

Reanalysis and lingering misinterpretation of linguistic dependencies in native and non-native sentence comprehension

Article

Accepted Version

Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Fujita, H. and Cunnings, I. (2020) Reanalysis and lingering misinterpretation of linguistic dependencies in native and non-native sentence comprehension. *Journal of Memory and Language*, 115. 104154. ISSN 0749-596X doi: <https://doi.org/10.1016/j.jml.2020.104154> Available at <http://centaur.reading.ac.uk/92359/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1016/j.jml.2020.104154>

Publisher: Elsevier

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in

the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

**Reanalysis and lingering misinterpretation of linguistic dependencies in
native and non-native sentence comprehension**

Hiroki Fujita & Ian Cunnings

University of Reading

Address for correspondence:

Dr Hiroki Fujita

School of Psychology and Clinical Language Sciences

The University of Reading

Whiteknights Campus

Reading

RG6 7BE

Abstract

Research on temporarily ambiguous “garden path” sentences (e.g., After Mary dressed the baby laughed) has shown that initially assigned misinterpretations linger after reanalysis of the temporarily ambiguous phrase in both native (L1) and non-native (L2) readers. L2 speakers have particular difficulty with reanalysis, but the source of this L1/L2 difference is debated. Furthermore, how lingering misinterpretation may influence other aspects of language processing has 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). Our results showed that L1 and L2 speakers are prone to lingering misinterpretation during dependency resolution. L1/L2 differences were observed such that L2 speakers had increased difficulty reanalysing some filler-gap dependencies, however this was dependent on how the dependency was disambiguated. These results are compatible with the “good enough” approach to language processing, and suggest that L1/L2 differences are more likely when reanalysis is particularly difficult.

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

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., Christianson, Hollingworth, Halliwell & Ferreira, 2001; Slattery, Sturt, Christianson, Yoshida, & Ferreira, 2013).

(1) After Mary dressed the baby laughed.

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 misinterpretation has 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 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, but

such constructions provide a novel way of teasing apart competing accounts of L1 and L2 processing. 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.

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 unambiguous sentences. 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 (e.g., Christianson et al., 2006; Christianson, Luke, Hussey, & Wochna, 2017; 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” approach to sentence processing accounts for the lingering effect from the perspective of how comprehenders process language (Ferreira, Bailey, & Ferraro, 2002; Ferreira, Christianson, & Hollingworth, 2001; Ferreira & Patson, 2007; Karimi & Ferreira, 2016; Slattery et al., 2013). Slattery et al. (2013) considered two 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 (see also Qian, Garnsey, & Christianson, 2018 for a similar claim). 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 father grew worried and gave himself approximately five days to reply.

(3b) After the bank manager telephoned(,) David’s mother 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 (see 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 in 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.

Like L1 speakers, L2 speakers also have difficulty reading garden-path sentences (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).

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) cause reading disruption at “us” compared to (6b). This suggests that readers initially misinterpret “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 this type of construction as a non-filled-gap sentence, as unlike filled-gap sentences like (6a), the temporarily ambiguous verb (“wrote”) does not have an overt direct object. 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.

The type and timing of the cue to disambiguation differs in filled-gap and non-filled-gap sentences. The disambiguating cue in filled-gap sentences like (6a) is the overt direct object (“us”), which appears directly adjacent to the temporarily ambiguous verb. For non-filled-gap sentences like (7), the disambiguating cue is the preposition (“about”), which may occur several words after the temporarily ambiguous verb. This means that readers may commit to the initial misinterpretation for longer in non-filled-gap than filled-gap constructions. Also, the temporarily ambiguous verb in filled-gap sentences (“bring” in 6a) remains transitive, while in non-filled-gap sentences (7b), it is initially misinterpreted as transitive (“wrote the book” in

7b) but is ultimately intransitive (“wrote about the book”). These differences may influence the success of reanalysis. For example, the overt direct object in filled-gap constructions may act as a better cue to disambiguation than merely having a later preposition in non-filled-gap constructions. Additionally, previous research on garden-path sentences has indicated that reanalysis is more difficult the longer a reader maintains an initial misinterpretation (e.g., Christianson et al., 2001; Ferreira & Henderson, 1991; Tabor & Hutchins, 2004). If this also applies to filler-gap dependencies, non-filled-gap sentences that are not disambiguated until several words later, may cause greater reanalysis difficulty than filled-gap sentences.

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 different cues influence the formation of filler-gap dependencies and whether misanalysed filler-gap dependencies affect language comprehension. We are aware of only a few studies that have examined these issues, which investigated L1 but not L2 comprehension, and utilised offline tasks only (Lassotta, Omaki & Franck, 2016; Omaki, Davidson White, Goro, Lidz & Phillips, 2014; Wagers, Borja, & Chung, 2015). Omaki et al. (2014) investigated the interpretation of Japanese filler-gap dependencies in sentences like “Doko-de Yukiko-chan-wa choucho-o tsukamaeru-to itteta-no?” (*Where was Yukiko telling someone that she would catch a butterfly?*), where the filler (“Doko-de” *Where*) can be associated with either the main or embedded clause verb (“itteta” *telling* or “tsukamaeru” *catch* respectively). Here, the embedded clause verb appears before the main clause verb, and thus the filler may be incrementally associated with the embedded clause verb due to active gap filling (Stowe, 1986). In an offline experiment, Omaki et al. examined how two different cues that either syntactically or semantically block the embedded clause verb interpretation affect active gap filling. They found that while adults utilised both cues to disambiguation, for children the syntactic cue did not effectively prevent the formation of filler-gap dependencies at the embedded clause verb.

In Lassotta et al. (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 allait 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.

The present study

Against this background, we examined the processing and interpretation of filler-gap dependencies in L1 and L2 speakers. 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 in testing lingering misinterpretation, as although offline methods can test 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 greatly than L1 speakers.

To explore these issues, the present study investigated whether and to what extent initially assigned misinterpretations linger for filler-gap dependencies in L1 and L2 speakers. 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 misinterpretation in sentences containing filled-gaps and non-filled-gaps, to examine whether and how different disambiguating cues influence lingering misinterpretation in L1 and L2 processing.

