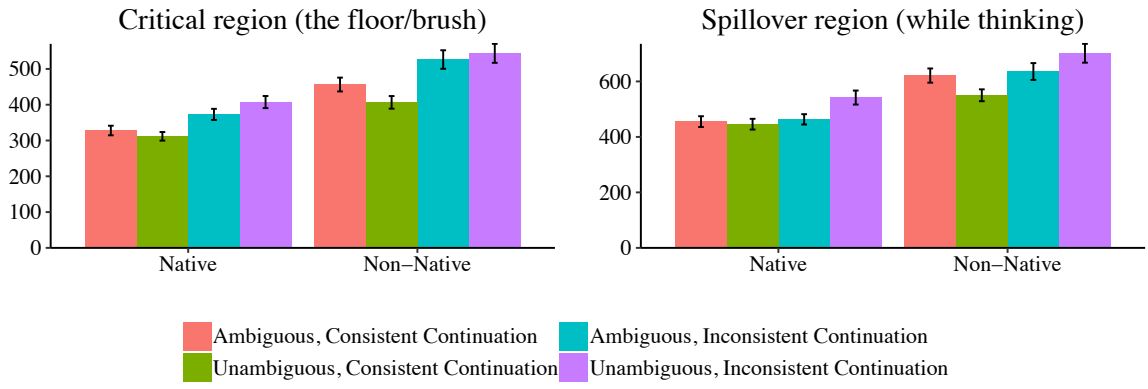


Figure 1. *Total viewing times in milliseconds at the critical and spillover regions in Experiment 1b.*

| | Disambiguating Region (the floor) | | Continuation Region (the floor/brush) | | Spillover Region (while thinking) | |
|---|---|---|---|---|---|---|
| | L1 | L2 | L1 | L2 | L1 | L2 |
| **First pass time** | | | | | | |
| Amb, CC | 344 (196) | 426 (203) | 254 (113) | 317 (134) | 306 (184) | 390 (198) |
| Unamb, CC | 318 (184) | 397 (182) | 242 (98) | 317 (125) | 310 (174) | 288 (208) |
| Amb, IC | 324 (194) | 426 (223) | 259 (108) | 335 (162) | 321 (169) | 379 (201) |
| Unamb, IC | 319 (193) | 384 (177) | 271 (135) | 323 (139) | 336 (192) | 406 (219) |
| **Regression path time** | | | | | | |
| Amb, CC | 625 (933) | 680 (951) | 278 (152) | 374 (634) | 513 (763) | 552 (849) |
| Unamb, CC | 499 (591) | 518 (573) | 332 (653) | 345 (219) | 449 (600) | 501 (560) |
| Amb, IC | 481 (439) | 597 (599) | 289 (205) | 372 (230) | 553 (1240) | 549 (769) |
| Unamb, IC | 456 (625) | 546 (606) | 304 (176) | 353 (217) | 684 (1234) | 615 (842) |
| **Total viewing time** | | | | | | |
| Amb, CC | 605 (448) | 908 (735) | 328 (207) | 456 (300) | 455 (301) | 621 (496) |
| Unamb, CC | 516 (372) | 667 (426) | 312 (186) | 407 (273) | 446 (300) | 550 (331) |
| Amb, IC | 692 (508) | 1014 (782) | 373 (240) | 526 (401) | 463 (288) | 636 (469) |
| Unamb, IC | 593 (421) | 796 (563) | 407 (262) | 544 (411) | 542 (391) | 702 (527) |

Table 3. *Reading times for three eye-movement measures at three regions of texts in Experiment 1b (SDs in parentheses). CC = Consistent Continuation, IC = Inconsistent Continuation, Amb = Ambiguous and Unamb = Unambiguous*.

| | First pass time | | | Regression path time | | | Total viewing time | | |
|---|---|---|---|---|---|---|---|---|---|
| | *E (SE)* | *t* | *p* | *E (SE)* | *t* | *p* | *E (SE)* | *t* | *p* |
| **Disambiguating region** | | | | | | | | | |
| Amb | 0.06 (0.03) | 2.04 | .053 | 0.12 (0.03) | 4.01 | < .001 | 0.18 (0.04) | 5.20 | < .001 |
| Group | 0.26 (0.06) | 4.56 | < .001 | 0.18 (0.07) | 2.51 | .014 | 0.32 (0.08) | 4.16 | < .001 |
| C | 0.02 (0.02) | 1.04 | .304 | 0.05 (0.03) | 1.72 | .087 | 0.12 (0.03) | 3.72 | < .001 |
| Amb:Group | 0.02 (0.05) | 0.47 | .640 | 0.03 (0.06) | 0.49 | .628 | 0.07 (0.05) | 1.39 | .169 |
| Amb:C | 0.01 (0.06) | 0.09 | .926 | 0.08 (0.07) | 1.17 | .254 | 0.01 (0.06) | 0.19 | .852 |
| Group:C | 0.003 (0.05) | 0.06 | .953 | 0.10 (0.06) | 1.63 | .103 | 0.01 (0.06) | 0.11 | .913 |
| Amb:Group:C | 0.08 (0.10) | 0.79 | .436 | 0.04 (0.12) | 0.35 | .731 | 0.01 (0.11) | 0.09 | .926 |
| **Critical / spillover regions** | | | | | | | | | |
| Amb | 0.01 (0.02) | 0.43 | .671 | 0.02 (0.02) | 0.83 | .417 | 0.01 (0.02) | 0.58 | .560 |
| Group | 0.23 (0.04) | 5.21 | < .001 | 0.18 (0.05) | 3.82 | < .001 | 0.27 (0.05) | 5.01 | < .001 |
| C | 0.04 (0.01) | 2.79 | .011 | 0.07 (0.02) | 2.98 | .008 | 0.13 (0.03) | 4.09 | < .001 |
| R | 0.14 (0.05) | 2.68 | .013 | 0.28 (0.07) | 3.76 | < .001 | 0.24 (0.06) | 4.01 | < .001 |
| Amb:Group | 0.004 (0.03) | 0.14 | .891 | 0.03 (0.04) | 0.86 | .401 | 0.05 (0.04) | 1.26 | .207 |
| Amb:C | 0.02 (0.03) | 0.75 | .456 | 0.08 (0.04) | 2.07 | .051 | 0.13 (0.05) | 2.95 | .007 |
| Group:C | 0.04 (0.03) | 1.74 | .099 | 0.04 (0.04) | 1.11 | .281 | 0.01 (0.04) | 0.19 | .855 |
| Amb:R | 0.04 (0.03) | 1.40 | .176 | 0.04 (0.03) | 1.22 | .223 | 0.03 (0.03) | 1.26 | .209 |
| Group:R | 0.01 (0.04) | 0.34 | .734 | 0.06 (0.06) | 0.95 | .348 | 0.003 (0.04) | 0.09 | .933 |
| C:Region | 0.0002 (0.03) | 0.01 | .994 | 0.05 (0.04) | 1.22 | .237 | 0.07 (0.04) | 1.74 | .096 |
| Amb:Group:C | 0.003 (0.05) | 0.06 | .955 | 0.05 (0.07) | 0.66 | .508 | 0.06 (0.08) | 0.74 | .460 |
| Amb:Group:R | 0.02 (0.05) | 0.36 | .724 | 0.01 (0.07) | 0.15 | .881 | 0.02 (0.05) | 0.30 | .764 |
| Amb:C:R | 0.01 (0.05) | 0.23 | .817 | 0.13 (0.09) | 1.44 | .165 | 0.02 (0.07) | 0.25 | .804 |
| Group:C:R | 0.03 (0.06) | 0.48 | .635 | 0.05 (0.08) | 0.72 | .479 | 0.003 (0.05) | 0.06 | .953 |
| Amb:Group:C:R | 0.16 (0.11) | 1.51 | .148 | 0.06 (0.15) | 0.42 | .679 | 0.01 (0.11) | 0.07 | .945 |

Table 4. *Summary of statistical analyses for Experiment 1b. Amb = Ambiguity, C = Consistency, and R = Region.*

### 4.3.3 Discussion

Both L1 and L2 participants had more difficulty reading filled-gap sentences than preposition-fronted, unambiguous sentences, indicating a filled-gap effect (Stowe, 1986).

Experiment 1b also provides clear evidence that the initially assigned misinterpretation lingers in both L1 and L2 participants. There was evidence of attenuated reading times for inconsistent continuation sentences and increased reading times for consistent continuation sentences after ambiguous filler-gap sentences. L1 and L2 participants' reading times at the critical and spillover regions were also generally longer for inconsistent than consistent sentences, an effect that was also observed in total viewing times at the disambiguating region of the filler-gap sentence. We interpret this finding as evidence that both groups generally succeeded in reanalysing the filled-gap sentences but that the initially assigned misinterpretation lingered some proportion of the time (Slattery et al. 2013). These results are also consistent with the comprehension accuracy data from Experiment 1a, which indicated lingering misinterpretation in ambiguous filled-gap sentences.

Experiments 2a and 2b aimed to extend these findings in two ways. Experiment 2a aimed to replicate the results of Experiment 1a using a sentence-picture matching task. To generalise the results of filled-gap sentences in Experiment 1b, Experiment 2b investigated lingering misinterpretation online following sentences containing non-filled-gaps.

## 4.4 Experiment 2a

Experiment 2a aimed to test the generalisability of our findings from Experiment 1a and tested for lingering misinterpretation in filler-gap dependencies using a different task, namely sentence-picture matching. One benefit of sentence-picture matching is that it avoids explicitly asking comprehension questions that repeat the ambiguous phrase, which

may prime or reactivate the initially assigned misinterpretation (Tabor et al., 2004; van Gompel et al., 2006). Using pictures may also reduce potential inferences that may have led to low accuracy overall for non-filled-gap sentences in Experiment 1a, given that even if participants make an inference, they still have to pick which picture they think provides the best match. Experiment 2a thus tested filled-gap and non-filled-gap sentences as in (13/14).

(13) *Filled-gap*

(a) *Ambiguous*

Anna looked at the table which the man carried the chair near quite hastily.

(b) *Unambiguous*

Anna looked at the table near which the man carried the chair quite hastily.

(14) *Non-filled gap*

(a) *Ambiguous*

Kevin saw the letter which the candle burnt very quickly beside last night.

(b) *Unambiguous*

Kevin saw the letter beside which the candle burnt very quickly last night.

(c) *Control*

Kevin saw the letter which the candle burnt very quickly last night.

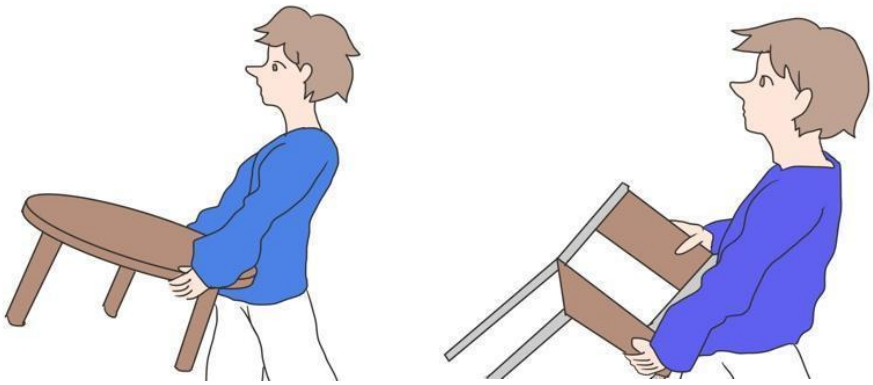

Figure 2. *Example picture pairs used for filled-gap sentences in Experiment 2a*.

Figure 3. *Example picture pairs used for non-filled-gap sentences in Experiment 2a.*

Experimental sentences were similar to those in Experiment 1a but modified to be depictable. For the non-filled-gap conditions, care was taken to attempt to minimise potential inferences by distinguishing between two different interpretations (e.g., "the candle burnt the letter" vs. "the candle burnt beside the letter"), and control sentences were created in addition to ambiguous and unambiguous sentences to test interpretation in sentences without a preposition. These additional controls were included to further test the success of reanalysis in temporarily ambiguous (14a), with a comparison that obviates the potential of inference in unambiguous (14b). That is, based on Experiment 1a, low

accuracy in unambiguous (14b) may obscure effects of lingering misinterpretation in the comparison to ambiguous (14a). However, comparing (14a) to (14c) allows us a further test of lingering misinterpretation, such that if the initially assigned misinterpretation in (14a) lingers, accuracy should be lower in (14a) than (14c).

The predictions for Experiment 2a were identical to Experiment 1a. If misinterpretation lingers, accuracy rates should be lower for ambiguous than unambiguous sentences and control sentences. Also, if L2 participants are more persistent with misinterpretation than L1 participants, accuracy rates should be lower for L2 participants in the ambiguous condition.

### 4.4.1 Method

*Participants*

40 L1 English speakers (7 males, mean age = 19; range = 18–23) and 40 L2 English speakers (14 males, mean age = 23; range = 18–47) from the University of Reading community, none of whom took part in Experiment 1a/b, completed Experiment 2a for course credit or payment. The L2 participants had various L1 backgrounds[14], started learning English from age eight onwards, and their performance on the OPT indicated that they were upper intermediate-advanced English language learners (mean = 76, SD = 10.6; range = 51–94).

*Materials*

[14] First languages of L2 participants were Greek (8), Bulgarian (5), Italian (5), French (3), Indonesian (3), Polish (3), Lithuanian (2), Spanish (2), Turkish (2), Arabic (1), Bangladeshi (1), Chinese (1), Dutch (1), Finnish (1), German (1), and Thai (1).

Experiment 2a employed 12 sets of filled-gap sentences (13) and 18 sets of non-filled-gap sentences (14) (see Appendix C). A pair of pictures was constructed for each filled-gap and non-filled-gap sentence as in Figures 1 and 2. In the filled-gap picture pair, the two pictures depicted the initial misinterpretation ("the man carried the table") and globally correct interpretation ("the man carried the chair"). The same was true of the non-filled-gap conditions ("the candle burnt the letter" vs. "the candle burnt beside the letter"). The experiment also contained 78 filler sentences, each with an accompanying picture pair.

*Procedure and data analysis*

The procedure was the same as in Experiment 1a, except that participants were instructed to choose which picture they felt best corresponded to the sentence rather than answer a comprehension question. The experiment was administered via the IBEX Farm web-based experimental presentation platform (Drummond, 2013). Participants however completed the experiment in a traditional lab setting. Similarly to Experiment 1, the order of correct and incorrect pictures was counterbalanced. The data analysis was the same as Experiment 1a. Regarding non-filled-gap sentences, we initially did not include control sentences in the statistical model but when there was an indication of lingering misinterpretation, we conducted an additional analysis to compare ambiguous and control sentences.

### 4.4.2 Results

A summary of comprehension accuracy rates and statistics is shown in Tables 1 and 2. Overall accuracy rates of the comprehension questions following fillers were 94% for L1 participants (range = 69–100) and 95% for L2 participants (range = 83–100).

For the filled-gap condition, the model showed a main effect of ambiguity due to lower accuracy rates for the ambiguous than unambiguous conditions, with no interaction by group.

The non-filled-gap condition similarly revealed a main effect of ambiguity in the same direction and a marginal effect of group with lower accuracy rates for L2 than L1 participants. As these effects were modulated by a significant interaction between ambiguity and group, we performed planned comparisons. This analysis showed significantly lower accuracy rates for the ambiguous than unambiguous condition for L2 participants only (L1: *estimate* = 0.099, *z* = 0.33, *SE* = 0.30, *p* = .738; L2: *estimate* = 1.109, *z* = 3.99, *SE* = 0.28, *p* < .001). Pairwise comparisons by ambiguity also revealed L1/L2 differences only in the ambiguous condition, due to lower accuracy rates for L2 than L1 participants (ambiguous: *estimate* = 1.088, *z* = 3.38, *SE* = 0.32, *p* < .001; unambiguous: *estimate* = 0.008, *z* = 0.02, *SE* = 0.35, *p* = .981).

To further analyse the non-filled-gap condition with the control condition, we compared the ambiguous condition with the control condition. This analysis showed main effects of ambiguity and group and a significant interaction between them (ambiguity: *estimate* = 1.658, *z* = 5.81, *SE* = 0.29, *p* < .001; group: *estimate* = 0.550, *z* = 2.03, *SE* = 0.27, *p* = .043: interaction: *estimate* = 1.004, *z* = 2.78, *SE* = 0.36, *p* = .005). Pairwise comparisons by group revealed for both participant groups higher accuracy rates for the control condition than the ambiguous condition (L1: *estimate* = 1.1470, *z* = 3.42, *SE* = 0.34, *p* < .001; L2: *estimate* = 2.199, *z* = 5.69, *SE* = 0.39, *p* < .001). The interaction appears to be driven by the lower accuracy in the ambiguous condition for L2 participants, discussed in the previous paragraph.

### 4.4.3 Discussion

The analyses revealed that both L1 and L2 participants were less accurate in their sentence-picture matching following ambiguous than unambiguous sentences in the filled-gap condition. Regarding the non-filled-gap condition, L2 but not L1 participants were also less accurate in the ambiguous than unambiguous condition. However, follow-up

analyses with the control conditions revealed that comprehension accuracy rates were lower for ambiguous than control sentences in both L1 and L2 participants. This may suggest that L1 participants sometimes persisted with the initial misinterpretation during offline language comprehension, but low accuracy in the unambiguous condition obscured this effect. We discuss this issue in more detail in the General Discussion.

