Interference and filler-gap dependency formation in native and non-native language comprehension

Journal of Experimental Psychology: Learning, Memory, and Cognition

Interference and Filler-Gap Dependency Formation in Native and Non-Native Language Comprehension

Hiroki Fujita and Ian Cunnings

Online First Publication, April 14, 2022. http://dx.doi.org/10.1037/xlm0001134

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: Learning, Memory, and Cognition

ISSN: 0278-7393

https://doi.org/10.1037/xlm0001134

Interference and Filler-Gap Dependency Formation in Native and Non-Native Language Comprehension

Hiroki Fujita and Ian Cunnings
School of Psychology and Clinical Language Sciences, University of Reading

The mechanisms underlying native (L1) and non-native (L2) sentence processing have been widely debated. One account of potential L1/L2 differences is that L2 sentence processing underuses syntactic information and relies heavily on semantic and surface cues. Recently, an alternative account has been proposed, which argues that the source of L1/L2 differences lies in how susceptible L1 and L2 speakers are to interference during memory retrieval operations. The present study tested these two accounts by investigating filler-gap dependency formation and susceptibility to similarity-based interference in L1 and L2 language comprehension. The results demonstrated that L1 and L2 speakers recover the information of the filler upon encountering a gap and are susceptible to similarity-based interference during filler-gap dependency formation. However, there was no significant evidence of L1/L2 differences. These findings suggest that L1 and L2 speakers similarly engage in cue-based memory retrieval operations during filler-gap dependency formation.

Keywords: filler-gap dependencies, interference, memory retrieval, second language processing

A central question in sentence processing research is how readers form linguistic dependencies. Some previous studies have argued that a cue-based memory retrieval mechanism plays a role in dependency formation (Lewis et al., 2006; Lewis & Vasishth, 2005; Van Dyke & Lewis, 2003). For example, consider sentences (1a/b) below, which contain a so-called filler-gap dependency (Fodor, 1978).

(1a) Mary saw the beer that the man with the wine drank __.

(1b) Mary saw the beer that the man with the food drank __.

In (1a/b), "the beer" is the object of the embedded clause verb "drank" but preposed to the front of the embedded clause. The present study refers to such a moved constituent as a filler and its original position as a gap. Successful comprehension of (1a/b) requires that when encountering the gap at "drank," readers recover the information of the filler. Cue-based memory retrieval accounts claim that readers retrieve the filler from memory using a set of retrieval cues such as [direct object] and [drinkable] that are available at the gap. A consequence is that similarity-based interference effects are predicted to arise during dependency formation. For example, (1a/b) contain another noun phrase, which matches the [drinkable] cue in (1a; "the wine") but mismatches in (1b; "the food"). Cue-based memory retrieval models predict that noun phrases that partially match a set of retrieval cues, such as "the wine" in (1a), interfere with dependency resolution during language comprehension.

Another key issue in sentence processing research is how native (L1) and non-native (L2) speakers process sentences. Filler-gap sentences such as (1a/b) provide an interesting test case about potential differences between L1 and L2 sentence processing. Although L2 processing is likely influenced by many factors (e.g., proficiency, age of acquisition, L1 background, amount and type of L2 exposure, among many others), a key debate in L2 sentence processing research has been the extent and nature of potential L1/L2 differences that may remain between L1 speakers and L2 speakers of even advanced proficiency (e.g., Clahsen & Felser, 2006a, 2006b, 2006c; Cunnings, 2017; Hopp, 2006, 2014; Jiang, 2004; McDonald, 2006). For example, in their Shallow Structure Hypothesis, Clahsen and Felser (2006c) argued that L2 speakers underuse syntactic information and rely heavily on semantic and surface cues (e.g., linear order) during nonlocal dependency formation. One potential reflex of shallow processing that we consider here is that L2 speakers may attempt to retrieve the intervening noun phrase "the wine/food" as a filler because "the wine/food" is linearly closer to the gap than the correct filler (see also Tanner et al., 2012 for discussion of a similar prediction based on linear order in subject-verb agreement). In this case, L2 speakers should encounter reading difficulty in (1b) owing to the implausibility of "the food." By contrast, Cunnings (2017) recently proposed that L1 and L2 speakers similarly parse filler-

Hiroki Fujita https://orcid.org/0000-0001-7649-9707
Ian Cunnings https://orcid.org/0000-0002-5318-0186
The research of this article was funded by a Leverhulme Trust Research Project Grant (RPG-2018-337) awarded to Ian Cunnings.

Data, analysis code, and experimental materials can be found at https://osf.io/vh4sm/.

Correspondence concerning this article should be addressed to Hiroki Fujita, School of Psychology and Clinical Language Sciences, University of Reading, Whiteknights Campus, Earley Gate, Reading, RG6 7BE, United Kingdom. Email: hiroki.fujita@reading.ac.uk

gap sentences, and the cause of L1/L2 differences is L2 speakers' increased susceptibility to interference effects that may arise during retrieval of the filler. Thus, Cunnings's account predicts that the size of interference effects predicted by cue-based memory retrieval models is larger for L2 speakers than L1 speakers.

The present study reports three preregistered experiments, which aimed to test these two accounts by investigating filler-gap dependency formation and susceptibility to similarity-based interference in L1 and L2 language comprehension. These experiments demonstrated that L1 and L2 speakers process filler-gap sentences and are susceptible to interference in a similar way. Before reporting the results in detail, we initially discuss previous studies on filler-gap dependency formation and interference effects, followed by a discussion on L1/L2 differences in sentence processing.

Filler-Gap Dependencies and Interference Effects

Filler-gap dependency formation requires that upon encountering a gap, readers recover the information of the filler from memory (Fodor, 1978; McElree & Bever, 1989). Many previous studies have investigated the mechanisms underlying such a gap-filling operation. These studies have demonstrated that because of the incremental nature of sentence processing (e.g., Crocker, 1996; Fujita, 2021; Kimball, 1973; Yoshida et al., 2013), readers either recover the information of the filler upon encountering a gap or predictively posit a gap (Aoshima et al., 2004; Cunnings & Sturt, 2018; Frazier & Clifton, 1989; Frazier & Flores d'Arcais, 1989; Kim et al., 2020; Lee, 2004; Nakano et al., 2002; Nicol & Swinney, 1989; Omaki et al., 2015; Parker, 2017; Phillips, 2006; Pickering & Guy, 1991; Pickering & Traxler, 2003; Stowe, 1986; Traxler & Pickering, 1996; Wagers & Phillips, 2014). For example, Traxler and Pickering (1996) tested sentences (2a/b) below.

(2a) We like **the book** that the author **wrote** unceasingly and with great dedication **about** while waiting for a contract.

(2b) We like **the city** that the author **wrote** unceasingly and with great dedication **about** while waiting for a contract.

(2a/b) contain a filler ("the book/city") in the main clause. The correct gap position appears immediately after the preposition "about," but readers can temporarily construe the filler as an object of the embedded clause verb "wrote". (2a/b) manipulate whether the filler is a plausible object of the verb (2a: "wrote the book") or an implausible object (2b: "wrote the city"). Traxler and Pickering predicted that if readers recover the information of the filler upon encountering a gap, (2b) should cause reading disruption at the verb due to the implausibility of the filler as an object of "wrote" (implausibility effects). Consistent with this prediction, Traxler and Pickering observed longer reading times at the gap in (2b) than (2a).

How do readers engage in gap-filling operations? Some studies have argued that a cue-based memory retrieval mechanism subserves them (Engelmann et al., 2019; Lewis et al., 2006; Lewis & Vasishth, 2005; Van Dyke & Lewis, 2003). According to cue-based memory retrieval accounts, when readers encounter a gap, all items in memory compete for activation, and the one that receives the highest activation will be retrieved. One consequence of such a cue-based memory retrieval mechanism is that when multiple items match a retrieval cue, similarity-based interference is predicted to occur. According to the activation-based memory retrieval model (Lewis & Vasishth, 2005; Vasishth et al., 2019), when memory retrieval is triggered, each retrieval cue distributes spreading activation to all items that match it. When the item to be retrieved has a higher level of activation, retrievals become faster and more accurate (Vasishth et al., 2019). Different patterns of (inhibitory and facilitatory) similarity-based interference are predicted based on whether the retrieval target fully matches the cues to retrieval or whether no item in memory provides a full match. When the target fully matches all retrieval cues and another item (a so-called *distractor*) matches one (or a subset) of them, the target receives reduced activation because each cue has a limited amount of spreading activation. In this case, retrieval times become slower and misretrievals occur, compared with when the distractor provides a poorer match. The present study refers to such increased retrieval times as *inhibitory interference* (Jäger et al., 2017). In the case that no item provides a full match to a set of retrieval cues, the activation model predicts decreased retrieval times, so-called *facilitatory interference* (Jäger et al., 2017). Specifically, when the target item and a distractor match two different retrieval cues, they each receive spreading activation from one cue. In this case their activation levels become similar, though activation fluctuates, as the model contains a random noise component. This means that the distractor gets retrieved with a probability of approximately 50% trial by trial. Facilitatory interference arises in such trials because their retrieval times become faster, compared with when the target is retrieved.[1]

Across various syntactic structures, facilitatory interference has been observed widely in previous studies, whereas evidence of inhibitory interference is less conclusive (Dillon et al., 2013; González Alonso et al., 2021; Jäger et al., 2015, 2020; Lago et al., 2015; Nicenboim et al., 2018; Van Dyke, 2007; Van Dyke & McElree, 2011; Wagers et al., 2009; for review, see Jäger et al., 2017). For example, Cunnings and Sturt (2018) investigated similarity-based interference using filler-gap sentences as below.

