Applied Psycholinguistics (2024), **45**, 81–109
doi:10.1017/S0142716423000498

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

Subject-verb dependency formation and semantic interference in native and non-native language comprehension

Hiroki Fujita[1] and Ian Cunnings[2]

[1]Department of Linguistics, University of Potsdam, Potsdam, Germany and [2]School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK
Corresponding author: Hiroki Fujita; Email: hiroki.fujita@uni-potsdam.de

(Received 17 February 2023; revised 3 November 2023; accepted 4 November 2023; first published online 18 December 2023)

Abstract

Differences between native (L1) and non-native (L2) comprehension have been debated. This study explores whether a source of potential L1/L2 differences lies in susceptibility to memory-based interference during dependency formation. Interference effects are known to occur in sentences like *The key to the cabinets were rusty*, where ungrammaticality results from a number mismatch between the sentence subject and verb. Such sentences are sometimes misperceived as grammatical due to the presence of a number-matching "distractor" ("the cabinets"). Interference has been well-examined in a number agreement. However, whether and how forming thematic relations is susceptible to interference remains underexplored in L1 and L2 language comprehension. In six preregistered experiments, we investigated semantic interference in language comprehension and explored whether potential L1/L2 differences can be attributed to different degrees of susceptibility to interference. The results did not show that L2 speakers are more susceptible to interference than L1 speakers. Also, the observed interference patterns were only partially consistent with existing theories of memory retrieval during comprehension. We discuss how these theories may be reconciled with our findings and argue our results suggest that similar processes are involved in L1 and L2 subject-verb dependency formation.

Keywords: interference; language comprehension; memory retrieval; second language processing; subject-verb dependency

Introduction

Why non-native (L2) speakers may process language differently from native (L1) speakers is debated. Some argue that L2 comprehension relies less on syntactic information than L1 comprehension (Clahsen & Felser, 2006), while others attribute L1/L2 differences to lexical processing ability or memory-based demands (Cunnings, 2017; Hopp, 2022). This study investigates the claim that potential

L1/L2 differences lie in memory retrieval processes during dependency formation (Cunnings, 2017), as in (1a/b) below.

(1a) The thief that was near the robber stole the diamond.

(1b) The thief that was near the locker stole the diamond.

In (1a/b), "The thief" is the subject NP, and "stole" serves as its predicate, but how is this *subject-verb dependency* established in comprehension? Cue-based models of memory retrieval posit that comprehenders retrieve the subject NP (*the target*) from memory upon encountering "stole" by utilizing information available at the verb as retrieval cues (Lewis & Vasishth, 2005). In (1a/b), a subject-verb dependency must be formed between structurally licensed elements, with the two elements being semantically appropriate. According to cue-based models, comprehenders use structure-based (but see Kush, 2013) and semantic-based cues to retrieve the target. Structural cues in this case may include [+SUBJECT], to guide retrieval to a structurally licensed element. Semantic cues may stem from the lexical properties of verbs, like [+STEALER] for "stole" (Cunnings & Sturt, 2018). A crucial assumption of this hypothesis is that any elements in memory that (partially) match retrieval cues interfere with target retrieval. In (1a), "the robber" (*a distractor*) matches the semantic cue, as it is a plausible (but not structurally licensed) actor of the verb. Therefore, the cue-based model predicts interference, resulting in processing difficulty.

Cunnings (2017) proposed that L2 speakers weight retrieval cues differently from L1 speakers and consequently become more susceptible to interference. Cunnings predicted that L2 speakers may weight structural cues less than L1 speakers, and conversely weight non-structural cues, such as semantic cues, higher. This would predict increased interference in sentences like (1a). We tested this proposal in six preregistered experiments. Our results showed semantic interference in L1 and L2 speakers, but no consistent evidence of increased interference in the L2. However, the observed interference patterns were not fully consistent with cue-based models. Below, we begin by discussing interference effects and then discuss potential similarities and differences between L1 and L2 comprehension.

