*Studies in Second Language Acquisition* (2026), 1–23
doi:10.1017/S0272263126101910

RESEARCH ARTICLE

# Structure building, memory retrieval, and structural prediction in second language sentence comprehension

Ian Cunnings[1], Shatha Alaskar[2] and Hiroki Fujita[3]

[1]Clinical Language Sciences, University of Reading, Reading, UK; [2]Department of English, College of Education, Majmaah University, Al Majmaah, Saudi Arabia and [3]Department of Linguistics, University of Potsdam, Potsdam, Germany
**Corresponding author:** Ian Cunnings; Email: i.cunnings@reading.ac.uk

(Received 05 September 2025 UTC; Revised 24 June 2026 UTC; Accepted 09 July 2026 UTC)

## Abstract

Potential difficulty in second language (L2) compared to first language (L1) sentence processing has been ascribed to various factors, including difficulty in building syntactic structure, difficulty in retrieving information from memory during comprehension, and difficulty in making predictions about upcoming elements. In this paper, we examine pronoun resolution in so-called *crossover* constructions (*The girl who she helped earlier today went home*), where comprehension requires the interaction of structure building, structural prediction, and memory retrieval in real-time, as a test case of these claims. In three preregistered experiments (N = 96 in Experiments 1a/b, and N = 160 in Experiment 2), we found that L1 and L2 real-time processing and final comprehension are constrained by crossover constraints. There was also no evidence for L1/L2 differences in the processing of crossover. These results suggest similar processes of structure building, structural prediction, and memory retrieval in L1 and L2 readers during sentence comprehension.

## Introduction

Successful second language (L2) acquisition is contingent on successful L2 sentence processing. The factors that influence L2 processing remain debated (e.g., Clahsen & Felser, 2006, 2018; Cunnings, 2017; Hopp, 2022). One influential claim is that L2 speakers have difficulty using syntactic information during processing (Clahsen & Felser, 2006, 2018). The role of working memory, broadly defined as the ability to maintain a limited amount of task-relevant information at one time, in L2 acquisition and processing is also contested (for discussion, see Cunnings, 2017, 2022). Cunnings (2017) proposed that differences in L1 and L2 processing may be explained by the memory retrieval processes that subserve comprehension. Results suggesting that L2 speakers may not predict linguistic material in the same way as L1 speakers have

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:44:42, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101910

also motivated research into L2 processing (for review, see Kaan & Grüter, 2021; Schlenter, 2023).

In this study, we test a linguistic phenomenon known as *crossover* to examine the roles of structure building, memory retrieval, and prediction in L2 processing. Consider (1a/b), where a reader must correctly interpret the pronoun "he" for successful sentence comprehension. In (1a), the pronoun can corefer with the preceding noun phrase (NP) "the man." However, in (1b), this coreference is ungrammatical. Coreference in (1b) is ruled out because although "the man" precedes the pronoun, "the man" is also interpreted as the direct object of the verb "phoned," which syntactically blocks coreference. This is known as crossover (Postal, 1971).

(1a) The man who said that he phoned yesterday was late.
(1b) The man who he phoned yesterday was late.

Anaphora resolution during real-time sentence processing involves retrieving a representation of an antecedent previously encoded in memory (Jäger et al., 2020; Parker et al., 2017). This antecedent memory retrieval is constrained by various factors, such as lexical features like gender, and also syntactic structure (Chow et al., 2014; Cunnings et al., 2015; Dillon et al., 2013; Fujita & Yoshida, 2024; Sturt, 2003). In (1b), restricting the sentence subject "the man" from antecedent retrieval at "he" requires sufficient structure building to have taken place, and importantly, a prediction that "the man" is also to be interpreted at a position later in the sentence, in this case the direct object of the upcoming, but as yet unencountered, verb. Thus, for crossover to apply during real-time processing at the pronoun, readers must predictively project sufficiently detailed syntactic structure about the role that previously encountered referents will play in the upcoming structure. Crossover thus provides a unique test case of structure building, memory retrieval, and a specific type of (structural) prediction in L2 processing. Whilst we do not attempt to tease apart different L2 accounts which predict difficulty in terms of shallow processing (Clahsen & Felser, 2006, 2018), memory retrieval (Cunnings, 2017), or prediction (Grüter et al., 2014), we utilize crossover as a way of testing claims about how structure-building, memory retrieval and structural prediction should influence L2 processing.

We report three preregistered experiments that examined the processing and interpretation of pronouns in crossover and non-crossover constructions. Our results indicated that L1 and high proficiency L2 readers respect crossover constraints during online processing, a finding that limits theoretical claims regarding L1/L2 differences in use of syntactic information, memory retrieval, and prediction during processing. Below, we discuss different accounts of L2 processing, before reviewing existing research on crossover in L1 and L2 comprehension.

## *Syntactic structure, memory retrieval, and prediction in L2 processing*

The Shallow Structure Hypothesis (*SSH*; Clahsen & Felser, 2006, 2018) has been influential in motivating research in L2 sentence processing. The SSH predicts that L2 readers have difficulty using syntactic information during processing and rely more on semantic or discourse-level information, or properties such as surface word order. Some evidence for the SSH comes from research on anaphora resolution, including reflexives (Felser et al., 2009; Felser & Cunnings, 2012). Felser et al. (2009), for example, monitored L1 and L2 readers' eye movements as they read texts such as (2a/b). In (2a/b), "Richard" is the only possible antecedent for the reflexive "himself" (Chomsky, 1981). We refer to antecedents that are structurally allowed to corefer with pronouns as

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:44:42, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101910

*accessible antecedents*. The sentences include another NP "John/Jane," which either matches or mismatches the reflexive in gender. This NP cannot be an accessible antecedent for the reflexive. We refer to this type of structurally inaccessible antecedent as *a distractor*.

(2a) John noticed that Richard had cut himself with a very sharp knife.
(2b) Jane noticed that Richard had cut himself with a very sharp knife.

L2 readers had longer first-pass reading times at the reflexive when the distractor matched the reflexive in gender. This difference was not observed in L1 readers. Also, this difficulty did not arise when the distractor was in a less discourse-salient position (*It was clear to John/Jane that Richard had cut himself …*). Felser et al. concluded that, during initial stages of anaphora resolution, L2 readers are guided more strongly by discourse salience than by structural constraints. They took this finding as evidence for the SSH.

Cunnings (2017), however, argued that such findings do not necessarily implicate shallow structures, and instead may be explained based on memory retrieval operations during sentence processing. To interpret anaphors correctly, a reader must retrieve information previously encoded in memory about their antecedents. According to cue-based models of memory retrieval during sentence processing (e.g., Lewis et al., 2006; Lewis & Vasishth, 2005), memory retrieval is achieved by matching a set of retrieval cues against all elements in memory, with the element that provides the best match subsequently being retrieved. In (2a), for example, retrieval cues at the reflexive may include gender information derived from the reflexive (e.g., [+MASC]) and structure-based cues derived from the syntactic structure. Simplifying somewhat, for reflexives, this may include a cue such as [+CLAUSEMATE]. In (2a/b), the structurally accessible antecedent "Richard" matches both cues, while the distractor only matches the gender cue ("John") or neither ("Jane"). A key prediction of cue-based models is interference. Because retrieval cues are matched against all elements in memory in parallel, elements that partially match a set of retrieval cues affect memory retrieval. Specifically, the cue-based model predicts processing difficulty (*inhibitory interference*) when the accessible antecedent and the distractor (i.e., the structurally inaccessible antecedent) match the same cue. That is, the cue-based model predicts longer reading times at the reflexive in (2a) than (2b). Cunnings (2017) claimed that the results from studies such as Felser et al. (2009) may indicate that L2 learners weight memory retrieval cues differently to L1 speakers, and as such L2 learners are more susceptible to retrieval interference, rather than implicating shallow L2 parsing. That is, L2 speakers may overweight non-structural cues during antecedent retrieval, which in turn leads to greater interference in L2 than L1 readers (Fujita, 2025). Note that while some studies have reported more interference during reflexive resolution for L2 than L1 speakers (Cunnings & Felser, 2012; Felser et al., 2009), some more recent work has indicated that this is not always the case, at least when a distractor antecedent is in a less discourse salient non-subject position (e.g., "The man near the policeman strangely isolated himself from society for many years"; see Alaskar & Cunnings, 2025).