(3a) Sue remembered **the plate** that the butler with **the cup** accidently **shattered** today in the dining room.

(3b) Sue remembered **the plate** that the butler with **the tie** accidently **shattered** today in the dining room.

(3c) Sue remembered **the letter** that the butler with **the cup** accidently **shattered** today in the dining room.

(3d) Sue remembered **the letter** that the butler with **the tie** accidently **shattered** today in the dining room.

(3a/b) are plausible sentences, whereas (3c/b) are implausible sentences because of the plausibility of the filler ("the plate/letter") as an object of the verb "shattered" at the gap. (3a–d) contain a

[1] Whereas we discuss our predictions in terms of the activation-based model (Lewis et al., 2006; Lewis & Vasishth, 2005), there are other implementations of cue-based memory retrieval (e.g., see Parker, 2017; Van Dyke, 2007; Van Dyke & McElree, 2011; Vasishth et al., 2019). We do not attempt to tease apart these two instantiations of cue-based memory retrieval here. For present purposes, both accounts predict inhibitory and facilitatory interference.

distractor "the cup/tie" between the filler and the gap and manipulate whether the distractor can be a plausible object of "shattered" (3a/c: "shattered the cup") or not (3b/d: "shattered the tie"). Under the assumption that readers use a structural cue associated with the filler (e.g., [direct object]) and a sematic cue associated with a lexical property of "shatter" (e.g., [shatterable]), the activation model predicts that (3a/b) and (3c/d) should cause inhibitory and facilitatory interference, respectively. In (3a/b) the filler matches both the structural cue and the semantic cue. In (3a), both the filler ("plate") and the distractor ("cup") match the semantic cue, whereas in (3b), only the filler matches it. Thus, the filler in (3a) should receive reduced activation from the semantic cue compared with the filler in (3b), leading to inhibitory interference and longer reading times in (3a) than (3b). In (3c/d) the filler matches the structural cue but not the semantic cue. The distractor matches the semantic cue (but not the structural cue) in (3c) but matches neither cue in (3d). Because the filler and distractor both match one cue in (3c), the distractor is predicted to be retrieved some proportion of the time. Thus, retrieval times are predicted to become faster on average in (3c) than (3d), as evidence of facilitatory interference. Cunnings and Sturt observed implausibility effects, with longer reading times in implausible (3c/d) than plausible (3a/b). Crucially, facilitatory interference was also observed, with shorter reading times in (3c) than (3d) (see also Laurinavichyute & von der Malsburg, 2020 for evidence of a similar semantic interference effect). However, there was no evidence of inhibitory interference in (3a/b). These findings are largely compatible with the activation model except for the absence of inhibitory interference. However, the lack of inhibitory effects may be because they are very small and hence difficult to detect without large samples (Nicenboim et al., 2018). The results are also consistent with Wagers et al. (2009), who claimed that similarity-based interferences arises only if processing fails when the target item does not fully match the cues at retrieval. Although Wagers et al. made their claims based on ungrammatical sentences in subject–verb agreement, the results of Cunnings and Sturt may suggest a similar mechanism in implausible sentences.

L1 and L2 Differences in Sentence Processing

Many previous studies have investigated the similarities and differences between L1 and L2 sentence processing. Some previous studies have shown similar sentence processing patterns between L1 and L2 speakers (Cheng et al., 2021; Cunnings & Fujita, 2021a, 2021b; Felser et al., 2012; Foote, 2011; Fujita & Cunnings, 2021a, 2021b; Lago & Felser, 2018; Lim & Christianson, 2015; Omaki & Schulz, 2011; Tanner et al., 2012), whereas others have observed different patterns (Felser et al., 2003, 2009; Felser & Cunnings, 2012; Fujita & Cunnings, 2020; Jiang, 2004; Keating, 2009; Marinis et al., 2005; Papadopoulou & Clahsen, 2003). Some previous studies have reported that interference affects L1 and L2 dependency formation differently. For example, Felser et al. (2009) investigated how L1 and L2 speakers process sentences with reflexive resolution such as "**John/Jane** noticed that **Richard** had cut **himself** with a very sharp knife." Here, the reflexive pronoun "himself" must corefer with "Richard" (Chomsky, 1981), and the sentence manipulates whether the main clause subject, a syntactically unlicensed antecedent, matches ("John") or mismatches ("Jane") the gender of the reflexive. Cue-based accounts predict longer reading times at the reflexive when the unlicensed antecedent matches the reflexive in gender, as a result of inhibitory interference. Felser et al. observed that when the main clause subject matched with the reflexive in gender, L2 speakers but not L1 speakers had longer reading times, suggesting increased L2 difficulty retrieving the antecedent as a result of inhibitory interference. Felser and Cunnings (2012) further examined L2 reflexive resolution using texts such as "James/Mary has worked at the army hospital for years. The soldier that **he or she** treated on the ward wounded **himself/herself** while on duty." Syntactically, the reflexive ("himself/herself") in this text must corefer with "The soldier." Felser and Cunnings manipulated whether the grammatical antecedent matched the reflexive's gender and also whether a local distractor ("he or she"), which corefer with the subject of the lead-in sentence ("James/Mary"), matched it. Felser and Cunnings found longer first pass reading times and regression path duration at the reflexive when the distractor mismatched the reflexive's gender than when it matched only for L2 participants. Crucially, in these measures, L2 participants did not show gender mismatch effects by the grammatical antecedent. Although this finding might suggest that L2 speakers focused on the distractor because it was linearly closer, in another experiment, Felser and Cunnings found similar results even when the distractor was linearly more distant to the reflexive than the grammatical antecedent. They interpreted this as indicating that L2 processing was influenced by the distractor due to its discourse saliency. By contrast, studies investigating morpho-syntactic agreement have not observed increased susceptibility to either interference or a local distractor in L2 speakers (e.g., Lago & Felser, 2018).

Regarding the processing of filler-gap sentences, L2 speakers recover the information of the filler upon encountering a gap (Felser et al., 2012; Williams et al., 2001) but sometimes show non-nativelike patterns (Fujita & Cunnings, 2020; Kim et al., 2015; Marinis et al., 2005). Marinis et al. (2005) argued that L2 speakers do not process sentences with filler-gap dependencies spanning several clauses in the same manner as L1 speakers (Gibson & Warren, 2004; but see Pliatsikas & Marinis, 2013, who reported that L2 speakers with extended naturalistic exposure to English, their L2, show a nativelike pattern). Fujita and Cunnings (2020) showed that, unlike L1 speakers, L2 speakers may fail to revise temporarily ambiguous filler-gap sentences when revision cost is high. Kim et al. (2015) found that L2 speakers whose L1 is a wh-in-situ language, where a wh-element does not overtly move to a higher position (Ross, 1967), violate syntactic constraints on when a dependency may be formed during sentence processing. However, it remains unexplored about whether and how syntactically unlicensed constituents such as distractors in (3a–d) affect L2 filler-gap dependency formation.

The nature of potential L1/L2 differences in sentence processing remains controversial. One account of L1/L2 differences in sentence processing is that L2 speakers underuse syntactic information and heavily rely on semantic and surface information during sentence processing (e.g., see Clahsen & Felser, 2006a). Here, we consider one type of surface information namely linear order (see also Tanner et al., 2012; for similar consideration of linear order in subject–verb agreement). For example, as discussed in the previous section, in filler-gap sentences, an element moves to a higher position, but syntactically, it must be analyzed in the original, lower position. Underuse of syntactic information and heavy reliance on surface information when encountering a gap may lead L2 speakers to attempt to analyze a noun phrase that is locally closer to the gap as the filler. We test whether this potential type of shallow processing modulates how L2 speakers form filler-gap dependencies. To our knowledge, whereas existing studies have examined how distractors influence L2 processing of anaphora and subject-verb agreement (e.g., Felser & Cunnings, 2012; Lago & Felser, 2018), no existing study has examined the relative roles of syntactic and semantic information, and linear order in the resolution of filler-gap dependencies.

Recently, Cunnings (2017) proposed that L2 speakers construct syntactic structures in the same manner as L1 speakers but weight retrieval cues differently, leading to an increased susceptibility to similarity-based interference. For example, if L2 speakers weight the semantic cue more heavily than the structural cue in sentences like (3a–d), they should show increased inhibitory and facilitatory interference effects. That is, if L2 speakers weight the semantic cue more heavily than L1 speakers, larger inhibitory interference effects should be observed because of increased difficulty in locating the structurally legitimate filler in sentences like (3a), when the distractor matches the semantic cue, compared with (3b), when it does not. Furthermore, in sentences like (3c), overweighting the semantic cue would lead L2 speakers to retrieve the distractor faster and more often than L1 speakers, leading to larger facilitatory interference effects in the comparison between (3c/d).