Unlike Experiment 1a, L1/L2 differences were observed in the non-filled-gap conditions in that L2 participants had more difficulty answering questions following ambiguous sentences than L1 participants. These results are compatible with previous studies indicating increased lingering misinterpretation following garden-path sentences in L2 speakers (e.g., Pozzan & Trueswell, 2016). However, given that no such differences were observed in filled-gap sentences, Experiment 2a may suggest that the nature of disambiguation influences L1/L2 differences, with differences appearing in non-filled-gap but not filled-gap sentences. We return to this issue in Experiment 3, which provides a more direct test of these potential L1/L2 differences.

Experiment 1a and 2a showed that the initially assigned misinterpretation lingers in both filled-gap and non-filled gap sentences for L2 speakers. For L1 speakers, while similar effects were observed for filled-gap sentences, the results for non-filled-gap sentences were less clear, especially in the ambiguous/unambiguous comparison. Nonetheless, as indicated by the ambiguous/control condition comparisons, this does not necessarily mean that L1 speakers can fully erase misinterpretation in the non-filled-gap conditions, given the relatively low accuracy rates for the ambiguous and unambiguous conditions compared to control sentences. To test lingering misinterpretation in non-filled-gap sentences more implicitly, Experiment 2b adopted eye-tracking to investigate lingering misinterpretation in a design similar to Experiment 1b.

### 4.5 Experiment 2b

To test how and whether misinterpretations of non-filled-gap sentences linger during online processing, in Experiment 2b participants read sentences like (15) while their eye-movements were monitored.

(15a) *Ambiguous, Consistent Continuation*

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

Alan was driving near the school bus very patiently on the road.

(15b) *Unambiguous, Consistent Continuation*

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

Alan was driving near the school bus very patiently on the road.

(15c) *Ambiguous, Inconsistent Continuation*

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

Alan was driving the school bus very patiently on the road.

(15d) *Unambiguous, Inconsistent Continuation*

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

Alan was driving the school bus very patiently on the road.

It was extremely crowded.

Each experimental set contained three sentences, which manipulated ambiguity and consistency as in Experiment 1b. The first sentence in (15a/15c) is temporarily ambiguous while (15b/15d) is unambiguous due to the fronted preposition. The second, continuation sentence in (15a/15b) is consistent with the globally correct interpretation of the filler-gap sentences ("Alan was driving near the school bus"), whereas the continuation sentence in

(15c/15d) is inconsistent with this analysis, but consistent with the initial misinterpretation ("Alan was driving the school bus").

The predictions of Experiment 2b are similar to Experiment 1b. If the initial misinterpretation of temporarily ambiguous filler-gap sentences is completely erased, longer reading times are expected in the continuation sentence in inconsistent (15c/15d) than consistent (15a/15b). If the initial misinterpretation lingers, reading times for inconsistent sentences should be attenuated following ambiguous filler-gap sentences. This would predict shorter reading times in the continuation sentence in (15c) than (15d). Lingering misinterpretation also predicts longer reading times for the ambiguous/consistent condition (15a) than the unambiguous/consistent condition (15b). Thus, the crucial prediction is whether a main effect of consistency is observed or an ambiguity by consistency interaction.

### 4.5.1 Method

*Participants*

The participants from Experiment 2a also took part in Experiment 2b. As in Experiments 1a/b, participants completed Experiment 2b first, at least one week before Experiment 2a.

*Materials*

24 sets of experimental texts as in (15) were created in a Latin square design with two levels of ambiguity (ambiguous/unambiguous) and consistency (consistent/inconsistent) (see appendix D for full list). The filler-gap sentences appeared on the first line onscreen, while the continuation sentence appeared on the second line. A wrap-up sentence followed, also on the second line. Experimental trials were interleaved with 72 additional filler texts. All experimental texts and two-third of filler texts were followed by a binary yes/no

comprehension question, which asked about different parts of text evenly but never referred to the interpretation of the temporary ambiguity in the experimental texts.

*Procedure and data analysis*

The procedure was the same as in Experiment 1b except that after the main experimental session, L2 participants were required to look through a vocabulary list containing words used for the filler ("the school bus") and the embedded clause verb ("driving") in the experimental texts and tick a box if they were unsure of its meaning.

For analysis, the experimental texts were divided into three regions. These include the disambiguating region in the first sentence ("earlier"), and the critical region ("the school bus"), and spillover region ("very patiently") in the second, continuation sentence. As in Experiment 1b, the spillover region contains the rest of the continuation sentence after the critical region except the last two words. Though the temporary ambiguity of the first sentence can potentially be disambiguated at the preposition ("near"), the following word was specified as the disambiguating region, given that the preposition does not appear in the unambiguous conditions, nor does it decisively disambiguate towards the globally correct interpretation (e.g., The girl was in the school bus which Alan was driving very slowly near the kindergarten.). Reported reading time measures, data exclusion criteria and the data analysis method were identical to Experiment 1b. Skipping rates were approximately 5% for the L1 data and 3% for the L2 data. Trials including words that L2 participants did not know the meaning of were also removed, which affected less than 0.1% of the L2 data.

### 4.5.2 Results

Mean accuracy rates to comprehension questions were 89% for L1 participants (range = 75–97%) and 88% for L2 participants (range = 76–97%). Raw reading time data and the

inferential analyses are summarised in Table 5 and 6. As in Experiment 1b, there was a main effect of group in most reported measures, as reading times were longer for L2 participants.

*Disambiguating region*

There was a significant main effect of ambiguity in regression path duration and total viewing times due to longer reading times for ambiguous than unambiguous sentences, with the effect for total viewing times being qualified by a marginal three-way interaction between ambiguity, group and consistency. To interpret the marginal three-way interaction, we conducted planned 2×2 analyses by group, which showed significant main effects of ambiguity for both groups (L1: *estimate* = 0.313, $t$ = 6.33, *SE* = 0.05, $p < .001$; L2: *estimate* = 0.266, $t$ = 5.17, *SE* = 0.05, $p < .001$) and a further marginal two-way interaction between ambiguity and consistency only for L1 participants (L1: *estimate* = 0.159, $t$ = 1.90, *SE* = 0.08, $p$ = .071; L2: *estimate* = 0.068, $t$ = 0.84, *SE* = 0.08, $p$ = .404). Planned comparisons comparing the two ambiguous and two unambiguous conditions did not show any significant differences however (ambiguous: *estimate* = 0.080, $t$ = 1.51, *SE* = 0.05, $p$ = .132; unambiguous: *estimate* = 0.079, $t$ = 1.32, *SE* = 0.06, $p$ = .200).

*Critical and spillover regions*

No measure showed significant main effects of ambiguity or consistency. In first pass reading times, there was a significant interaction between ambiguity and consistency. To interpret this interaction, we performed pairwise comparisons by consistency. This analysis showed significantly shorter reading times for the ambiguous than unambiguous condition in inconsistent sentences (*estimate* = 0.053, $t$ = 2.19, *SE* = 0.02, $p$ = .041), but no significant differences between consistent sentences (*estimate* = 0.037, $t$ = 1.34, *SE* = 0.03, $p$ = .194). This pattern of results is illustrated in Figure 4.

First pass times also showed a significant interaction between consistency, group and region. Analyses for each region showed a significant two-way interaction between consistency and group only in the critical region (*estimate* = 0.157, *t* = 2.43, *SE* = 0.06, *p* = .020). Analyses for each group showed significantly longer reading times for inconsistent than consistent sentences for L1 participants only (L1: *estimate* = 0.106, *t* = 2.26, *SE* = 0.05, *p* = .030; L2: *estimate* = 0.051, *t* = 1.03, *SE* = 0.05, *p* = .312). This consistency effect for L1 participants only in this measure may suggest that L1 participants were more successful in completing reanalysis than L2 participants.

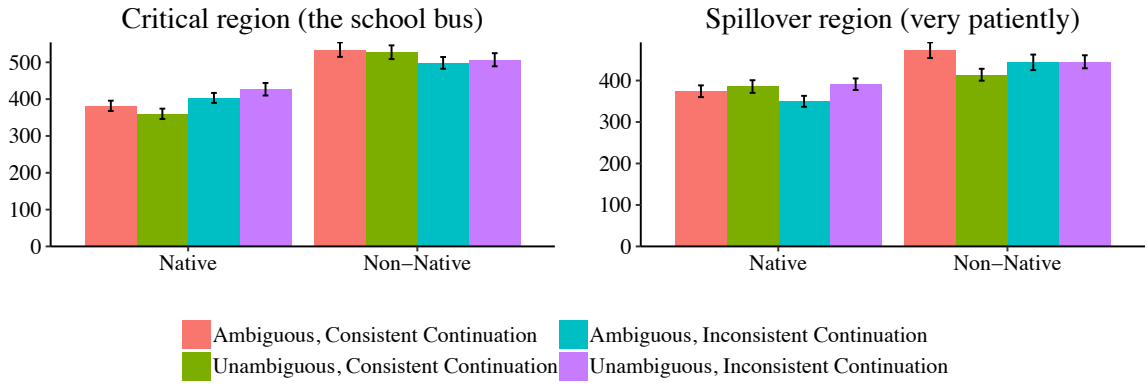


Figure 4. *First pass reading times in milliseconds at the critical and spillover regions in Experiment 2b*.

For regression path duration, as there was a significant four-way interaction, we analysed each region separately. For the critical region, this analysis showed a marginal interaction between consistency and group (*estimate* = 0.127; *t* = 2.04, *SE* = 0.06, *p* = .051). While L1 participants showed no significant differences, L2 participants had marginally longer reading times in consistent than inconsistent conditions (L1: *estimate* = 0.036, *t* = 0.76, *SE* = 0.05, *p* = .455; L2: *estimate* = 0.090, *t* = 1.91, *SE* = 0.05, *p* = .068). Although we are cautious in interpreting this marginal effect, it may be consistent with L2 participants not always completing reanalysis of the temporary ambiguity in the filler-gap sentences. For the spillover region, there was a significant main effect of consistency (*estimate* = 0.077; *t* = 2.50, *SE* = 0.03, *p* = .023), a marginal ambiguity by consistency

interaction (*estimate* = 0.114; *t* = 1.84, *SE* = 0.06, *p* = .082), and a significant three-way interaction between ambiguity, consistency and group (*estimate* = 0.300; *t* = 2.64, *SE* = 0.11, *p* = .010). 2x2 analyses by group revealed no main effects nor interactions for L1 participants (all effects: *estimate* < 0.058, *t* < 1.07, *SE* > 0.04, *p* > .291). For L2 participants, there was a significant main effect of consistency, with longer reading times for inconsistent than consistent conditions (*estimate* = 0.105, *t* = 2.56, *SE* = 0.04, *p* = .012), and a significant two-way interaction between ambiguity and consistency (*estimate* = 0.256, *t* = 2.92, *SE* = 0.09, *p* = .005). Pairwise comparisons indicated significantly shorter reading times in the ambiguous than unambiguous condition for inconsistent sentences (*estimate* = 0.157, *t* = 2.52, *SE* = 0.06, *p* = .013). Although there was a numerical trend in the opposite direction for consistent sentences, the difference between ambiguous and unambiguous conditions was not significant (*estimate* = 0.099, *t* = 1.68, *SE* = 0.06, *p* = .112).

Total viewing times showed a significant interaction between ambiguity and consistency. Follow-up analyses showed an effect of ambiguity in inconsistent sentences due to significantly reduced reading times in the ambiguous condition (*estimate* = 0.072, *t* = 2.19, *SE* = 0.03, *p* = .039), but this difference was not significant in consistent sentences (*estimate* = 0.035, *t* = 1.14, *SE* = 0.03, *p* = .257). There was also a significant three-way interaction between consistency, group and region. Analysis for each region showed a significant two-way interaction between consistency and group in the critical region (*estimate* = 0.168, *t* = 2.46, *SE* = 0.07, *p* = .017). Pairwise comparisons by group at the critical region revealed a significant effect of consistency for L1 participants only, with longer reading times for inconsistent than consistent sentences (L1: *estimate* = 0.135, *t* = 2.50, *SE* = 0.05, *p* = .019; L2: *estimate* = 0.033, *t* = 0.56, *SE* = 0.06, *p* = .578).

| | Disambiguating Region (earlier) | | Critical Region (the school bus) | | Spillover Region (very patiently) | |
|---|---|---|---|---|---|---|
| | L1 | L2 | L1 | L2 | L1 | L2 |
| **First pass time** | | | | | | |
| Ambiguous, CC | 333 (203) | 371 (208) | 382 (216) | 535 (305) | 374 (219) | 473 (294) |
| Unambiguous, CC | 340 (185) | 343 (179) | 360 (217) | 527 (288) | 386 (236) | 414 (221) |
| Ambiguous, IC | 339 (201) | 334 (156) | 403 (209) | 498 (248) | 350 (205) | 444 (290) |
| Unambiguous, IC | 310 (152) | 322 (128) | 427 (264) | 507 (276) | 391 (216) | 445 (243) |
| **Regression path time** | | | | | | |
| Ambiguous, CC | 899 (1002) | 832 (1200) | 538 (440) | 721 (791) | 470 (467) | 656 (847) |
| Unambiguous, CC | 778 (941) | 689 (932) | 547 (499) | 695 (626) | 521 (657) | 529 (508) |
| Ambiguous, IC | 963 (1763) | 716 (812) | 534 (502) | 676 (800) | 534 (726) | 621 (785) |
| Unambiguous, IC | 845 (1325) | 758 (1178) | 647 (980) | 614 (466) | 545 (717) | 780 (1025) |
| **Total viewing time** | | | | | | |
| Ambiguous, CC | 749 (483) | 943 (619) | 590 (420) | 993 (737) | 573 (375) | 775 (530) |
| Unambiguous, CC | 595 (376) | 701 (456) | 608 (456) | 897 (650) | 595 (416) | 761 (709) |
| Ambiguous, IC | 815 (601) | 909 (723) | 665 (462) | 822 (508) | 569 (396) | 741 (527) |
| Unambiguous, IC | 544 (361) | 690 (460) | 661 (431) | 910 (539) | 603 (387) | 774 (463) |

Table 5. *Reading times for three eye-movement measures at three regions of texts in Experiment 2b (SDs in parentheses). CC = Consistent Continuation and IC = Inconsistent Continuation.*

| | First pass time | | | Regression path time | | | Total viewing time | | |
|---|---|---|---|---|---|---|---|---|---|
| | *E (SE)* | *t* | *p* | *E (SE)* | *t* | *p* | *E (SE)* | *t* | *p* |
| **Disambiguating region** | | | | | | | | | |
| Amb | 0.02 (0.03) | 0.67 | .502 | 0.13 (0.04) | 3.32 | .0014 | 0.29 (0.04) | 7.70 | < .001 |
| Group | 0.06 (0.04) | 1.52 | .133 | 0.13 (0.08) | 1.69 | .095 | 0.18 (0.08) | 2.22 | .030 |
| C | 0.04 (0.02) | 1.78 | .076 | 0.01 (0.04) | 0.37 | .716 | 0.03 (0.03) | 1.00 | .322 |
| Amb:Group | 0.03 (0.05) | 0.52 | .607 | 0.05 (0.08) | 0.59 | .559 | 0.05 (0.07) | 0.70 | .487 |
| Amb:C | 0.03 (0.04) | 0.65 | .517 | 0.02 (0.07) | 0.23 | .818 | 0.05 (0.05) | 0.88 | .379 |
| Group:C | 0.02 (0.04) | 0.51 | .612 | 0.02 (0.07) | 0.31 | .759 | 0.06 (0.06) | 1.08 | .296 |
| Amb:Group:C | 0.12 (0.09) | 1.43 | .172 | 0.02 (0.15) | 0.14 | .888 | 0.22 (0.12) | 1.83 | .081 |
| **Critical / spillover regions** | | | | | | | | | |
| Amb | 0.01 (0.02) | 0.39 | .701 | 0.02 (0.02) | 1.17 | .246 | 0.02 (0.02) | 0.87 | .397 |
| Group | 0.23 (0.05) | 4.54 | < .001 | 0.21 (0.06) | 3.56 | < .001 | 0.31 (0.06) | 4.88 | < .001 |
| C | 0.01 (0.02) | 0.45 | .654 | 0.03 (0.02) | 1.19 | .245 | 0.04 (0.04) | 1.06 | .298 |
| R | 0.13 (0.05) | 2.46 | .020 | 0.16 (0.06) | 2.54 | .017 | 0.15 (0.06) | 2.57 | .016 |
| Amb:Group | 0.05 (0.03) | 1.54 | .127 | 0.04 (0.04) | 0.97 | .342 | 0.02 (0.05) | 0.50 | .620 |
| Amb:C | 0.09 (0.03) | 2.48 | .015 | 0.08 (0.04) | 2.03 | .055 | 0.11 (0.05) | 2.17 | .034 |
| Group:C | 0.07 (0.04) | 1.72 | .096 | 0.03 (0.04) | -0.81 | .422 | 0.07 (0.05) | 1.33 | .187 |
| Amb:R | 0.04 (0.04) | 1.06 | .301 | 0.04 (0.04) | 1.02 | .308 | 0.02 (0.04) | 0.58 | .571 |
| Group:R | 0.11 (0.05) | 2.46 | .016 | 0.02 (0.05) | 0.50 | .621 | 0.11 (0.04) | 2.66 | .011 |
| C:R | 0.04 (0.04) | 0.89 | .382 | 0.10 (0.05) | 2.07 | .049 | 0.02 (0.04) | 0.55 | .590 |
| Amb:Group:C | 0.00 (0.07) | 0.03 | .979 | 0.10 (0.07) | 1.38 | .169 | 0.15 (0.10) | 1.57 | .120 |
| Amb:Group:R | 0.12 (0.07) | 1.83 | .071 | 0.02 (0.08) | 0.20 | .841 | 0.07 (0.05) | 1.20 | .231 |
| Amb:C:R | 0.04 (0.07) | 0.56 | .580 | 0.07 (0.11) | 0.63 | .536 | 0.01 (0.06) | 0.15 | .881 |
| Group:C:R | 0.18 (0.08) | 2.33 | .023 | 0.19 (0.08) | 2.23 | .028 | 0.20 (0.06) | 3.15 | .002 |
| Amb:Group:C:R | 0.10 (0.12) | 0.79 | .429 | 0.40 (0.15) | 2.61 | .009 | 0.05 (0.14) | 0.40 | .692 |

Table 6. *Summary of statistical analyses for Experiment 2b. Amb = Ambiguity, C: Consistency, R: Region.*

### 4.5.3 Discussion

Experiment 2b showed several important findings. First, participants took more time to read ambiguous than unambiguous sentences. Although the lexical material preceding the disambiguating region differs from the unambiguous condition's counterpart, this reading difference may indicate reanalysis cost.