Other potential differences between L1 and L2 sentence processing have been described in terms of prediction (e.g., Kaan, 2014; Kaan & Grüter, 2021; Schlenter, 2023). For example, Grüter et al. (2014) reported in a story continuation experiment that L1 but not L2 readers could anticipate which referent would be mentioned next in a discourse based on event structure. Based on such findings, Grüter et al. claimed that L2 learners have a reduced ability to generate expectations (*RAGE*) during processing. While RAGE focuses on L2 ability to make discourse-level predictions, the extent to

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:44:42, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101910

which L2 learners can predict more widely is contested. Although more recent instantiations of the RAGE hypothesis focus not on whether L2 learners can or cannot predict, but on the utility of prediction in L2 processing, when and how L2 learners make predictions is still debated (for discussion, see Kaan & Grüter, 2021).

Against the backdrop of debate regarding sentence structure, memory retrieval, and prediction, crossover provides a unique way of assessing these different issues in L2 processing. While anaphora resolution in crossover constructions like (1b) triggers the retrieval of an antecedent from memory, this retrieval can only be achieved correctly if it is adequately constrained by syntactic structure, and the structural prediction that the sentence subject is also analyzed as the direct object of the upcoming verb.

## Crossover in L1 and L2 processing

Crossover phenomena have been widely discussed in linguistics. Although different accounts have been proposed (e.g., Chomsky, 1981; Lasnik & Funakoshi, 2017; Postal, 1971; Wasow, 1972), an influential theory is that coreference in crossover is structurally ruled out by Binding Principle C (Chomsky, 1981), which states that a referential expression cannot corefer with a pronoun that c-commands it.[1,2] For example, in (3a), the pronoun "he" c-commands "the man," and thus they cannot corefer. In (1b), repeated below as (3b), the pronoun does not c-command the sentence subject "the man." However, the pronoun c-commands the object position of the verb "phoned," where "the man" is interpreted, via the relative pronoun "who," as the direct object of "phoned" (denoted as t in [3b]). Consequently, coreference between "he" and "the man" is prohibited in (3b). We refer to this as the crossover constraint.

(3a) He phoned the man yesterday afternoon.
(3b) The man who he phoned t yesterday afternoon was late.

Research in L1 processing has investigated whether crossover constrains anaphora resolution during processing. Kush et al. (2017) tested sentences as in (4), which use the gender mismatch paradigm, to examine the online application of crossover constraints. In non-crossover sentences like (4a), coreference between the pronoun ("he") and the preceding wh-phrase is structurally permissible, while in crossover sentences, like (4b), this is impermissible because of the crossover constraint. In a self-paced reading experiment, Kush et al. reported longer reading times shortly after the pronoun when the wh-phrase mismatched it in gender compared to when it matched in non-crossover sentences like (4a) (*gender mismatch effects*). No such difference was observed in crossover sentences like (4b). Kush et al. interpreted this interaction between crossover and gender mismatch effects as evidence that readers made sufficiently detailed structural predictions to restrict pronoun resolution in crossover constructions.

[1]The present study examines so-called *strong crossover*, while there is another type of crossover known as *weak crossover* (e.g., The man who his friend phoned yesterday afternoon was late). The distinction between strong and weak crossover has been important in characterizing syntactic constraints in the linguistics literature. As we focus on strong crossover, we do not discuss weak crossover. Unless otherwise indicated, when we use the term crossover in this paper, we are referring to strong crossover.

[2]C-command refers to a relationship between two nodes within a syntactic structure, such that a node c-commands its sister node and any constituents that this sister node dominates (Reinhart, 1976).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:44:42, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101910

(4a) Jane asked [which maintenance man]/[which lunch lady] had said that he already spoke …
(4b) Jane asked [which maintenance man]/[which lunch lady] it appeared that he already spoke …

We are aware of only one existing study that examined crossover in L2 processing. Felser and Drummer (2017) examined crossover in L1 German readers and L1 Russian, L2 German learners, using sentences like (5). In an eye-movement experiment, similar reading times were observed at the pronoun between the gender match and gender mismatch conditions for both L1 and L2 readers. Additionally, in an offline task testing similar materials but only including gender-matching wh-phrases, both L1 and L2 readers rejected coreference between the pronoun and wh-phrase, indicating knowledge of crossover.

(5) Neben welchem Experten/welcher Expertin aus Dänemark er unbedingt sitzen sollte, das wusste Daniel genau.
"Next to which $\text{expert}_{\text{masc/fem}}$ from Denmark he absolutely should sit, Daniel knew exactly."

Felser and Drummer concluded that both L1 and L2 readers respect crossover during processing. While these results suggest L2 readers respect crossover during processing, we are cautious in drawing strong conclusions from this study. Firstly, these conclusions are drawn from the lack of gender mismatch effects, and Felser and Drummer did not test for an interaction between crossover and a non-crossover control condition, as was tested by Kush et al. (2017). To draw stronger conclusions about the use of crossover constraints during processing, an interaction design is required.[3] Additionally, Felser and Drummer only tested crossover sentences in which there was no grammatically accessible antecedent for the pronoun at the point it was encountered. Testing sentences where there is a grammatically accessible antecedent is important for fully assessing the interaction between structure building, memory retrieval and prediction during anaphora resolution, and is key to testing an interference-based account of L1 and L2 processing (Cunnings, 2017). Given the unique way in which crossover allows examination of structure building, memory retrieval and prediction during L2 comprehension, our study aimed to address these gaps in the existing literature.

## The present study

We report three preregistered experiments on crossover in L1 and L2 processing. Experiments 1a and 1b tested the extent to which L1 and L2 readers respect constraints on anaphora resolution in crossover as compared to non-crossover constructions. Experiment 1a was an offline task to examine knowledge of crossover constraints, while in Experiment 1b we monitored participants' eye-movements during reading to examine real-time sentence processing. More specifically, the aim of Experiment 1b was to assess whether L2 readers build sufficient sentence structure and make appropriate structural predictions to constrain anaphora resolution in crossover constructions.

[3]Felser and Drummer (2017) tested weak crossover in addition to strong crossover, but analyzed them separately.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:44:42, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101910

Experiment 2 aimed to further test these issues, using a design that also allows for direct testing of interference effects in anaphora resolution, to provide a test of the claim that L2 readers are more susceptible to interference during memory retrieval than L1 speakers. Because of the apparent universal nature of crossover, we tested L2 readers from a variety of different L1 backgrounds. In all experiments, we tested highly proficient L2 readers as we intended to test residual L1/L2 differences that may persist at high levels of proficiency, which are of theoretical interest in the L2 processing theories described in the introduction. We do not intend to draw conclusions about other bilingual learner groups (e.g., low proficiency L2 learners), who do not fit into this category.

Our research designs, sampling method, and data analysis plans for each experiment were preregistered before data collection. Our preregistrations, materials, raw data and analysis code, along with summaries of our L2 participants' language background information, can be found at the Open Science Framework website (https://osf.io/yrdm7/).

## Experiments 1a and 1b

In Experiment 1a, participants completed an offline task in which they read sentences containing pronouns and gender-matching antecedents in crossover and non-crossover constructions, as in (6a) and (6b), respectively. Participants simply read each sentence and had to decide who the pronoun referred to, with options being either the mentioned antecedent or "someone else."

(6a) Crossover
The boy who he helped yesterday afternoon was very busy.

(6b) Non-Crossover
The boy who claimed that he helped yesterday afternoon was very busy.
Who does "HE" refer to?
1. The boy 2. Someone else

We expected both L1 and L2 readers to pick the sentence internal antecedent more often in non-crossover than crossover constructions. If L2 learners have difficulty applying crossover offline, they may allow sentence internal antecedents in (6a) more often than L1 readers.

In Experiment 1b, participants read sentences as in (7a–d) while their eye movements were monitored. (7a/b) include crossover constructions where coreference between the pronoun "he" and the preceding NP "the man/the lady" is structurally prohibited. In the non-crossover conditions (7c/d), "the man/the lady" is a structurally accessible antecedent for the pronoun. In (7a/c), the pronoun matches in gender with the preceding NP, while in (7b/d), there is a gender mismatch. To ensure pronouns across all experimental conditions ultimately had a structurally accessible and gender-matching antecedent in each sentence, all sentences also included another gender-matching NP later in the sentence ("Tom").