The present study aims to tease these two accounts apart by testing filler-gap sentences as in (3a–d). As illustrated in the previous section, these sentences ("Sue remembered the plate/letter that the butler with the cup/tie accidently shattered today in the dining room") contain a distractor ("the cup/tie") between the filler ("the plate/letter") and the gap (immediately following "shattered") and manipulate the plausibility of the filler and the distractor as an object of "shattered." If L2 sentence processing relies on surface cues, L2 speakers may attempt to analyze the distractor as an object of "shattered," given that the distractor is linearly closer to the gap than the filler (e.g., the plate/letter$_{[\text{filler}]}$ . . . the cup/tie$_{[\text{distractor}]}$ . . . shattered$_{[\text{gap}]}$). By contrast, Cunnings' account assumes that L2 speakers attempt retrieval of the filler at the gap but are more susceptible to inhibitory and facilitatory interference caused by the distractor than L1 speakers. Thus, if this account is correct, L2 speakers should demonstrate implausibility effects as observed in Traxler and Pickering (1996) and Cunnings and Sturt (2018), but interference effects of a larger magnitude than L1 speakers.

The Present Study

Below, we report three preregistered experiments, which examined potential L1/L2 differences in the processing of filler-gap sentences and susceptibility to similarity-based interference. Experiment 1 recorded participants' eye movements during reading to investigate filler-gap dependency formation and similarity-based interference during sentence processing. Experiment 2 further investigated L1/L2 differences in similarity-based interference using an offline comprehension question task. Experiment 3 aimed to conceptually replicate the findings of Experiment 1 using a speeded judgment task. L2 speakers in our study were defined as having begun learning the L2 after their L1 (after at least age 5) and were required to have at least intermediate proficiency (see the Method section for further details). We thus do not intend to draw any conclusions about L1/L2 processing at lower levels of proficiency, or in other types of bilingual speakers.

We preregistered the research designs, sampling method, and data analysis plan in Experiments 1–3 on the Open Science Framework website (https://osf.io/5up4f for Experiments 1 and 2 and https://osf.io/nvxam for Experiment 3). We recruited participants and conducted data analysis as in the preregistrations. Our materials, data, and analysis code can be found at https://osf.io/vh4sm/.

Experiment 1

Method

Experiment 1 investigated filler-gap dependency formation and similarity-based interference in L1 and L2 sentence processing. Experiment 1 adopted the research design used in Cunnings and Sturt (2018), in which participants read filler-gap sentences as in (4a–b).

(4a) *Plausible filler*, *Plausible distractor*
Mary saw the beer that the man with the wine very happily drank during the party. The night was fun.

(4b) *Plausible filler*, *Implausible distractor*
Mary saw the beer that the man with the food very happily drank during the party. The night was fun.

(4c) *Implausible filler*, *Plausible distractor*
Mary saw the cake that the man with the wine very happily drank during the party. The night was fun.

(4d) *Implausible filler, Implausible distractor*
Mary saw the cake that the man with the food very happily drank during the party. The night was fun.

(4a–d) contain a filler ("the beer/cake") in the main clause and its gap at the embedded clause verb ("drank"). The filler is either a plausible object of "drank" (4a/b: "drank the beer") or an implausible object (4c/d: "drank the cake"). (4a–d) also contain a distractor ("the wine/food"), which is either plausible (4a/c) or implausible (4b/d) as an object of the embedded clause verb (4a/c: "drank the wine"; 4b/d: "drank the food").

For L1 speakers, we predicted longer reading times at "drank" in (4c/d) than (4a/b) owing to implausibility effects (Traxler & Pickering, 1996). If semantic interpretation during filler-gap dependency formation is subject to facilitatory interference (Cunnings & Sturt, 2018), the plausible distractor condition (4c) should elicit reduced reading times compared with the implausible distractor condition (4d). Also, if inhibitory interference arises (Lewis & Vasishth, 2005), reading times should be longer in (4a) than (4b).

Regarding predictions for L2 speakers, unlike L1 processing, L2 processing may rely heavily on semantic and surface information during filler-gap dependency formation. At the surface level in (4a–d), the distractor is linearly closer to the gap than the filler (e.g., the beer/cake$_{[\text{filler}]}$ . . . the wine/food$_{[\text{distractor}]}$ . . . drank$_{[\text{gap}]}$). Thus, if L2 speakers are prone to such surface information, they should attempt to analyze the distractor as an object of "drank." Consequently, reading times should be longer in (4b/d) than (4a/c) due to the implausibility of the distractor. By contrast, Cunnings (2017) assumes that L2 speakers process filler-gap sentences in the same manner as L1 speakers but are more susceptible to similarity-based interference. If this assumption is correct, L2 speakers should show implausibility effects in the implausible filler condition as predicted for L1 speakers but a larger size of interference effects than L1 speakers.

Participants

Eighty L1 English speakers (mean age 21; range 18–60) and 80 L2 English speakers (mean age 25; range 18–54) from the University of Reading community participated in Experiment 1. The L1 participants identified English as their native language. The L2 participants claimed that they spoke English as their non-native language and had different L1 backgrounds.[2] They started learning English in a school environment after age five and had on average spent approximately 3 years in an English-speaking country. After Experiment 1, both L1 and L2 participants completed a lexical-decision task (Hopp, 2014). The L2 participants additionally completed the Quick Placement test (Quick placement test: Version 1., 2004).[3] Their average score on the Quick Placement test was 46 of 60 (range 30–60).[4]

Materials

We initially created 40 experimental sentences as in (4a–d) and conducted a pretest to ensure that materials displayed the intended range of plausibility. For the pretest, we recruited 40 L1 English speakers via Prolific (https://prolific.co/). None of these participants took part in the main experiments reported in the present study. In the pretest, we created four conditions to test the plausibility of each filler and distractor as a displaced direct object of the embedded clause verb (e.g., Mary saw the beer/wine/cake/food that the man very happily drank during the party). The order of the experimental sentences was pseudorandomized in a Latin square design with 20 plausible and 20 implausible filler sentences. Participants rated the plausibility of each sentence on a scale from 1 (*highly implausible*) to 7 (*highly plausible*). Based on the results of the pretest, we selected 24 sentences. The average ratings of these 24 sentences were 6.28 (range 5.4–6.9) and 6.39 (range 5.4–6.9) for the plausible filler and distractor and 1.50 (range 1.1–2.3) and 1.51 (range 1.1–2.4) for the implausible filler and distractor. We added a wrap-up sentence to each trial (e.g., "The night was fun," as in [4a–d]), to minimize end-of-trial artefacts from influencing reading times of the critical sentence.

For the main experiment, we created 72 filler texts.[5] These filler texts consisted of various syntactic structures. A yes/no comprehension question followed each experimental and filler text. The questions for experimental texts queried different parts of the texts but did not probe the interpretation of the filler-gap dependency (for example, "Did the man drink happily?" for [4a–d]). Both experimental and filler texts always appeared on one line.

Procedure

We recorded eye movements while participants read texts, using an SR Research Eyelink 1000 at a sample rate of 1000 Hz. Although viewing was binocular, we sampled eye movements from participants' right eye only. Each session began with calibration of the eye-tracker on a nine-point grid. Before each trial, a gaze trigger appeared above the first word of the text. Upon fixation on the gaze trigger, the text appeared. After reading each text, participants answered a comprehension question by pressing a button on a gamepad. The experiment lasted approximately 30 minutes for L1 participants and 50 minutes for L2 participants.

Preregistered Data Analysis

We analyzed the eye-movement data in R (R Core Team, 2020). For data analysis, we calculated three reading time measures from participants' eye movements: First pass times, regression path duration and total viewing times. First pass times are the sum of fixations within a region entered from the left up until an eye movement away from the region. Regression path duration is the summed duration of all fixations measured from when a region has the first fixation from the left, up until but not including the first fixation in a region to the right. Thus, on trials where readers regress from the target region before moving on, regression path duration includes fixations to previous regions. Total viewing times are the summed duration of all fixations in a region. Before calculating these reading time measures, we merged fixations shorter than 80 milliseconds that were within one degree of visual arc of another fixation. We then removed any other fixations shorter than 80 milliseconds or longer than 800 milliseconds. Trials with excessive track loss were also removed from data analysis, which affected less than 0.1% of the L1 and L2 data.

To analyze these reading times, we divided the experimental texts into two regions. One is the embedded clause verb ("drank"), the critical region. The other is the rest of the critical sentence ("during the party."), a spillover region. We fit linear mixed-effect models to each reading time measure at these regions, using the package, *lme4* (Bates et al., 2015). When fitting the models, we log-transformed reading times to reduce the skew of residuals. Following Cunnings and Sturt (2018), we analyzed the critical and spillover regions in the same model. Thus, each model contained fixed effects of sum-coded group (L1/L2), region (critical region/spillover region), filler (plausible/implausible), distractor (plausible/implausible), and their interactions. The models also included random intercepts for participants, items and trials, random slopes for each within-participants, within-items, and within-trials fixed effect and random correlations between random intercepts and slopes (Barr et al., 2013). When this maximal model failed to converge, we initially removed the random correlations. If the model

[2] First languages of the L2 participants were Chinese (25), Japanese (8), Korean (7), French (16), Greek (8), Spanish (6), Bulgarian (4), Romanian (3) and German (3).

[3] L2 participants completed the Quick Placement Test to assess their proficiency. The lexical decision task was conducted to investigate whether lexical automaticity explains any L1/L2 differences, assuming such differences were observed (see the pre-registration). Because we did not find clear differences between L1 and L2 speakers, this additional analysis was not conducted.

[4] In the preregistration, we specified that we would continue participant recruitment until we had 80 L1 speakers and 80 L2 speakers who correctly answered at least 70% of the comprehension questions in the eye-movement study, as an index that they paid attention. Because we were interested in testing proficient rather than novice L2 English speakers, the L2 participants additionally needed to score 30/60 on the on the Quick Placement Test. None of the participants failed to fulfil these criteria.