Similarity-based interference

Research has indicated that linguistic dependency formation is susceptible to interference (e.g., Wagers et al., 2009; for review, see Jäger et al., 2017), and according to the cue-based model (Lewis & Vasishth, 2005), interference results from the similarity between elements in memory to a set of retrieval cues (*similarity-based interference*). The cue-based model postulates that during memory retrieval, elements that match a set of retrieval cues cause interference. This predicts two types of interference, dependent on how well the intended retrieval target matches the retrieval cues: *inhibitory interference* and *facilitatory interference* (Jäger et al., 2017). Inhibitory interference is predicted in sentences like (1a/b) (so-called *target-match* sentences). Here, the target ("the thief") fully matches the retrieval cues at the verb, in that it is both [+SUBJECT] and [+STEALER]. However, in (1a), the distractor

https://doi.org/10.1017/S0142716423000498 Published online by Cambridge University Press

("the robber") also matches the semantic cue [+STEALER], which, according to the cue-based model, should lead to longer processing times compared to (1b), where the distractor ("the locker") does not match any cues.

Facilitatory interference is predicted in *target-mismatch* sentences, when the retrieval target does not fully match the retrieval cues, as in (2a/b).

(2a) The table that was near the robber stole the diamond.

(2b) The table that was near the locker stole the diamond.

In (2a/b), the target is an implausible actor of the critical verb ("The table . . . stole") and thus mismatches the semantic cue. Such sentences cause processing difficulty and are judged as implausible (*implausibility effects*; Fujita & Cunnings, 2022). Crucially, the distractor matches the semantic cue in (2a) but mismatches in (2b). According to the cue-based model, the partial match in (2a) results in the potential retrieval of the distractor, leading to shorter reading times after "stole" is encountered in (2a) than (2b).

Similarity-based interference has been widely investigated (Fujita & Cunnings, 2023; González Alonso et al., 2021; Jäger et al., 2020; Kim et al., 2020; Lago et al., 2015; Van Dyke, 2007). However, several unresolved issues remain. One such issue is that while facilitatory interference has been observed frequently, evidence of inhibitory interference is inconclusive (Nicenboim et al., 2018; Wagers et al., 2009). We refer to this as the *target-mismatch asymmetry*. Wagers et al. claimed that similarity-based interference only arises during revision, when comprehenders have difficulty forming grammatical dependencies and so check the target's features. Since this difficulty occurs in target-mismatch but not target-match sentences, Wagers et al. argued that dependency formation is susceptible to only facilitatory interference. This argument is also supported by the finding that interference effects are delayed (Sturt, 2003). Sturt, for example, observed interference only in eye-tracking measures indicating later stages of processing. Alternatively, Nicenboim et al. (2018) argued that many studies are underpowered and that large samples are required to observe inhibitory interference.

There has also been discussion on what factors affect interference, such as a distractor's argument status (Parker & An, 2018; Van Dyke & McElree, 2011). Parker and An (2018) tested sentences like (3a/b).

(3a) The waitress who sat the girl(s) unsurprisingly was unhappy about the noise.

(3b) The waitress who sat near the girl(s) unsurprisingly were unhappy about the noise.

(3a/b) are ungrammatical because the subject NP and verb disagree in number. (3a/b) also manipulate the number of the distractor. Thus, the cue-based model predicts facilitatory interference in (3a/b) when the distractor matches the verb's number. Consistent with this prediction, Parker and An observed facilitatory interference in (3b). However, (3a) did not show facilitatory interference. Parker and An attributed this finding to the distractor being a direct object of a verb (*a core argument*) in (3a)

https://doi.org/10.1017/S0142716423000498 Published online by Cambridge University Press

but inside a prepositional phrase (*an oblique argument*) in (3b). According to Parker and An, core arguments are encoded distinctly in memory and consequently, easily rejected as retrieval candidates, nullifying interference.

A distractor's position may also influence interference in various ways. For example, research has reported poorer sentence comprehension when a distractor intervened between the target and retrieval probe compared to when it preceded the target (e.g., Van Dyke & McElree, 2011). A distractor's position may also contribute to its saliency, which may influence memory retrieval. As discussed earlier, the cue-based model predicts inhibitory interference in target-match conditions when a distractor matches a retrieval cue. However, some studies have observed interference patterns opposite to inhibitory interference, especially in reflexive resolution (i.e., processing difficulty in target-match/distractor-mismatch conditions; Cunnings & Felser, 2013; Sturt, 2003). Crucially, these studies had a distractor in a subject position and as a discourse topic. Based on these studies, Engelmann et al. (2019) revised the original cue-based model by Lewis and Vasishth (2005), implementing saliency of elements. Their model predicted interference opposite to inhibitory interference when a distractor's saliency is high enough, as in a subject position. According to Engelmann et al., inhibitory interference in target-match sentences gets attenuated because, as the distractor's saliency increases, it is more likely to get retrieved than the target, counteracting inhibitory interference. In this paper, we use *the cue-based model* to cover both the Lewis and Vasishth and Engelmann et al. models, and where appropriate refer to each model separately, when their predictions diverge.