Experiment 1

The aim of Experiment 1 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 host wanted the beer which the guest drank very quickly near during the party.

(b) *Unambiguous*

The host wanted the beer near which the guest drank very quickly during the party.

(c) Did the guest drink the beer?

1. Yes
2. No

In (10a/11a), the gap is ultimately at the preposition “near”, but there is an earlier possible gap following the verbs “stopped” and “drank”. 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 (Pozzan & Trueswell, 2016), accuracy rates should be lower for L2 than L1 participants in the ambiguous conditions.

Participants

Forty 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 backgrounds¹, from the University of Reading community, took part in Experiment 1. 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 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

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 from each experiment reported here is available at the first author's Open Science Framework (OSF) website (<https://osf.io/6uz95/>).

Seventy-two filler sentences were also constructed of which two-thirds were accompanied by a binary comprehension question. Half of the filler questions included two options (two nouns) as possible answers as in (10c), while the other half were yes/no questions like (11c). Half of these fillers required “yes” and the other half “no” answers. Experimental items were presented with four counterbalanced lists in a Latin Square design.

¹ 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).

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 with a binomial distribution, 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).² 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 OSF webpage (<https://osf.io/6uz95/>).

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

* INSERT TABLES 1 AND 2 HERE *

Discussion

Though the results were clearest in the filled-gap conditions, they 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

² 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 convergence was achieved.

& 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 that 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 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; see also Christianson et al., 2017 for similar results in main verb/reduced relative garden-path sentences). Another possibility is that the form of the question (“Did the girl read the novel?”) may have biased towards the misinterpretation. Note that despite this potential inference, we found suggestive evidence that the initially assigned misinterpretation lingers in non-filled-gap sentences.

Experiment 3 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 2 however, which tested misinterpretation of filled-gap sentences during online reading.

Experiment 2

Experiment 2 tested texts in four conditions like (12a–d) to investigate whether L1 and L2 speakers persist with initial misinterpretations during sentence processing following filled-gaps. 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 2 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 direct object (“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 would suggest readers generally computed the correct structure for the temporarily ambiguous conditions. If reanalysis is complete but the initially assigned interpretation lingers, as predicted by good-enough processing (Slattery et al., 2013), consistency effects should interact with ambiguity. 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.

Participants

The participants from Experiment 1 took part in Experiment 2. 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 1 first, Experiment 2 was completed by participants before Experiment 1, 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 <https://osf.io/6uz95/> 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 at a sample rate of 1000 Hz. 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 (gaze duration), 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 to minimise skew and to ensure that model residuals were normally distributed (see Vasishth & Nicenboim, 2016). At the request of the editor, we also analysed the non-

transformed data, which led to broadly similar results and did not alter our conclusions. The inferential statistics tables below report the log-transformed analyses, while we also include the non-transformed analyses as Online Supplement A. We note in footnotes any discrepancies between the two analyses.

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 1 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.

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%). A summary of the reading time data and inferential statistics are provided in Tables 3 and 4.³

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 in 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

³ The comparisons of theoretical interest in Experiment 2 were similar in the analysis of non-transformed reading times. At the disambiguating region, there was a significant ambiguity by group interaction in non-transformed total viewing times that was not significant in the log-transformed analysis. Here, L2ers had numerically larger filled-gap effects in non-transformed reading times than L1ers, but this interaction was not significant in the log analysis due to L2ers being generally slower. The ambiguity by consistency interaction at the critical and spillover regions was significant in both analyses of total viewing times.

There was a significant main effect of consistency in all measures due to longer reading times in 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.

* INSERT TABLES 3 AND 4 HERE *

INSERT FIGURE 1 HERE

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 2 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 suggesting that both groups generally succeeded in reanalysing the filled-gap sentences but that the initially assigned misinterpretation lingered some proportion of the time, consistent with good-enough processing (Slattery et al. 2013). These results are also consistent with the comprehension accuracy data from Experiment 1, which indicated relatively high accuracy but lingering misinterpretation in ambiguous filled-gap sentences.

Experiments 3 and 4 aimed to extend these findings in two ways. Experiment 3 aimed to replicate the results of Experiment 1 using a sentence-picture matching task. To further explore how reanalysis of filler-gap dependencies affects L1 and L2 sentence processing, Experiment 4 investigated lingering misinterpretation online following sentences containing non-filled-gaps.

Experiment 3

Experiment 3 aimed to test the generalisability of our findings from Experiment 1 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 1, given that even if participants make an inference, they still have to pick which picture they think provides the best match. Experiment 3 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.

INSERT FIGURES 2 AND 3 HERE

Experimental sentences were similar to those in Experiment 1 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 1, 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 3 were identical to Experiment 1. 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.

Participants

Forty 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 Experiments 1/2, completed Experiment 3 for course credit or payment. The L2 participants had various L1 backgrounds,⁴ 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

Experiment 3 employed 12 sets of filled-gap sentences (13) and 18 sets of non-filled-gap sentences (14) (see <https://osf.io/6uz95/>). A pair of pictures was constructed for each filled-gap

⁴ 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).

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 1, except that participants were instructed to choose which picture they felt best corresponded to the sentence rather than answer a comprehension question. The correct response was counterbalanced across experimental and filler items to be on either side of the screen an equal number of times. 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. The data analysis was the same as Experiment 1. 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.

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 conditions, the model showed a significant main effect of ambiguity due to lower accuracy rates in the ambiguous than unambiguous conditions, with no interaction by group.

The non-filled-gap conditions similarly revealed a significant 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 in 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 sentences, we compared the ambiguous condition with the control condition. This analysis showed significant 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 is driven by the lower accuracy in the ambiguous condition for L2 participants, discussed in the previous paragraph.

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 conditions. However, follow-up analyses with the control

condition 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 1, 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 reanalysis difficulty 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 3 may suggest that the nature of disambiguation modulates L1/L2 differences in reanalysis. We return to this issue in Experiment 5, which provides a more direct test of these potential L1/L2 differences.