Second, there was evidence that misinterpretation lingers for L1 and L2 speakers. However, this lingering effect was observed significantly only in the inconsistent condition, and no effects were significant in the consistent condition. This result is partly consistent with Experiment 1b, which showed lingering misinterpretation in both consistent and inconsistent conditions.

Also, main effects of consistency, with longer reading times for inconsistent than consistent sentences, were more elusive in Experiment 2b than in Experiment 1b, especially for L2 speakers. The only measure in Experiment 2b that showed this effect in L2 participants was regression path. However, here the effect was driven by the unambiguous conditions. L2, but not L1, participants also showed a trend for a reversed inconsistency effect in one measure. Together, this may indicate that L2 participants were less successful at reanalysing the temporary ambiguity than L1 participants.

Experiment 2b suggests L2 speakers may be less likely to derive the correct interpretation of temporarily ambiguous non-filled-gap sentences than L1 participants. We did not find significant evidence that L2 participants were less successful than L1 participants in Experiment 1b however, which tested filled-gap sentences. This pattern of results is compatible with Experiment 2a, which showed that L2 participants were more persistent with misinterpretation than L1 participants in non-filled-gap sentences, but not filled-gap sentences. However, given that the filled-gap and non-filled-gap manipulations differed in lexical material here, and given that L1/L2 differences were not observed in

Experiment 1a, we decided to conduct one final offline experiment to more directly compare filled-gap and non-filled-gap sentences, using a within-sentence manipulation.

### 4.6 Experiment 3

Experiment 3 aimed to replicate the offline results from Experiments 1a/2a with an additional aim to directly compare filled-gap and non-filled-gap sentences, using sentences like (16).

(16a) *Filled-gap*, *Ambiguous*

Elisa noticed the truck which the policeman watched the car from earlier that morning.

(16b) *Filled-gap*, *Unambiguous*

Elisa noticed the truck from which the policeman watched the car earlier that morning.

(16c) *Non-filled-gap*, *Ambiguous*

Elisa noticed the truck which the policeman watched very quietly from earlier that morning.

(16d) *Non-filled-gap*, *Unambiguous*

Elisa noticed the truck from which the policeman watched very quietly earlier that morning.

Question: Did the policeman watch the truck?

(16a/b) are filled-gap sentences while (16c/d) are non-filled-gap sentences. As in Experiments 1a/2a, (16) manipulates ambiguity such that (16a/c) are temporarily ambiguous sentences and (16b/d) are unambiguous controls. One crucial difference from Experiments 1a/2a is that the lexical material used in (16) is the same between filled-gap and non-filled-gap sentences up to the disambiguating region, and the same question form, which refers to the initially assigned misinterpretation, is adopted for both sentence types to directly compare them.

Similarly to Experiments 1a/2a, we expected more incorrect responses for ambiguous than unambiguous sentences in both sentence types if misinterpretation lingers. It was also expected that L2 participants would have lower accuracy than L1 participants in ambiguous conditions, especially in the non-filled-gap sentences. Given the numerical differences in accuracy rates observed in Experiments 1a/2a, we also expected lower accuracy rates for non-filled-gap than filled-gap sentences in Experiment 3.

### 4.6.1 Method

*Participants*

48 L1 English speakers (9 males, mean age = 20; range = 18–48) and 48 L2 English speakers (12 males, mean age = 21; range = 17–36) of various L1 backgrounds[15] from the University of Reading community, who did not complete Experiments 1/2, participated in Experiment 3 for either course credit or payment. The L2 participants started learning English in a school environment after age five onwards. L2 participants completed the Quick Placement test after the experiment. This test indicated an average score 80% (48 out of 60; SD = 0.4; range = 52–98%), showing that the L2 participants were upper intermediate to advanced English language learners as in Experiments 1/2.

*Materials*

Experimental materials consisted of 24 sets of sentences as in (16a/b), which manipulated ambiguity and filler-gap type (see Appendix E). Each sentence was followed by a comprehension question referring to the initially assigned misinterpretation. Experiment 3

[15] First languages of the L2 participants were Greek (11), Italian (6), Bulgarian (3), German (3), Romance (3), Cantonese (2), Danish (2), French (2), Polish (2), Russian (2), Slovak (2), Spanish (2), Bahasa (1), Chinese (1), Croatian (1), Dutch (1), Lithuanian (1), Malay (1), Portuguese (1), Sinhala (1).

also contained 48 filler sentences, all of which were accompanied by a yes/no comprehension question. Experimental items were presented with four counterbalanced lists in a Latin Square design.

*Procedure and data analysis*

The procedure was identical to Experiment 1a except that Experiment 3 used IBEX farm as conducted in Experiment 2a. Data analysis was similar to Experiments 1a/2a, but additionally included a sum-coded fixed effect of filler-gap type (filled-gap/non-filled-gap), along with the relevant interactions.

### 4.6.2 Results

Mean accuracy rates of filler sentences was 93% for L1 participants (range = 67–100) and 95% for L2 participants (range = 75–100). Comprehension accuracy rates of experimental sentences and inferential statistics are summarised in Tables 1 and 7.

The analysis showed significant main effects of ambiguity and filler-gap type due to lower accuracy rates for ambiguous than unambiguous sentences, and non-filled-gap than filled-gap sentences respectively. There were also significant and marginal interactions between ambiguity and group and between ambiguity, filler-gap type and group, respectively. To interpret the three-way interaction, we conducted follow-up analysis on filler-gap type.

Filled-gap sentences revealed a main effect of ambiguity with no interaction by group, as accuracy rates were lower for ambiguous than unambiguous sentences (*estimate* = 0.428, $z = 2.89$, $SE = 0.15$, $p = .004$).

Non-filled-gap sentences similarly showed a main effect of ambiguity (*estimate* = 0.729, $z = 4.42$, $SE = 0.16$, $p < .001$), but this effect was qualified by an ambiguity by group interaction (*estimate* = 1.058, $z = 2.86$, $SE = 0.37$, $p = .004$). Pairwise comparisons

indicated that accuracy rates were lower for L2 than L1 participants only in the ambiguous conditions (ambiguous: *estimate* = 0.702, *z* = 2.02, *SE* = 0.35, *p* = .043; unambiguous: *estimate* = 0.500, *z* = 1.26, *SE* = 0.40, *p* = .208).

| | Experiment 3 | | |
|---|---|---|---|
| | *Estimate (SE)* | *z value* | *p* value |
| Ambiguity | 0.54 (0.11) | 4.72 | < .001 |
| Filler gap type | 1.30 (0.13) | 10.03 | < .001 |
| Group | 0.09 (0.22) | 0.43 | .671 |
| Ambiguity:Filler gap type | 0.27 (0.20) | 1.35 | .178 |
| Ambiguity:Group | 0.61 (0.24) | 2.53 | .011 |
| Filler gap type:Group | 0.30 (0.25) | 1.18 | .240 |
| Ambiguity:Filler gap type:Group | 0.74 (0.40) | 1.87 | .062 |

Table 7. *Summary of statistical analyses for Experiment 3*.

### 4.6.3 Discussion

Experiment 3 suggested lingering misinterpretation in both L1 and L2 speakers and crucially provided further evidence that L2 speakers persist with misinterpretation more strongly than L1 speakers but only in non-filled-gap sentences, a finding consistent with Experiment 2a. Regarding comparisons of filled-gap and non-filled-gap sentences, there was clear evidence that non-filled-gap sentences had lower accuracy than filled-gap sentences, irrespective of ambiguity. Why accuracy rates are lower for L2 than L1 speakers only in the non-filled-gap construction, along with the overall differences between non-filled-gap and filled-gap sentences, are discussed in detail below.

## 4.7 General Discussion

The present study examined whether initially assigned misinterpretations linger in sentences with filler-gap dependencies. We compared offline and online tasks and potential differences between L1 and L2 speakers in filled-gap and non-filled-gap constructions. The results of Experiments 1a, 2a and 3 showed that initially assigned misinterpretations linger offline in both L1 and L2 speakers. Experiment 2a and 3 also suggested that L2 speakers are more persistent with misinterpretation than L1 speakers in

non-filled-gap sentences but not in filled-gap sentences. Experiments 1b and 2b indicated that the initially assigned misinterpretation of filler-gap sentences persists in memory and interferes with subsequent language processing in both L1 and L2 speakers. Experiment 2b also suggested L2 speakers may more frequently fail to complete reanalysis in ambiguous non-filled-gap sentences than L1 speakers. Below, implications of these results are discussed in turn.

*Lingering misinterpretation caused by filler-gap dependencies*

The lingering effects observed for both L1 and L2 speakers in the filled-gap construction are consistent with previous studies showing that initially assigned misinterpretations are not completely erased after reanalysis in L1 and L2 processing (Christianson et al., 2001, 2006; Jacob & Felser, 2016; Patson et al., 2009; Slattery et al., 2013; Sturt, 2007; van Gompel et al., 2006). While these studies tested classic garden-path sentences, the present study is the first to show that misinterpretation also lingers in filler-gap dependencies at the online level. Our offline and online results for filled-gap sentences are broadly compatible with the Good-Enough Sentence Processing Account, which assumes that readers often create imperfect representations during language comprehension (e.g., Christianson et al., 2001, 2006; Ferreira et al., 2002; Ferreira et al., 2001; Ferreira & Patson, 2007; Karimi & Ferreira, 2016; Slattery et al., 2013). Recently, Slattery et al. reported that for garden-path sentences, L1 speakers perform syntactic reanalysis and lingering effects result from failures to erase the initial misinterpretation completely. This account assumes that the globally correct interpretation is overlaid on the initial misinterpretation after reanalysis. Our filled-gap results are consistent with this account for L1 and L2 speakers. The results of Experiment 1b suggest both groups succeeded in reanalysis of filled-gap sentences, given that reading times at the critical region were generally longer for inconsistent than consistent continuation sentences. The effects of ambiguity on consistency effects

observed in total viewing times at this region in Experiment 1b, along with our offline results, suggest that despite this, the initial misinterpretation was not fully erased, as predicted by good enough processing.

Although we believe our results of non-filled-gap sentences also indicate lingering misinterpretation, the pattern of effects here was a little more complex. Experiment 1a provided only suggestive evidence of lingering misinterpretation in non-filled-gaps, while comparisons between the ambiguous and unambiguous conditions in Experiments 2a and 3 showed lingering misinterpretation in L2 speakers only. However, interpretation of our results here is complicated by the generally low accuracy in the unambiguous condition that included a fronted preposition. In Experiment 2a, where we tested an additional control condition without a preposition, we found that this control condition received higher accuracy than the ambiguous condition in both L1 and L2 speakers. This might be taken as evidence of lingering misinterpretation in the ambiguous condition for both groups. Experiment 2b also provided evidence of lingering misinterpretation of the initially assigned gap during L1 and L2 processing. The clearest evidence of this came from reading times for inconstant continuations, where reading times were shorter following ambiguous than unambiguous sentences. However, unlike Experiment 1b, lingering effects were found in inconsistent but not consistent continuations in Experiment 2b.

We believe these differences may be due to processing and interpretation of the preposition-fronted unambiguous condition. We argued that readers may make an interference in this condition that may have lowered accuracy. For example, in “The boy bought the novel about which the girl read”, the reader may infer that “the novel” was both “read about” and subsequently “read”. However, we cannot rule out that both groups may not have always constructed the globally correct structure, such that the disambiguating cue used for the non-filled-gap construction did not disambiguate as strongly as intended. Although we expected the fronted preposition to block the temporary ambiguity, Radford

(2009) demonstrates based on corpus data that L1 speakers sometimes engage in "preposition copying", as in (17).

(17) *Elisa noticed the truck from which the policeman watched from.

(17) is ungrammatical, as although the prepositional wh-phrase ("from which") is moved to the front of the clause, a copy of the preposition remains in its canonical position. Preposition copying may occur such that during wh-dependency formation, speakers spell out a preposition in front of the wh-constituent (e.g., "from which"). In wh-dependency, this preposition-fronted wh-filler needs to be stored in memory until the gap ("watched") appears. However, processing resource limitations may prevent speakers from recalling the already spelled-out preposition. As a result, speakers may sometimes spell out the preposition redundantly in the gap position. Potentially, similar mechanisms may underlie language comprehension, which may make the fronted prepositional phrase an unreliable disambiguating cue as a result of a failure to recall the presence of the fronted preposition. Radford, Felser and Boxell (2012) further observed that L1 speakers judge sentences like (17) to be as acceptable as sentences without the copied preposition, similar to the unambiguous non-filled-gap sentences used in the present study. Although how sentences like (17) are processed during online reading has not been tested, given Radford et al.'s results, it is possible that readers sometimes forget the fronted preposition, and when no copy of the preposition appears in the clause, compute a technically ungrammatical thematic relation between the filler and gap.

Note that this may not apply to the filled-gap construction, where the globally correct interpretation is clearly indicated by the direct object noun phrase immediately following the gap position (e.g., Elisa noticed the truck from which the policeman watched the car.). Thus, the difference in results between the filled-gap and non-filled-gap

construction may be attributed to the low diagnosticity of the prepositional wh-phrase for disambiguation in non-filled-gap sentences, compared to the overt noun phrase in filled-gap structures (for similar claims relating to cue diagnosticity in garden-path sentences, see Martin & McElree, 2018).

Another potential difference between filled-gap and non-filled-gap sentences, especially in the ambiguous conditions, is that the timing of disambiguation may have affected the reanalysis difficulty. In the non-filled-gap construction used in the present study, readers need to commit to the misanalysis longer than the filled-gap construction ("Elisa noticed the truck which the policeman watched the car from earlier that morning." vs. "Elisa noticed the truck which the policeman watched very quietly from earlier that morning."). Several studies report that reanalysis cost increases as the ambiguous region is made longer (e.g., Ferreira & Henderson, 1991; Tabor & Hutchins, 2004). It is difficult to tease these issues apart in the current study, but comparing the roles of diagnosticity and length would be an interesting avenue of further research. Despite these differences between filled-gap and non-filled-gap constructions, we maintain that our results show that readers sometimes persist with misinterpretation caused by filler-gap dependencies in both constructions, a finding that is consistent with the predictions of good enough processing (Christianson et al., 2001; Slattery et al., 2013).

*L1 and L2 differences in lingering misinterpretation*

Our results showed some evidence that L1 and L2 participants were prone to misinterpretation differently. Experiment 2a and 3 showed that L2 participants answered questions following ambiguous non-filled-gap sentences less accurately than L1 participants. This is compatible with previous studies reporting that L2 speakers have more difficulty recovering from garden paths than L1 speakers (e.g., Gerth et al., 2017; Jacob & Felser, 2016; Pozzan & Trueswell, 2016). However, comprehension accuracy rates for

filled-gap sentences did not differ significantly between L1 and L2 participants. Additionally, in our online experiments, although we found no significant differences in the size of inconsistency effects in L1 and L2 participants for filled-gap constructions in Experiment 1b, L1 participants showed clearer inconsistency effects than L2 participants in Experiment 2b, which tested non-filled-gap sentences. These reduced inconsistency effects may suggest reanalysis was not as successful for L2 speakers in non-filled-gap sentences.

One potential account of this L1/L2 difference may be that L2 speakers are more prone to increased reanalysis cost than L1 speakers when reanalysis is difficult. As discussed above, we argued that non-filled-gap sentences are more difficult to reanalyse than filled-gap sentences due to either the long ambiguous phrase or insufficient amount of information (diagnosticity) provided for disambiguation. The increased reanalysis difficulty for non-filled-gaps may explain the L1/L2 differences observed in this construction. This interpretation is consistent with previous studies on garden-path sentences. Roberts and Felser (2011) examined L1 and L2 reanalysis using sentences like (18a/b).

(18a) The woman read the magazine had shocked the university staff.

(18b) While the band played the song pleased all the customers.

(18a) requires reanalysis at "had", as "the magazine" is initially interpreted as the direct object of the main clause verb ("read"). (18b) is thought to cause greater reanalysis cost than (18a) for several reasons. For example, (18b) involves more drastic deconstruction of syntactic structure than (18a) (e.g., Gorrell, 1995; Marcus, Hindle, & Fleck, 1983; Sturt & Crocker, 1996, 1997; Sturt, Pickering, & Crocker, 1999). Roberts and Felser reported that L2 participants answered comprehension questions less accurately than

L1 participants when following difficult-to-reanalyse sentences like (18b). However, when questions followed ambiguous sentences causing little reanalysis like (18a), L2 participants outperformed L1 participants. Thus, it is possible that L2 speakers suffer from increased reanalysis cost more severely than L1 speakers, while reanalysis processes are similar when cost is low.