(7a) Crossover, Gender match

It was the man in the office who he phoned yesterday afternoon, but Tom didn't talk very much.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:44:42, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101910

(7b) Crossover, Gender mismatch
It was the lady in the office who he phoned yesterday afternoon, but Tom didn't talk very much.

(7c) Non-crossover, Gender match
It was the man in the office who said that he phoned yesterday afternoon, but Tom didn't talk very much.

(7d) Non-crossover, Gender mismatch
It was the lady in the office who said that he phoned yesterday afternoon, but Tom didn't talk very much.

When the pronoun "he" is encountered during incremental reading in (7a/b), application of the crossover constraint requires that the reader predicts that the prior noun (man/lady), via the relative pronoun "who," is also predicted to be interpreted as the direct object of the upcoming but not yet encountered verb ("phoned"). Application of the crossover constraint at the pronoun is thus only possible if readers make sufficiently detailed structural predictions about what role the previously encountered referent has in the upcoming, but not yet encountered, syntactic structure of the sentence. Note that other continuations are possible. For example, the relative pronoun could ultimately be interpreted as the complement of a preposition (e.g., "It was the man in the office who he was phoned by yesterday afternoon"). However, irrespective of other potential continuations, for crossover to apply at the pronoun, the reader must predict that "who" is to be interpreted at an as yet unencountered position later in the sentence that is c-commanded by the pronoun.

In (7c/d), although the relative pronoun "who" still refers to the prior noun (man/lady), it is no longer an argument of the subsequent verb "phoned." As such, the structural prediction required to respect crossover in (7a/b) is not required in (7c/d), and coreference between the pronoun and prior antecedent is possible. We expected L1 readers to respect crossover during processing (Kush et al., 2017), such that gender mismatch effects should be observed in non-crossover conditions, with longer reading times at or shortly after the pronoun in (7d) than (7c), but not in crossover conditions (7a/b). It is this interaction between crossover and gender mismatch effects that we take as providing evidence of readers making structural predictions during processing.

If L2 learners do not build sufficient syntactic structure to guide anaphora resolution during crossover, as might be predicted under a strong version of the SSH (Clahsen & Felser, 2006), L2 readers may show gender mismatch effects in both non-crossover (7c/d) and crossover constructions (7a/b). If L2 learners build sufficient structure, but fail to predict in (7a/b) that the relative pronoun is also the direct object of the upcoming verb, this would again predict gender mismatch effects in both (7c/d) and (7a/b), but for different reasons. This finding would provide general support for claims about prediction causing difficulty in L2 processing (see Kaan & Grüter, 2021). If, however, L2 readers respect crossover constraints, they should show gender mismatch effects in (7c/d) only, like L1 readers. Thus, the key issue is whether we observe a main effect of gender or a crossover by gender interaction, that may be further qualified by an interaction with group.

### *Participants*

Experiments 1a/b involved 48 L1 English speakers (mean age 21, range 18–49) and 48 L2 English speakers (mean age 25, range 19–47). We recruited L2 speakers from

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:44:42, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101910

various L1 backgrounds from the University of Reading community.[4] The participants initially completed the eye-movement task (Experiment 1b) and then the offline task (Experiment 1a). We followed our preregistration and continued participant recruitment until we had 48 L1 speakers and 48 L2 speakers who correctly answered more than 70% of comprehension questions in the experimental trials in the eye-movement task, as an index that they paid attention during the task. L2 participants also needed to score more than 40 out of 60 on the Quick Placement Test (mean 49, range 40–58), placing them within upper intermediate to advanced proficiency. L2 participants started learning English in school settings in a non-English-speaking country (mean age of onset 9; range 7–15).[5] All participants were tested in the UK in an immersion environment, though they had varying degrees of immersive English experience.

### Materials

Experiment 1a involved 16 sets of sentences accompanied by comprehension questions, as shown in (6a/b). The pronoun and the antecedent always matched in gender. Half of the experimental sentences used the masculine pronoun and half the feminine pronoun. There were also 32 filler sentences, which always required pronoun resolution, as in the experimental trials, but none of which contained the crossover construction. The experimental and filler trials were pseudo-randomly distributed across two lists in a Latin-square design.

Experiment 1b involved 24 sets of experimental sentences, as exemplified in (7a–d). The first clause was always an expletive cleft, followed by a prepositional phrase and a relative clause introduced by "who." Each experimental sentence was shown across two lines, with the line break being inserted after the comma. Half of the experimental sentences had a masculine pronoun and half a feminine pronoun. There were also 72 fillers with a variety of syntactic structures. These fillers were presented across either two or three lines on the screen. A yes/no comprehension question followed all experimental sentences and two-thirds of the fillers. Filler and experimental items were pseudo-randomized across four presentation lists in a Latin-square design, with 96 items in total.

### Procedure

Participants first completed Experiment 1b and then Experiment 1a. In the eye-movement task (Experiment 1b), while viewing was binocular, participants' eye movements were recorded from the right eye only, using an SR Research Eyelink 1000. Before the experiment began, calibration of the eye-tracker was conducted on a nine-point grid. Recalibration was conducted during the experiment when needed. Before each sentence was shown, participants fixated on a black square above the first word of the sentence. Upon fixation, the sentence appeared. After finishing reading each text, participants pressed a button on a game pad to either move to the next trial or answer

[4]Thirteen participants identified either Chinese or Mandarin as their L1. Arabic and Russian were identified as L1s by four participants each, while Cantonese, Thai, Turkish and Greek were each identified as L1s by two participants each. All other L1s were reported by only one participant.

[5]In the preregistration, we specified that L2 participants must have started learning English at age 8, but five participants did not meet this criterion (all at age 7). We included these participants because, despite our best recruitment and outreach efforts, the number of eligible participants fell short of our expectations.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:44:42, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101910

a comprehension question. Participants answered questions by pressing an appropriate button on a gamepad. The experiment began with four practice trials.

Experiment 1a was administered in the Ibex Farm web-based experiment presentation platform (https://github.com/addrummond/ibexfarm). During the task, each sentence and its comprehension question were presented simultaneously on the screen. Participants answered the questions by pressing either the "1" key for the first option (e.g., "The boy") or the "2" key for the second option (i.e., "Someone else"). The experimental and filler trials were pseudo-randomized so that some filler sentences always appeared between each experimental trial. The experiment began with four practice trials.

After Experiments 1a/b, L2 participants completed the Quick Placement Test as a measure of proficiency, and a vocabulary screening test to ensure they knew the gender of the critical nouns in both experiments (e.g., man/lady). L2 participants simply looked through a list of vocabulary containing the critical nouns and were asked to mark any that they did not know the meaning of.

### *Preregistered data analysis*

For Experiment 1a, the outcome variable was antecedent responses (1 = antecedent, 0 = someone else), to which we fit a mixed-effects logistic regression model using the lme4 package (Bates et al., 2015) in R (R Core Team, 2022). The model included sum-coded fixed effects of crossover (crossover/no-crossover coded −1/1, respectively), group (L1/L2 coded −1/1) and their interaction. Random effects included random intercepts for subjects and items, and all possible random slopes for each repeated measures fixed effect. If this maximal model did not converge, we initially removed random effect correlations and then iteratively removed the random slope accounting for the least variance until the model converged (Barr et al., 2013).

For Experiment 1b, we analyzed three measures at the critical region (the pronoun and the preceding word, e.g., "who/that he") and at the spillover region (the subsequent words up to and including the comma, e.g., "phoned yesterday afternoon,"). These measures were first pass times, the summed duration of all fixations within a region until an eye-movement away from the region. Regression path times were calculated by summing fixations within a region when it was first fixated from the left, up to but not including the first fixation in a region to the right. Regression path times thus include regressive fixations to previous regions of text. We also calculated total viewing times, the summed duration of all fixations within a region. Fixations shorter than 80 ms within one degree of visual arc of another fixation were merged, and any other fixations shorter than 80 ms or longer than 800 ms were removed. For first-pass times and regression path times, regions that were initially skipped during reading were treated as missing data. For total viewing times, regions that were not fixated at all were treated as missing data. No trials were removed due to the vocabulary screening test. We fit linear mixed-effects models to log-transformed reading times. Log-transformations were conducted to remove skew. The models included sum-coded fixed effects of crossover (crossover/no-crossover, coded −1/1 respectively), gender (gender match/mismatch, coded −1/1), group (L1/L2, coded −1/1), region (critical/spillover region, coded −1/1) and their interactions. Region was included as a fixed effect in the analysis to examine potential differences in the time-course of effects across the critical and spillover regions. The random effects included random intercepts for subjects and items. As including region as a fixed effect involves inclusion of two non-independent datapoints from the same trial, we also included a random effect of trial, defined as the unique

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:44:42, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101910

pairing of each subject and item (see Cunnings & Sturt, 2018). Random slopes for all repeated measures fixed effects were fit using the maximal model that converged, using the same procedure for non-convergence as in Experiment 1a. For each fixed effect, $p$ values were estimated from the $t$ distribution (Baayen, 2008).