Experiments 1 and 3 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 4 adopted eye-tracking to investigate lingering misinterpretation in a design similar to Experiment 2.

Experiment 4

To test how and whether misinterpretations of non-filled-gap sentences linger during online processing, in Experiment 4 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 2. The first sentence in (15a/15c) is temporarily ambiguous while in (15b/15d) it 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 4 are similar to Experiment 2. 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.

Participants

The participants from Experiment 3 also took part in Experiment 4. As in Experiments 1/2, participants completed Experiment 4 first, at least one week before Experiment 3.

Materials

Twenty-four 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 <https://osf.io/6uz95/> 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-thirds 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 2 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 2, 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 2. 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.

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 2, there was a significant 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 by consistency did not show any significant differences however (ambiguous conditions: *estimate* = 0.080, *t* = 1.51, *SE* = 0.05, *p* = .132; unambiguous conditions: *estimate* = 0.079, *t* = 1.32, *SE* = 0.06, *p* = .200).⁵

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.

⁵ In the non-transformed analysis, the main effect of ambiguity was only significant in total viewing times, and the 3-way interaction was not significant in any measure.

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 may suggest that L1 participants were more successful in completing reanalysis than L2 participants.

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 non-filled-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 significant 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). This consistency effect for L1ers may suggest they were more successful at completing reanalysis than L2ers.⁶

* INSERT TABLES 5 AND 6 HERE *

INSERT FIGURE 4 HERE

⁶ The ambiguity by consistency interaction was significant in first-pass reading times in both the non-transformed and log-transformed analyses. The group by consistency by region interaction was significant in all three measures in each analysis, as was the four-way interaction in regression path times.

Discussion

Experiment 4 showed several important findings. First, participants took more time to read ambiguous than unambiguous sentences. This might suggest reanalysis cost. However, here any differences between ambiguous and unambiguous conditions at the disambiguating region need to be taken with caution, given that the immediately preceding lexical material differed between ambiguous (“very slowly near”) and unambiguous conditions (“very slowly”).

Second, there was evidence that misinterpretation lingers for L1 and L2 speakers. However, this lingering effect was observed significantly only in the inconsistent conditions. In total viewing times, inconsistent continuation sentences had shorter reading times following ambiguous rather than unambiguous sentences, while no significant differences were found in the consistent conditions. This result is partly consistent with Experiment 2, 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 4 than in Experiment 2, especially for L2 speakers. While L1 speakers had significantly longer first-pass and total times for inconsistent than consistent sentences, L2 speakers did not show significant main effects of consistency in these measures. Indeed, the only measure in Experiment 4 that showed this consistency effect in L2 participants was regression path. However, here the effect was driven by the unambiguous conditions. Also, unlike L1 participants, L2 participants showed a trend for increased reading times for consistent sentences in one measure. Together, these results may indicate that L2 participants were less successful at reanalysing the temporary ambiguity than L1 participants.

In summary, these results suggest that both L1 and L2 speakers persist with initial misinterpretations in non-filled-gap sentences, but that reanalysis may have been more

successful for L1 speakers. While Experiment 4 thus 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 2 however, which tested filled-gap sentences. This pattern of results across Experiments 2 and 4 may be compatible with Experiment 3, which showed that L2 participants were more persistent with misinterpretation than L1 participants in non-filled-gap sentences, but not in 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 1, we decided to conduct one final offline experiment to more directly compare filled-gap and non-filled-gap sentences, using a within-sentence manipulation.

Experiment 5

Experiment 5 aimed to replicate the offline results from Experiments 1/3 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 1/3, (16) manipulates ambiguity such that (16a/c) are temporarily ambiguous sentences and (16b/d) are unambiguous controls. One crucial difference from Experiments 1/3 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 1/3, 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 1/3, we also expected lower accuracy rates for non-filled-gap than filled-gap sentences in Experiment 5.

Participants

Forty-eight 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 backgrounds⁷ from the University of Reading community, who did not complete Experiments 1/2, participated in Experiment 5 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

⁷ 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).

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 <https://osf.io/6uz95/>). Each sentence was followed by a comprehension question referring to the initially assigned misinterpretation. Experiment 5 also contained 48 filler sentences, all of which were accompanied by a yes/no comprehension question. Half of the filler questions required “yes” responses and the other half “no” responses. Experimental items were presented with four counterbalanced lists in a Latin Square design.

Procedure and data analysis

The procedure was identical to Experiment 1 except that Experiment 5 used IBEX farm as conducted in Experiment 3. Data analysis was similar to Experiments 1/3, but additionally included a sum-coded fixed effect of filler-gap type (filled-gap/non-filled-gap), along with the relevant interactions.

Results

Mean accuracy rates of filler sentences were 93% for L1 participants (range = 67–100) and 95% for L2 participants (range = 75–100).⁸ Comprehension accuracy rates of experimental sentences and inferential statistics are summarised in Tables 1 and 7.

⁸ A reviewer queried whether base-rate effects in the yes/no filler questions may have led to response bias in Experiments 1 and 5. Fillers were generally answered accurately irrespective

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 significant 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 significant main effect of ambiguity (*estimate* = 0.729, $z = 4.42$, $SE = 0.16$, $p < .001$). 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, $z = 1.26$, $SE = 0.40$, $p = .208$).

* INSERT TABLE 7 HERE *

Discussion

Experiment 5 suggested lingering misinterpretation in both L1 and L2 speakers and crucially provided further evidence that L2 speakers persist with misinterpretation more greatly than L1 speakers but only in non-filled-gap sentences, a finding consistent with Experiment 3.

of response. In Experiment 1, “yes” fillers had numerically higher accuracy (96%) than “no” fillers (88%), while the opposite was found in Experiment 5 (yes = 90%, no = 96%). Thus, we do not believe our findings can be reduced to response bias.

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.