In the Introduction, we considered two accounts of L1/L2 differences in reanalysis. One account was that L2 speakers may succeed in syntactic reanalysis, but have increased persistence of initial misinterpretations (Cunnings, 2017). Alternatively, the shallow structure hypothesis (Clahsen & Felser, 2006, 2017) might predict that L2 speakers do not complete syntactic reanalysis. For filled-gap sentences, although we did not find significant evidence of increased persistence in L2 participants, we believe our results are most compatible with the idea that lingering misinterpretation, rather than failed reanalysis, is the main cause of difficulty in both groups. For non-filled-gap sentences, our results may indicate that L2 participants were less likely to complete syntactic reanalysis than L1 participants, as suggestive of shallow L2 processing, although whether L1 speakers always computed the correct interpretation of non-filled-gap sentences is debateable. Given the dearth of studies examining how differing degrees of reanalysis difficulty may influence the success of reanalysis in L1 and L2 processing, further research on this issue will be a fruitful avenue of future research.

### 4.8 Conclusion

The present study examined whether initially assigned misinterpretations influence the processing and interpretation of filler-gap dependencies. The reported experiments showed that L1 and L2 speakers persist with initially assigned misinterpretations both offline and during online processing. Some L1/L2 differences were observed, which suggested that L2 speakers have more difficulty reanalysing non-filled-gap sentences than L1 speakers. To

conclude, expanding on previous research on garden-path sentences, the present study revealed lingering misinterpretation in filler-gap dependencies in L1 and L2 sentence comprehension. These results suggest that good-enough processing is a general property of language comprehension.

## *Chapter 5. General discussion and conclusion*

This thesis examined properties of L1 and L2 reanalysis in garden-path and filler-gap sentences. Previous studies have shown that both L1 and L2 speakers initially misanalyse temporarily ambiguous sentences (e.g., Frazier & Rayner, 1982; Juffs & Harrington, 1996) and persist with initial misinterpretations after reanalysis (e.g., Christianson et al., 2001; Jacob & Felser, 2016). While a growing number of studies examine reanalysis of garden-path sentences, little has been explored regarding reanalysis of filler-gap dependencies. I hypothesised that if lingering effects are a general property of reanalysis, such effects should be observed in filler-gap dependencies as well, where reanalysis is required as in garden-path sentences. With regard to L2 reanalysis, some studies show that L2 speakers persist with misinterpretations more strongly than L1 speakers (e.g., Pozzan & Trueswell, 2016). However, the mechanisms underlying the L1 and L2 differences were previously unexplored. Potentially, the increased L2 reanalysis difficulty may result from a failure to complete reanalysis. Alternatively, it may relate to the possibility that L2 speakers are more susceptible to interference from initial misinterpretations than L1 speakers, as discussed in Cunnings (2017).

Study 1 showed that L2 speakers have increased reanalysis difficulty, compared with L1 speakers and importantly revealed that L2 speakers complete reanalysis of the temporarily ambiguous noun phrase during online reading as L1 speakers do. Study 1 also showed that lingering misinterpretations in L2 are at least partially ascribed to a failure to discard initial misinterpretations.

Study 2, which used the structural priming paradigm, showed that misanalysis of the subordinate clause verb is not fully discarded after reanalysis, and this lingering misinterpretation affects subsequent sentence processing in both L1 and L2.

Study 3 showed that both L1 and L2 speakers persist with misinterpretations in filler-gap dependencies, with L2 speakers having particular reanalysis difficulty only in

non-filled-gap sentences. The online data also showed that reanalysis of non-filled-gap sentences may be less successfully completed for L2 than L1 speakers. Below, I discuss implications for the results of Study 1–3 more broadly.

### *5.1. Similarities and dissimilarities between L1 and L2 reanalysis*

As discussed above, I considered two different accounts of L1 and L2 reanalysis processes. One potential account of this is that unlike L1 speakers, L2 speakers cannot complete syntactic reanalysis, which might be compatible with the Shallow Structure Hypothesis (Clahsen & Felser, 2006). Alternatively, the particular L2 reanalysis difficulty may relate to difficulty discarding initial misinterpretations (Cunnings, 2017).

In garden-path sentences, the offline experiments of Study 1 and 2 suggested that L2 speakers may have particular reanalysis difficulty, compared with L1 speakers. However, Study 1 also showed that L2 speakers conduct syntactic reanalysis during online reading and that they have difficulty erasing the memory trace of initial interpretations. Study 2 further showed that there were no significant L1/L2 differences in terms of reanalysis of the subordinate clause verb during online reading. The findings of Study 1 and 2 are more consistent with the Interference model rather than the incomplete hypothesis, given these studies suggested that L2 reanalysis difficulty resulted from failures to discard initial misinterpretations rather than failures to complete syntactic reanalysis.

Regarding filler-gap sentences, Study 3 showed evidence of lingering misinterpretation in L1 and L2 speakers during online reading and offline comprehension. However, the offline experiments suggested that L2 speakers have particular reanalysis difficulty in non-filled-gap sentences but not in filled-gap sentences. The eye-tracking experiments of this study also showed a sign of less complete reanalysis for L2 speakers than L1 speakers in non-filled-gap sentences, while reanalysis seemed to be successful for

both L1 and L2 speakers in filled-gap sentences. The sign of less complete reanalysis for L2 speakers in non-filed-gap sentences is more consistent with the Shallow Structure Hypothesis (Clahsen & Felser, 2006), which presupposes reduced L2 parsing ability.

Regarding filler-gap sentences, as discussed in Study 3, reanalysis is potentially difficult in the non-filled-gap sentences used in this thesis but less so in filled-gap sentences as a result of differences of ambiguity length (Ferreira & Henderson, 1991a) and diagnosticity of disambiguating cues (Martin & McElree, 2018). The possibility that increased reanalysis difficulty affects L1/L2 reanalysis differently may also account for why only L2 speakers showed a sign of incomplete reanalysis only in non-filled-gap sentences but not in garden-path and filled-gap sentences during online reading. Specifically, it is possible that when reanalysis is extremely difficult, L2 speakers may not complete syntactic reanalysis, under the assumption that ambiguous non-filled-gap sentences tested in this thesis are more difficult to reanalyse than garden-path sentences. I assume this is the case, given that comprehension accuracy rates for ambiguous non-filled-gap sentences were numerically lower than those for garden-path sentences. Given no study systematically examines L2 reanalysis processes with different degrees of syntactic complexity, this would be a promising avenue of future research.

Other models of L2 processing assume that L2 comprehension can become nativelike (e.g. Hopp, 2006; McDonald, 2006). Such models would predict that with increased L2 proficiency, L2 processing becomes nativelike (e.g., Hopp, 2006). To explore this possibility, I reanalysed the L2 data of Study 1–3 by fitting (generalised) linear mixed-effect models with centred Oxford Placement Test scores as a continuous predictor. Across the offline experiments of Study 1 and 2 (garden path sentences), I found one main effect of proficiency as a result of increased comprehension accuracy rates for more proficient L2 participants out of three comparisons. In terms of offline comprehension of filler-gap sentences, out of five comparisons, there was one main effect of proficiency for filled-gap

sentences and two main effects of proficiency for non-filled-gap sentences due to positive correlations between accuracy rates and proficiency as in Study 1 and 2. There was no evidence of any significant ambiguity by proficiency interactions for comprehension accuracy across Study 1–3.

Regarding online data, there was no significant evidence in Study 1 that proficiency affected reanalysis processes as measured by gender mismatch effects during online reading (Experiment 1b), or in terms of the continuation manipulation to test lingering misinterpretation (Experiment 2b). In Study 2, there was a positive correlation between priming and proficiency such that more proficient L2 participants showed larger priming effects during online reading. Regarding the size of garden-path effects during online reading, across all disambiguating and spillover regions (eight in total), each with three measures, in Study 1 and 2, there were four marginal and significant positive correlations between ambiguity and proficiency, suggesting that garden-path effects were larger for more proficient L2 participants during online reading. However, there were also two negative correlations showing larger garden-path effects for less proficient L2 participants.

For lingering misinterpretation in Study 3, reading times were longer in inconsistent than consistent sentences with increased L2 proficiency at the critical continuation region in Experiment 2b. However, its spillover region showed the opposite reading time pattern; i.e., reading times were longer in consistent than inconsistent conditions for more proficiency L2 participants. There were no significant effects of L2 proficiency on the size of filled-gap effects.

The reading times showing larger garden-path effects for more proficient L2 participants are consistent with the L2 processing models of Hopp (2006) and McDonald (2006), given that some previous studies have shown that the size of the garden-path effect is larger for L1 than L2 speakers (e.g., Jacob & Felser, 2016). However, this needs to be taken with caution, considering that some eye-tracking measures showed larger garden-

path effects for less proficient L2 participants as well. Also, the finding that reading times were longer for inconsistent sentences in more proficient L2 participants, observed at the critical region in Experiment 2b of Study 3, may suggest that complete reanalysis depends on L2 proficiency. However, its spillover region showed the totally opposite reading time pattern. Thus, some of the findings are compatible with the idea that L2 processing becomes nativelike as L2 proficiency improves but it does not fully account for the results observed in Study 1–3. Future research may need to test a wider range of L2 proficiency to further investigate any potential interaction between reanalysis and L2 proficiency.

### *5.2. Persistence of misinterpretation*

Previous studies have shown that readers often persist with initial misinterpretations (e.g., Christianson et al., 2001). While several ideas have been proposed to account for persistence of misinterpretation, Slattery et al. (2013) recently provided evidence that L1 speakers complete reanalysis but fail to discard initial misinterpretations. Based on their results, Slattery et al. refined the Good Enough Language Processing account, which presupposes that readers often hesitate to create fully specified representations to keep processing cost minimal. The online experiments of Study 1 corroborated the main findings of Slattery et al.: L1 participants showed gender mismatch effects in both ambiguous and unambiguous conditions. Study 2 further showed that the initially assigned transitive misinterpretation is not fully discarded after reanalysis in L1. Study 3 also suggested that initial misinterpretations linger in filler-gap dependencies in L1 as well. These findings are largely consistent with Slattery et al. and thus can be accounted for by the Good Enough Language Processing model.

One issue that the Good Enough Language Processing account may not fully explain is that L1 participants in Study 1 sometimes answered the ambiguous main clause questions wrongly in the offline experiments (e.g., “Who laughed very happily? 1. The mother 2. The baby” for “After the mother dressed the baby in the living room laughed

very happily."). As Slattery et al. and Experiment 1b of Study 1 suggested, if the temporarily ambiguous noun phrase is successfully reanalysed as the main clause subject, the subordinate clause subject should not be picked when the question asks about the actor of the main clause predicate. Christianson et al. (2001) tested interpretation of the main clause in a similar way to Study 1. Although they found slightly higher error rates for ambiguous than unambiguous sentences in main clause questions, they concluded given high comprehension accuracy that readers successfully reanalyse the temporarily ambiguous noun phrase as the main clause subject. Study 1 found larger error rates for ambiguous main clause questions even in L1 participants however.

In the General Discussion of Study 1 I considered that these numerical differences may result from differences in how the questions were presented across studies. However, another possibility is that difficulty in main clause questions results from persistence of the initially assigned direct object analysis (e.g., "the mother dressed the baby"). If this lingering direct object analysis influences subsequent language comprehension, readers may pick the subordinate clause subject ("the mother") as the correct answer, given that the subordinate clause subject may have a stronger agent representation in memory than the temporarily ambiguous noun phrase, if the memory trace of the temporarily ambiguous noun phrase as the direct object lingers. Further research is required to test this hypothesis.

Regarding L2 reanalysis, Study 1–3 also generally showed lingering misinterpretations in L2 comprehension. However, L2 participants also showed an indication of reanalysis failures only in non-filled-gap sentences. Further, as observed in L1 participants, L2 participants sometimes failed to answer main clause questions correctly in Study 1. Thus, although the core principle of the Good Enough Language Processing account appears to apply to both L1 and L2 processing, future studies will need to further test the extent to which lingering misinterpretation and incomplete reanalysis account for reanalysis difficulty in L1 and L2 processing.

### *5.3. Limitations and Future Directions*

While this thesis shed light on L1 and L2 reanalysis, there are some limitations. First, Study 1 revealed that L2 speakers reanalyse garden-path sentences successfully but fail to discard the initially assigned misinterpretation. However, Study 1 did not test if reanalysis of garden-path sentences is syntactically complete during online reading even when reanalysis difficulty is increased. That is, unlike the offline experiments, where ambiguity length was made long in order to increase the chance of observing lingering misinterpretation (Christianson et al., 2001), Experiment 1b kept ambiguity length to be minimal to make it consistent with Slattery et al., (2013). As discussed previously, ambiguity length potentially increases reanalysis difficulty, and importantly this thesis showed that completeness of L2 reanalysis may depend on reanalysis difficulty. Thus, future research may need to test how increased reanalysis difficulty affects L1 and L2 reanalysis processes during online reading.

Additionally, the first online experiment in Study 1, that used gender mismatch effects with reflexives, showed successful reanalysis during online reading but the offline experiments indicated the possibility that reanalysis of the temporarily ambiguous noun phrase is not always successful in both L1 and L2. As discussed in the General Discussion of Study 1, one difference between the offline tasks and this online experiment is the reflexive. It might be that reflexive resolution facilitated reanalysis, given that at the reflexive, participants needed to search for its antecedent. This may explain why only the offline experiments showed that reanalysis of the temporarily ambiguous noun phrase is sometimes incomplete, given that the materials used in these experiments did not require reflexive resolution. Thus, future research may need to test if reflexive resolution helps to reanalyse the temporarily ambiguous noun phrase, using sentences like (1).

(1a) After Tom visited the man decided to prepare himself a cold drink.

(1b) After Tom visited the man decided to prepare a cold drink.

Question: Who prepared a cold drink? “Tom” or “the man”

(1a/b) are both ambiguous sentences. However only (1a) contains a reflexive with the temporarily ambiguous noun phrase being its antecedent. The following comprehension question probes the actor of the main clause predicate. If reflexive resolution facilitates reanalysis, comprehension accuracy should be higher in (1a) than (1b).

Another possibility is to test reanalysis of the temporarily ambiguous noun phrase, using subject-verb agreement dependencies like (2).

(2a) When Tom visited the lady looked tired and was going to consult the doctor.

(2b) When Tom visited the ladies looked tired and was going to consult the doctor.

(2a/b) is similar to the experimental materials used in Experiment 1b of Study 1 except that (2a/b) contain subject-verb agreement dependencies rather than reflexive resolution. (2a/b) manipulate the temporarily ambiguous noun phrase such that it is either singular (“the lady”) or plural (“the ladies”). As the main clause contains a verb form (“was”) that requires a singular main clause subject, (2a) is grammatical but (2b) is ungrammatical. Thus, reading times should be longer at “was” in (2b) than (2a) if the temporarily ambiguous noun phrase is reanalysed as the main clause subject. However, if this noun phrase is misinterpreted as the subordinate clause direct object after reanalysis as a result of incomplete reanalysis, similar reading times may be observed between (2a/b). Given some studies argue that anaphor resolution is rigorously guided by syntactic cues during online reading (e.g., Dillon, Mishler, Sloggett & Phillips, 2013), which may

potentially help to conduct reanalysis as discussed above, using subject-verb agreement dependencies may provide further evidence of whether reanalysis is complete in L1 and L2 comprehension.

Study 2 showed that the initially assigned transitive misinterpretation is not fully discarded in L1 and L2 comprehension. However, why the transitive misinterpretation lingers is not entirely clear, especially in L2. Study 2 cannot clarify whether the observed priming effects result from the unerased memory trace of the misinterpretation or incomplete reanalysis of the subordinate clause verb. Traxler (2015) suggested that L1 speakers reanalyse the subordinate clause verb as intransitive at least to some degree, but the design used by Traxler has not been tested with L2 speakers. As such, applying the research design of Traxler (2015) to L2 speakers like (3) can further explore L2 reanalysis of the subordinate clause verb.

Prime

(3a) When Mary dressed the baby laughed happily.

(3b) When Mary dressed the baby the boy laughed happily.

Target

(3c) After Tom dressed the boy had some lunch in the living room.

As illustrated previously, Traxler showed, using this manipulation, that temporarily ambiguous sentences like (3a) reduced garden-path effects of ambiguous target sentences like (3c), relative to unambiguous prime sentences like (3b), as a result of (3a) facilitating the globally correct intransitive interpretation. If L2 speakers fail to reanalyse the subordinate clause verb as intransitive more frequently than L1 speakers, they may show weaker or null facilitatory priming effects, compared with L1 speakers. Thus, testing

sentences like (3) in the structural priming paradigm may help to examine which representation, the initially assigned misinterpretation or the globally correct interpretation, lingers more strongly in L2 processing.

Also, this thesis showed that initial misinterpretations linger in filler-gap dependencies. However, I questioned whether readers always complete syntactic reanalysis of filler-gap dependencies. This can be tested, using non-filled-gap sentences like (4), which manipulates plausibility between the filler and the first/second gaps following verbs.

(4a) It was the book which the boy found in the library was trying to read very quickly about last weekend.

(4b) It was the watch which the boy found in the library was trying to read very quickly about last weekend.

(4c) It was the book which the boy who was found in the library was trying to read very quickly about last weekend.