[5] Twenty-four fillers in both Experiments 1 and 2 were from a different manipulation not reported here.

still did not converge, we iteratively removed the random effect accounting for the least variance until the model converged (Barr et al., 2013). We estimated p values from the t distribution (Baayen, 2008) and interpreted p values smaller than .05 as significant. When a significant interaction appeared, follow-up analyses examined nested simple effects of theoretical interest within a single model (Schad et al., 2021).

Results

Mean accuracy rates to comprehension questions were 94% (range 81–99%) for L1 participants and 93% (range 71–100%) for L2 participants. Table 1 reports raw reading times for the critical and spillover regions. Table 2 provides a summary of inferential statistics. For brevity, all models showed a significant main effect of group, with longer reading times for L2 than L1 participants. Below, we do not discuss main effects of region and group by region interactions, as these effects, unless they interact further with other fixed effects, are irrelevant to our research questions.

First pass times did not show any significant effects of theoretical interest.

In regression path duration, there was a significant main effect of filler, which shows longer reading times in the implausible filler than plausible filler conditions. This indicates implausibility effects. This was modified by a significant region by filler interaction, and follow-up analyses indicated significant implausibility effects at the spillover region only (critical region: *estimate* = .022, SE = .01, t = 1.62, p = .105; spillover region: *estimate* = .145, SE = .02, t = 8.87, $p < .001$). There was also a significant region by distractor interaction, with longer reading times in the implausible distractor than plausible distractor conditions at the spillover region (critical region: *estimate* = .000, SE = .01, t = .01, p = .988; spillover region: *estimate* = .036, SE = .01, t = 2.51, p = .012), and a significant group by distractor interaction. However, a follow-up analysis examining the effect of distractor for each group but did not show any significant effects (L1: *estimate* = .036, SE = .01, t = 1.94, p = .053; L2: *estimate* = .000, SE = .01, $t = -.04$, p = .968). Importantly, there was a significant filler by distractor interaction that did not further interact with group. A follow-up analysis, which examined the effect of distractor at the two levels of filler, showed that for the implausible filler conditions, implausibility effects were significantly attenuated when the distractor was a plausible object (*estimate* = .044, SE = .01, t = 3.39, $p < .001$). This effect indicates facilitatory interference and is illustrated in Figure 1. By contrast, the follow-up analysis did not show a significant effect of distractor in the plausible filler conditions (*estimate* = −.008, SE = .01, $t = -.61$, p = .543).

Total viewing times showed a significant main effect of filler with longer reading times in the implausible filler than plausible filler conditions due to implausibility effects. There was also a significant main effect of distractor, which shows longer reading times in the implausible distractor than plausible distractor conditions. This main effect of distractor however interacted with region and group, and follow-up analyses indicated that it was present only at the spillover region in the L1 group (L1/critical region: *estimate* = .005, SE = .01, t = .34, p = .731; L1/spillover region: *estimate* = .044, SE = .01, t = 3.70, $p < .001$; L2/critical region: *estimate* = .015, SE = .01, t = 1.19, p = .234; L2/spillover region: *estimate* = .002, SE = .01, t = .18, p = .858).[6]

Discussion

Both L1 and L2 participants showed implausibility effects when the filler was an implausible object. This finding demonstrates that L1 and L2 participants recovered the information of the filler upon encountering the gap (Traxler & Pickering, 1996; Williams et al., 2001).

There was also evidence that the implausibility effects were attenuated in the plausible distractor compared with implausible distractor conditions. This was most clearly found in regression path duration. Such an attenuation indicates that our participants were susceptible to facilitatory interference (Wagers et al., 2009). However, we did not find evidence of inhibitory interference. The absence of inhibitory interference is consistent with Cunnings and Sturt (2018), who used the same research design as in Experiment 1 and observed only facilitatory interference.

Crucially, we did not find evidence of increased L2 difficulty in filler-gap dependency formation. Although group by distractor interactions appeared in regression path duration and total viewing times, they were mainly because L1 participants had increased difficulty in the implausible distractor conditions, not the L2 speakers. The direction of the effect here, with L1 but not L2 speakers showing larger effects of distractor plausibility, is the opposite of what would be expected from shallow L2 processing. Also, there was no evidence that L2 participants were more susceptible to similarity-based interference than L1 speakers (cf. Cunnings, 2017).

In sum, Experiment 1 demonstrated facilitatory interference in L1 and L2 speakers, suggesting both groups use cue-based memory retrieval during processing. In Experiment 2, we further investigated similarity-based interference in L1 and L2 speakers using an offline task.

Experiment 2

Method

Experiment 2 investigated whether L2 speakers are more susceptible to similarity-based interference than L1 speakers at the offline level, using sentences as in (5a–b).

(5a) *Filler-Gap dependency*, *Plausible distractor*
Kevin saw the sandwich that the boy by the cake quickly ate during lunch.

(5b) *Filler-Gap dependency*, *Implausible distractor*
Kevin saw the sandwich that the boy by the milk quickly ate during lunch.

(5c) *No dependency*, *Plausible distractor*

[6] As exploratory analyses, we also investigated whether L1 background (wh-in-situ/wh-movement), L2 proficiency and lexical automaticity affected the results of the L2 participants. However, we did not find evidence that these factors significantly interacted with the effects of theoretical interest.

Table 1
Mean Raw Reading Times in Milliseconds and Standard Errors (SE) at the Critical and Spillover Regions in Experiments 1

	Critical Region (drank)				Spillover Region (during the party)			
	Native Speakers		Non-Native Speakers		Native Speakers		Non-Native Speakers	
Condition	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
First pass time								
Plausible filler, Plausible distractor	271	5.4	320	7.0	418	11.1	602	17.4
Plausible filler, Implausible distractor	265	5.1	334	8.9	446	13.6	584	17.1
Implausible filler, Plausible distractor	269	5.0	331	8.2	446	12.7	606	18.6
Implausible filler, Implausible distractor	272	5.2	326	6.9	457	13.8	607	17.7
Regression path duration								
Plausible filler, Plausible distractor	403	28.9	423	19.3	905	53.4	1,299	70.8
Plausible filler, Implausible distractor	374	23.6	404	19.1	960	49.8	1,088	51.0
Implausible filler, Plausible distractor	419	34.7	498	33.5	1,233	72.0	1,697	91.8
Implausible filler, Implausible distractor	456	29.3	438	21.4	1,440	71.4	1,858	99.4
Total viewing time								
Plausible filler, Plausible distractor	382	10.8	554	18.9	683	19.9	1,153	36.9
Plausible filler, Implausible distractor	385	11.6	547	18.8	733	22.3	1,178	40.5
Implausible filler, Plausible distractor	464	13.7	656	20.8	826	25.6	1,442	45.5
Implausible filler, Implausible distractor	474	14.1	696	23.1	912	25.1	1,443	46.2

Kevin saw the boy by the cake who quickly ate the sandwich during lunch.

(5d) *No dependency*, *Implausible distractor*
Kevin saw the boy by the milk who quickly ate the sandwich during lunch.

Question: What did the boy eat during lunch? 1 The sandwich 2 The cake

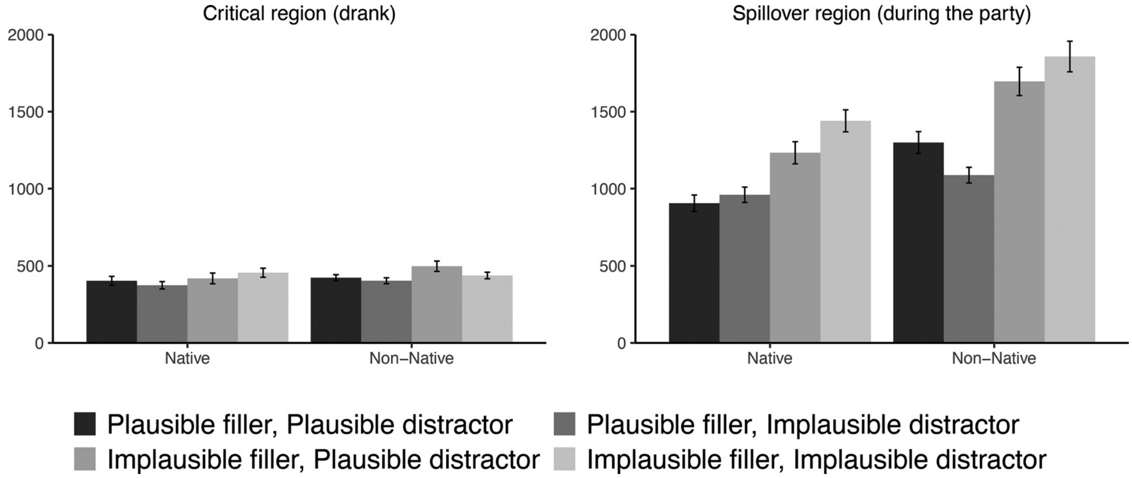
(5a/b) contain a filler "the sandwich" in the main clause and its gap at the embedded clause verb "ate." (5c/d) do not involve such a filler-gap dependency because "the sandwich" is adjacent to "ate" in these conditions. Thus, (5c/d) do not require direct-object filler-gap dependency formation at "ate." (5a–d) contain a distractor and manipulate whether it is a plausible object of the embedded clause verb (5a/c; "ate the cake") or not (5b/d; "ate the milk"). Note that, unlike Experiment 1, this additional experiment contains only plausible sentences. (5a–d) are followed by a comprehension question with two answer options, which queries the object of the embedded clause verb. The two answer options were the same across conditions, both being plausible direct objects of the verb, as including the target and distractor in implausible conditions (e.g., What did the boy eat during lunch? The sandwich/The milk) would have meant that participants could answer the questions without reading the critical sentence.