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 1, 3 and 5 showed that initially assigned misinterpretations linger offline in both L1 and L2 speakers. Experiments 3 and 5 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 2 and 4 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 4 also suggested that 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 discarded 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 predictions of good enough processing, 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; Qian et al., 2018). Recently, Slattery et al. reported that for garden-path sentences, L1 speakers perform syntactic reanalysis and lingering effects result from failures to discard 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. In Experiment 2, we observed filled-gap effects in the temporarily ambiguous sentence, along with generally longer reading times in the continuation sentence for inconsistent than consistent conditions. Together, we believe these results suggest that reanalysis was largely successful in both groups. Specifically, in filled-gap sentences (e.g., Elisa noticed the truck which the policeman watched the car from) we contend that readers largely completed reanalysis, such that the overt direct object of the verb (“the car”) was interpreted as its theme. However, the effects of ambiguity on consistency observed in total viewing times in the continuation sentence in Experiment 2, along with our offline results, suggest that despite this, the initial misinterpretation was not fully erased from memory, as predicted by good enough processing (e.g., Qian et al., 2018; Slattery et al., 2013).

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 1 provided only suggestive evidence of lingering misinterpretation in non-filled-gap sentences, while comparisons between the ambiguous and unambiguous conditions in Experiment 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 3, 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 4 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 2, lingering effects were found in inconsistent but not consistent continuations in Experiment 4. While these results are thus indicative of lingering misinterpretation in non-filled-gap sentences following attempts at reanalysis, as would be expected under good enough processing, the extent to which reanalysis was always complete, especially in L2 speakers, in non-filled-gap sentences is less clear than was the case for filled-gap sentences. That is, in non-filled-gap sentences (e.g., Elisa noticed the truck which the policeman watched very quietly from), L2 speakers in particular may not have always successfully reinterpreted “watched” as intransitive.

We believe these differences may be at least partially due to the processing and interpretation of the preposition-fronted unambiguous condition. We argued that readers may make an inference 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, we hypothesise that it may not have disambiguated as expected as a result of a type of structural forgetting (Chacón, 2019; Gibson & Thomas, 1999; Vasishth, Suckow, Lewis, & Kern, 2010).

Although we are not aware of any previous work examining preposition fronting from the perspective of structural forgetting, we believe some existing evidence is compatible with this idea. Consider Radford (2009), who 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 here if during production the fronted preposition has an impoverished representation in memory, such that the producer repeats it to ensure the intended meaning is conveyed. Something similar may also occur in comprehension, such that the fronted preposition may be encoded in memory in an impoverished manner. Indeed, Radford, Felser and Boxell (2012) observed that L1 speakers judge sentences like (17) to be as acceptable as sentences without the copied preposition, suggesting the fronted preposition may not always be remembered accurately. Applying this to our own data, if readers sometimes had an impoverished representation of the fronted preposition, they may have sometimes computed a technically ungrammatical thematic relation between the wh-filler and the gap. Note that it cannot be the case that readers always completely forgot the fronted preposition, as Experiment 2 showed clear evidence of filled-gap effects (see also Wagers & Phillips, 2014), which indicates that the fronted preposition must have disambiguated to at least some degree.

If forgetting of the fronted preposition plays a role in the low accuracy for unambiguous non-filled-gap sentences, the question arises as to why unambiguous filled-gap sentences had higher accuracy, as these also include fronted prepositions. However, filled-gap sentences may have higher accuracy because the globally correct interpretation is more 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). That is, even if readers sometimes forget the fronted preposition, the filled direct object (“the car”) provides an overt cue that disambiguates against the initially assigned thematic interpretation between the wh-filler and the gap (“watched the truck”). Thus, the difference in results between the filled-gap and non-filled-gap constructions may be partially attributed to the low diagnosticity of the fronted preposition for disambiguation in non-filled-gap sentences, compared to the overt noun phrase in filled-gap structures (see also Omaki et al., 2014, and Martin & McElree, 2018, for similar claims relating to cue diagnosticity in garden-path sentences).

Another potential account of the results from the non-filled-gap sentences is that comprehension accuracy rates to unambiguous non-filled-gap sentences (e.g., “Kevin saw the letter beside which the candle burn very quickly last night”) were low because the correct interpretation of these sentences (e.g., “the candle burnt beside the letter”) was less plausible than the incorrect one (e.g., “the candle burnt the letter”). Noisy channel accounts of comprehension (e.g., Gibson et al., 2013; Levy et al., 2009), for example, might predict the incorrect interpretation is preferred if it is more plausible, even though it is technically incompatible with the input.⁹ To explore this possibility, we conducted three separate plausibility judgement studies that tested the plausibility of the correct and incorrect interpretations of each unambiguous non-filled-gap sentence used in Experiments 1, 3 and 5 respectively (for full details, see Online Supplement B). The results showed higher plausibility ratings to the correct than incorrect interpretations in Experiments 3 (correct interpretation: mean = 5.6, SD = 1.8; incorrect interpretation: mean = 3.3, SD = 2.2) and 5 (correct interpretation: mean = 5.3, SD = 2; incorrect interpretation: mean = 4, SD = 2.2) but

⁹ We thank the editor for pointing this out.

lower plausibility ratings to correct interpretations in Experiment 1 (correct interpretation: mean = 4.9, SD = 2; incorrect interpretation: mean = 5.5, SD = 2). The results for the items used in Experiment 1 may be, as mentioned in the Discussion section of that experiment, due to some items in which the incorrect interpretation could be plausibly inferred. However, given the results for the items in Experiments 3 and 5, we maintain that across experiments our results cannot be explained by differences in plausibility between the two (correct and incorrect) interpretations of the unambiguous non-filled-gap constructions.

Another potential difference between filled-gap and non-filled-gap sentences in the ambiguous conditions is that the timing of disambiguation may have also affected reanalysis. In the non-filled-gap construction used in the present study, readers need to commit to the initial misanalysis for longer than in 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). As such, the length of the ambiguity, rather than or in addition to diagnosticity, may have led to more successful reanalysis in filled-gap rather than non-filled-gap sentences. It is difficult to tease apart these different accounts based on ambiguity length and cue diagnosticity in the current study, but comparing the roles these two factors may play in reanalysis 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 the initially assigned 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. Experiments 3 and 5 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 across the offline experiments reported. 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 2, L1 participants showed clearer inconsistency effects than L2 participants in Experiment 4, which tested non-filled-gap sentences. These reduced inconsistency effects may suggest that reanalysis was not as successful for L2 speakers in non-filled-gap sentences.