A final controversial issue is that evidence of similarity-based interference mainly comes from certain types of agreement, and semantic interference as expected in sentences like (1/2) remains underexplored (Cunnings & Sturt, 2018; Laurinavichyute & von der Malsburg, 2022; Van Dyke, 2007).

L1/L2 language comprehension

How L2 speakers process sentences compared to L1 speakers has been debated (Clahsen & Felser, 2006, 2018; Cunnings, 2017; Hopp, 2022). Some studies have observed similar processing patterns between L1 and L2 speakers (Cunnings & Fujita, 2021, 2023; Fujita & Cunnings, 2021a, 2021b, 2022; Ito et al., 2018; Lago & Felser, 2018; Omaki & Schulz, 2011; Tanner et al., 2012; Williams et al., 2001), whereas others have observed different patterns (Deniz, 2022; Felser et al., 2003, 2009; Felser & Cunnings, 2012; Fujita & Cunnings, 2020; Jacob & Felser, 2016; Jiang, 2004; Keating, 2009; Marinis et al., 2005; Papadopoulou & Clahsen, 2003; Patterson et al., 2014; Roberts & Felser, 2011). Although different accounts of potential L1/L2 differences exist (Clahsen & Felser, 2006; Cunnings, 2017; Hopp, 2014), crucial to the present study is Cunnings (2017), who proposed that L2 speakers may be particularly susceptible to memory-based interference. Cunnings argued that L2 speakers weight retrieval cues differently than L1 speakers, leading to larger interference effects for L2 speakers. Specifically, if L2 speakers underweight structural cues and instead heavily weight non-structural cues, such as the semantic cue in (1a/b), interference effects in (1a) should become larger for L2 speakers than L1 speakers. That is, if L2 speakers heavily weight the semantic cue, distinguishing between the target and distractor in (1a) should become particularly difficult.

https://doi.org/10.1017/S0142716423000498 Published online by Cambridge University Press

Although few studies have investigated L1/L2 differences in semantic interference, some research is compatible with the idea that L1/L2 speakers weigh retrieval cues differently. For example, Felser et al. (2009) examined reflexive resolution using sentences like (4a/b).

(4a) John noticed that Richard had cut himself with a very sharp knife.

(4b) Jane noticed that Richard had cut himself with a very sharp knife.

(4a/b) contain "himself," which must corefer with a masculine NP in a structurally appropriate position (Chomsky, 1981; Dillon et al., 2013; Fujita, 2021b, 2023a, 2023b; Sturt, 2003; Yoshida et al., 2013). Here, the cue-based model posits that structure-based and gender-based cues are used for retrieving the reflexive's antecedent. The grammatical antecedent is "Richard." (4a/b) also have a distractor in the matrix clause and manipulate whether its gender matches (4a: "John") or mismatches (4b: "Jane") the reflexive. The Lewis and Vasishth (2005) model predicts inhibitory interference in (4a) due to the gender match between the distractor and reflexive. In an eye-movement-during-reading task, Felser et al. observed inhibitory interference for L2 speakers but not L1 speakers, suggesting that L2 speakers give particular weight to non-structural cues (see also Felser & Cunnings, 2012 who observed L1/L2 differences in reflexive resolution). However, such L1/L2 differences did not arise when the distractor was in a less salient position (e.g., *It was clear to John/Jane that Richard had cut himself*...). These findings suggest increased interference in L2 speakers when distractors are salient.

Fujita and Cunnings (2022) investigated semantic interference in L1/L2 sentence processing. In their study, L1 and L2 speakers read so-called *filler-gap* sentences (Fodor, 1978), as in (5a/b), where "the beer"/"the cake" must serve as the object of "drank."

(5a) Mary saw the beer that the man with the wine/the food very happily drank at the party.

(5b) Mary saw the cake that the man with the wine/the food very happily drank at the party.