If similarity-based interference arises during filler-gap dependency formation and affects offline language comprehension, comprehension accuracy should be lower in (5a) than (5b). We also predicted that such differences should not arise between (5c) and (5d), where no filler-gap dependency formation is required to retrieve the object of "ate." Crucial evidence of

Table 2
Summary of Statistical Analyses for Reading Times in Experiment 1

	First pass time				Regression path duration				Total viewing time			
Fixed effects	Estimate	*SE*	*t* value	*p* value	Estimate	*SE*	*t* value	*p* value	Estimate	*SE*	*t* value	*p* value
Intercept	5.811	0.02	254.64	<.001	6.245	0.03	215.74	<.001	6.346	0.03	197.32	<.001
Group	0.110	0.02	6.50	<.001	0.102	0.02	4.48	<.001	0.192	0.03	7.44	<.001
Region	0.241	0.03	8.45	<.001	0.492	0.03	14.16	<.001	0.358	0.04	9.08	<.001
Filler	0.006	0.01	0.90	.370	0.083	0.01	8.01	<.001	0.099	0.01	10.85	<.001
Distractor	0.004	0.01	0.55	.583	0.017	0.01	1.56	.118	0.017	0.01	2.21	.027
Group:Region	0.035	0.01	3.10	.002	0.054	0.02	3.44	<.001	0.044	0.01	3.68	<.001
Group:Filler	−0.006	0.01	−0.85	.394	−0.003	0.01	−0.24	.810	0.001	0.01	0.16	.872
Region:Filler	0.004	0.01	0.52	.600	0.061	0.01	6.18	<.001	0.006	0.01	1.05	.292
Group:Distractor	0.002	0.01	0.28	.783	−0.019	0.01	−2.09	.037	−0.008	0.01	−1.13	.260
Region:Distractor	−0.002	0.01	−0.36	.721	0.018	0.01	2.08	.038	0.006	0.01	1.19	.235
Filler:Distractor	−0.001	0.01	−0.13	.898	0.026	0.01	2.94	.003	0.010	0.01	1.41	.157
Group:Region:Filler	−0.002	0.01	−0.29	.771	0.000	0.01	−0.04	.965	0.002	0.01	0.30	.761
Group:Region:Distractor	−0.007	0.01	−0.98	.328	−0.014	0.01	−1.58	.115	−0.013	0.01	−2.37	.018
Group:Filler:Distractor	0.001	0.01	0.16	.873	0.003	0.01	0.32	.747	0.001	0.01	0.10	.921
Region:Filler:Distractor	0.001	0.01	0.11	.910	0.019	0.01	1.92	.054	−0.002	0.01	−0.39	.693
Group:Region:Filler:Distractor	0.009	0.01	1.35	.176	0.017	0.01	1.24	.215	−0.004	0.01	−0.57	.570

Figure 1
Regression Path Duration in Milliseconds at the Critical and Spillover Regions in Experiment 1



interference in this case is thus the interaction between distractor plausibility and sentence type. If L2 speakers are more susceptible to similarity-based interference than L1 speakers (Cunnings, 2017), they should show a larger difference between (5a/b) than L1 speakers.

Participants

The participants from Experiment 1 took part in this experiment in a separate session. The participants completed the task in a lab setting at least 1 week after Experiment 1. We tested the same participants because in L2 research, it is typical to test online processing and offline comprehension in the same group of learners.

Materials

The experiment contained 24 sets of experimental texts as in (5a–d) and 72 filler sentences with various syntactic structures. These experimental and filler sentences were different from those used in Experiment 1. A comprehension question that had two options as possible answers followed each experimental and filler sentence.

Procedure

Participants completed the experiment using IbexFarm (http://spellout.net/ibexfarm) in a traditional lab setting. At the start of each trial, an underline that masked the whole sentence appeared onscreen. When participants pressed the space bar, the sentence appeared in full. When participants pressed the space bar again, the sentence disappeared, and a comprehension question containing two options appeared. Participants answered each question by pressing an appropriate key on the keyboard. Half of the correct answers were on the left side and half on the right side. The position of the two options and the order of the sentences were pseudorandomized in a Latin square design.[7] The experiment began with four practice trials. The experiment took approximately 20 minutes for L1 participants and 30 for L2 participants. The materials, data and analysis code for this experiment are available on the OSF (https://osf.io/vh4sm/).

Preregistered Data Analysis

We analyzed comprehension accuracy rates as the dependent variable by a fitting logistic regression using generalized linear mixed-effects models. The models included sum-coded fixed effects of group (L1/L2), dependency (filler-gap dependency/no dependency), distractor (plausible/implausible), and their interactions. Also, the models consisted of the maximal random effects structure that converged.

Results

Mean comprehension accuracy rates to filler questions were 90% (range 54–100%) for L1 participants and 93% range 73–100%) for L2 participants.[8] Table 3 reports descriptive statistics, and Table 4 summarizes inferential statistics.

There was a significant main effect of group with lower comprehension accuracy rates for L1 than L2 participants. This finding may simply result from L1 speakers having been less attentive to the task, given lower comprehension accuracy rates and a wider range to filler materials for L1 speakers. There was also a significant main effect of distractor, which shows lower comprehension accuracy rates in the plausible distractor than implausible distractor conditions. The models did not show any significant interactions.

Discussion

The results indicate that comprehension accuracy was influenced by the plausibility of the distractor. However, we did not find a significant interaction between dependency and distractor,

[7] Owing to technical error, the Latin-square for the L1 data was not fully balanced, with 27, 14, 25 and 14 completing lists 1–4, respectively. For the L2 data, each list was completed by 20 participants.

[8] Five L1 participants scored below 70%, but we included these participants in data analysis because we did not pre-register any threshold for the inclusion of participants in data analysis in Experiment 2. Note that we conducted an exploratory analysis to assess the effect of these participants, but similar results obtained.

Table 3
Mean Comprehension Accuracy Rates and Standard Errors (SE) in Experiments 2

Condition	Native Speakers		Non-Native Speakers	
	M	*SE*	*M*	*SE*
Filler-gap dependency, Plausible distractor	0.760	0.02	0.890	0.01
Filler-gap dependency, Implausible distractor	0.944	0.01	0.962	0.01
No dependency, Plausible distractor	0.802	0.02	0.852	0.02
No dependency, Implausible distractor	0.933	0.01	0.954	0.01

which would have been predicted if interference arose during filler-gap dependency formation. These results were unexpected, but we contend that they may indicate interference during the question-response phase, where participants may have had to recall the sentence, rather than indexing interference during incremental sentence processing (see also Meng & Bader, 2021, who queried whether offline comprehension questions tap misinterpretation during sentence processing). Crucially, we did not find any significant evidence of L2 speakers being more susceptible to interference than L1 speakers. Owing to the lack of an interaction between dependency and distractor in Experiment 2, and in an attempt to replicate the main findings of facilitatory interference in Experiment 1, we conducted Experiment 3, which used a speeded judgment task. Speeded judgment tasks have been employed in several previous studies where facilitatory interference was observed consistently (González Alonso et al., 2021; Schlueter et al., 2018; Wagers et al., 2009). Thus, speeded judgment tasks are a good testing method to assess the robustness of the facilitatory interference effects observed in Experiment 1.

Experiment 3

Method

Experiment 3 investigated L1/L2 differences in the susceptibility to similarity-based interference using a speeded judgment task. In this task, participants read sentences as in (6a–d)

(6a) *Plausible filler*, *Plausible distractor*
Mary saw the beer that the man with the wine very happily drank.

(6b) *Plausible filler*, *Implausible distractor*
Mary saw the beer that the man with the food very happily drank.

(6c) *Implausible filler*, *Plausible distractor*
Mary saw the cake that the man with the wine very happily drank.

(6d) Implausible filler, Implausible distractor
Mary saw the cake that the man with the food very happily drank.

(6a–d) are identical to (4a–d) without the spillover region and wrap-up sentence. Thus, as in (4a/d), (6a–d) manipulate whether the sentence is plausible (6a/b) or not (6c/d) and whether the distractor is a plausible object of the embedded clause verb (6a/c) or not (6b/d).

We predicted that L1 speakers should judge (6a/b) to be plausible and (6c/d) to be implausible, but that (6c) should elicit more incorrect responses than (6d) due to facilitatory interference. If L2 speakers are more susceptible to interference than L1 speakers (Cunnings, 2017), they should show larger facilitatory interference effects.

Participants

Ninety-six L1 English speakers (mean age 27; range 19–49) and 96 L2 English speakers (mean age 30; range 21–46) participated in Experiment 3. We recruited the participants via Prolific. The L1 participants were university students, had U.K. nationality, resided in the U.K. and spoke English as their L1. The L2 participants had either an undergraduate, graduate, or doctoral degree, resided in either the U.K. or the U.S., identified as having been raised as monolinguals, spoke English as their L2 language, and had different L1 backgrounds.[9] The L2 participants started learning English after age 5 and had on average spent approximately 8 years in an English-speaking country. After the judgment task, the L2 participants completed the Quick Placement Test. Their average score was 50 (range 34–60).[10]

Materials

The materials were the same as those used in Experiment 1 except that they ended at the embedded clause verb and did not have wrap-up sentences. Experiment 3 also contained 60 filler sentences with a variety of syntactic structures. A comprehension question followed eight experimental sentences and 20 filler sentences to check that participants paid attention.