(4d) It was the watch which the boy who was found in the library was trying to read very quickly about last weekend.

(4a/b) contain a temporarily available gap at "found" due to the reduced relative clause. On the other hand, there is no such a gap in (4c/d), as the relative clause is marked overtly by "who". (4) also contain a temporarily available gap at "read" and manipulates plausibility between this verb and the filler such that in (4a/c), the gap is a plausible direct object ("the book") for "read" while in (4b/d), it is an implausible theme ("the watch"). The globally correct gap position appears at "about" to minimise the overall sentence implausibility. In (4a/b), readers are expected to incrementally associate the filler ("the book/watch") with the first gap position ("found") according to the active gap filling strategy. However, this analysis turns out to be incorrect at "was" following "library",

which should inform readers that “found” is in the past participle form rather than in the past form. If readers conduct reanalyse at “was”, attempting to detach the filler from the temporarily available first gap and seek out the globally correct gap position, reading times should be longer at “read” in (4b/d) than (4a/c) due to implausibility effects, as “read” is the second earliest gap position for (4a/b) and “the watch” is not a semantically appropriate direct object for “read”. However, if readers are satisfied with the temporarily completed gap filling at “found” and hesitate to seek out the globally correct gap, there should be no such implausibility effects, or at least reduced implausibility effects are expected in (4b), compared with (4d). If only L2 speakers fail to reanalyse non-filled-gap sentences, such reduced implausibility effects may be observed only for L2 speakers. Testing sentences like (4) may further reveal the cause of lingering misinterpretations in filler-gap dependencies in L1 and L2 speakers.

Finally, this thesis tested subject-object ambiguities and filler-gap dependencies that involve reanalysis of a verb. However, I did not consider the possibility that how probabilistic information of verbs may affect reanalysis. This is partly because while creating the experimental sentences, I tried to adopt the verbs that are likely to cause reanalysis (e.g., more transitive verbs for subject-object ambiguities). However, as pointed out by the examiners, it is possible that during reanalysis, L2 speakers make use of verb transitivity information as a disambiguating cue less efficiently than L1 speakers. Examining this issue may reveal why L2 speakers have increased reanalysis difficulty, compared with L1 speakers.

### *5.4 Conclusion*

The similarities and differences between L1 and L2 processing have been widely debated. Within the field of L2 sentence processing, syntactic ambiguity resolution has been widely examined. Recent studies on syntactic reanalysis show that L1 and L2 speakers often

persist with initial misinterpretations with L2 speakers having particular reanalysis difficulty. This thesis aimed to examine the cause of the increased reanalysis difficulty for L2 speakers and additionally examined reanalysis of filler-gap dependencies in L1 and L2 comprehension. Three studies reported in this thesis provided evidence that L1 and L2 speakers persist with initial misinterpretations in both garden-path and filler-gap sentences. As in previous studies, L2 participants persisted with misinterpretations more strongly than L1 participants in garden-path sentences and some filler-gap dependencies. Study 1–3 showed that a failure to erase the memory trace of initially assigned misinterpretations relates to reanalysis difficulty. Although I have argued that lingering misinterpretation, rather than incomplete reanalysis, causes difficulty in both garden-path sentences and filler-gap dependencies, there was some evidence that L2ers in particular may sometimes have difficulty completing reanalysis, particularly in certain types of filler-gap dependencies which do not contain a filled-gap. Nonetheless, Study 1 and 3 showed that L2 speakers complete reanalysis of garden-path sentences and filler-gap dependencies containing filled-gaps similarly to L1 speakers. Taken together, the results reported in this thesis provide evidence that persistence of misinterpretation is a general property of reanalysis in both L1 and L2 language comprehension.

***References***

Altmann, G. T. M., Garnham, A., & Dennis, Y. (1992). Avoiding the garden path: Eye movements in context. *Journal of Memory and Language*, *31*(5), 685–712. https://doi.org/10.1016/0749-596X(92)90035-V

Altmann, G. T. M., Garnham, A., & Henstra, J.-A. (1994). Effects of syntax in human sentence parsing: Evidence against a structure-based proposal mechanism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(1), 209–216. http://dx.doi.org/10.1037/0278-7393.20.1.209

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*(3), 191–238. https://doi.org/10.1016/0010-0277(88)90020-0

Altmann, G. T. M., van Nice, K. Y., Garnham, A., & Henstra, J.-A. (1998). Late closure in context. *Journal of Memory and Language*, *38*(4), 459–484. https://doi.org/10.1006/jmla.1997.2562

Aoshima, S., Phillips, C., & Weinberg, A. (2004). Processing filler-gap dependencies in a headfinal language. *Journal of Memory and Language*, *51(1)*, 23–54. https://doi.org/10.1016/j.jml.2004.03.001

Arai, M., & Nakamura, C. (2016). It's Harder to Break a Relationship When you Commit Long. *PLoS ONE*, *11*(6), e0156482. https://doi.org/10.1371/journal.pone.0156482

Arai, M., van Gompel, R. P. G., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology*, *54*(3), 218–250. https://doi.org/10.1016/j.cogpsych.2006.07.001

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossedrandom effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459. https://doi.org/10.3758/BF03193014

Barr, D., Levy, R., Scheepers, C., & Tily, H. (2013). Random-effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. http://dx.doi.org/10.18637/jss.v067.i01

Bever, T. G. (1970). The cognitive basis for linguistic structures. In: J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–352). New York, NY: Wiley.

Blodgett, A. (2004). *The interaction of prosodic phrasing, verb bias, and plausibility during spoken sentence comprehension*. Unpublished doctoral thesis, Ohio State University.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387. https://doi.org/10.1016/0010-0285(86)90004-6

Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, *129*(2), 177–192. http://dx.doi.org/10.1037/0096-3445.129.2.177

Branigan, H., & Pickering, M. (2017). An experimental approach to linguistic representation. Behavioral and Brain Sciences, *40*, E282. https://doi.org/10.1017/S0140525X16002028

Branigan, H. P., Pickering, M. J., & McLean, J. F. (2005). Priming prepositional-phrase attachment during comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 468–481. http://dx.doi.org/10.1037/0278-7393.31.3.468

Chacón, D. A., Imtiaz, M., Dasgupta, S., Murshed, S. M., Dan, M., & Phillips, C. (2016). Locality and word order in active dependency formation in Bangla. *Frontiers in Psychology*, *7*. http://dx.doi.org/10.3389/fpsyg.2016.01235.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272. http://dx.doi.org/10.1037/0033-295X.113.2.234

Chomsky, N. (1981). Lectures on government and binding. Dordrecht: Foris.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic Roles Assigned along the Garden Path Linger. *Cognitive Psychology*, *42*(4), 368–407. https://doi.org/10.1006/cogp.2001.0752

Christianson, K., Luke, S. G., & Ferreira, F. (2010). Effects of plausibility on structural priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 538–544. http://dx.doi.org/10.1037/a0018027

Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006). Younger and older adults' good enough interpretations of garden path sentences. *Discourse Processes*, *42*(2), 205–238. https://doi.org/10.1207/s15326950dp4202_6

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27(1)*, 3–42. https://doi.org/10.1017/S0142716406060024

Clahsen, H., & Felser, C. (2017). Some Notes on The Shallow Structure Hypothesis. *Studies in Second Language Acquisition*, 1–14. https://doi.org/10.1017/S0272263117000250

Cleland, A. A., & Pickering, M. J. (2003). The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language*, *49*(2), 214–230. https://doi.org/10.1016/S0749-596X(03)00060-3

Cleland, A. A., & Pickering, M. J. (2006). Do writing and speaking employ the same syntactic representations? *Journal of Memory and Language*, *54*(2), 185–198.

https://doi.org/10.1016/j.jml.2005.10.003

Clifton, C., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, *49*(3), 317–334. https://doi.org/10.1016/S0749-596X(03)00070-6

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological syntax processor. In: D. R. Dowty, L. Karttunen, & A. M. Zwicky (Eds), *Natural language parsing: Psychological, computational and theoretical perspectives* (pp. 320–358). Cambridge, England: Cambridge University Press.

Cunnings, I. (2017). Parsing and Working Memory in Bilingual Sentence Processing. *Bilingualism: Language and Cognition*, *20*(4), 659–678. https://doi.org/10.1017/S1366728916000675

Cunnings, I., Fotiadou, G., & Tsimpli, I. (2017). Anaphora Resolution and Reanalysis During L2 Sentence Processing: Evidence from the Visual World Paradigm. *Studies in Second Language Acquisition*, *39*(4), 621–652. https://doi.org/10.1017/S0272263116000292

Cunnings, I., & Sturt, P. (2018) Coargumenthood and the processing of pronouns, *Language, Cognition and Neuroscience*, *33*, 1235–1251. https://doi.org/10.1080/23273798.2018.1465188

Dallas, A., DeDe, G., & Nicol, J. (2013). An event-related potential (ERP) investigation of filler-gap processing in native and second language speakers. *Language Learning*, *63*(4), 766–799. https://doi.org/10.1111/lang.12026

Dallas, A., & Kaan, E. (2008). Second language processing of filler-gap dependencies by late learners. *Language and Linguistics Compass*, *2*(3), 372–388. https://doi.org/10.1111/j.1749-818X.2008.00056.x

Dekydtspotter, L., Donaldson, B., Edmonds, A., Fultz, A., & Petrush, R. (2008). Syntactic and prosodic computations in the resolution of relative clause attachment ambiguity by English-French earners. *Studies in Second Language Acquisition*, *30*(4), 453–480. https://doi.org/10.1017/S0272263108080728

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. http://dx.doi.org/10.1016/j.jml.2013.04.003.

Drummond, A. (2013). IBEX Farm. Available online, http://spellout.net/ibexfarm/

Dussias, P. E., & Cramer Scaltz, T. R. (2008). Spanish–English L2 speakers' use of subcategorization bias information in the resolution of temporary ambiguity during second language reading. *Acta Psychologica*, *128*(3), 501–513. https://doi.org/10.1016/j.actpsy.2007.09.004

Dussias, P.E., Marful, A., Gerfen, C., & Bajo Molina, M.T. (2010). Usage frequencies of complement-taking verbs in Spanish and English: Data from Spanish monolinguals and Spanish-English bilinguals. *Behavior Research Methods*, *42*(4), 1004–1011. https://doi.org/10.3758/BRM.42.4.1004

Dussias, P., & Piñar, P. (2012). Effects of reading span and plausibility in the reanalysis of wh-gaps by Chinese-English second language speakers. *Second Language Research*, *26*(4), 443–472. https://doi.org/10.1177/0267658310373326

Felser, C., Cunnings, I., Batterham, C. and Clahsen, H. (2012). The timing of island effects in nonnative sentence processing. *Studies in Second Language Acquisition*, *34*(1), 67–98. https://doi.org/10.1017/S0272263111000507

Felser, C. & Cunnings, I. (2012). Processing reflexives in English as a second language: The role of structural and discourse-level constraints. *Applied Psycholinguistics*, *33*(3), 571–603. https://doi.org/10.1017/S0142716411000488

Felser, C., & Roberts, L. (2007). Processing wh-dependencies in a second language: A cross-modal priming study. *Second Language Research*, *23*(1), 9–36. https://doi.org/10.1177/0267658307071600

Felser, C., Roberts, L., Marinis, T. & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, *24*(3), 453–489. https://doi.org/10.1017/S0142716403000237

Felser, C., Sato, M., & Bertenshaw, N. (2009). The on-line application of Binding Principle A in English as a second language. *Bilingualism: Language and Cognition*, *12*(4), 485–502. https://doi.org/10.1017/S1366728909990228

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. https://doi.org/10.1111/1467-8721.00158

Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: Implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research, 30*(1), 3–20. https://doi.org/10.1023/A:1005290706460

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*(3), 75–87. https://doi.org/10.1016/0749-596X(86)90006-9

Ferreira, F., & Henderson, J. M. (1991a). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, *30*(6), 725–745. https://doi.org/10.1016/0749-596X(91)90034-H

Ferreira, F., & Henderson, J. M. (1991b). How is verb information used during syntactic parsing? In G. B. Simpson (Ed.), *Understanding word and sentence* (pp. 305–330). Amsterdam: North-Holland.

Ferreira, F., & Henderson, J. M. (1998). Syntactic reanalysis, thematic processing, and sentence comprehension. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 73–100). Dordrecht: Kluwer Academic Publishers.

Ferreira, F., Lau, E. F., & Bailey, K. G. D. (2004). Disfluencies, language comprehension, and tree adjoining grammars. *Cognitive Science*, *28*(5), 721–749. https://doi.org/10.1207/s15516709cog2805_5

Ferreira, F., & McClure, K. K. (1997). Parsing of garden-path sentences with reciprocal verbs. *Language and Cognitive Processes*, *12*(2), 273–306. https://doi.org/10.1080/016909697386862

Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, *1*(1–2), 71–83. https://doi.org/10.1111/j.1749-818X.2007.00007.x

Fodor, J. D., Bever, T. G. & Garrett, M. F. (1974). *The psychology of language*. London, England: McGraw-Hill.

Fodor, J. D., & Inoue, A. (1998). Attach anyway. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 101–141). Dordrecht: Kluwer Academic Publishers.

Fodor, J. D., & Frazier, L. (1980). Is the human sentence parsing mechanism an ATN? *Cognition*, *8*(4), 417–459. https://doi.org/10.1016/0010-0277(80)90003-7

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

Frazier, L. (1987). Syntactic Processing: Evidence from Dutch. *Natural Language & Linguistic Theory*, *5*(4), 519–559. Retrieved from

http://www.jstor.org/stable/4047505

Frazier, L. (1990). Parsing modifiers: Special purpose routines in the human sentence processing mechanism? In: D. A. Balota, G. B. Flores d Arcais, & K. Rayner (Eds), *Comprehension processes in reading* (pp. 303–330). Hillsdale, NJ: Erlbaum.

Frazier, L., & Clifton, C., Jr. (1998) Sentence reanalysis, and visibility. In J. D. Fodor & F. Ferreira (eds.), *Reanalysis in sentence processing* (pp 143-176). Amsterdam: Kluwer.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291–325. http://doi.org/10.1016/0010- 0277(78)90002-1

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178–210. https://doi.org/10.1016/0010-0285(82)90008-1

Frenck-Mestre, C., & Pynte, J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *50A*(1), 119–148. http://dx.doi.org/10.1080/027249897392251

Fujita, H., & Cunnings, I. (under review). Reanalysis processes in non-native sentence comprehension. Ms, University of Reading.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, *37*(1), 58–93. https://doi.org/10.1006/jmla.1997.2512

Garnsey, S. M., Tanenhaus, M. K., and Chapman, R. M. (1989). Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research*, *18*(1), 51–60. https://doi.org/10.1007/BF01069046

Gerth, S., Otto, C., Nam, Y., & Felser, C. (2017). Strength of garden-path effects in native œand non-native speakers' processing of subject-object ambiguities. *International Journal of Bilingualism 21*(2), 125–144. https://doi.org/10.1177/1367006915604401

Gibson, E. (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown* (Unpublished doctoral thesis). School of Computer Science, Carnegie Mellon University, PA.

Gorrell, P. (1995). *Syntax and parsing*. Cambridge, England: Cambridge University

Gorrell P. (1998) Syntactic Analysis and Reanalysis in Sentence Processing. In: Fodor J.D., Ferreira F. (eds) *Reanalysis in Sentence Processing. Studies in Theoretical Psycholinguistics*, vol 21. Springer, Dordrecht

Grodner, D., Gibson, E., Argaman, V., & Babyonyshev, M. (2003). Against Repair-Based Reanalysis in Sentence Comprehension. *Journal of Psycholinguistic Research*, *32*(2), 141–166. https://doi.org/10.1023/A:1022496223965

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, *15*(6), 883–899. https://doi.org/10.1162/089892903322370807

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In A. Kehler, L. Levin, & D. Marcu (Eds.), *Proceedings of NAACL 2001* (pp. 1–8). Morristown, NJ: Association for Computational Linguistics.

Hare, M., McRae, K., & Elman, J. L. (2003). Sense and structure: Meaning as a determinant of verb subcategorization preferences. *Journal of Memory and Language*, *48*(2), 281–303. https://doi.org/10.1016/S0749-596X(02)00516-8

Havik, E., Roberts, L., van Hout, R., Schreuder, R., & Haverkort, M. (2009). Processing subject–object ambiguities in the L2: A self-paced reading study with German L2 learners of Dutch. *Language Learning*, *59*(1), 73–112.

https://doi.org/10.1111/j.1467-9922.2009.00501.x

Hopp, H. (2015). Individual differences in the second language processing of object–subject ambiguities. *Applied Psycholinguistics*, *36*(2), 129–173. https://doi.org/10.1017/S0142716413000180

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research, 22*(3), 369–397. https://doi.org/10.1191/0267658306sr272oa

Hopp, H. (2014). Working memory effects on the L2 processing of ambiguous relative clauses. *Language Acquisition*, *21*(3), 250–278. https://doi.org/10.1080/10489223.2014.892943

Hwang, H. & Schafer, A.J. (2006). Prosodic effects in parsing early vs. late closure sentences by second language learners and native speakers. In Hoffmann, R. & Mixdorff (Eds.), *Proceedings of Speech Prosody 2006* (pp. 585–588). Dresden, Germany.

Inoue, A., & Fodor, J. D. (1995) Information-paced parsing of Japanese. In: R. Mazuka and N. Nagai (Eds.), *Japanese sentence processing*, (pp. 9–63). Hillside, NJ: Lawrence Erlbaum Associates.