## Results

### *Experiment 1a*

Mean accuracy rates of filler trials were 96% for L1 participants (range = 75–100) and 94% for L2 participants (range = 78–100). Response patterns are shown in Figure 1.

There was a significant main effect of crossover (Estimate = –1.867, SE = 0.17, $z$ = –11.25, $p < .001$), showing that both L1 and L2 readers chose the sentence internal antecedent more often in the non-crossover conditions than in the crossover conditions. There was also a significant group by crossover interaction (Estimate = 0.311, SE = 0.12, $z$ = 2.49, $p$ = .013). A follow-up analysis tested group effects within each level of crossover. This analysis showed no significant differences in the crossover condition (Estimate = 0.079, SE = 0.20, $z$ = 0.39, $p$ = .699), but more sentence internal antecedent responses for L1 than L2 speakers in the non-crossover condition (Estimate = –0.535, SE = 0.20, $z$ = –2.718, $p$ = .007).

### *Experiment 1b*

Average comprehension accuracy of the experimental trials was 92% for L1 participants (range 81–100) and 92% for L2 participants (range 73–100), indicating that participants paid attention during the task. A summary of the statistical analysis is provided in Table 1, and total viewing times at the critical and spillover regions are shown in Figures 2 and 3, respectively.

Main effects of group were found in each measure, with longer reading times in the L2 group. As this main effect is not of primary theoretical interest, we do not discuss it

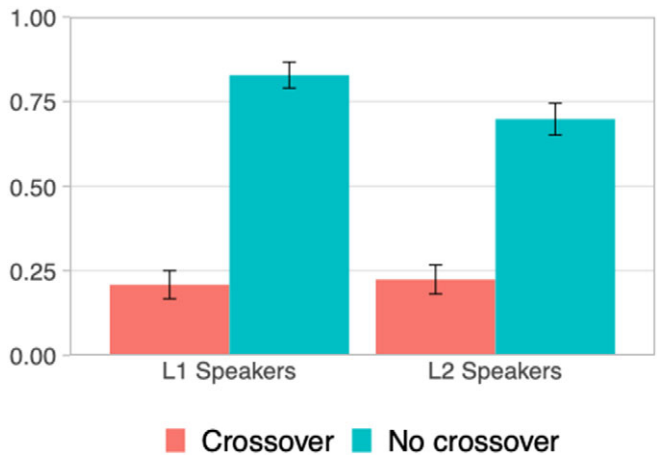



**Figure 1.** Proportion of sentence internal antecedent responses in Experiment 1a (error bars show two standard errors).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:44:42, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101910

**Table 1.** Summary of the statistical analysis for Experiment 1b.

| | First pass times | | | | Regression path times | | | | Total viewing times | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Estimate | SE | *t* | *p* | Estimate | SE | *t* | *p* | Estimate | SE | *t* | *p* |
| Intercept | 5.883 | 0.02 | 265.23 | <.001 | 6.399 | 0.03 | 207.56 | <.001 | 6.819 | 0.04 | 164.86 | <.001 |
| group | 0.114 | 0.02 | 5.73 | <.001 | 0.105 | 0.03 | 3.87 | <.001 | 0.196 | 0.04 | 5.05 | <.001 |
| region | 0.344 | 0.02 | 16.18 | <.001 | 0.775 | 0.03 | 29.86 | <.001 | 0.405 | 0.02 | 17.74 | <.001 |
| crossover | –0.046 | 0.01 | –4.20 | <.001 | –0.022 | 0.01 | –2.21 | .027 | –0.027 | 0.01 | –2.18 | .029 |
| gender | –0.010 | 0.01 | –1.18 | .237 | 0.012 | 0.01 | 1.41 | .160 | 0.004 | 0.01 | 0.36 | .716 |
| group:region | 0.044 | 0.02 | 2.48 | .013 | 0.036 | 0.02 | 1.92 | .055 | 0.037 | 0.01 | 2.59 | .010 |
| group:crossover | –0.004 | 0.01 | –0.35 | .725 | –0.003 | 0.01 | –0.28 | .777 | –0.014 | 0.01 | –1.37 | .171 |
| region:crossover | –0.024 | 0.01 | –1.87 | .061 | –0.013 | 0.01 | –1.04 | .298 | –0.010 | 0.01 | –0.93 | .351 |
| group:gender | 0.012 | 0.01 | 1.36 | .175 | 0.018 | 0.01 | 1.90 | .057 | 0.007 | 0.01 | 0.75 | .451 |
| region:gender | –0.001 | 0.01 | –0.12 | .904 | 0.020 | 0.01 | 1.58 | .114 | –0.007 | 0.01 | –1.26 | .210 |
| crossover:gender | 0.004 | 0.01 | 0.49 | .624 | –0.021 | 0.01 | –2.33 | .020 | –0.037 | 0.01 | –3.53 | <.001 |
| group:region:crossover | –0.002 | 0.01 | –0.18 | .861 | 0.004 | 0.01 | 0.36 | .719 | 0.001 | 0.01 | 0.06 | .949 |
| group:region:gender | –0.003 | 0.01 | –0.32 | .750 | –0.004 | 0.01 | –0.41 | .682 | –0.005 | 0.01 | –0.66 | .508 |
| group:crossover:gender | –0.001 | 0.01 | –0.14 | .892 | –0.002 | 0.01 | –0.27 | .790 | –0.007 | 0.01 | –0.72 | .474 |
| region:crossover:gender | 0.006 | 0.01 | 0.60 | .550 | –0.021 | 0.01 | –2.55 | .011 | 0.003 | 0.01 | 0.38 | .706 |
| group:region:crossover:gender | –0.007 | 0.01 | –0.91 | .361 | –0.001 | 0.01 | –0.15 | .879 | –0.007 | 0.01 | –1.31 | .191 |

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:44:42, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101910

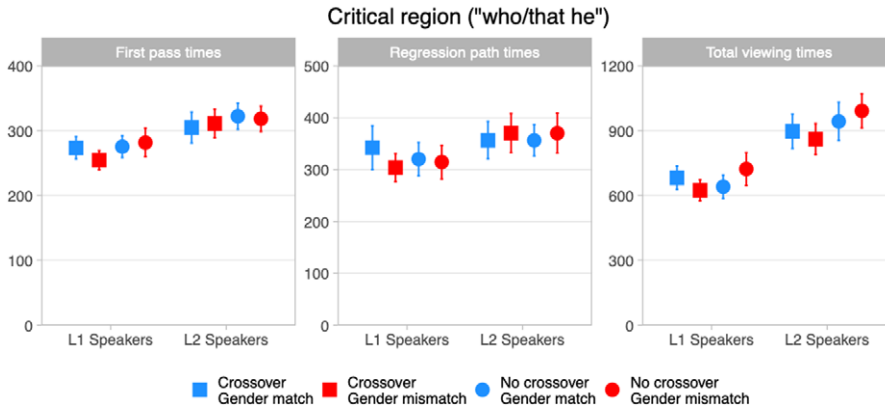



**Figure 2.** Reading times at the critical region in Experiment 1b (error bars show two standard errors).

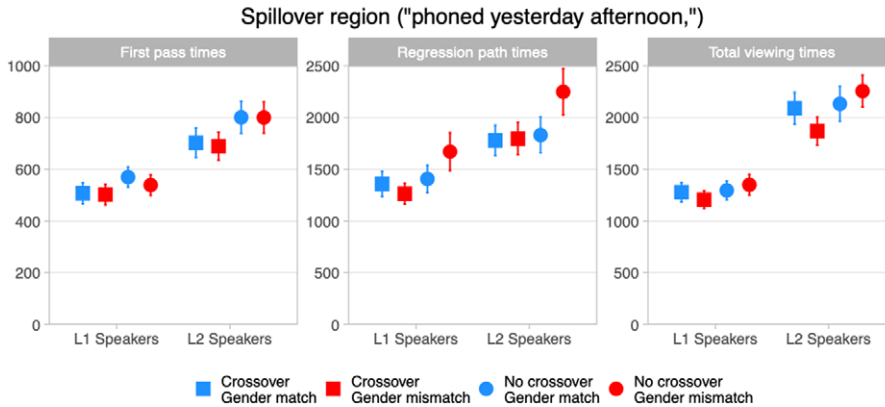



**Figure 3.** Reading times at the spillover region in Experiment 1b (error bars show two standard errors).

further. We also do not discuss below main effects of region, crossover, group by region or region by crossover interactions, as these effects on their own are not of theoretical interest. We instead focus on gender mismatch effects, and their potential interactions with crossover, group and/or region.