Procedure

Participants completed the experiment online in their own time via IbexFarm (http://spellout.net/ibexfarm). At the start of each trial, a cross appeared at the center of the screen. When participants pressed the space bar, the cross disappeared, and the sentence was presented word by word at the center of the screen. The pacing was 500 milliseconds per word. After the last word, a question mark appeared, and participants judged whether the sentence

[9] First languages of the L2 participants were Chinese (37), Japanese (3), Korean (8), French (5), Greek (6), Spanish (18), Bulgarian (7), Romanian (9) and German (3).

[10] Following the preregistration, we continued participant recruitment until we had 96 L1 speakers and 96 L2 speakers who scored at least 75% correct on post-trial comprehension questions, as an index that they were paying attention to the task. L2 participants also had to score 30/60 on the Quick Placement Test. Twenty-seven additional participants took part, but their data were not included in the analysis because they failed to meet these criteria.

Table 4
Summary of Statistical Analyses for Comprehension Accuracy Rates in Experiment 2

Fixed effects	Estimate	*SE*	*z* value	*p* value
Intercept	2.554	0.13	20.07	< .001
Group	0.267	0.09	3.14	.002
Dependency	−0.064	0.07	−0.93	.352
Distractor	0.701	0.08	8.68	< .001
Group:Dependency	−0.079	0.07	−1.14	.253
Group:Distractor	−0.055	0.06	−0.86	.388
Dependency:Distractor	−0.036	0.06	−0.57	.571
Group:Dependency:Distractor	0.077	0.06	1.22	.222

was plausible or implausible within a timeout of 1,500 milliseconds. Feedback was provided if participants missed the timeout, but participants were not given feedback about the correctness of their responses. After making a judgment, participants sometimes answered a comprehension question by pressing an appropriate key on the keyboard. The experiment lasted approximately 25 minutes for L1 participants and 40 minutes for L2 participants.

Preregistered Data Analysis

We analyzed judgment responses in R by fitting a mixed-effects logistic regression. The models contained the same fixed and random effects as those in Experiment 1 except that they did not include a region variable. The dependent variable was judgment accuracy (correct/incorrect), with the correct response being coded as "plausible" in the plausible conditions and "implausible" in the implausible conditions. Time-outs accounted for less than 1% of the data in each group and were removed before data analysis.

Results

Mean comprehension accuracy rates to comprehension questions were 91% (range 77–100) for L1 participants and 88% (range 75–100) for L2 participants. Tables 5 and 6 show descriptive and inferential statistics, respectively.

There was a significant main effect of filler, with lower judgment accuracy rates in the plausible filler than implausible filler conditions. There was also a significant main effect of distractor, which indicated lower judgment accuracy rates when the distractor was a plausible object than when it was an implausible object. Crucially, the model showed a significant filler by distractor interaction. A follow-up analysis examined the effect of distractor at the two levels of filler. This analysis showed that for the implausible filler condition, judgment accuracy rates were significantly lower in the plausible distractor condition (*estimate* = .540, $SE = .07$, $t = 7.45$, $p < .001$). This effect indicates facilitatory interference, as participants sometimes considered implausible sentences as plausible. For the plausible filler conditions, the follow-up analysis indicated significantly lower judgment accuracy rates in the implausible distractor condition (*estimate* = −.135, $SE = .06$, $t = -2.26$, $p = .024$). Note that the size of the effect in the plausible filler conditions is much smaller than in the implausible filler conditions.

Discussion

Experiment 3 demonstrated that participants misjudged implausible sentences to be plausible more often when the distractor was plausible than implausible. This finding is compatible with previous studies, which have observed facilitatory interference in ungrammatical sentences (González Alonso et al., 2021; Schlueter et al., 2018; Wagers et al., 2009). However, we also found higher accuracy rates in the plausible filler conditions when the distractor was plausible. There was also no significant evidence of L1/L2 differences in the susceptibility to similarity-based interference. We discuss these findings from Experiment 3, along with Experiments 1 and 2, in more detail below.

General Discussion

The present study investigated potential L1/L2 differences in memory retrieval operations during filler-gap dependency formation, using three different tasks. Experiment 1 demonstrated that L1 and L2 speakers retrieve the filler upon encountering a gap and that facilitatory interference influences this gap-filling operation.

Table 5
Mean Judgement Scores and Standard Errors (SE) in Experiments 3

Condition	Native Speakers		Non-Native Speakers	
	M	*SE*	*M*	*SE*
Plausible filler, Plausible distractor	0.75	0.02	0.74	0.02
Plausible filler, Implausible distractor	0.74	0.02	0.67	0.02
Implausible filler, Plausible distractor	0.82	0.02	0.79	0.02
Implausible filler, Implausible distractor	0.92	0.01	0.90	0.01

Table 6
Summary of Statistical Analyses for Judgement Scores in Experiment 3

Fixed effects	Estimate	*SE*	*z* value	*p* value
Intercept	1.913	0.10	19.84	< .001
Group	−0.134	0.08	−1.64	.101
Filler	0.556	0.10	5.48	< .001
Distractor	0.210	0.09	2.43	.015
Group: Filler	0.022	0.09	0.23	.817
Group:Distractor	−0.040	0.05	−0.75	.454
Filler:Distractor	0.347	0.05	6.79	< .001
Group: Filler:Distractor	0.062	0.06	1.10	.27

Experiment 3 also showed that L1 and L2 offline language comprehension are subject to similarity-based interference. We did not find evidence of L1/L2 differences in filler-gap dependency formation or in susceptibility to similarity-based interference. Below, we discuss the implications of these results.

Interference in Language Comprehension

Experiments 1 and 3 observed facilitatory interference for the implausible filler conditions in online and offline tasks. These findings are consistent with many previous studies on cue-based memory retrieval (e.g., Jäger et al., 2017, 2020; Wagers et al., 2009) and replicate the finding of Cunnings and Sturt (2018) that similarity-based interference affects semantic interpretation during filler-gap dependency formation. While many studies that have reported facilitatory interference tested sentence grammaticality, our results, along with Cunnings and Sturt (2018), indicate that similar effects arise because of plausibility. Cunnings and Sturt claimed that semantic interference arises because readers utilize the lexical properties of verbs, such as [drinkable] for the verb *drink*, as retrieval cues (see also Smith & Vasishth, 2020; Van Dyke & McElree, 2006). Our results are compatible with this claim for L1 language comprehension and extend it to L2 comprehension.

Although we found evidence of facilitatory interference in the implausible filler conditions in Experiment 1, we did not find evidence of inhibitory interference in the plausible conditions. This finding is compatible with many previous studies including Cunnings and Sturt (2018), which similarly found facilitatory interference only. Recall from the Introduction that Wagers et al. (2009) argued that cue-based memory retrieval may only occur as a result of revision, when readers recognize that the retrieval target does not match the properties of the relevant verb. They argued for this account based on their finding of facilitatory interference in ungrammatical sentences and a lack of interference in grammatical sentences. Their account could also explain our findings and may suggest that cue-based memory retrieval is initiated when the sentence is implausible.

Alternatively, the absence of inhibitory interference may be due to the present study lacking adequate statistical power. Although our eye-tracking experiment contained 80 L1 speakers and 80 L2 speakers, which is larger than most previous studies on similarity-based interference, Nicenboim et al. (2018) recently suggested that the size of inhibitory interference is smaller than that of facilitatory interference. Thus, the detection of inhibitory interference requires high statistical power. Such an account may explain why the present study failed to find evidence of inhibitory interference. It is also possible that the absence of inhibitory interference is due to our manipulation not being strong enough to elicit inhibitory effects (Van Dyke & McElree, 2011). In either case, replicating the results of the present study with higher statistical power and different lexical materials would be an interesting avenue of future research.

The offline results from Experiment 2 could be taken as evidence of inhibitory interference, but we are cautious in drawing strong conclusions here. In Experiment 2, accuracy for plausible sentences was lower when a distractor was also a plausible object of a verb. However, this finding was observed in both filler-gap and nonfiller-gap sentences, making it difficult to conclude that this difference arises due to similarity-based interference during retrieval at the verb. As mentioned in the discussion for Experiment 2, these results may instead indicate interference during the posttrial comprehension question phase, where participants may have needed to recall the critical sentences. Alternatively, as noted by a reviewer, lower accuracy in the plausible distractor conditions here could be an artifact of our design. Specifically, participants were always provided with two plausible answer options to our questions (e.g., “The sandwich” and “The cake” for “What did the boy eat during lunch?”). In plausible conditions, both answer options appeared in the critical sentences, whereas in the implausible distractor conditions, only the correct answer (the plausible filler) appeared in the sentence, while the implausible distractor (e.g., “the milk”) was not an answer option. As such, the main effect of distractor may simply indicate that the question was easier to answer in the implausible distractor conditions, because only one noun out of the two answer options appeared in the sentence. We adopted this design because a dependency by distractor interaction would still be indicative of interference, however we acknowledge that it does complicate interpretation of the main effect of distractor (in the absence of an interaction) that we did observe. As such, while we do not draw strong conclusions from Experiment 2, the results highlight the importance of including “no dependency” baseline conditions when assessing interference in comprehension tasks.