Jacob, G., & Felser, C. (2016). Reanalysis and semantic persistence in native and non-native garden-path recovery. *The Quarterly Journal of Experimental Psychology*, *69*(5), 907–925. https://doi.org/10.1080/17470218.2014.984231

Jegerski, J. (2012). The processing of temporary subject–object ambiguities in native and near-native Mexican Spanish. *Bilingualism: Language and Cognition, 15*(4), 721–735. https://doi.org/10.1017/S1366728911000654

Jessen, A., & Felser, C. (2018). Reanalysing object gaps during non-native sentence processing: Evidence from ERPs. *Second Language Research*. https://doi.org/10.1177/0267658317753030

Jessen, A., Festman, J., Boxell, O., & Felser, C. (2017). Native and non-native speakers' brain responses to filled indirect object gaps. *Journal of Psycholinguistic Research*, *46*(5), 1319–1338. https://doi.org/10.1007/s1093

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, *25*(4) 603–634. https://doi.org/10.1017/S0142716404001298

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning*, *57*(1), 1-33. https://doi.org/10.1111/j.1467-9922.2007.00397.x

Johnson, A., Fiorentino, R., & Gabriele, A. (2016). Syntactic constraints and individual differences in native and non-native processing of Wh-movement. *Frontiers in Psychology*, *7*. http://dx.doi.org/10.3389/fpsyg.2016.00549.

Juffs, A., & Harrington, M. (1996). Garden path sentences and error data in second language processing research. *Language Learning*, *46*(2), 283–323. https://doi.org/10.1111/j.1467-1770.1996.tb01237.x

Juffs, A., & Harrington, M. (1995). Parsing Effects in Second Language Sentence Processing: Subject and Object Asymmetries in wh-Extraction. *Studies in Second Language Acquisition*, *17*(4), 483–516. https://doi.org/10.1017/S027226310001442X

Juffs, A. (2004). Representation, processing, and working memory in a second language. *Transactions of the Philological Society, 102*(2), 199–225. https://doi.org/10.1111/j.0079-1636.2004.00135.x

Juffs, A. (2005). The influence of first language on the processing of wh-movement in English as a second language. *Second Language Research*, *21*(2), 121–51. https://doi.org/10.1191/0267658305sr255oa

Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *The Quarterly Journal of Experimental Psychology*, *69*(5), 1013–1040. http://dx.doi.org/10.1080/17470218.2015.1053951

Kaschak, M. P., & Glenberg, A. M. (2004). This construction needs learned. *Journal of Experimental Psychology: General*, *133*(3), 450–467. http://dx.doi.org/10.1037/0096-3445.133.3.450

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*(2), 205–225. https://doi.org/10.1016/j.jml.2004.10.002

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, *2*(1), 15–47. https://doi.org/10.1016/0010-0277(72)90028-5

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, *40*(2), 153–194. https://doi.org/10.1006/jmla.1998.2620

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lassotta, R., Omaki, A., & Franck, J. (2016). Developmental changes in the misinterpretation of garden-path wh-questions in French. *Quarterly Journal of Experimental Psychology*, *69*(5), 829–854. https://doi.org/10.1080/17470218.2015.1054845

Lau, E. F., & Ferreira, F. (2005). Lingering effects of disfluent material on comprehension of garden path sentences. *Language and Cognitive Processes*, *20*(5), 633–666. https://doi.org/10.1080/01690960444000142

Lee, E., Lu, D., & Garnsey, S. (2013). L1 word order and sensitivity to verb bias in L2 processing. *Bilingualism: Language and Cognition*, *16*(4), 761–775. https://doi.org/10.1017/S1366728912000776

Ledoux, K., Traxler, M. J., & Swaab, T. Y. (2007). Syntactic priming in comprehension: evidence from event-related potentials. *Psychological science*, *18*(2), 135–143. https://doi.org/10.1111/j.1467-9280.2007.01863.x

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Lewis, R., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375–419. https://doi.org/10.1207/s15516709cog0000_25

Lewis, R., Vasishth, S., & Van, Dyke, J. (2006). Computation principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454. https://doi.org/10.1016/j.tics.2006.08.007

Lim, J. H., & Christianson, K. (2013a). Second language sentence processing in reading for comprehension and translation. *Bilingualism: Language and Cognition*, *16*(3), 518–537. https://doi.org/10.1017/S1366728912000351

Lim, J. H., & Christianson, K. (2013b). Integrating meaning and structure in L1–L2 and L2–L1 translations. *Second Language Research*, *29*(3), 233–256. https://doi.org/10.1177/0267658312462019

Malyutina, S., & den Ouden, D. B. (2016). What is it that lingers? Garden-path (mis)interpretations in younger and older adults. *The Quarterly Journal of Experimental Psychology*, *69*(5), 880–906. https://doi.org/10.1080/17470218.2015.1045530

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676–703.

http://dx.doi.org/10.1037/0033-295X.101.4.676

Marcus, M., Hindle, D., & Fleck, M. (1983). D-theory: Talking about talking about trees. *In Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics* (pp. 129–136). Cambridge, MA.

Marinis, T., Roberts, L, Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, *27*(1), 53–78. https://doi.org/10.1017/S0272263105050035

Martin, A. E., & McElree, B. (2018). Retrieval cues and syntactic ambiguity resolution: speed-accuracy tradeoff evidence, *Language, Cognition and Neuroscience*, *33*(6), 769–783. https://doi.org/10.1080/23273798.2018.1427877

McDonald, J. (2006). Beyond the critical period: Processing- based explanations for poor grammaticality judgement performance by late second language learners. *Journal of Memory and Language*, *55*(3), 381–401. https://doi.org/10.1016/j.jml.2006.06.006

McDonough, K. (2006). INTERACTION AND SYNTACTIC PRIMING: English L2 Speakers' Production of Dative Constructions. *Studies in Second Language Acquisition*, *28*(2), 179–207. http://dx.doi.org/10.1017/S0272263106060098

McDonough, K., & Chaikitmongkol, W. (2010). Collaborative syntactic priming activities and EFL learners' production of wh-questions. *Canadian Modern Language Review*, *66*(6), 817–841. http://dx.doi.org/10.3138/cmlr.66.6.817

McDonough, K., & Kim, Y. (2009). Syntactic priming, type frequency, and EFL learners' production of wh-questions. *Modern Language Journal*, *93*(3), 386 –398. https://doi.org/10.1111/j.1540-4781.2009.00897.x

McElree, B., & Griffith, T. (1995). Syntactic and thematic processing in sentence comprehension: Evidence for a temporal dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 134–157. http://dx.doi.org/10.1037/0278-7393.21.1.134

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, *38*(3), 283–312. https://doi.org/10.1006/jmla.1997.2543

Mitchell, D. C., Corley, M. M. B., & Garnham, A. (1992). Effects of context in human sentence parsing: Evidence against a discourse-based proposal mechanism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 69–88. http://dx.doi.org/10.1037/0278-7393.18.1.69

Murray, W. S., & Liversedge, S. P. (1994). Referential context effects on syntactic processing. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 359–388). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Nakamura, C., & Arai, M. (2016). Persistence of initial misanalysis with no referential ambiguity. *Cognitive Science*, *40*(4), 909– 940. https://doi.org/10.1111/cogs.12266

Nakano, Y., Felser, C., and Clahsen, H. (2002). Antecedent priming at trace positions in Japanese long-distance scrambling. *Journal of Psycholinguistic Research*, *31*(5), 531–571. https://doi.org/10.1023/A:1021260920232

Nickels, S., Opitz, B., & Steinhauer, K. (2013). ERPs show that classroom-instructed late second language learners rely on the same prosodic cues in syntactic parsing as native speakers. *Neuroscience Letters*, *557*, 107–11. https://doi.org/10.1016/j.neulet.2013.10.019.

Nicol, J., and Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, *18*(1), 5–19. https://doi.org/10.1007/BF01069043

Nitschke, S., Serratrice, L., & Kidd, E. (2014). The effect of linguistic nativeness in structural priming in comprehension. *Language, Cognition and Neuroscience*, *29*(5), 525-542. https://doi.org/10.1080/01690965.2013.766355

Nitschke, S., Kidd, E., & Serratrice, L. (2010). First language transfer and long-term structural priming in comprehension. *Language and Cognitive Processes*, *25*(1), 94–114. https://doi.org/10.1080/01690960902872793

Omaki, A., Lau, E. F., Davidson White, I., Dakan, M. L., Apple, A., and Phillips, C. (2015). Hyper-active gap filling. *Frontiers in Psychology*, *6*(384). https://doi.org/10.3389/fpsyg.2015.00384

Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island constraints in second language sentence processing. *Studies in Second Language Acquisition*, *33*(4), 563–588. https://doi.org/10.1017/S0272263111000313

Omaki, A., White, I. D., Goro, T., Lidz, J., & Phillips, C. (2014). No fear of commitment: Children's incremental interpretation in English and Japanese Wh-Questions. *Language Learning and Development*, *10*(3), 206–233. https://doi.org/10.1080/15475441.2013.844048

Paape, D., Nicenboim, B., & Vasishth, S. (2017) Does antecedent complexity affect ellipsis processing? An empirical investigation. *Glossa*, *2*(1), 77. http://doi.org/10.5334/gjgl.290

Pan, H., & Felser, C. (2011). Referential context effects in L2 ambiguity resolution: Evidence from self-paced reading. *Lingua*, *121*(2), 221–236. https://doi.org/10.1016/j.lingua.2010.08.003

Pan, H., Schimke, S. & Felser, C. (2015). Referential context effects in non-native relative clause ambiguity resolution. *International Journal of Bilingualism*, *19*(3), 298–313. https://doi.org/10.1177/1367006913515769

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition*, *25*(4), 501–528. https://doi.org/10.1017/S0272263103000214

Parker, D. (2017). Processing multiple gap dependencies: Forewarned is forearmed. *Journal of Memory and Language*, *97*, 175–186. http://dx.doi.org/10.1016/j.jml.2017. 08.003

Patson, N. D., Darowski, E. S., Moon, N., & Ferreira, F. (2009). Lingering misinterpretations in garden-path sentences: Evidence from a paraphrasing task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 280–285. http://dx.doi.org/10.1037/a0014276

Patson, N. D., & Ferreira, F. (2009). Conceptual plural information is used to guide early parsing decisions: Evidence from garden-path sentences with reciprocal verbs. *Journal of Memory and Language*, *60*(4), 464–486. https://doi.org/10.1016/j.jml.2009.02.003

Phillips, C. (1995). Right association in parsing and grammar. Schütze, Carson; Ganger, Jennifer; Broihier, Keven (Ed.): *Papers on Language Acquisition and Processing* (pp. 37–93). Department of Linguistics and Philosophy, Massachusetts Institute of Technology.

Phillips, C., & Gibson, E. (1997). On the strength of the local attachment preference. *Journal of Psycholinguistic Research*, *26*(3), 323–346. https://doi.org/10.1023/A:1025028725042

Piaget, J. (1952). The origins of intelligence. New York: International University Press.

Piaget, J. (1977). The development of thought: Equilibration of cognitive structures. New York: Viking.

Piaget, J. (1985). The equilibrium of cognitive structures. Chicago: University of Chicago Press.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, *39*(4), 633–651. https://doi.org/10.1006/jmla.1998.2592

Pickering, M. J., & Traxler, M. J. (2003). Evidence against the use of subcategorisation frequencies in the processing of unbounded dependencies. *Language and Cognitive Processes*, *18*(4), 469–503. https://doi.org/10.1080/01690960344000017

Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(4), 940–961. http://dx.doi.org/10.1037/0278-7393.24.4.940

Pickering, M. J., & Traxler, M. J. (2001). Strategies for processing unbounded dependencies: Lexical information and verb–argument assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1401–1410. http://dx.doi.org/10.1037/0278-7393.27.6.1401

Pliatsikas, C., & Marinis, T. (2013). Processing of regular and irregular past tense morphology in highly proficient second language learners of English: a self-paced reading study. *Applied Psycholinguistics*, *34*(5), 943–970. https://doi.org/10.1017/S0142716412000082

Pozzan, L., & Trueswell, J. (2016). Second language processing and revision of garden-path sentences: A visual world study. *Bilingualism: Language and Cognition*, *19*(3), 636–643. https://doi.org/10.1017/S1366728915000838

Pritchett, B. (1988). Garden Path Phenomena and the Grammatical Basis of Language Processing. *Language*, *64*(3), 539-576. Retrieved from

https://www.jstor.org/stable/414532?seq=1#page_scan_tab_contents

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

Qian, Z., Garnsey, S., & Christianson, K. (2018) A comparison of online and offline measures of good-enough processing in garden-path sentences. *Language, Cognition and Neuroscience*, *33*(2), 227–254. https://doi.org/10.1080/23273798.2017.1379606

Quick placement test: Version 1. (2004). Oxford: Oxford Univ. Pr.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Radford, A. (2009). Wh-movement. In *Analysing English Sentences: A Minimalist Approach* (Cambridge Textbooks in Linguistics, pp.183–237). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511801617.006

Radford, A., Felser, C., & Boxell, O. (2012). Preposition copying and pruning in present-day English. *English Language and Linguistics*, *16*(3), 403-426. https://doi.org/10.1017/S1360674312000172

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 358–374. https://doi.org/10.1016/S0022-5371(83)90236-0

Rayner, K., Garrod, S., & Perfetti, C. A. (1992). Discourse influences during parsing are delayed. *Cognition*, *45*(2), 109–139. https://doi.org/10.1016/0010-0277(92)90026-E

Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics*, *32*(2), 299–331.

https://doi.org/10.1017/S0142716410000421

Rohde, D. (2010). *Linger*. Available online, https*://tedlab.mit.edu/~dr/Linger/*

Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics*, *32*(2), 299–331. https://doi.org/10.1017/S0142716410000421

Ross, J. R. (1967). *Constraints on variables in syntax*. Doctoral dissertation, Massachusetts Institute of Technology.

Schepman, A., & Rodway, P. (2000). Prosody and parsing in coordination structures. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *53A*(2), 377–396. http://dx.doi.org/10.1080/027249800390538

Schneider, D., & Phillips, C. (2001). Grammatical search and reanalysis. *Journal of Memory and Language*, *45*(2), 308–336. https://doi.org/10.1006/jmla.2001.2777

Shin, J., & Christianson, K. (2012). Structural priming and second language learning. *Language Learning*, *62*(3), 931–964. https://doi.org/10.1111/j.1467-9922.2011.00657.x

Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, *69*(2), 104–120. http://dx.doi.org/10.1016/j.jml.2013.04.001

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238–299. https://doi.org/10.1016/j.cogpsych.2004.03.001

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, *48*(1), 103–130. https://doi.org/10.1016/S0749-596X(02)00519-3

Sorace, A. (2011). Pinning down the concept of 'interface' in bilingualism. *Linguistic Approaches to Bilingualism*, *1*(1), 1–33. https://doi.org/10.1075/lab.1.1.01sor

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1521–1543. http://dx.doi.org/10.1037/0278-7393.24.6.1521

Staub, A. (2007a). The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 550–569. http://dx.doi.org/10.1037/0278-7393.33.3.550

Staub, A. (2007b). The return of the repressed: Abandoned parses facilitate syntactic reanalysis. *Journal of Memory and Language, 57*(2), 299–323. https://doi.org/10.1016/j.jml.2006.09.001

Stevenson, S., & Merlo, P. (1997). Lexical structure and parsing complexity. *Language & Cognitive Processes*, *12*(2-3), 349–399. https://doi.org/10.1080/016909697386880

Stowe, L. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, *1*(3), 227–245. https://doi.org/10.1080/01690968608407062

Sturt, P., & Crocker, M. (1995). Monotonic parsing and reanalysis. Paper presented at *8th Annual CUNY Conference on Human Sentence Processing*, Tucson, AZ.

Sturt, P., & Crocker, M. W. (1996). Monotonic syntactic processing: A cross-linguistic study of attachment and reanalysis. *Language and Cognitive Processes*, *11*(5), 449 – 494. https://doi.org/10.1080/016909696387123

Sturt, P., & Crocker, M. W. (1997). Thematic monotonicity. *Journal of Psycholinguistic Research*, *26*(3), 297–322. https://doi.org/10.1023/A:1025076608204

Sturt, P., & Crocker, M. W. (1998). Generalized monotonicity for reanalysis models. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing*. Dordrecht, The Netherlands: Kluwer.