Fujita and Cunnings observed longer reading times at the verb "drank" in implausible (5b) than plausible (5a). This implausibility effect was attenuated in (5b) when the distractor was a plausible object of the verb ("the wine") compared to when it was not ("the food"). The distractor did not however influence reading times of plausible sentences, suggesting facilitatory interference in L1 and L2 speakers in the absence of inhibitory interference. Crucially, no significant L1/L2 differences in the size of interference were observed.

While Fujita and Cunnings (2022) examined semantic interference in filler-gap dependencies, we are unaware of any existing studies on semantic interference in L2 processing of subject-verb dependencies. Numerous studies have examined interference in subject-verb number agreement, with mixed results (Armstrong et al., 2018; Cai et al., 2007; Foote, 2011; Jiang, 2004; Lago & Felser, 2018; Lee & Phillips, 2023; Tanner et al., 2012, 2013). In general, these studies indicate that both

https://doi.org/10.1017/S0142716423000498 Published online by Cambridge University Press

L1 and L2 speakers are susceptible to morphosyntactic facilitatory interference (e.g., Lago & Felser, 2018; Tanner et al., 2012), but the extent to which subject-verb dependency formation is susceptible to semantic interference has not been examined.

In summary, why L1/L2 speakers may show different processing patterns remains controversial, and the extent to which memory-based interference may differ between L1 and L2 speakers has not been systematically examined. Also, no studies have systematically explored whether a distractor's argument status and saliency influence L2 memory retrieval.

The present study

We report six preregistered experiments examining potential L1/L2 differences in memory retrieval (Cunnings, 2017) by investigating semantic interference and the potential effects of a distractor's argument status on memory retrieval. For these aims, we employed self-paced reading (Experiments 1/3), offline comprehension (Experiments 2/4), and speeded judgment (Experiments 5/6) tasks. L1 participants acquired English since birth, while L2 participants started learning English in a school setting after age 5. All L2 participants had high English proficiency, as we intended to test theoretical claims about potential L1/L2 differences that remain at high levels of proficiency, and spoke German as their L1. We do not intend to necessarily draw any conclusions about L2 speakers who do not fit into these categories. All L1 participants spoke English as their first language and were born and raised in the UK.

The research designs, sampling method, and data analysis plan in Experiments 1–6 were preregistered. We recruited participants and conducted data analysis following the preregistrations. Our materials, data, analysis code, and preregistrations are available at https://osf.io/f7uwe/.

Experiment 1

Experiment 1 investigated interference in sentences like (6a–d).

(6a) *Plausible target, Plausible distractor*
The thief that the robber was near in the bank calmly stole the diamond last night.

(6b) *Plausible target, Implausible distractor*
The thief that the locker was near in the bank calmly stole the diamond last night.

(6c) *Implausible target, Plausible distractor*
The table that the robber was near in the bank calmly stole the diamond last night.

(6d) *Implausible target, Implausible distractor*
The table that the locker was near in the bank calmly stole the diamond last night.

https://doi.org/10.1017/S0142716423000498 Published online by Cambridge University Press

Regions

The thief/that the robber was near/in the bank/calmly stole/the diamond/last night.

In (6a–d), "The thief/table" is the retrieval target, and "stole" is the retrieval probe. (6a–d) contain a distractor ("the robber/locker") in a subject position embedded within the target. The target in (6a/b) is an animate NP and a plausible actor of the critical verb, whereas, in (6c/d), it is an inanimate NP and an implausible actor. The distractor is also either animate (6a/c) or inanimate (6b/d), thereby creating a plausibility dichotomy as an actor of the verb. We used the predicate "was near" across all items. We made this design choice to ensure that, across animate and inanimate conditions, sentences were equally felicitous.

If L1 speakers retrieve the target at the verb, their reading times should be longer at "stole" in (6c/d) than (6a/b) due to implausibility effects. Regarding interference, the original Lewis and Vasishth model predicts longer reading times at "stole" in (6a) than (6b) due to inhibitory interference and shorter reading times at "stole" in (6c) than (6d) because of facilitatory interference. However, subjecthood may increase saliency. If saliency influences memory retrieval, as predicted by the revised Engelmann et al. (2019) model, reading times may be longer in implausible-distractor than plausible-distractor conditions, irrespective of the target's plausibility. Alternatively, if a distractor resists interference when it is a core argument (Parker & An, 2018), interference effects should be absent, given that the distractor is a core argument. L2 speakers are expected to show implausibility effects (Fujita & Cunnings, 2022), and if they are more susceptible to interference, they should show larger interference effects than L1 speakers (Cunnings, 2017).