First pass times did not show any significant effects of theoretical interest. For regression path times, there was a crossover by gender interaction and a three-way interaction between crossover, gender and region. A follow-up analysis examined gender effects within the crossover and non-crossover conditions for each region. This analysis showed significant gender effects in the non-crossover conditions at the spillover region, with longer reading times in the gender-mismatch than gender-match condition (Estimate = 0.075, SE = 0.02, $t$ = 3.15, $p$ = .002). This effect was not statistically significant in the crossover conditions (Estimate = −0.010, SE = 0.02, $t$ = −0.55, $p$ = .579). Also, the critical region did not show significant gender effects for either crossover (Estimate = –0.008, SE = 0.02, $t$ = –0.46, $p$ = .648) or non-crossover (Estimate = −0.008, SE = 0.02, $t$ = −0.48, $p$ = .630) conditions.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:44:42, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101910

Total viewing times showed a significant crossover by gender interaction, and we examined this interaction by testing gender effects within each level of crossover. This analysis showed significant gender effects in both non-crossover conditions (Estimate = 0.041, SE = 0.02, $t = 2.55$, $p = .011$) and crossover conditions (Estimate = –0.033, SE = 0.01, $t = –2.25$, $p = .025$), with the effects going in different directions. That is, for the non-crossover conditions, reading times were longer for the gender-mismatch than gender-match condition (i.e., gender mismatch effects), while in the crossover conditions, reading times were longer in the gender-match than gender-mismatch conditions (i.e., opposite to gender mismatch effects). We did not find any significant interactions with group of theoretical interest in any measure.

## Discussion

In Experiment 1a, we did find differences between L1 and L2 speakers, but only in the non-crossover conditions, not in the crossover conditions. This suggests that our L1 and L2 participants had similar knowledge of crossover and applied it in a similar way at the offline level. The findings in the non-crossover conditions suggest L2 participants were more willing to consider a sentence external antecedent for the pronoun than L1 participants. Importantly, however, the results from the crossover condition did not indicate that L2 readers were more likely to violate the crossover constraint than L1 readers.

In Experiment 1b, we found gender mismatch effects in the non-crossover conditions in regression path times and total viewing times, suggesting that when L1 and L2 readers encountered the pronoun, they attempted to establish a coreference relation between it and the preceding NP. Crucially, there were no gender mismatch effects in the crossover conditions for either L1 or L2 speakers. Instead, total viewing times were longer in the gender-match than gender-mismatch conditions. These findings in the crossover condition are not what we expected to observe. However, the contrasting effects observed in non-crossover and crossover conditions indicate that the pronoun was processed differently in these two constructions. In the non-crossover conditions, we contend that readers freely attempted to corefer the pronoun with the preceding NP, leading to gender mismatch effects. In the crossover conditions, we argue that L1 and L2 readers considered the NP as a potential antecedent for the pronoun in the gender-matching conditions, in at least some proportion of trials, but rejected this interpretation due to the crossover constraint, leading to longer reading times in the gender match than mismatch condition. We discuss this issue in more detail in the general discussion.

In sum, the results from Experiments 1a/b suggest that L1 and L2 readers build sufficiently detailed syntactic structure for the crossover constraint to apply during online processing, contra what might be expected under a strong version of the claim that L2 learners adopt shallow syntactic processing (Clahsen & Felser, 2006). Experiments 1a/b, however, do not test whether both facilitatory and inhibitory interference during memory retrieval influence how crossover is applied, which is crucial for testing the interference account of L2 processing (Cunnings, 2017). Experiment 2 was conducted for this purpose.

## Experiment 2

In Experiment 2, participants' eye-movements were monitored in sentences like (8a–d). These sentences all contained crossover constructions. In (8a–d), the only structurally accessible antecedent for the pronoun ("he") is the sentence subject ("John/Lisa"),

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:44:42, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101910

which matches this pronoun in gender (8a/b) but mismatches in (8c/d). There is also a distractor in the second clause ("the man/the woman"), which is linearly closer to the pronoun but is a structurally inaccessible antecedent for it due to crossover.

(8a) Subject gender match, Distractor gender match
John said that it was the man in the office who he phoned yesterday afternoon after the meeting.

(8b) Subject gender match, Distractor gender mismatch
John said that it was the woman in the office who he phoned yesterday afternoon after the meeting.

(8c) Subject gender mismatch, Distractor gender match
Lisa said that it was the man in the office who he phoned yesterday afternoon after the meeting.

(8d) Subject gender mismatch, Distractor gender mismatch
Lisa said that it was the woman in the office who he phoned yesterday afternoon after the meeting.

We expected L1 readers to have longer reading times for the subject gender mismatch conditions (8c/d) than for subject gender match conditions (8a/b). If L1 anaphora resolution resists interference (e.g., Cunnings et al., 2015; Dillon et al., 2013), the distractor should not affect reading times. Cue-based models, however, predict that the distractor should cause interference (Jäger et al., 2020). Specifically in the subject gender mismatch conditions (8c/d), these models predict facilitatory interference. That is, reading times should be shorter when the distractor matches the pronoun in gender (8c) compared to when it mismatches (8d). In the subject gender match conditions, inhibitory interference is predicted, such that reading times should be longer in (8a), when both the preceding NPs match the gender of the pronoun, compared to (8b), when the distractor mismatches.

Following Experiment 1b, we expected L2 readers to also show subject gender mismatch effects, with longer reading times in (8c/d) than (8a/b). If L2 learners are more susceptible to interference than L1 readers (Cunnings, 2017), L2 readers should show inhibitory and facilitatory interference effects of a greater magnitude than L1 readers.

### *Participants*

Experiment 2 included 80 L1 English speakers and 80 L2 English speakers from various L1 backgrounds from the University of Reading community.[6,7] We continued participant recruitment until we had 80 L1 speakers and 80 L2 speakers who correctly

[6]Given that facilitatory interference effects are typically smaller in magnitude than gender mismatch effects, as by definition facilitatory interference is an attenuation of gender mismatch effects in this case, we increased the sample size in Experiment 2 compared to Experiment 1b.

[7]L1s included Arabic (n = 20); Italian or Spanish (both n = 11); Polish, Greek or Portuguese (each n = 6); Bulgarian, French or Russian (each n = 5). Other L1s were spoken by only one participant. Note this adds up to more than 80 participants in total, as some participants reported more than one L1.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:44:42, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101910

answered more than 70% of comprehension questions. L2 participants had an average English age of onset of 8 (range 5–18) and needed to score at least 30 on the Quick Placement Test (mean 51, range 33–60), placing them within intermediate to advanced proficiency. All L2 participants were tested in the UK but had been in an immersive environment for varying lengths.

### Materials

We prepared 24 sets of experimental sentences like those in (8a–d). There were 72 filler sentences of various syntactic structures, and a yes/no comprehension question followed all experimental and filler trials. All sentences appeared on one line onscreen. Experimental and filler items were pseudo-randomized across four lists in a Latin-square design.

### Procedure and preregistered data analysis

The procedure and data analysis were the same as in Experiment 1b, except that fixed effects were subject gender (match/mismatch, coded −1/1), distractor gender (match/mismatch, coded −1/1), group (L1/L2, coded −1/1), and region (critical/spillover region, coded −1/1). The critical region was defined as the pronoun and preceding relative pronoun ("who he"), while the spillover region contained the rest of the sentence, up to but not including the final three words ("phoned yesterday afternoon").