Sturt, P., Pickering, M. J., & Crocker, M. W. (1999). Structural change and reanalysis difficulty. *Journal of Memory and Language*, *40*(1), 136–150. https://doi.org/10.1006/jmla.1998.2606

Sturt, P., Pickering, M. J., Scheepers, C., & Crocker, M. W. (2001). The preservation of structure in language comprehension: Is reanalysis the last resort? *Journal of Memory & Language*, *45*(2), 283–307. https://doi.org/10.1006/jmla.2001.2776

Sturt, P. (2007). Semantic re-interpretation and garden-path recovery. *Cognition*, *105*(2), 477–488. https://doi.org/10.1016/j.cognition.2006.10.009

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, *48*(3), 542–562. https://doi.org/10.1016/S0749-596X(02)00536-3

Sussman, R. S., & Sedivy, J. C. (2003). The time-course of processing syntactic dependencies: Evidence from eye movements. *Language and Cognitive Processes*, *18*(2), 143–163. https://doi.org/10.1080/01690960143000498

Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*, *36*(1), 201–216. https://doi.org/10.3758/MC.36.1.201

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, *50*(4), 355–370. https://doi.org/10.1016/j.jml.2004.01.001

Tabor, W., & Hutchins, S. (2004). Evidence for self-organized sentence processing: Digging-in effects. *Journal of experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 431–450. http://dx.doi.org/10.1037/0278-7393.30.2.431

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634. http://science.sciencemag.org/content/268/5217/1632#BIBL

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. https://doi.org/10.1016/0749-596X(88)90011-3

Tooley, K, M., & Bock, K. (2014). On the parity of structural persistence in language production and comprehension. *Cognition*, *132*(2), 101–36. https://doi.org/10.1016/j.cognition.2014.04.002

Tooley K, M., Swaab, T, Y., Boudewyn, M, A., Zirnstein, M., & Traxler, M, J. (2014). Evidence for priming across intervening sentences during on-line sentence comprehension. *Language, Cognition and Neuroscience*, *29*(3), 289–311. https://doi.org/10.1080/01690965.2013.770892

Tooley, K. M., Traxler, M. J., & Swaab, T. Y. (2009). Electrophysiological and behavioural evidence of syntactic priming in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, Cognition*, *35*(1), 19–45 http://dx.doi.org/10.1037/a0013984

Traxler, M. J. (2008). Lexically independent priming in online sentence comprehension. *Psychonomic Bulletin and Review*, *15*(1), 149–155. https://doi.org/10.3758/PBR.15.1.149

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, *35*(3), 454–475. https://doi.org/10.1006/jmla.1996.0025

Traxler, M. J., & Tooley, K. M. (2008). Priming in sentence comprehension: Strategic or syntactic? *Language and Cognitive Processes*, *23*(5), 609–645. https://doi.org/10.1080/01690960701639898

Traxler, M. J., Tooley, K. M., & Pickering, M. J. (2014). Syntactic priming during sentence comprehension: Evidence for the lexical boost. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 905–918. http://dx.doi.org/10.1037/a0036377

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. https://doi.org/10.1006/jmla.1994.1014

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 528–553. http://dx.doi.org/10.1037/0278-7393.19.3.528

Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, *35*(4), 566–585. https://doi.org/10.1006/jmla.1996.0030

Tyler, L. K., & Marslen-Wilson, W. D. (1977). The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(6), 683–692. https://doi.org/10.1016/S0022-5371(77)80027-3

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: the declarative / procedural model. *Bilingualism: Language and Cognition*, *4*(2), 105–122. http://doi.org/10.1017/S1366728901000220

Van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, *41*(2), 147–182. https://doi.org/10.1006/jmla.1999.2641

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 407–430. http://dx.doi.org/10.1037/0278-7393.33.2.407

van Gompel, R. P., Pickering, M. J., Pearson, J., & Jacob, G. (2006). The activation of inappropriate analyses in garden-path sentences: Evidence from structural priming. *Journal of Memory and Language*, *55*(3), 335–362. https://doi.org/10.1016/j.jml.2006.06.004

van Gompel, R. P. G., Pickering, M. J., & Traxler, M. J. (2000). Unrestricted race: A new model of syntactic ambiguity resolution. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 621–648). Amsterdam, Netherlands: North Holland/Elsevier Science Publishers. http://dx.doi.org/10.1016/B978-008043642-5/50029-2

von der Malsburg, T., & Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of Memory and Language*, *94*, 119–133. https://doi.org/10.1016/j.jml.2016.10.003

Wagers, M. W., and Phillips, C. (2009). Multiple dependencies and the role of the grammar in real-time comprehension. *Journal of Linguistics*, *45*(2), 395-433. https://doi.org/10.1017/S0022226709005726

Wagers, M. W., and Phillips, C. (2014). Going the Distance: Memory and Control Processes in Active Dependency Construction. *Quarterly Journal of Experimental Psychology*, *67*(7), 1274–1304. https://doi.org/10.1080/17470218.2013.858363

Watson, D., & Gibson, E. (2005). Intonational phrasing and constituency in language production and comprehension. *Studia Linguistica*, *59*(2–3), 279–300. https://doi.org/10.1111/j.1467-9582.2005.00130.x

Weber, K., & Indefrey, P. (2009). Syntactic priming in German–English bilinguals during sentence comprehension. *NeuroImage*, *46*(4), 1164–1172. https://doi.org/10.1016/j.neuroimage.2009.03.040

Wei, H., Boland, J., & Brennan, J. (2018). Lexicalized structural priming in second language online sentence comprehension. *Second Language Research*, *34*(3), 395–416. https://doi.org/10.1177/0267658317723684

Wei, H., Boland, J. E., Cai, Z. G., Yuan, F., & Wang, M. (2019). Persistent structural priming during online second-language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(2), 349–359. http://dx.doi.org/10.1037/xlm0000584

Weinberg, A. (1993). Parameters in the theory of sentence processing: Minimal commitment theory goes East. *Journal of Psycholinguistic Research*, *22*(3), 339–364. https://doi.org/10.1007/BF01068016

White, L.A., Goad, H., Goodhue, D., Hwang, H., & Lieberman, M.W. (2013). Syntactic Ambiguity Resolution in L2 Parsing: Effects of Prosodic Boundaries and Constituent Length. *In Proceedings of the 28th West Coast Conference on Formal Linguistics, ed*. Mary Byram Washburn et al., 267–274. Somerville, MA: Cascadilla Proceedings Project.

Williams, J. (2006). Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition*, *9*(1), 71–88. https://doi.org/10.1017/S1366728905002385

Williams, J., Möbius, P., & Kim, C. (2001). Native and non-native processing of English wh-questions: Parsing strategies and plausibility constraints. *Applied Psycholinguistics*, *22*(4), 509–540. https://doi.org/10.1017/S0142716401004027

*Appendices for Study 1*

**Appendix A**

1 After the lawyer calmed down(,) the man in the office took some time to rest.

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

3 After the two journalists met(,) the president in the office smoked some cigarettes.

4 When the two children hugged(,) their friend at the station smiled very happily.

5 After the father washed(,) the boy in the bathroom went to the garden.

6 When the two kids fought(,) their neighbour in the kitchen started to get very irritated.

7 After the man woke up(,) the lady in the bedroom started watching the news.

8 After Mary dressed(,) the boy in the house started cleaning the room.

9 While the pianist calmed(,) down the singer in the hall prepared for the concert.

10 While the two soldiers fought(,) the General on the battlefield received some new orders.

11 After the lady woke up(,) her friend in the flat started cooking breakfast.

12 While the nurse washed(,) the baby at the hospital seemed to be very sleepy.

13 After the girl calmed down(,) the boy in the classroom started studying for the exam.

14 After the doctor woke up(,) the nurse in the hospital started working very hard.

15 After the aunt dressed(,) the boy in the bedroom looked at the clock.

16 After the researcher woke up(,) the assistant in the lab decided to go back home.

17 While the two men fought(,) the terrorist in the street tried to escape.

18 After the father washed(,) the baby at home went to sleep.

19 After the coach calmed down(,) the basketball player in the gym started thinking about the game.

20 When the two friends met(,) the man in the mall ate a hamburger.

21 After Anna dressed the child(,) in the bedroom started watching television.

22 After the two footballers fought(,) the coach at the school cleaned up the gym.

23 After the mother washed(,) the child at home started to get dressed.

24 When the two actresses met(,) the director in the studio went for lunch.

**Appendix B**

1 Some people had a party at a friend's house at the weekend. After the neighbour visited(,) Ken's dad/mum decided to prepare himself a cold drink. It was very tasty.

2 Many people from the local school attended the sports day. While the student watched Tom's brother/sister started to prepare himself for the big race. It ended up being a great success.

3 Dinner was finished and it was time to finish some chores. While the parent helped Roger's son/daughter hoped to allow himself some time to relax. It had been a long day.

4 Everybody was getting ready for the highs school graduation ceremony. Before the student called Ben's son/daughter started to dress himself in the locker room. Many people were graduating that day.

5 The local school held an art contest a couple of weeks ago. While the student sketched Julia's sister/brother tried to enjoy herself during the contest. The winner got a great prize.

6 A huge party was taking place close to the university. After the guest visited Sarah's sister/brother started to enjoy herself at the party. It lasted until midnight.

7 There was a big argument between a family and a friend. After the friend called Mary's mother/father attempted to calm herself down very quickly. They apologised to each other the next day.

8 There was a big problem that needed to be dealt with quickly. After the manager called Lisa's mother/father started to focus herself on the problem. It was quickly solved after that.

9 A family and their gym instructor were getting exhausted due to exercise. After the instructor trained David's nephew/niece decided to allow himself time to rest. Everybody agreed to change the exercise plan in the end.

10 The drawing had already taken a long time and everyone was tired. While the painter sketched Kevin's father/mother hoped to give himself time to relax. Hopefully it wouldn't take much longer.

11 There was going to have to be a meeting about the bad behaviour. After the teacher called Paul's father/mother started to prepare himself for the meeting. The school was busy that day.

12 Everybody was told to be very quiet during the ceremony. As the child watched George's uncle/aunt tried to entertain himself while keeping quiet. Some famous people were at the event.

13 The rich family always travelled with an assistant. While the assistant helped Sarah's mum/dad started to prepare herself for the dinner. They had booked an expensive restaurant.

14 There was a big party at a house in the neighbourhood. After the guest visited Jessica's aunt/uncle decided to prepare herself a large cocktail. It tasted really nice.

15 There was an art event at a local school on the weekend. While the artist sketched Emily's daughter/brother started to enjoy herself at the event. It was a very successful afternoon.

16 There was a large dinner in town for some local celebrities. After the manager called Alice's daughter/brother started to prepare herself for the ceremony. It was going to be great.

17 The research assistant started an experiment with a volunteer. While the assistant watched Mark's nephew/niece tried to entertain himself for a few minutes. The experiment was quite boring in the end.

18 It was early morning and everyone at the hotel was sleeping. When the staff woke up John's brother/sister decided to prepare himself a strong coffee. It helped the day get off to a good start.

19 The children needed some help to get ready for the holiday. While the child asked Robert's dad/mum started to prepare himself for the trip. There was a lot to get ready.

20 Some friends were having a lunch party at the weekend. After the neighbour visited Bob's uncle/aunt started to introduce himself to the guests. A variety of food was served at the party.

21 Some friends were having some fun drawing pictures. While the child sketched Rose's mum/dad decided to prepare herself a small snack. They all went to the playground afterwards.

22 The parents helped the kids clean after they all got very muddy. As the adult washed Elena's niece/nephew wanted to warm herself with a hot drink. After taking a bath, it was time for bed.

23 It was a really busy day at the large company. Before the staff called Diana's aunt/uncle started to prepare herself for some work. Everyone got really tired that day.

24 Many of the pupils competed in the annual sports day. After the child watched Anna's niece/nephew began to prepare herself for a race. Everyone was exhausted by the end of the day.

**Appendix C**

1 After the lady woke up(,) the child in the room ate a sandwich.

2 After the woman dressed(,) the boy in the living room read a book.

3 After the bear scratched(,) the lion in the forest climbed the tree.

4 After the parents hugged(,) the child in the park sat on the bench.

5 After the lady washed(,) the boy in the house cleaned the plate.

6 After the parents kissed(,) the child in the kitchen ate some chocolate.

7 After the mouse scratched(,) the researcher in the university drank some water.

8 After Emily dressed(,) the boy in the house watched the television.

9 After the lady woke up(,) her husband in the apartment drank some coffee.

10 After the parents kissed(,) their son in the dining room ate some bread.

11 When the two students hugged(,) the teacher at the graduation ceremony cried very happily.

12 After the nurse washed(,) the child in the hospital sat on the chair.

13 After the two brothers hugged(,) their mother at the station ate some cookies.

14 After the nurse woke up(,) the patient in the hospital knocked on the door.

15 After the grandmother dressed(,) the child in the bedroom looked at the clock.

16 After the two girls hugged(,) their father at the station drank some lemonade.

17 After the cat scratched(,) the girl in the garden sat on the blanket.

18 After the grandmother washed(,) the boy at home listened to some music.

19 After the lady woke up(,) her boyfriend in the flat drank a cup of tea.

20 After the parents kissed(,) their child in the bedroom ate some biscuits.

21 After the old lady dressed(,) the boy in the bedroom checked the time.

22 After the dog scratched(,) the boy in the house drank some milk.

23 After the mother washed(,) the child in the house watched the television.

24 After the parents kissed(,) their daughter in the hospital sat down on the sofa.

**Appendix D**

1 When the husband woke up(,) his wife in the bedroom looked at the time. It seemed that the husband was waking himself/his wife up to prepare for a trip. It was going to be fun.

2 When the mother dressed(,) her son at home called the dog. It was clear that the mother was dressing herself/her son formally for an important ceremony. It was tiring.

3 When the lawyer calmed down(,) the lady in the office looked very tired. It seemed that the lawyer was calming himself/the lady down after having a long talk. It was very serious.

4 When the boy hid(,) the toy in the room fell to the floor. It seemed that the boy was hiding himself/the toy while playing a fun game. It was all very silly.

5 When Jonathan washed(,) the dog in the garden looked very sleepy. It was clear that Jonathan was washing himself/the dog very quickly with some soap. It was quite tiring.

6 When the two men fought(,) the criminal in the city looked very nervous. It seemed that the two men were fighting each other/the criminal for a very long time. It was extremely noisy.

7 When the king calmed down(,) the queen in the castle looked out the window. It appeared that the king was calming himself/the queen down after hearing an explosion. It was rather scary.

8 While Rebecca dressed(,) the boy in the room wanted to have breakfast. It appeared that Rebecca was dressing herself/the boy after cleaning the kitchen. It was a busy day.

9 When Mary woke up(,) her grandson in the bedroom looked very hungry. It seemed that Mary was waking herself/her grandson up on the bed very quickly. It was only 6:00am.

10 While the two waiters fought(,) the guest in the bar became very annoyed. It looked like the two waiters were fighting each other/the guest loudly for no reason. It was very annoying.

11 When the captain hid(,) the weapon in the ship dropped from the shelf. It seemed that the captain was hiding himself/the weapon in a small back room. It had been a tough day.

12 While Lisa washed(,) the patient at the care home looked a little sleepy. It seemed that Lisa was washing herself/the patient before preparing for dinner. It was a very busy day.

13 When the manager calmed down(,) the singer in the studio looked a bit nervous. It appeared that the manager was calming himself/the singer down quietly behind the stage. It had been a hard day.

14 While the maid woke up(,) the man at the hotel wanted some tea. It seemed that the maid was waking herself/the man up much earlier than expected. It was only 6:00am.

15 While the babysitter dressed(,) the baby in the room looked quite happy. It seemed that the babysitter was dressing herself/the baby after cooking the lunch. There was still a lot to do.

16 When the student hid(,) the textbook in the classroom fell from the desk. It seemed that the student was hiding himself/the textbook after the class finished. It was lots of fun.

17 When some citizens fought(,) the politician in the street became very irritated. It seemed that the citizens were fighting each other/the politician angrily for no apparent reason. The police intervened.

18 While the farmer washed(,) the horse in the backyard seemed to be sleepy. It looked like the farmer was washing himself/the horse with some very cold water. It was a sunny day.

19 When the flight attendant woke up(,) the pilot in the airport prepared for the flight. It was clear that the flight attendant was waking herself/the pilot up early in the morning. It would be a long flight.

20 While the director calmed down(,) the newscaster in the studio tried to relax. It appeared that the director was calming himself/the newscaster down while checking the news. There was lots to do.

21 When the nurse dressed(,) the boy in the hospital room wanted to sleep. It appeared that the nurse was dressing herself/the boy very quickly and casually. There was work to be done.

22 When the two customers fought(,) the assistant in the shop looked rather upset. It appeared that the two customers were fighting each other/the assistant in front of the till. It was very loud.

23 While Jessica washed(,) the cat in the back garden yawned very sleepily. It looked like Jessica was washing herself/the cat after working very hard. It had been a long day.

24 When the parents hugged(,) the baby in the room smiled very happily. It seemed that the parents were hugging each other after/the baby watching a movie. It was very good.

*Appendices for Study 2*

## Appendix A

1

After the doctor woke up/called(,) the nurse worked for hours in the hospital.

When Richard woke up(,) his wife looked very sleepy and tired.

2

When Mary calmed down/watched(,), the child sat quietly on the sofa.

Although the man calmed down(,) his wife looked quite nervous that day.

3

While James washed/called(,) his child waited very quietly in the bathroom.

After the lady washed(,) the dog started eating some food quickly.

4

Before the small cat scratched/watched(,) the boy played very happily in the park.

While the dog scratched(,) the lady started to prepare for dinner.

5

When the children hugged/watched(,) the baby laughed very happily by the bed.

After the tourists hugged(,) the tour guide decided to relax on the bench.

6

After the grandparents kissed/helped(,) the small child watched the television in the lounge.

When the babies kissed(,) their mother smiled very happily on the chair.

7

Although the boxer fought/visited(,) the coach didn't give any advice at all.

While the men fought(,) the criminal tried to escape down the street.

8

As Tom dressed/helped(,) his son began cooking lunch in the kitchen.

When the parents dressed(,) their baby smiled very happily in the bed.

9

When Eva woke up/asked(,) her husband started to cook breakfast.

After the secretary woke up(,) the politician prepared some documents for the meeting.

10

While Brian washed/cleaned(,) the small towel fell to the floor and got dirty.

After the woman washed(,) her daughter decided to arrange a plan for the tri

11

While the coach calmed down/helped(,) the baseball players prepared for the big game.

When Lily calmed down(,) her boyfriend watched the television very contently.

12

While the black cat scratched/watched(,) Thomas hid under the blanket.