Note however that we found some findings that actually go in the opposite direction to inhibitory interference. In total viewing times in Experiment 1, L1 speakers had longer reading times in the implausible distractor conditions, irrespective of the plausibility of the filler. This finding in the plausible filler conditions is the opposite of inhibitory interference. In Experiment 3, we also found significantly lower accuracy in the plausible filler conditions when the distractor was implausible, an effect that did not differ significantly by group. These findings are inconsistent with cue-based memory retrieval models. While we do not draw strong conclusions about potential L1/L2 differences here, given significant L1/L2 differences were not observed across experiments, these findings warrant further discussion.

The pattern of results here in the plausible filler conditions is potentially compatible with so-called representational accounts of interference in subject-verb agreement (Hammerly et al., 2019; Pearlmutter et al., 1999). Whereas cue-based memory retrieval models explain interference based on the degree of match between the properties of a verb and previous items in a sentence, representational accounts of subject-verb agreement argue that such effects result from how the relevant features of a complex noun phrase are encoded in memory. For example, such accounts predict that in a complex noun phrase such as "The key to the cabinets," the plural noun "cabinets" increases the plurality of the complex noun phrase as a whole. This predicts longer reading times at its predicate in both grammatical ("The key to the cabinets was rusty") and ungrammatical sentences ("The keys to the cabinets was rusty"). However, as discussed previously, although results have been mixed, many studies (e.g., Dillon et al., 2013; Jäger et al., 2020; Wagers et al., 2009) have observed such effects in ungrammatical but not grammatical sentences, a finding which has been taken as evidence of cue-based memory retrieval rather than representational accounts.

Although our findings for the plausible filler conditions in Experiment 1 (total viewing times) and Experiment 3 may be compatible with representational accounts, we are cautious of drawing strong conclusions here for several reasons. First, representational accounts typically make predictions about morphosyntactic agreement features, and it is not clear that they readily predict interference as a result of semantic features (see also Laurinavichyute & von der Malsburg, 2020). Second, representational accounts predict processing difficulty based on how information is encoded within a complex noun phrase. Given that the distractor is not immediately embedded within the filler noun phrase in our materials (e.g., "Mary saw the beer$_{[\text{filler}]}$ that the man with the wine$_{[\text{distractor}]}$ very happily drank"), it is not clear how such accounts could explain interference in filler-gap dependencies. It is conceivable that an extension of representational accounts could explain our findings, though the extent to which such an extension can be made on independent theoretical grounds is not clear.

We also note that the size of interference was larger in the implausible filler than plausible filler conditions. Indeed, in regression path times in Experiment 1, although we observed significant facilitatory interference in the implausible filler conditions, the plausible filler conditions did not differ significantly in this measure. In Experiment 3, the effect size for facilitatory interference in the implausible filler conditions was larger than the difference observed in the plausible filler conditions. Traditional representational accounts (Pearlmutter et al., 1999) would predict an effect of a similar size in both the plausible and implausible filler conditions, but this was not observed across our experiments. For subject–verb agreement, Hammerly et al. (2019) recently argued that smaller effects may appear in grammatical than ungrammatical conditions as a result of response bias. They noted that in speeded judgment studies on subject–verb agreement, participants often respond overall less accurately to ungrammatical than grammatical sentences. This response bias they argued can explain the larger effects often observed in ungrammatical than grammatical sentences, and that when this bias is accounted for, a pattern of results more compatible with representational accounts emerges.

It is not clear that this account can explain our findings, however. In Experiment 3, participants were actually more accurate in the implausible filler conditions than the plausible filler conditions. Yet, we still observed a larger effect of facilitatory interference in the implausible filler conditions than the effect observed in the plausible filler conditions. This pattern of results is the opposite to what would be predicted based on response bias. That is, if response bias influenced our results, we would expect a larger difference between the two plausible filler conditions because these conditions were judged on the whole less accurately than the two implausible filler conditions, but the opposite pattern was observed.

In sum, although we acknowledge that the extent to which representational accounts can explain our findings is a matter of debate and future research, we argue that our results are compatible with the prediction that filler-gap dependency formation involves a cue-based memory retrieval mechanism, perhaps as a last resort following revision as assumed by Wagers et al. (2009).

Filler-Gap Dependency Formation and Memory Retrieval Operations in L1 and L2 Processing

The present study did not find significant evidence of L1/L2 differences in filler-gap dependency formation or susceptibility to similarity-based interference. The filler plausibility effects observed in Experiment 1 demonstrated that when L2 participants encountered the gap during sentence processing, they immediately recovered the information of the filler. This finding is inconsistent with an account of L2 processing that relies on linear surface distance, which would have predicted that L2 speakers should be influenced more by the plausibility of the distractor than the filler. In implausible filler-gap sentences, L1 and L2 participants showed attenuated implausibility effects when the distractor was a plausible object compared with when it was an implausible object. We also observed this facilitatory interference in Experiment 3 as well, where L1 and L2 participants judged implausible filler-gap sentences to be plausible more often in the plausible distractor than implausible distractor conditions. These findings suggest that a cue-based memory retrieval mechanism plays a role in both L1 and L2 language comprehension. However, there was no evidence that L2 speakers are more susceptible to similarity-based interference than L1 speakers (cf. Cunnings, 2017), nor did we find any evidence that would be indicative of shallow L2 processing as operationalized in terms of linear distance in the present study.

There are several possible accounts of why the present study did not observe L1/L2 differences while some previous studies did. One is that L1/L2 differences in filler-gap dependencies arise only in certain circumstances, such as in sentences with filler-gap dependencies spanning several clauses or with increased processing cost due to revision (Fujita & Cunnings, 2020; Marinis et al., 2005). Another is that L2 speakers are susceptible to discourse saliency as suggested by Felser and Cunnings (2012). Felser and Cunnings claimed that in their materials, the distractor was salient at the discourse level because it was in a subject position. By contrast, in our materials, the distractor is embedded within a prepositional phrase. Thus, the distractor is likely to be in a less salient state (Engelmann et al., 2019; Lowder & Gordon, 2015). Such differences may account for the different L2 processing patterns observed between previous L2 sentence processing studies and the present study. It is also possible however that previous studies have overestimated L1/L2 differences owing to small sample sizes. Although we acknowledge that the absence of L1/L2 differences in the present study may be attributable to the lack of statistical power, our sample size is larger than the previous L2 studies that have observed L1/L2 differences in filler-gap dependency formation (e.g., Fujita & Cunnings, 2020; Marinis et al., 2005) and interference in reflexive resolution (Felser et al., 2009; Felser & Cunnings, 2012). Given that a small sample size potentially leads to overestimates of effect sizes (Vasishth et al., 2018), future research will need to test the replicability of the present study and previous ones. Given the heterogeneous language experiences of L2 speakers, we also do not draw conclusions about how various factors, such as L2 proficiency, age of acquisition, and L1 background, among others, may influence interference and filler-gap dependency resolution in L2 speakers. We leave this issue also as an avenue for future research.

Conclusion

The present study reported three experiments investigating filler-gap dependency formation and similarity-based interference in L1 and L2 language comprehension. These experiments demonstrated that L1 and L2 speakers retrieve the filler upon encountering a gap and are susceptible to facilitatory interference. None of the experiments observed significant L1/L2 differences that would be suggestive of either shallow parsing or increased susceptibility to interference in L2 speakers. We interpret these results as indicating that a cue-based memory retrieval mechanism operates during filler-gap dependency formation in L1 and L2 sentence processing, especially when the filler does not fully match the set of retrieval cues at the verb.

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. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press. 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). https://doi.org/10.18637/jss.v067.i01

Cheng, Y., Rothman, J., & Cunnings, I. (2021). Parsing preferences and individual differences in nonnative sentence processing: Evidence from eye movements. *Applied Psycholinguistics*, *42*(1), 129–151. https://doi.org/10.1017/S014271642000065X

Chomsky, N. (1981). *Lectures on government and binding: The Pisa lectures*. Foris.

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

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

Clahsen, H., & Felser, C. (2006c). How native-like is non-native language processing? *Trends in Cognitive Sciences*, *10*(12), 564–570. https://doi.org/10.1016/j.tics.2006.10.002

Crocker, M. W. (1996). *Computational psycholinguistics: An interdisciplinary approach to the study of language* (Vol. 20). Springer. https://doi.org/10.1007/978-94-009-1600-5

Cunnings, I. (2017). Parsing and working memory in bilingual sentence. *Bilingualism: Language and Cognition*, *20*(4), 659–678. https://doi.org/10.1017/S1366728916000675

Cunnings, I., & Fujita, H. (2021a). Similarity-based interference and relative clauses in second language processing. *Second Language Research*. Advance online publication. https://doi.org/10.1177/02676583211063534

Cunnings, I., & Fujita, H. (2021b). Quantifying individual differences in native and nonnative sentence processing. *Applied Psycholinguistics*, *42*(3), 579–599. https://doi.org/10.1017/S0142716420000648

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

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

Engelmann, F., Jäger, L. A., & Vasishth, S. (2019). The effect of prominence and cue association on retrieval processes: A computational account. *Cognitive Science*, *43*(12), e12800. https://doi.org/10.1111/cogs.12800

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

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

Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry*, *9*(3), 427–473.