## Results

Average comprehension accuracy of experimental trials was 88% for L1 participants (range 73–98) and 86% for L2 participants (range 72–98). A summary of the statistical analysis is shown in Table 2, and reading times at the critical and spillover regions are visualized in Figures 4 and 5, respectively.

All measures showed significantly slower reading times for L2 readers than L1 readers. First-pass times also showed a significant main effect of subject gender, with longer reading times in the gender-match than gender-mismatch conditions. Although this effect was in the opposite direction to what was predicted, shorter first-pass times can sometimes be observed in difficult conditions that are more likely to trigger a regression (see, e.g., Sturt, 2007).

Regression path times also showed a significant main effect of subject gender, with effects in the predicted direction, such that reading times were longer in gender mismatch than gender match conditions. That we observed clear gender mismatch effects in regression path times is consistent with the idea that the shorter first pass reading times for gender mismatching subjects indicate participants were more likely to make a regression in these conditions. There was also a significant interaction between this effect and region. A follow-up analysis tested effects of subject gender within each level of region. This analysis showed a significant effect of subject gender at the spillover region (Estimate = 0.089, SE = 0.01, $t$ = 8.97, $p$ < .001), but not at the critical region (Estimate = –0.002, SE = 0.01, $t$ = –0.19, $p$ = .848).

For total viewing times, there was a significant main effect of subject gender and a significant region by subject gender interaction. Subject gender mismatch effects were found at both regions, but were larger at the critical region (Estimate = 0.117, SE = 0.01, $t$ = 11.82, $p$ < .001) than the spillover region (Estimate = 0.095, SE = 0.01, $t$ = 10.77,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:44:42, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101910

**Table 2.** Summary of the statistical analysis for Experiment 2.

| | First pass times | | | | Regression path times | | | | Total viewing times | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Estimate | SE | *t* | *p* | Estimate | SE | *t* | *p* | Estimate | SE | *t* | *p* |
| Intercept | 5.897 | 0.02 | 297.88 | <.001 | 6.318 | 0.03 | 236.67 | <.001 | 6.974 | 0.03 | 204.05 | <.001 |
| group | 0.107 | 0.02 | 5.88 | <.001 | 0.080 | 0.02 | 3.51 | <.001 | 0.145 | 0.03 | 4.90 | <.001 |
| region | 0.281 | 0.02 | 14.53 | <.001 | 0.579 | 0.02 | 28.03 | <.001 | 0.455 | 0.02 | 21.56 | <.001 |
| subject | –0.017 | 0.01 | –2.32 | .021 | 0.044 | 0.01 | 6.26 | <.001 | 0.106 | 0.01 | 12.74 | <.001 |
| distractor | 0.008 | 0.01 | 1.15 | .252 | 0.010 | 0.01 | 1.48 | .138 | –0.017 | 0.01 | –2.23 | .026 |
| group:region | 0.048 | 0.01 | 3.42 | <.001 | 0.023 | 0.01 | 1.83 | .068 | 0.007 | 0.01 | 0.64 | .520 |
| group:subject | 0.000 | 0.01 | –0.03 | .979 | 0.003 | 0.01 | 0.47 | .640 | 0.011 | 0.01 | 1.11 | .266 |
| region:subject | –0.004 | 0.01 | –0.60 | .546 | 0.046 | 0.01 | 6.45 | <.001 | –0.011 | 0.00 | –2.39 | .017 |
| group:distractor | –0.004 | 0.01 | –0.62 | .537 | 0.004 | 0.01 | 0.61 | .541 | 0.005 | 0.01 | 0.54 | .588 |
| region:distractor | –0.007 | 0.01 | –1.05 | .295 | –0.001 | 0.01 | –0.13 | .896 | –0.002 | 0.00 | –0.42 | .674 |
| subject:distractor | –0.010 | 0.01 | –1.39 | .166 | –0.006 | 0.01 | –0.91 | .364 | 0.006 | 0.01 | 0.84 | .398 |
| group:region:subject | 0.005 | 0.01 | 0.74 | .458 | 0.003 | 0.01 | 0.33 | .744 | 0.001 | 0.00 | 0.23 | .817 |
| group:region:distractor | –0.006 | 0.01 | –0.95 | .344 | –0.002 | 0.01 | –0.30 | .763 | –0.001 | 0.00 | –0.30 | .766 |
| group: subject:distractor | –0.009 | 0.01 | –1.32 | .187 | 0.004 | 0.01 | 0.61 | .544 | 0.007 | 0.01 | 0.86 | .389 |
| region: subject:distractor | –0.005 | 0.01 | –0.73 | .466 | 0.009 | 0.01 | 1.19 | .233 | –0.005 | 0.00 | –1.10 | .272 |
| group:region: subject:distractor | –0.008 | 0.01 | –1.29 | .198 | 0.003 | 0.01 | 0.49 | .625 | –0.005 | 0.01 | –0.97 | .334 |

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:44:42, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101910

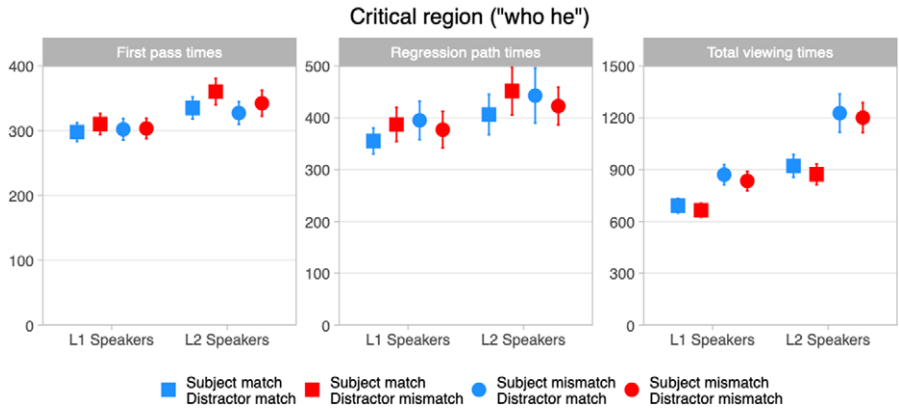



**Figure 4.** Reading times at the critical region in Experiment 2 (error bars show two standard errors).

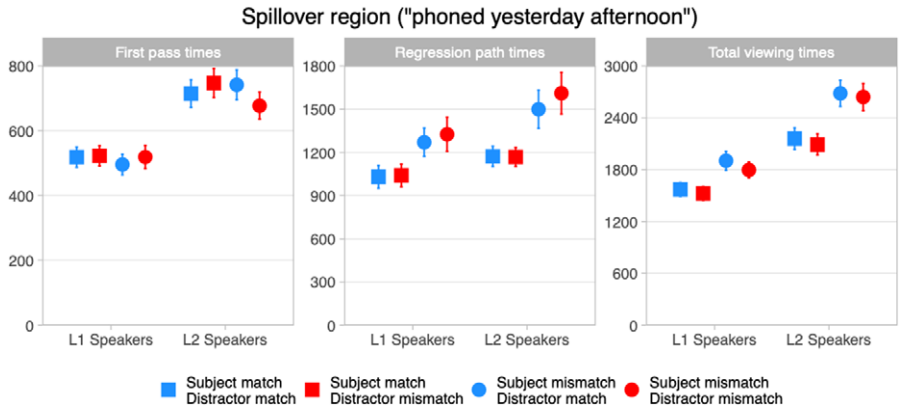



**Figure 5.** Reading times at the spillover region in Experiment 2 (error bars show two standard errors).

$p < .001$). There was also a significant main effect of distractor gender, with longer reading times in the gender-match than gender-mismatch conditions, in the absence of an interaction with subject gender. No significant interactions with group of theoretical interest were found in any measure.

## Discussion

The results showed gender mismatch effects of the sentence subject, suggesting that L1 and L2 readers attempted to resolve the pronoun after encountering it by retrieving its structurally accessible antecedent. In total viewing times, there was evidence of distractor effects, with increased reading times when the distractor matched the pronoun in gender. In subject gender match conditions, observing distractor gender match effects is consistent with the predictions of cue-based memory retrieval (Jäger et al., 2020; Lewis et al., 2006; Lewis & Vasishth, 2005), which predicts inhibitory interference

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:44:42, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101910

here. However, the direction of the effect in subject gender mismatch conditions is not predicted by this model, which would predict facilitatory interference (shorter reading times for distractor match than mismatch conditions), which is opposite to what we found. We discuss this finding in detail below. Most importantly, we did not find any significant evidence that L2 readers were more susceptible to interference than L1 readers. Additionally, both groups showed interference effects in the same direction, and as such we did not find any significant evidence, nor any trends, in our data of qualitatively different interference patterns between the two groups. These findings are consistent with Experiment 1b and suggest that memory retrieval processes are similar between L1 and L2 readers in crossover constructions.