When the mouse scratched(,) the researcher checked the monitor very carefully.

13

When the girls hugged/met(,) the large dog yawned very sleepily in the garden.

After the students hugged(,) their teacher smiled quite happily at school.

14

When the children kissed/helped(,) their cousin had some lunch at the table.

After the neighbours kissed(,) their baby started crying aloud for milk.

15

When the soldiers fought/stopped(,) the enemy tried to retreat from the battlefield.

As the people fought(,) the politician seemed quite irritated in the street.

16

Before the grandmother dressed/telephoned(,) her grandchild drank some water in the room.

After the parents dressed(,) their children began to prepare for dinner.

17

As Helen woke up/asked(,) her old friend cleaned the room very quickly.

After the researcher woke up(,) the assistant started working on the project.

18

When the farmer washed/approached(,) the horse jumped suddenly at the farm.

After Mark washed(,) his little son waited very quietly in the bathroom.

19

Before the owner calmed down/approached(,) the large dog behaved very badly in the garden.

Although Linda calmed down(,) the young actor looked quite nervous until the show ended.

20

While the lady dressed/called(,) the little girl waited very quietly in the room.

After the parents dressed(,) their children enjoyed the television show quietly.

21

Before the waiters fought/asked(,) the customers checked the menu in the restaurant.

When the two boys fought(,) their mother started to become very annoyed.

22

When the brothers met/called(,) their friends ran together slowly around the park.

While the politicians met(,) the president considered the important project quietly.

23

After the dog hid/left(,) the food washed away down the sink.

When Jessica hid(,) the small key dropped suddenly to the floor.

24

After the trainers hugged/watched(,) the runner prepared for the important race.

While the hosts hugged(,) the guest drank some wine in the kitchen.

25

When the workmen met/called(,) the employer looked exhausted after a busy day.

After the soldiers met(,) the king rested very quietly in the room.

26

When Steven hid/approached(,) the painting fell down in the hallway.

While the captain hid(,) the weapon dropped accidentally from the shelf.

27

When the customers met/visited(,) the shop assistant tried to sell some new products.

After the artists met(,) the director worked very hard in the studio.

28

While Anna hid/read(,) some books fell to the floor in the living room.

After the student hid(,) the pencil dropped quietly from the table.

29

After the relatives kissed/hugged(,) the young boy became very embarrassed quite quickly.

When the parents kissed(,) the little girl looked really happy in the room.

30

While the maid woke up/hurried(,) the woman looked very annoyed in the hotel.

After Judy woke up(,) the pilot started preparing for the long flight.

31

While the engineer washed/watched(,) the bicycle stopped in front of the window.

After Julia washed(,) the little girl played with friends in the garden.

32

Before the manager calmed down/visited(,) the singer got worried about the concert.

After the king calmed down(,) the queen ordered the guards to be alert.

33

While the fans hugged/approached(,) the basketball player smiled very happily on the court.

When the little girls hugged, the big teddy bear dropped down from the bed.

34

As Angela dressed/helped, the child stayed very quiet in the room.

After Michael dressed(,) his little brother began watching the television show.

35

As the two friends fought/approached(,) the guest started to get annoyed.

After the children fought(,) the neighbour cleaned the street very quietly.

36

When the pupils met/called(,) their teacher had lunch quietly in the cafeteria.

After the professors met(,) the young researcher analysed some data very quietly.

*Appendices for Study 3*

**Appendix A**

1 John saw the car (near) which the officer stopped the bicycle (near) earlier today.

2 The man saw the wall (beside) which the artist painted the truck (beside) last night.

3 The captain sailed the ship (from) which the tourist watched the island (from) last weekend.

4 The parent checked the window (near) which the child broke the door (near) last night.

5 Ann noticed the taxi which the man drove the car (near) during the journey.

6 Helen liked the movie (in) which the actor watched the race (in) very happily.

7 The tourist visited the tower (from) which the artist sketched the sky (from) during the vacation.

8 Mike saw the bridge (near) which the workers built the church (near) last week.

9 The cook stood by the chair (near) which the waiters carried the table (near) that morning.

10 The man saw the tool (with) which the mechanic repaired the car (with) a couple of days ago.

11 Julia entered the room (in) which the maid cleaned the table (in) a little while ago.

12 The child looked at the cake (beside) which the lady baked the bread (beside) late at night.

13 The girl waited for the book (about) which the author wrote very passionately (about) last summer.

14 The boy bought the novel (about) which the girl read very happily (about) last night.

15 The host wanted the beer (near) which the guest drank very quickly (near) during the party.

16 The commander saw the helicopter (from) which the soldier attacked very swiftly (from) during the war.

17 The musician brought the guitar (near) which the boy played very excitingly (near) that evening.

18 The teacher saw the desk (around) which the student moved very quickly (around) during the lunch break.

19 The boy saw the plates (near) which the lady washed very quickly (near) that morning.

20 The girl saw the bus (behind) which the man drove very fast (behind) this morning.

21 Michael liked the story (about) which the journalist wrote very briefly (about) last summer.

22 The neighbour saw the bag (near) which the boy hid very quietly (near) during lunch.

23 The sister looked at the piano (near) which the girl played very happily (near) last night.

24 The child looked through the window (near) which the mother washed very patiently (near) at the weekend.

**Appendix B**

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

2 It was a nice, sunny day. John wanted to use the camera (near) which the man was carrying the chair (near) quite hastily. It was clear that the man was carrying the chair/camera into the living room.

3 It was dark outside. The reporter photographed the captain (for) who(m) the soldier was shooting the enemy (for) very carefully. It appeared that the soldier was shooting the enemy/captain from a great distance.

4 It was very early morning. Jonathan walked past the house (near) which the workmen were building the hotel (near) surprisingly quickly. It appeared that the workmen were building the hotel/house as quickly as possible.

5 There was some free time at work. The editor looked for the book (about) which the journalist was writing the report (about) that afternoon. It was likely that the journalist was writing the report/book quickly but very carefully.

6 It was busy in the park. The child sat by the tree (near) which the girl was sketching the rose (near) very happily. It was clear that the girl was sketching the rose/tree because it was beautiful.

7 It was a very relaxing weekend. Thomas saw the piano (near) which the girl was playing the game (near) quite quietly. It was clear that the girl was playing the game/piano happily on her own.

8 It had been a very busy day. Matthew noticed the child (about) who(m) the lady was calling the teacher (about) rather unhappily. It seemed that the lady was calling the teacher/child about trouble at school.

9 It was very hot that day. The lady saw the hammer (with) which the carpenter was fixing the table (with) very loudly. It seemed that the carpenter was fixing the table/hammer very quickly and easily.

10 There were some problems that needed fixing. Carol held the ladder (with) which the workman was repairing the window (with) very carefully. It seemed that the workman was repairing the window/ladder slowly and very calmly.

11 It was a very rainy day. Kevin noticed the novel (near) which the girl was reading the letter (near) that morning. It appeared that the girl was reading the letter/novel before going to school.

12 It was very busy in the hospital. Jill saw the parent (for) who(m) the doctor was examining the child (for) very carefully. It seemed that the doctor was examining the child/parent after an unfortunate accident.

13 It was a very cold day. The lady visited the office (behind) which the decorator was painting the tower (behind) quite carefully. It was likely that the decorator was painting the tower/office to make it more attractive.

14 The art class was a lot of fun. The teacher looked at the flower (beside) which the student was drawing the fruit (beside) really carefully. It was clear that the student was drawing the fruit/flower very happily in class.

15 There was a day off from work. Caroline noticed the car (near) which the man was cleaning the bike (near) quite carelessly. It was obvious that the man was cleaning the bike/car very happily in the garage.

16 The jewellery shop in town was popular. Laura saw the watch (near) which the lady was purchasing the ring (near) very happily. It was clear that the lady was purchasing the ring/watch for a close friend.

17 The town was very busy that day. The student noticed the bus (behind) which the man was driving the taxi (behind) earlier that day. It was clear that the man was driving the taxi/bus into the town centre.

18 It was a quiet Sunday afternoon. The child saw the map (beside) which the cleaner was wiping the picture (beside) that day. It seemed that the cleaner was wiping the picture/map without paying much attention.

19 Everyone was at home that day. The lady sat at the desk (beside) which the man was cleaning the computer (beside) quite hastily. It seemed that the man was cleaning the computer/desk because it was dirty.

20 It was a clear and sunny day. Kelly saw the car (near) which the young man was washing the bike (near) very happily. It seemed that the young man was washing the bike/car for a family friend.

21 The weather was very good. Susan looked at the building (around) which the workers were constructing the fence (around) that morning. It appeared that the workers were constructing the fence/building very quickly and carefully.

22 The road wasn't very busy. The girl saw the truck (behind) which Jeff was driving the car (behind) very fast. It seemed that Jeff was driving the car/truck quickly towards the city.

23 It was a day off. The child looked at the chair (beside) which Monica was painting the toy (beside) earlier that day. It seemed that Monica was painting the toy/chair with lots of colours.

24 The road was very crowded. Frank noticed the bike (near) which the policeman was stopping the bus (near) very calmly. It was clear that the policeman was stopping the bus/bike to avoid an accident.

**Appendix C**

1 The teacher saw the apple (beside) which the student drew the banana (beside) earlier that day.

2 The girl looked at the truck (beside) which the neighbour washed the bike (beside) last night.

3 The father found the piano (beside) which the child played the guitar (beside) earlier today.

4 Luke saw the flower (near) which the girl drew the tree (near) very happily.

5 Anna looked at the table (near) which the man carried the chair (near) quite hastily.

6 Katie saw the window (beside) which the boy broke the toy (beside) during the weekend.

7 The nephew saw the door (beside) which the mother locked the suitcase (beside) before sleeping.

8 Alex looked at the cake (beside) which the grandmother baked the bread rolls (beside) late at night.

9 The pilot landed the airplane (near) which the tourist watched the balloon (near) late at night.

10 Jack found the bicycle (beside) which the mechanic repaired the car (beside) during the morning.

11 Sophia saw the cup (beside) which the boy dropped the book (beside) that day.

12 Monica saw the wall (beside) which the child touched the chair (beside) earlier that day.

13 The cameraman saw the helicopter (from) which the soldier attacked very swiftly (from) during the war

14 The teacher saw the chair (from) which the student moved very quickly (from) before class began.

15 The lady looked at the bench (from) which the workman moved very slowly (from) during lunch.

16 Kevin saw the letter (beside) which the candle burnt very quickly (beside) last night.

17 Jacob saw the desk (from) which the lady moved quite carefully (from) after writing an email.

18 The daughter saw the towel (near) which the father washed very quickly (near) that morning.

19 The parent noticed the paper (beside) which the fire burnt very intensely (beside) last weekend.

20 Ellie observed the dog (towards) which the neighbour walked really happily (towards) during the morning.

21 Elisa noticed the truck (from) which the policeman watched very quietly (from) earlier that morning.

22 The criminal noticed the car (from) which the guard watched very carefully (from) during the evening.

23 The reporter watched the ship (from) which the sniper shot very carefully (from) during the battle.

24 Rose saw the truck (from) which the man shot very quickly (from) during the robbery.

25 The taxi driver saw the newspaper (near) which the cigarette burnt very slowly (near) during lunch.

26 The king saw the castle (from) which the men attacked very bravely (from) during the battle.

27 The captain noticed the ship (from) which the enemy attacked very swiftly (from) during the war.

28 The lady saw the bus (from) which the spy watched completely silently (from) during the evening.

29 The commander saw the tank (from) which the soldier shot very quickly (from) during the weekend.

30 The lady saw the horse (near) which the owner washed very quickly (near) later that morning.

**Appendix D**

1 Ann found the red box (in) which the boy was hiding very quietly (in) during lunch. The boy was hiding (in) the red box behind the big sofa. It was lots of fun.

2 Mary saw the large bed (from) which the man was moving quite sleepily (from) before lunch. The man was moving (from) the large bed to start cleaning the room. It was very messy.

3 Amy saw the red airplane (from) which the soldier was attacking very quickly (from) during the war. The soldier was attacking (from) the red airplane very quickly and aggressively. It was very scary.

4 The lady watched the white horse (with) which Ben was hunting very quietly (with) earlier today. Ben was hunting (with) the white horse even though it was raining. It was really hard work.

5 The girl was in the school bus (near) which Alan was driving very slowly (near) this morning. Alan was driving (near) the school bus very patiently on the road. It was extremely crowded.

6 Judy saw the small key (near) which the boy was hiding quite quietly (near) during dinner. The boy was hiding (near) the small key behind the old closet. Everyone thought it was funny.

7 Anna watched the grey wolf (with) which the hunter was hunting quite carefully (with) that day. The hunter was hunting (with) the grey wolf near the small village. It was late at night.

8 Joe saw the small castle (from) which the army was attacking very quickly (from) during the battle. The army was attacking (from) the small castle very late into the evening. It had been a terrible year.

9 Beth saw the big truck (near) which the man was driving very safely (near) after dinner. The man was driving (near) the big truck on the very busy road. It was getting very late.

10 John saw the old piano (from) which the pianist was moving very slowly (from) during practice. The pianist was moving (from) the old piano while drinking some water. It was hot that day.

11 Tom saw the small table (from) which the boy was moving very quickly (from) after breakfast. The boy was moving (from) the small table to prepare for school. It would be a good day.

12 Eva saw the black car (from) which the terrorist was attacking very quickly (from) during the afternoon. The terrorist was attacking (from) the black car even though nobody was near. It was very scary.

13 The man saw the old car (near) which Lisa was driving very slowly (near) during the holiday. Lisa was driving (near) the old car towards the city centre. It was a good day.

14 Kim liked the new helicopter (from) which the man was watching very excitedly (from) before lunch. The man was watching (from) the new helicopter while eating some fruit. It was a nice day.

15 Lisa saw the small ship (near) which the captain was sailing very swiftly (near) during the trip. The captain was sailing (near) the small ship very carefully but quickly. It was really exciting.

16 Rose liked the old boat (from) which the tourist was watching very calmly (from) while on holiday. The tourist was watching (from) the old boat on the large boating lake. The view was very nice.

17 Maria saw the huge truck (from) which the guard was shooting quite quickly (from) during the weekend. The guard was shooting (from) the huge truck to protect the president. The gunshots rang out loudly.

18 Paul noticed the small ship (in) which the lady was washing really slowly (in) while on vacation. The lady was washing (in) the small ship for a very long time. It was very relaxing.

19 Emma saw the new brush (with) which the housemaid was cleaning very carefully (with) that day. The housemaid was cleaning (with) the new brush in the small bathroom. It was hard work.

20 Sarah saw the large truck (from) which the officer was shooting very carefully (from) late at night. The officer was shooting (from) the city bus with an automatic pistol. There had been a bad crime.

21 The mother saw the soft blanket (in) which the boy was hiding very secretly (in) earlier that day. The boy was hiding (in) the soft blanket before eating some food. It was cold that day.

22 Kevin saw the old towel (with) which Julia was cleaning very quickly (with) before bed. Julia was cleaning (with) the old towel even though it was late. There was still lots to do.

23 The man saw the white cloth (with) which Mary was cleaning very carefully (with) late at night. Mary was cleaning (with) the white cloth with some new soap. It had been a good day.

24 Alice saw the big dog (with) which the man was hunting very carefully (with) that day. The man was hunting (with) the big dog after taking some rest. It was dark outside.

**Appendix E**

1 The commander saw the helicopter (from) which the soldier attacked the tank/very swiftly (from) during the war.

2 The teacher saw the desk (around) which the student moved the chair/very quickly (around) during the lunch break.

3 The girl saw the bus (behind) which the man drove the car/quite fast (behind) this morning.

4 John saw the car (near) which the officer stopped the bicycle/very slowly (near) earlier today.

5 Mike saw the bridge (near) which the workers built the road/really patiently (near) last weekend.

6 Kevin saw the letter (beside) which the candle burnt the paper/quite quickly (beside) last night.

7 Elisa noticed the truck (from) which the policeman watched the car/very quietly (from) earlier that morning.

8 The reporter watched the ship (from) which the sniper shot the helicopter/very carefully (from) during the battle.

9 The lady watched the horse (with) which Ben hunted the deer/very quietly (with) earlier today.

10 The man saw the cloth (with) which Mary cleaned the window/really carefully (with) late at night.

11 Linda saw the towel (near) which the father washed the plate/very quickly (near) that morning.

12 Ann found the box (in) which the boy hid the wallet/very quietly (in) during lunch.

13 Joe saw the castle (from) which the army attacked the airplane/very quickly (from) during the battle.

14 The lady looked at the bench (from) which the workman moved the helmet/very slowly (from) during lunch.

15 Ann noticed the taxi (behind) which the man drove the car/quite carelessly (behind) during the journey.

16 Frank noticed the bike (behind) which the policeman stopped the taxi/very quickly (behind) late at night.

17 Susan looked at the house (around) which the workers built the fence/quickly early (around) that morning.

18 The child noticed the book (beside) which the fire burnt the newspaper/very intensely (beside) last weekend.

19 The lady saw the bus (from) which the spy watched the car/completely silently (from) during the evening.

20 The commander saw the tank (from) which the soldier shot the airplane/very quickly (from) during the weekend.

21 Anna watched the wolf (with) which the hunter hunted the rabbit/quite carefully (with) that day.

22 Emma saw the brush (with) which the housemaid cleaned the floor/very carefully (with) that day.

23 The girl looked at the truck (beside) which the neighbour washed the bike/really quickly (beside) last night.

24 The mother saw the blanket (under) which the boy hid the key/very secretly (under) earlier that day.