Participants

Ninety-six L1 English (mean age 21, range 18–35) and 96 L1 German, L2 English (mean age 27, range 18–40) speakers, recruited via Prolific (https://www.prolific.com), participated in Experiment 1.[1] L2 participants had a mean age of English onset of 9 (range 5–14). Some L2 participants reported knowledge of other languages, but we did not use this as an exclusionary criterion. After Experiment 1, the L2 participants completed the Quick Placement Test (QPT). Their average score was 53 (range 34–60).[2]

Materials

Experiment 1 had 24 sets of experimental materials like (6a–d) and 72 fillers, which consisted of various syntactic structures. Four experimental lists were constructed such that each participant saw six experimental sentences from each condition in a Latin-square design. A yes/no question, which did not probe the dependency, followed all sentences to ensure participant attention.

Procedure

We administered a non-cumulative phrase-by-phrase self-paced reading task in PCIbex Farm (Zehr & Schwarz, 2018), using code provided online (Fujita, 2021a). Before each sentence was presented, a sequence of dashes masking a whole sentence appeared. Participants pressed the space bar to read each phrase. Phrases for

https://doi.org/10.1017/S0142716423000498 Published online by Cambridge University Press

experimental sentences are illustrated under (6a–d). A question appeared after participants read the last phrase. The experiment began with four practice trials. Experimental sentences and fillers for all experiments in the present study were pseudo-randomized in a different order for each participant.

Preregistered data analysis

Following the preregistration, we analyzed reciprocally transformed reading times at the critical ("calmly stole") and spillover ("the money") regions as the dependent variable in R (R Core Team, 2020). We fit linear mixed models using the lme4 package (Bates et al., 2015). Before data analysis, reading times shorter than 100ms or longer than 10s at each segment were excluded, as these likely index lapses of attention. The fixed effects were sum-coded (1/–1) main effects of group (L1/L2), region (critical/spillover regions), target (plausible/implausible), distractor (plausible/implausible), and their interactions. Random effects included random intercepts and all relevant slopes for participants and materials. The models also included a random intercept for trial to account for the two non-independent data points from each trial because of the fixed effect of region. When this maximal model (Barr et al., 2013) failed to converge, we simplified it by initially removing the random correlations and then iteratively removing the random effect with the smallest variance until the model converged. We estimated p values from the t distribution (Baayen, 2008) and interpreted p values smaller than .05 as significant. Interactions were examined with additional models containing nested contrasts.

Results

Average comprehension accuracy was 93% (range 76–100) for both L1 and L2 participants, indicating that participants paid attention during the task. Reading times are shown in Figure 1, and the statistical analysis is provided in Table 1.[3]

Critical/spillover regions

Analysis showed a significant main effect of group, with shorter reading times for L1 than L2 participants. There was also a significant main effect of target indicating implausibility effects, with longer reading times in (6c/d) than (6a/b). This main effect of target significantly interacted with region due to the presence of implausibility effects only at the spillover region (critical region: Estimate = –0.008, SE = 0.01, t = –0.98, p = .328; spillover region: Estimate = –0.063, SE = 0.01, t = –7.63, $p < .001$). There was also a significant four-way interaction, but a follow-up analysis examining the effect of distractor at each level of group, region and target did not show any significant effects for either L1 participants (all Estimate < 0.010, SE = 0.01 or 0.02, all $t < 0.60$, all $p > .545$) or L2 participants (all Estimate < 0.025, SE = 0.01 or 0.02, all $t < 1.69$, all $p > .092$).

https://doi.org/10.1017/S0142716423000498 Published online by Cambridge University Press

Table 1. Statistical analysis for Experiment 1 (critical/spillover regions; "calmly stole/the diamond")