## General discussion

We conducted three experiments to investigate structure building, structural prediction, and memory retrieval in L1 and L2 processing, using crossover constructions. Experiment 1a indicated that L2 readers had knowledge of crossover like L1 readers. Experiments 1b and 2 showed that both L1 and L2 readers construct appropriate syntactic structures and make appropriate structural predictions to constrain anaphora resolution in crossover constructions during processing. Experiment 2 also showed that both L1 and L2 processing are susceptible to interference from the distractor, although the precise pattern of results was not fully consistent with cue-based models. Below, we discuss the implications of these findings.

### *Structure building, prediction, and memory retrieval in L2 sentence processing*

Our study found similar processing patterns between L1 and L2 readers. The fact that we observed gender mismatch effects in non-crossover but not crossover conditions in Experiment 1b and gender mismatch effects for the subject noun in Experiment 2 indicates that crossover guided anaphora resolution for both L1 and L2 readers. These findings are compatible with Felser and Drummer (2017), but provide a stronger test of how crossover affects L2 sentence processing. One difference between Felser and Drummer and our study is that while Felser and Drummer showed similar processing times between gender-match and gender-mismatch conditions in crossover constructions, our Experiment 1b observed longer total viewing times in the gender-match than gender-mismatch crossover conditions. As we discussed earlier, this might indicate consideration of the inaccessible antecedent in crossover conditions when it matched in gender, with readers experiencing processing difficulty as they found coreference here infelicitous because of the crossover constraint (for a similar interpretation of gender match effects indicating consideration and retraction of potential coreference, see Kazanina & Phillips, 2010). Whilst this might suggest temporary violation of the crossover constraint during processing, on at least some proportion of trials, we emphasize for present purposes that we did not find any significant differences in the size of this effect between L1 and L2 readers. In particular, we didn't find significant evidence that L2 readers were more likely to violate crossover than L1 speakers. However, the question arises as to why different processing patterns emerged between our study, in both L1 and L2 readers, and Felser and Drummer.

Although we are cautious in drawing conclusions across studies, one difference is that while our crossover conditions tested cleft constructions containing a relative pronoun, Felser and Drummer tested wh-phrases in prepositional phrases. It might be

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:44:42, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101910

that subtle differences in discourse salience between these two constructions might explain this discrepancy. In our Experiment 1b, for example, the structurally inaccessible antecedent in crossover constructions was put into a discourse-prominent focus position, while in Felser and Drummer's study the inaccessible antecedent was likely less prominent as it appeared inside a prepositional phrase.[8] Note also that in our Experiment 1b, we tested more participants (n = 96) than in Felser and Drummer (n = 66), and so differences in power may also explain the discrepancy across studies. We are thus cautious in drawing strong conclusions here, but future research is required to examine how different constructions may influence how crossover constraints are applied during processing.

Our findings that L1 and L2 readers showed similar sensitivity to crossover constraints during processing, with no significant differences in time-course or interference, have important implications for accounts of L2 processing in terms of shallow parsing (Clahsen & Felser, 2006, 2018), interference during memory retrieval (Cunnings, 2017), and prediction (Grüter et al., 2014). In terms of the SSH (Clahsen & Felser, 2006, 2018), our results suggest L2 speakers are able to build sufficient syntactic structure during online processing to restrict pronoun resolution in a similar way to L1 speakers. With regards to interference (Cunnings, 2017), we found no evidence that L2 readers are more susceptible to interference during the memory retrieval operations that subserve pronoun resolution than L1 readers. Whilst our results do not rule out the possibility that L2 learners parse certain complex structures in a "shallow" manner, or that L2 learners may be particularly susceptible to interference in certain situations, our results restrict the extent to which this is the case.

A factor relevant to both the SSH and Cunnings' interference account is discourse salience. In our Experiment 2, the distractor may not have been discourse salient enough to influence processing more in L2 learners than L1 speakers. In the two previous studies that reported that L2 speakers are more likely to violate binding constraints on anaphor resolution than L1 speakers (Felser et al., 2009; Felser & Cunnings, 2012), the distractor was placed in a discourse-salient topic position (for discussion, see Alaskar & Cunnings, 2025). It should be noted, however, that the distractor in our two eye-movement experiments was in a focused, cleft construction that should make it salient. Nevertheless, it might be that discourse salience, rather than focus, is important in considering potential L1/L2 differences. How discourse salience may interact with crossover during anaphora resolution in L1 and L2 readers is an avenue for future research.

In terms of prediction, our study focuses on predictive structure building. Our results suggest L2 speakers can make such structural predictions in a similar way to L1 speakers. However, some previous studies showing L1/L2 differences in prediction tested prediction based on other types of information (e.g., lexical information such as case, gender, phonological form, etc.), and the extent to which L2 prediction is affected by other factors, such as crosslinguistic influence or methodological issues, remains controversial (Kaan & Gruter, 2021; Schlenter, 2023). Our study thus suggests that structural predictions may be similar during L1 and L2 processing, but we leave open

[8]As noted by a reviewer, in our crossover conditions ("It was the man in the office who he phoned …") there is a potential filled-gap effect, such that "who" may initially be interpreted as a subject, which is then overtly filled by "he." This filled-gap-effect is absent from the stimuli tested by Felser and Drummer (2017). Whilst this is another difference between studies that may contribute to the conflicting results, we are cautious in drawing strong conclusions here until direct comparisons between constructions are made.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:44:42, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101910

the question of whether prediction more generally in L1 and L2 speakers diverges with respect to the use of other types of linguistic information.

We also note that our results are specific to proficient L2 learners tested in an immersion environment. As we tested L2 learners from various L1 backgrounds, this also leaves open the possibility that an L2 learner's L1 may influence L2 processing of crossover. Testing how factors such as L2 proficiency and L1 background may influence L2 sensitivity to crossover would be a fruitful avenue for future research. Regardless, our study suggests that proficient L2 speakers are able to predictively build sufficient syntactic structures to constrain memory retrieval in crossover constructions in a similar way to L1 speakers.

A final point we consider in relation to how predictability influences language processing is the distinction between prediction and integration. That is, predictable material may be predicted *before* it is reached, or alternatively predictable material may be more easily integrated with the preceding context, once it is encountered (for discussion, see Staub, 2015; Wong et al., 2025). Finding evidence of an experimental manipulation *before* the predicted element is encountered provides key evidence of prediction, whilst similar evidence at the predictable element itself may suggest prediction, but can also be explained in terms of integration.

In our experiments, we argue that evidence of crossover effects at the pronoun itself is indicative of structural prediction. For example, in our crossover conditions in Experiment 1b, repeated in (9), finding evidence of the crossover constraint influencing processing at *he* would indicate that the constraint is enforced before the verb (*phoned*), which provides bottom-up confirmation that the previously encountered referent (*man*/*woman*) occupies its object position. By contrast, effects at the verb itself (*phoned*) could suggest either prediction or integration.

(9) It was the man/woman in the office who he phoned yesterday afternoon …

In Experiments 1b and 2, we found significant effects consistent with crossover being considered in total viewing times at the critical pronoun region, and also in reading times at the spillover region that contained the verb. Effects at the spillover region could thus be interpreted in terms of either prediction or integration. Effects at the pronoun region might be taken as evidence of structural prediction, though we note that as total viewing times include regressive fixations from later regions of text, we cannot rule out the possibility that the effects here in part reflect processing difficulty incurred after the verb was encountered. As such, though we interpret our results in terms of structural prediction, we acknowledge further research clearly demonstrating crossover effects *before* the critical verb is encountered is required to fully tease apart a structural prediction versus integration account of our findings.