These results suggest that it is not simply the case that reanalysis is always more difficult in L2 compared to L1 processing. 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. We argued that non-filled-gap sentences are more difficult to reanalyse than filled-gap sentences due to either the long ambiguous phrase or the 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 assumed 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 et al., 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 cost 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.

Another possibility to be considered is that L2 speakers may not be as sensitive to verb transitivity biases as L1 speakers.¹⁰ Research in L1 processing has examined how active gap filling is influenced by how frequently a verb is used transitively or intransitively (e.g. Omaki et al., 2015; Pickering & Traxler, 2003; Staub, 2007; Stowe, Tanenhaus & Carlson, 1991). Transitivity biases may have influenced L1 and L2 processing in our study, and may have influenced either the degree to which gaps were filled actively, or how successfully initially assigned transitive misinterpretations were reanalysed as intransitive in the non-filled-gap constructions. Subtle differences in sensitivity to these transitivity biases could provide an alternative explanation of the observed L1/L2 differences.

To explore this possibility, we conducted an additional analysis of the ambiguous non-filled-gap condition from Experiments 3 and 5, where we observed L1/L2 differences. Here, for each verb we calculated verb transitivity bias by obtaining a random sample of sentences (100 for each verb) that used the verbs in the past tense from the British National Corpus

¹⁰ We thank an anonymous reviewer for pointing this out.

<http://www.natcorp.ox.ac.uk/>). Transitivity bias was calculated by dividing the number of transitive usages by the total number of transitive and intransitive usages for each verb. We then included this transitivity bias in a generalised linear mixed effect model of the combined data from Experiments 3 and 5. The model included a sum-coded fixed effect of group (L1/L2) and a centred, continuous predictor for transitivity bias. Random effects included by-subject, by-item and by-experiment (Experiment 3/Experiment 5) intercepts, and random slopes were included with the maximal random effect structure that converged.

This model showed significant main effects of group ($estimate = -0.710, z = -3.33, SE = 0.21, p < .001$) and transitivity bias ($estimate = -0.010, z = -2.36, SE = 0.04, p = .002$), with higher accuracy rates for L1 speakers and more intransitive verbs respectively. The group by transitivity bias interaction was not significant ($estimate = 0.009, z = 1.50, SE = 0.06, p = .134$). Thus, while transitivity biases appear to have influenced our results, this analysis does not support the idea that the observed L1/L2 differences can be reduced to differences in sensitivity to verb bias. Further research that systematically manipulates verb bias is required here to further examine how transitivity influences misinterpretation in L1 and L2 speakers.

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, L2 speakers may not complete syntactic reanalysis, a finding which may be compatible with the shallow structure hypothesis (Clahsen & Felser, 2006, 2017). Our results suggest both accounts might be too strong if they do not take reanalysis difficulty into account. 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 (Clahsen & Felser, 2006), 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. Irrespective of this issue, the current results indicate that reanalysis difficulty is an important factor in teasing apart different accounts of L1/L2 processing.

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. L1/L2 differences were observed, which indicated that L2 speakers have more difficulty reanalysing filler-gap dependencies than L1 speakers in non-filled-gap sentences but not in filled-gap sentences. These results indicate that theoretical accounts of L2 processing need to take reanalysis difficulty into account when explaining the similarities and differences between L1 and L2 sentence processing. To conclude, expanding on previous research on garden-path sentences, the present study provides novel evidence that initial misinterpretations linger in filler-gap dependencies in L1 and L2 sentence comprehension. In extending previous results from garden-path sentences to filler-gap dependencies, our results are broadly compatible with the predictions of good-enough language processing.

Table 1. Accuracy rates for comprehension questions following ambiguous and unambiguous sentences in Experiments 1, 3 and 5 (SDs in parentheses).

	Native Speakers					Non-Native Speakers				
	Filled-gap		Non-filled-gap			Filled-gap		Non-filled-gap		
	Ambiguous	Unambiguous	Ambiguous	Unambiguous	Control	Ambiguous	Unambiguous	Ambiguous	Unambiguous	Control
Experiment 1	71 (45)	76 (43)	34 (48)	39 (49)	-	65 (48)	78 (41)	23 (42)	35 (48)	-
Experiment 3	71 (45)	82 (38)	70 (46)	68 (47)	86 (35)	70 (46)	77 (42)	50 (50)	69 (46)	86 (35)
Experiment 5	68 (47)	73 (45)	47 (50)	50 (50)	-	69 (46)	78 (42)	36 (48)	59 (49)	-

Table 2. Summary of statistical analyses for Experiments 1 and 3

	Experiment 1			Experiment 3		
	<i>Estimate (SE)</i>	<i>z value</i>	<i>p value</i>	<i>Estimate (SE)</i>	<i>z value</i>	<i>p value</i>
Filled-gap						
Intercept	1.27 (0.21)	5.99	< .001	1.39 (0.20)	6.85	< .001
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						
Intercept	1.10 (0.39)	2.85	.004	0.77 (0.19)	4.00	< .001
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 3. Reading times for three eye-movement measures at three regions of texts in Experiment 2 (SDs in parentheses).

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

Table 4. Summary of statistical analyses for Experiment 2.