Foote, R. (2011). Integrated knowledge of agreement in early and late English–Spanish bilinguals. *Applied Psycholinguistics*, *32*(1), 187–220. https://doi.org/10.1017/S0142716410000342

Frazier, L., & Clifton, C., Jr. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, *4*(2), 93–126. https://doi.org/10.1080/01690968908406359

Frazier, L., & Flores d'Arcais, G. B. (1989). Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language*, *28*(3), 331–344. https://doi.org/10.1016/0749-596X(89)90037-5

Fujita, H. (2021). On the parsing of garden-path sentences. *Language, Cognition and Neuroscience*, *36*(10), 1234–1245. https://doi.org/10.1080/23273798.2021.1922727

Fujita, H., & Cunnings, I. (2020). Reanalysis and lingering misinterpretation of linguistic dependencies in native and non-native sentence comprehension. *Journal of Memory and Language*, *115*, 104154. https://doi.org/10.1016/j.jml.2020.104154

Fujita, H., & Cunnings, I. (2021a). Lingering misinterpretation in native and nonnative sentence processing: Evidence from structural priming. *Applied Psycholinguistics*, *42*(2), 475–504. https://doi.org/10.1017/S0142716420000351

Fujita, H., & Cunnings, I. (2021b). Reanalysis processes in non-native sentence comprehension. *Bilingualism: Language and Cognition*, *24*(4), 628–641. https://doi.org/10.1017/S1366728921000195

Gibson, E. A. F., & Warren, T. (2004). Reading-time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax*, *7*(1), 55–78. https://doi.org/10.1111/j.1368-0005.2004.00065.x

González Alonso, J., Cunnings, I., Fujita, H., Miller, D., & Rothman, J. (2021). Gender attraction in sentence comprehension. *Glossa: A Journal of General Linguistics*, *6*(1), 20. https://doi.org/10.5334/gjgl.1300

Hammerly, C., Staub, A., & Dillon, B. (2019). The grammaticality asymmetry in agreement attraction reflects response bias: Experimental and modeling evidence. *Cognitive Psychology*, *110*, 70–104. https://doi.org/10.1016/j.cogpsych.2019.01.001

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

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

Jäger, L. A., Engelmann, F., & Vasishth, S. (2015). Retrieval interference in reflexive processing: Experimental evidence from Mandarin, and computational modeling. *Frontiers in Psychology*, *6*, 617. https://doi.org/10.3389/fpsyg.2015.00617

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

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

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, *25*(4), 603–634. https://doi.org/10.1017/S0142716404001298

Keating, G. D. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning*, *59*(3), 503–535. https://doi.org/10.1111/j.1467-9922.2009.00516.x

Kim, E., Baek, S., & Tremblay, A. (2015). The role of island constraints in second language sentence processing. *Language Acquisition*, *22*(4), 384–416. https://doi.org/10.1080/10489223.2015.1028630

Kim, N., Brehm, L., Sturt, P., & Yoshida, M. (2020). How long can you hold the filler: Maintenance and retrieval. *Language, Cognition and Neuroscience*, *35*(1), 17–42. https://doi.org/10.1080/23273798.2019.1626456

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, *2*(1), 15–47. https://doi.org/10.1016/0010-0277(72)90028-5

Lago, S., & Felser, C. (2018). Agreement attraction in native and nonnative speakers of German. *Applied Psycholinguistics*, *39*(3), 619–647. https://doi.org/10.1017/S0142716417000601

Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement attraction in Spanish comprehension. *Journal of Memory and Language*, *82*, 133–149. https://doi.org/10.1016/j.jml.2015.02.002

Laurinavichyute, A., & von der Malsburg, T. (2020). *Semantic attraction in sentence comprehension*. https://doi.org/10.31234/osf.io/hk9nc

Lee, M.-W. (2004). Another look at the role of empty categories in sentence processing (and grammar). *Journal of Psycholinguistic Research*, *33*(1), 51–73. https://doi.org/10.1023/B:JOPR.0000010514.50468.30

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

Lim, J. H., & Christianson, K. (2015). Second language sensitivity to agreement errors: Evidence from eye movements during comprehension and translation. *Applied Psycholinguistics*, *36*(6), 1283–1315. https://doi.org/10.1017/S0142716414000290

Lowder, M. W., & Gordon, P. C. (2015). Focus takes time: Structural effects on reading. *Psychonomic Bulletin & Review*, *22*(6), 1733–1738. https://doi.org/10.3758/s13423-015-0843-2

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, *27*(01), 53–78. https://doi.org/10.1017/S0272263105050035

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, *55*(3), 381–401. https://doi.org/10.1016/j.jml.2006.06.006

McElree, B., & Bever, T. G. (1989). The psychological reality of linguistically defined gaps. *Journal of Psycholinguistic Research*, *18*(1), 21–35. https://doi.org/10.1007/BF01069044

Meng, M., & Bader, M. (2021). Does comprehension (sometimes) go wrong for noncanonical sentences? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *74*(1), 1–28. https://doi.org/10.1177/1747021820947940

Nakano, Y., Felser, C., & Clahsen, H. (2002). Antecedent priming at trace positions in Japanese long-distance scrambling. *Journal of Psycholinguistic Research*, *31*(5), 531–571. https://doi.org/10.1023/A:1021260920232

Nicenboim, B., Vasishth, S., Engelmann, F., & Suckow, K. (2018). Exploratory and confirmatory analyses in sentence processing: A case study of number interference in German. *Cognitive Science*, *42*(S4), 1075–1100. https://doi.org/10.1111/cogs.12589

Nicol, J., & 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., & 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., 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

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in l1 and l2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition*, *25*(4), 501–528. https://doi.org/10.1017/S0272263103000214

Parker, D. (2017). Processing multiple gap dependencies: Forewarned is forearmed. *Journal of Memory and Language*, *97*, 175–186. https://doi.org/10.1016/j.jml.2017.08.003

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, *41*(3), 427–456. https://doi.org/10.1006/jmla.1999.2653

Phillips, C. (2006). The real-time status of island phenomena. *Language*, *82*(4), 795–823. https://doi.org/10.1353/lan.2006.0217

Pickering, M. J., & Guy, B. (1991). Sentence processing without empty categories. *Language and Cognitive Processes*, *6*(3), 229–259. https://doi.org/10.1080/01690969108406944

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

Pliatsikas, C., & Marinis, T. (2013). Processing empty categories in a second language: When naturalistic exposure fills the (intermediate) gap. *Bilingualism: Language and Cognition*, *16*(1), 167–182. https://doi.org/10.1017/S136672891200017X

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

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ross, J. R. (1967). *Constraints on variables in syntax* [PhD Thesis, Massachusetts Institute of Technology]. http://hdl.handle.net/1721.1/15166

Schad, D. J., Betancourt, M., & Vasishth, S. (2020). Toward a principled Bayesian workflow in cognitive science. *Psychological Methods*, *26*(1), 103–126. https://doi.org/10.1037/met0000275

Schlueter, Z., Williams, A., & Lau, E. (2018). Exploring the abstractness of number retrieval cues in the computation of subject-verb agreement in comprehension. *Journal of Memory and Language*, *99*, 74–89. https://doi.org/10.1016/j.jml.2017.10.002

Smith, G., & Vasishth, S. (2020). A principled approach to feature selection in models of sentence processing. *Cognitive Science*, *44*(12), e12918. https://doi.org/10.1111/cogs.12918

Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, *3*(1), 227–245. https://doi.org/10.1080/01690968608407062

Tanner, D., Nicol, J., Herschensohn, J., & Osterhout, L. (2012). Electrophysiological markers of interference and structural facilitation in native and nonnative agreement processing. In A. K. Biller, E. Y. Chung & A. E. Kimball (Eds.), *Proceedings of the 36th Boston University Conference on Language Development* (pp. 594–606). Cascadilla.

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, *35*(3), 454–475. https://doi.org/10.1006/jmla.1996.0025

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 407–430. https://doi.org/10.1037/0278-7393.33.2.407

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, *49*(3), 285–316. https://doi.org/10.1016/S0749-596X(03)00081-0

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, *55*(2), 157–166. https://doi.org/10.1016/j.jml.2006.03.007

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, *65*(3), 247–263. https://doi.org/10.1016/j.jml.2011.05.002

Vasishth, S., Mertzen, D., Jäger, L. A., & Gelman, A. (2018). The statistical significance filter leads to overoptimistic expectations of replicability. *Journal of Memory and Language*, *103*, 151–175. https://doi.org/10.1016/j.jml.2018.07.004

Vasishth, S., Nicenboim, B., Engelmann, F., & Burchert, F. (2019). Computational models of retrieval processes in sentence processing. *Trends in Cognitive Sciences*, *23*(11), 968–982. https://doi.org/10.1016/j.tics.2019.09.003

Wagers, M. W., & 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., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*(2), 206–237. https://doi.org/10.1016/j.jml.2009.04.002

Williams, J. N., Möbius, P., & Kim, C. (2001). Native and non-native processing of English wh - questions: Parsing strategies and plausibility constraints. *Applied Psycholinguistics*, *22*(4), 509–540. https://doi.org/10.1017/S0142716401004027

Yoshida, M., Dickey, M. W., & Sturt, P. (2013). Predictive processing of syntactic structure: Sluicing and ellipsis in real-time sentence processing. *Language and Cognitive Processes*, *28*(3), 272–302. https://doi.org/10.1080/01690965.2011.622905

Received July 2, 2021
Revision received January 20, 2022
Accepted January 25, 2022 ■