### *Cue-based retrieval in L1 and L2 sentence processing*

In Experiment 1b, both groups showed longer total reading times in crossover conditions when the pronoun's antecedent matched in gender compared to when it mismatched. In Experiment 2, we also observed longer total reading times when the distractor matched the gender of the pronoun compared to when it mismatched. Importantly, in both cases, both groups showed interference effects in the same direction. As such, we did not find any significant evidence, nor trends, in our data of qualitatively different interference patterns between the two groups.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:44:42, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101910

The interference patterns we observed, especially in Experiment 2, suggest that anaphora resolution is susceptible to interference. However, the direction of these effects was not fully consistent with cue-based models of sentence processing. As discussed previously, for subject-match conditions ("John said that it was the man/woman in the office who he phoned …") cue-based models predict increased difficulty (inhibitory interference) for the distractor-match than distractor-mismatch condition, while for the subject-mismatch conditions ("Lisa said that it was the man/woman in the office who he phoned …") it predicts reduced difficulty (facilitatory interference) in the distractor-match than distractor-mismatch condition. The interference patterns observed in Experiment 2 are consistent with the predicted inhibitory interference in subject-match conditions, but inconsistent with the predicted facilitatory interference in subject-mismatch conditions that has been observed in several studies (for review, see Jäger et al., 2017).

We argue that the unexpected interference patterns observed in Experiment 2 may be accounted for by what we discussed for Experiment 1b. That is, readers may have attempted but retracted coreference between the pronoun and gender-matching, inaccessible antecedent. From this perspective, readers are not simply ignoring the crossover constraint, as might be the case on at least some proportion of trials if we observed facilitatory interference in the subject mismatch conditions, but are instead coercing a potential coreference relation they consider to be infelicitous. Future research is required to clarify this possibility, but for present purposes, our most important finding is that although the observed pattern of interference was not fully compatible with cue-based models, our results suggest similar memory retrieval processes between L1 and L2 processing.

## Conclusion

Across three experiments, we investigated the processing and interpretation of crossover constraints in L1 and L2 readers. Our results suggest that both groups are similarly sensitive to crossover constraints during real-time sentence processing. There was also evidence for interference, but both L1 and L2 groups showed similar interference effects. The roles of shallow structure, memory retrieval, and prediction have been influential in describing potential differences between L1 speakers and L2 learners during sentence processing. However, our results suggest that both L1 and (proficient) L2 readers predictively build syntactic structure and engage in memory retrieval in similar structure-sensitive ways during online sentence processing.

**Data availability statement.** Open data, materials, analysis code and links to our pre-registrations are available at https://osf.io/yrdm7

**Acknowledgments.** We thank the editor and reviewers for their comments on earlier drafts of this paper. The research reported here was supported by a Leverhulme Trust Research Project Grant (RPG-2018-337) awarded to Ian Cunnings. Hiroki Fujita was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project number 500540359. Shatha Alaskar extends her appreciation to the Deanship of Postgraduate Studies and Scientific Research at Majmaah University for funding this research work through project number (R-2026-269).

**Competing interests.** The authors declare no competing interests.

## References

Alaskar, S., & Cunnings, I. (2025). Agreement and reflexives in non-native sentence processing. *Bilingualism: Language and Cognition*, *28*(4), 897–910. Cambridge Core. https://doi.org/10.1017/S136672892400049X

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:44:42, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101910

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press; Cambridge Core. https://doi.org/10.1017/CBO9780511801686

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

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Chomsky, N. (1981). *Lectures on government and binding: The Pisa lectures* (7th ed.). Mouton de Gruyter.

Chow, W.-Y., Lewis, S., & Phillips, C. (2014). Immediate sensitivity to structural constraints in pronoun resolution. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00630

Clahsen, H., & Felser, C. (2006). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, *27*(1), 107–126. https://doi.org/10.1017/S0142716406060206

Clahsen, H., & Felser, C. (2018). Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition*, *40*(3), 693–706. https://doi.org/10.1017/S0272263117000250

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., Patterson, C., & Felser, C. (2015). Structural constraints on pronoun binding and coreference: Evidence from eye movements during reading. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00840

Cunnings, I. & Sturt, P. (2018). Retrieval interference and semantic interpretation. *Journal of Memory and Language*, *102*, 16–27. https://doi.org/10.1016/j.jml.2018.05.001

Cunnings, I. (2022). 'Working memory and L2 sentence processing' in Schwieter, J. & Wen, Z. (eds.), *The Cambridge Handbook of Working Memory and Language*, Cambridge: Cambridge University Press, 593–612. https://doi.org/10.1017/9781108955638.033

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, *69*(2), 85–103. https://doi.org/10.1016/j.jml.2013.04.003

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

Felser, C., & Drummer, J.-D. (2017). Sensitivity to crossover constraints during native and non-native pronoun resolution. *Journal of Psycholinguistic Research*, *46*(3), 771–789. https://doi.org/10.1007/s10936-016-9465-8

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

Fujita, H. (2025). Formal computational modelling in second language sentence processing research. *Studies in Second Language Acquisition*. https://doi.org/10.1017/S0272263125101009

Fujita, H., & Yoshida, M. (2024). Online reflexive resolution and interference. *Language, Cognition and Neuroscience*, *39*(4), 513–526. https://doi.org/10.1080/23273798.2024.2329269

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

Hopp, H. (2022). Second language sentence processing. *Annual Review of Linguistics*, *8*(1), 235–256. https://doi.org/10.1146/annurev-linguistics-030821-054113

Jäger, L. A., Mertzen, D., Van Dyke, J. A., & Vasishth, S. (2020). Interference patterns in subject-verb agreement and reflexives revisited: A large-sample study. *Journal of Memory and Language*, *111*, 104063. https://doi.org/10.1016/j.jml.2019.104063

Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, *94*, 316–339. https://doi.org/10.1016/j.jml.2017.01.004

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism*, *4*(2), 257–282. https://doi.org/10.1075/lab.4.2.05kaa

Kaan, E., & Grüter, T. (Eds.). (2021). *Prediction in second language processing and learning* (Vol. *12*). John Benjamins Publishing Company. https://doi.org/10.1075/bpa.12

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:44:42, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101910

Kazanina, N., & Phillips, C. (2010). Differential effects of constraints in the processing of Russian cataphora. *Quarterly Journal of Experimental Psychology*, *63*(2), 371–400. https://doi.org/10.1080/17470210902974120

Kush, D., Lidz, J., & Phillips, C. (2017). Looking forwards and backwards: The real-time processing of Strong and Weak Crossover. *Glossa: A Journal of General Linguistics*, *2*(1). https://doi.org/10.5334/gjgl.280

Lasnik, H., & Funakoshi, K. (2017). Condition C Violations and Strong Crossover. In *The Wiley Blackwell companion to syntax, Second Edition* (pp. 1–27). https://doi.org/10.1002/9781118358733.wbsyncom021

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375–419. https://doi.org/10.1207/s15516709cog0000_25

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454. https://doi.org/10.1016/j.tics.2006.08.007

Parker, D., Shvartsman, M., & Van Dyke, J. A. (2017). The cue-based retrieval theory of sentence comprehension: New findings and new challenges. In L. Escobar, V. Torrens, & T. Parodi (Eds.), *Language processing and disorders* (pp. 121–144). Cambridge Scholars Publishing.

Postal, P. M. (1971). *Cross-over phenomena*. Holt, Rinehart and Winston.

R Core Team. (2022). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Reinhart, T. (1976). *The syntactic domain of anaphora.* PhD Thesis, Massachusetts Institute of Technology.

Schlenter, J. (2023). Prediction in bilingual sentence processing: How prediction differs in a later learned language from a first language. *Bilingualism: Language and Cognition*, *26*(2), 253–267. Cambridge Core. https://doi.org/10.1017/S1366728922000736

Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass*, *9*, 311–327. https://doi.org/10.1111/lnc3.12151

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

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

Wasow, T. (1972). *Anaphoric Relations in English.* PhD Thesis, Massachusetts Institute of Technology.

Wong, R., Reichle, E. & Veldre, A. (2025). Prediction in reading: A review of predictability effects, their theoretical implications, and beyond. *Psychonomic Bulletin & Review*, *32*, 973–1006. https://doi.org/10.3758/s13423-024-02588-z

**Cite this article:** Cunnings, I., Alaskar, S., & Fujita, H. (2026). Structure building, memory retrieval, and structural prediction in second language sentence comprehension. *Studies in Second Language Acquisition,* 1–23. https://doi.org/10.1017/S0272263126101910

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:44:42, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263126101910