	First pass reading time			Regression path duration			Total viewing time		
	<i>Estimate (SE)</i>	<i>t value</i>	<i>p value</i>	<i>Estimate (SE)</i>	<i>t value</i>	<i>p value</i>	<i>Estimate (SE)</i>	<i>t value</i>	<i>p value</i>
Disambiguating region									
Intercept	5.76 (0.03)	186.64	< .001	5.99 (0.04)	142.58	< .001	6.33 (0.05)	132.62	< .001
Ambiguity	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
Consistency	0.02 (0.02)	1.04	.304	0.05 (0.03)	1.72	.087	0.12 (0.03)	3.72	< .001
Ambiguity:Group	0.02 (0.05)	0.47	.640	0.03 (0.06)	0.49	.628	0.07 (0.05)	1.39	.169
Ambiguity:Consistency	0.01 (0.06)	0.09	.926	0.08 (0.07)	1.17	.254	0.01 (0.06)	0.19	.852
Group:Consistency	0.003 (0.05)	0.06	.953	0.10 (0.06)	1.63	.103	0.01 (0.06)	0.11	.913
Ambiguity:Group:Consistency	0.08 (0.10)	0.79	.436	0.04 (0.12)	0.35	.731	0.01 (0.11)	0.09	.926
Critical / spillover regions									
Intercept	5.65 (0.03)	193.11	< .001	5.78 (0.04)	152.90	< .001	5.98 (0.04)	163.86	< .001
Ambiguity	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
Consistency	0.04 (0.01)	2.79	.011	0.07 (0.02)	2.98	.008	0.13 (0.03)	4.09	< .001
Region	0.14 (0.05)	2.68	.013	0.28 (0.07)	3.76	< .001	0.24 (0.06)	4.01	< .001
Ambiguity:Group	0.004 (0.03)	0.14	.891	0.03 (0.04)	0.86	.401	0.05 (0.04)	1.26	.207
Ambiguity:Consistency	0.02 (0.03)	0.75	.456	0.08 (0.04)	2.07	.051	0.13 (0.05)	2.95	.007
Group:Consistency	0.04 (0.03)	1.74	.099	0.04 (0.04)	1.11	.281	0.01 (0.04)	0.19	.855
Ambiguity:Region	0.04 (0.03)	1.40	.176	0.04 (0.03)	1.22	.223	0.03 (0.03)	1.26	.209
Group:Region	0.01 (0.04)	0.34	.734	0.06 (0.06)	0.95	.348	0.003 (0.04)	0.09	.933
Consistency:Region	0.0002 (0.03)	0.01	.994	0.05 (0.04)	1.22	.237	0.07 (0.04)	1.74	.096
Ambiguity:Group:Consistency	0.003 (0.05)	0.06	.955	0.05 (0.07)	0.66	.508	0.06 (0.08)	0.74	.460
Ambiguity:Group:Region	0.02 (0.05)	0.36	.724	0.01 (0.07)	0.15	.881	0.02 (0.05)	0.30	.764
Ambiguity:Consistency:Region	0.01 (0.05)	0.23	.817	0.13 (0.09)	1.44	.165	0.02 (0.07)	0.25	.804
Group:Consistency:Region	0.03 (0.06)	0.48	.635	0.05 (0.08)	0.72	.479	0.003 (0.05)	0.06	.953
Ambiguity:Group:Consistency:Region	0.16 (0.11)	1.51	.148	0.06 (0.15)	0.42	.679	0.01 (0.11)	0.07	.945

Table 5. Reading times for three eye-movement measures at three regions of texts in Experiment 4 (SDs in parentheses).

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

Table 6. Summary of statistical analyses for Experiment 4.

	First pass reading time			Regression path duration			Total viewing time		
	<i>Estimate (SE)</i>	<i>t value</i>	<i>p value</i>	<i>Estimate (SE)</i>	<i>t value</i>	<i>p value</i>	<i>Estimate (SE)</i>	<i>t value</i>	<i>p value</i>
Disambiguating region									
Intercept	5.70 (0.02)	262.32	< .001	6.22 (0.07)	92.08	< .001	6.37 (0.05)	135.24	< .001
Ambiguity	0.02 (0.03)	0.67	.502	0.13 (0.04)	3.32	.001	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
Consistency	0.04 (0.02)	1.78	.076	0.01 (0.04)	0.37	.716	0.03 (0.03)	1.00	.322
Ambiguity:Group	0.03 (0.05)	0.52	.607	0.05 (0.08)	0.59	.559	0.05 (0.07)	0.70	.487
Ambiguity:Consistency	0.03 (0.04)	0.65	.517	0.02 (0.07)	0.23	.818	0.05 (0.05)	0.88	.379
Group:Consistency	0.02 (0.04)	0.51	.612	0.02 (0.07)	0.31	.759	0.06 (0.06)	1.08	.296
Ambiguity:Group:Consistency	0.12 (0.09)	1.43	.172	0.02 (0.15)	0.14	.888	0.22 (0.12)	1.83	.081
Critical / spillover regions									
Intercept	5.90 (0.04)	164.82	< .001	6.12 (0.04)	149.38	< .001	6.35 (0.05)	137.70	< .001
Ambiguity	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
Consistency	0.01 (0.02)	0.45	.654	0.03 (0.02)	1.19	.245	0.04 (0.04)	1.06	.298
Region	0.13 (0.05)	2.46	.020	0.16 (0.06)	2.54	.017	0.15 (0.06)	2.57	.016
Ambiguity:Group	0.05 (0.03)	1.54	.127	0.04 (0.04)	0.97	.342	0.02 (0.05)	0.50	.620
Ambiguity:Consistency	0.09 (0.03)	2.48	.015	0.08 (0.04)	2.03	.055	0.11 (0.05)	2.17	.034
Group:Consistency	0.07 (0.04)	1.72	.096	0.03 (0.04)	-0.81	.422	0.07 (0.05)	1.33	.187
Ambiguity:Region	0.04 (0.04)	1.06	.301	0.04 (0.04)	1.02	.308	0.02 (0.04)	0.58	.571
Group:Region	0.11 (0.05)	2.46	.016	0.02 (0.05)	0.50	.621	0.11 (0.04)	2.66	.011
Consistency:Region	0.04 (0.04)	0.89	.382	0.10 (0.05)	2.07	.049	0.02 (0.04)	0.55	.590
Ambiguity:Group:Consistency	0.00 (0.07)	0.03	.979	0.10 (0.07)	1.38	.169	0.15 (0.10)	1.57	.120
Ambiguity:Group:Region	0.12 (0.07)	1.83	.071	0.02 (0.08)	0.20	.841	0.07 (0.05)	1.20	.231
Ambiguity:Consistency:Region	0.04 (0.07)	0.56	.580	0.07 (0.11)	0.63	.536	0.01 (0.06)	0.15	.881
Group:Consistency:Region	0.18 (0.08)	2.33	.023	0.19 (0.08)	2.23	.028	0.20 (0.06)	3.15	.002
Ambiguity:Group:Consistency:Region	0.10 (0.12)	0.79	.429	0.40 (0.15)	2.61	.009	0.05 (0.14)	0.40	.692

Table 7. Summary of statistical analyses for Experiment 5.

	Experiment 5		
	<i>Estimate (SE)</i>	<i>z value</i>	<i>p value</i>
Intercept	0.55 (0.14)	3.83	< .001
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

Figure 1. Total viewing times in milliseconds at the critical and spillover regions in Experiment 2.

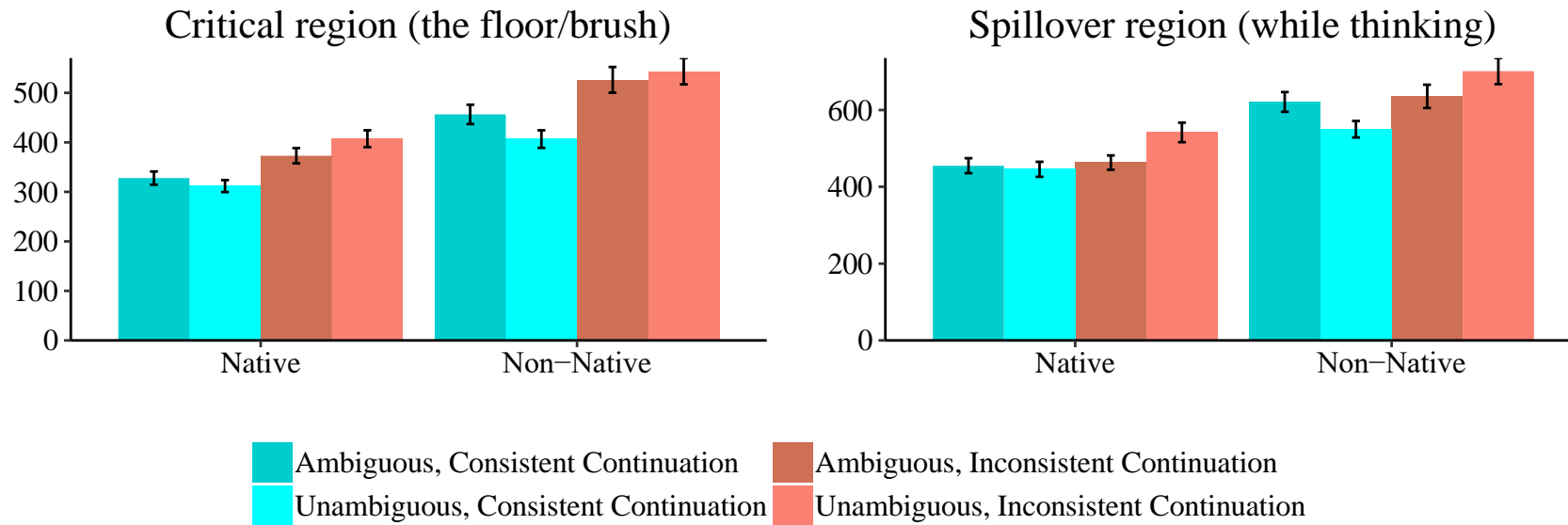


Figure 2. Example picture pairs used for filled-gap sentences in Experiment 3.

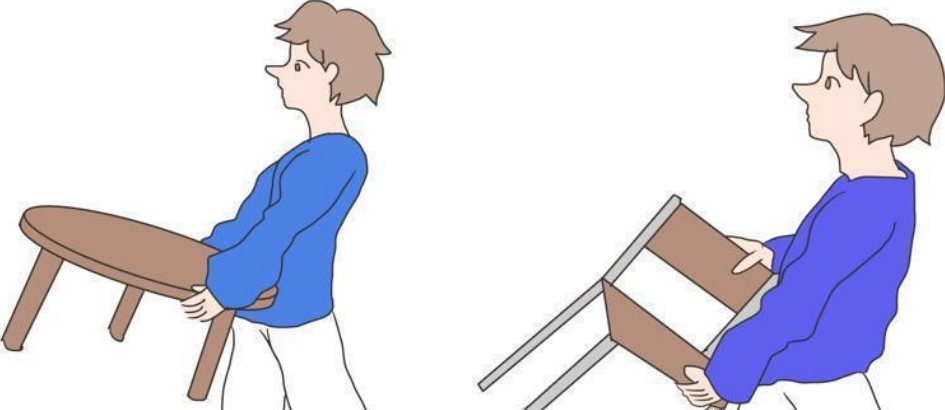
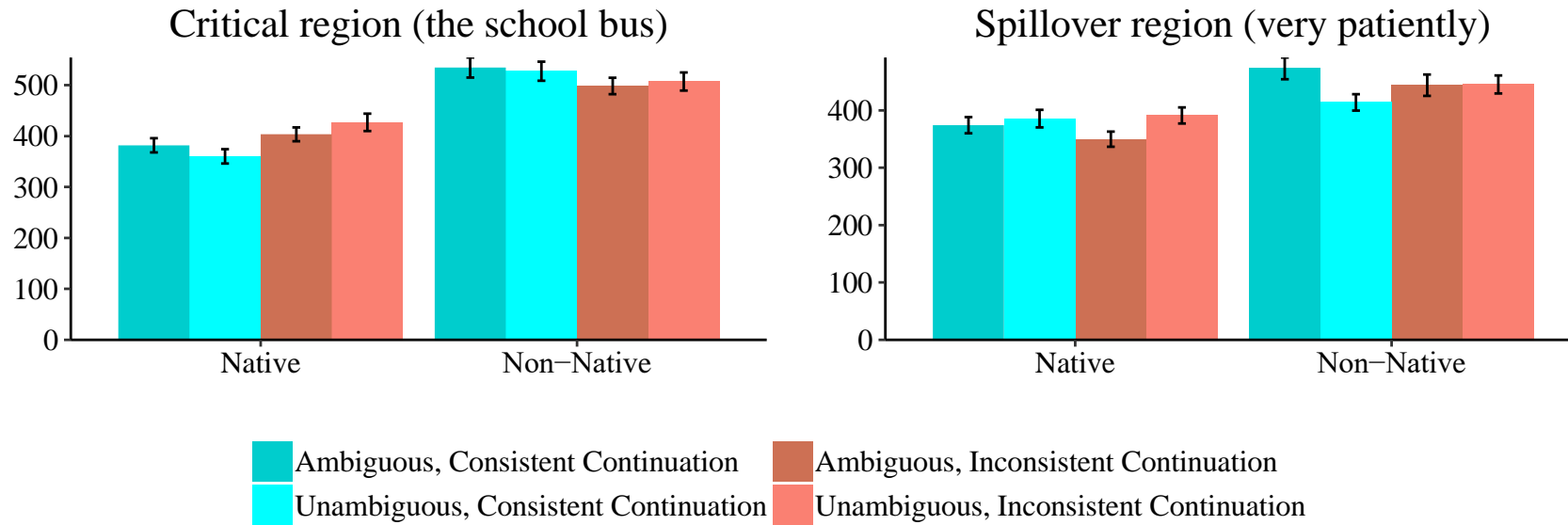


Figure 3. Example picture pairs used for non-filled-gap sentences in Experiment 3.



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



References

- Aoshima, S., Phillips, C., & Weinberg, A. (2004). Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language*, 51(1), 23–54. <https://doi.org/10.1016/j.jml.2004.03.001>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Balota, D. A., Yap, M. J., Cortese, M. J., et al. (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>
- Chacón, D. (2019). Minding the gap?: Mechanisms underlying resumption in English. *Glossa: A Journal of General Linguistics*, 4, 68. <http://doi.org/10.5334/gjgl.839>
- 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>.
- 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., Hussey, E. K., & Wochna, K. (2017). Why reread? Evidence from garden-path and local coherence structures. *Quarterly Journal of Experimental Psychology*, 70(7), 1380-1405. <https://doi.org/10.1080/17470218.2016.1186200>
- Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006). Younger and older adults' good enough interpretations of garden path sentences. *Discourse Processes*, 42(2), 205–238. https://doi.org/10.1207/s15326950dp4202_6
- 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>
- 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., & 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>
- Felser, C., Cunnings, I., Batterham, C., & Clahsen, H. (2012). The timing of island effects in nonnative sentence processing. *Studies in Second Language Acquisition*, 34(1), 67–98. <https://doi.org/10.1017/S0272263111000507>

- Felser, C., & Roberts, L. (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>
- 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., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, 30(6), 725–745.
[https://doi.org/10.1016/0749-596X\(91\)90034-H](https://doi.org/10.1016/0749-596X(91)90034-H)
- Ferreira, F., 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., & 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>
- Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, 4(2), 93–126.
<http://dx.doi.org/10.1080/01690968908406359>
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14(2), 178–210.
[https://doi.org/10.1016/0010-0285\(82\)90008-1](https://doi.org/10.1016/0010-0285(82)90008-1)

- Frazier, L. (1987). Syntactic Processing: Evidence from Dutch. *Natural Language & Linguistic Theory*, 5(4), 519–559. <http://www.jstor.org/stable/4047505>
- 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 and non-native speakers' processing of subject-object ambiguities. *International Journal of Bilingualism* 21(2), 125–144. <https://doi.org/10.1177/1367006915604401>
- Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A Noisy-Channel Account of Crosslinguistic Word-Order Variation. *Psychological Science*, 24(7), 1079–1088. <https://doi.org/10.1177/0956797612463705>
- Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognition Processes*, 14, 225-248.
- Gorrell, P. (1995). *Syntax and parsing*. Cambridge, England: Cambridge University Press.
- 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>
- 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>
- 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>
- Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). ImerTest 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>
- Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences*, 106(50), 21086–21090. <https://doi.org/10.1073/pnas.0907664106>
- 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>
- 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.

- 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>
- 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>
- 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>
- 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>
- Omaki, A., Lau, E. F., Davidson White, I., Dakan, M. L., Apple, A., & 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>
- 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). Lingerings 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>
- Pickering, M. J., & Traxler, M. J. (2003). Evidence against the use of subcategorisation frequency 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>
- 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>
- 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.
- R Core Team (2018). 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>
- 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>
- 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>
- Staub, A. (2007). The parser doesn't ignore transitivity, after all. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 33(3), 550–569.
<https://doi.org/10.1037/0278-7393.33.3.550>
- 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>
- Stowe, L., Tanenhaus, M. & Carlson, G. (1991). Filling gaps on-line: Use of lexical and semantic information in sentence processing. *Language and Speech*, 34(4), 319–340.
<https://doi.org/10.1177/002383099103400402>
- 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., Pickering, M. J., & Crocker, M. W. (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language*, 40(1), 136–150.
<https://doi.org/10.1006/jmla.1998.2606>
- 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](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>
- 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>
- Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies. *Journal of Memory and Language*, 35(3), 454–475.
<https://doi.org/10.1006/jmla.1996.0025>
- van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Jacob, G. (2006). The activation of inappropriate analyses in garden-path sentences: Evidence from structural priming. *Journal of Memory and Language*, 55(3), 335–362.
<https://doi.org/10.1016/j.jml.2006.06.004>

- Vasishth, S. & Nicenboim, B. (2016). Statistical methods for linguistic research: Foundational Ideas – Part I. *Language and Linguistics Compass*, 10(8), 349–369.
<https://doi.org/10.1111/lnc3.12201>
- Vasishth, S., Suckow, K., Lewis, R., & Kern, S. (2010). Short-term forgetting in sentence comprehension: Crosslinguistic evidence from verb-final structures. *Language and Cognitive Processes*, 25(4), 533–567.
<https://doi.org/10.1080/01690960903310587>
- Wagers, M. & Borja, M. F. & Chung, S. (2015). The real-time comprehension of WH-dependencies in a WH-agreement language. *Language*, 91, 109–144.
<https://doi.org/10.1353/lan.2015.0001>
- 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>
- 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>
- 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>