	Estimate	SE	*t*	*p*
Intercept	1.729	0.03	52.95	<.001
Group	−0.133	0.03	−4.54	<.001
Region	0.131	0.01	10.13	<.001
Target	−0.035	0.01	−6.23	<.001
Distractor	0.004	0.01	0.60	.546
Group:Region	0.012	0.01	1.77	.077
Group:Target	−0.004	0.01	−0.64	.522
Group:Distractor	0.003	0.01	0.55	.585
Region:Target	−0.027	0.01	−4.60	<.001
Region:Distractor	0.000	0.01	0.02	.981
Target:Distractor	−0.002	0.01	−0.36	.717
Group:Region:Target	0.003	0.00	0.59	.556
Group:Region:Distractor	−0.002	0.00	−0.47	.636
Group:Target:Distractor	0.000	0.01	−0.03	.973
Region:Target:Distractor	0.004	0.00	0.88	.380
Group:Region:Target:Distractor	0.009	0.00	1.99	.047

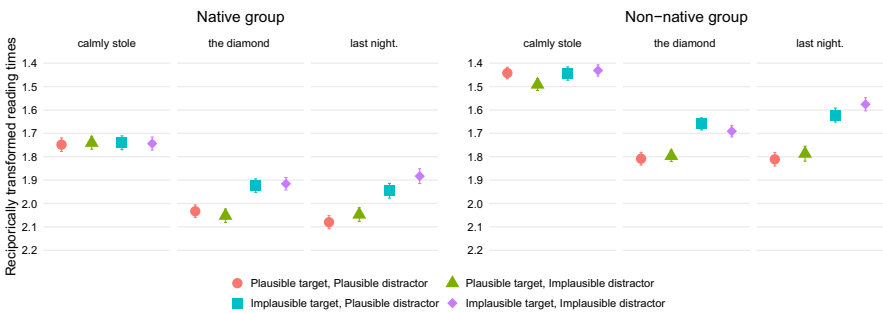

Figure 1. Reciprocally transformed reading times at the critical, spillover, and sentence-final regions in Experiment 1. The y axes are flipped for illustrative purposes.

Exploratory analysis: sentence-final region

Our preregistered analysis at the critical/spillover regions did not show clear interference effects. However, this might be because interference effects are delayed (Sturt, 2003). Hence, we exploratorily analyzed reading times at the sentence-final region. In addition to a significant main effect of target (Estimate = −0.087, SE = 0.01, t = −8.41, p < .001), this analysis suggested a significant main effect of distractor (Estimate = −0.021, SE = 0.01, t = −2.58, p = .010), indicating longer

https://doi.org/10.1017/S0142716423000498 Published online by Cambridge University Press

reading times in implausible-distractor than plausible-distractor conditions, irrespective of the target's plausibility. The target by distractor interaction was not significant (Estimate = –0.006, SE = 0.01, t = –0.54, p = .590), nor were any interactions with group (all Estimate < 0.013, SE = 0.01, t < 1.06, p > .288).

Discussion

The observed implausibility effects indicate retrieval of the target after the retrieval probe. The preregistered analysis did not show interference, which may be due to the distractor's argument status (Parker & An, 2018). However, the exploratory analysis suggested interference from an implausible distractor at the sentence-final region in both plausible and implausible sentences. Observing this effect in implausible conditions is compatible with the Lewis and Vasishth model, but finding it in plausible conditions is not. It may however be consistent with the Engelmann et al. model, given that the distractor was in a subject position, thereby increasing its saliency. We do not draw strong conclusions from this analysis however, given that it was not part of the preregistered analysis, and the effect appeared only at the sentence-final region. We further examine this issue in Experiments 3, 5 and 6. Crucially, we did not find significant L1/L2 differences in interference, suggesting that L1 and L2 speakers weight retrieval cues similarly. Below we report Experiment 2, where we further investigated interference using an offline task.

Experiment 2

Experiment 2 examined interference in offline comprehension using sentences like (7a–d).

(7a) *Long-distance dependency, Plausible distractor*
The girl that the boy was near played with friends.

(7b) *Long-distance dependency, Implausible distractor*
The girl that the car was near played with friends.

(7c) *Short-distance dependency, Plausible distractor*
The boy was near the girl that played with friends.

(7d) *Short-distance dependency, Implausible distractor*
The car was near the girl that played with friends.

Question: Who played with friends? (The girl/The boy)

In (7a–d), the verb phrase "played with friends" forms a dependency with the target, "the girl," which is a plausible actor for all conditions. Each condition has a distractor, which is a plausible actor of the verb in (7a/c) and an implausible actor in (7b/d). Thus, (7a/c) contains two plausible NPs for the verb, whereas there is only one plausible NP in (7b/d), and as such, we can expect similarity-based interference in (7a/c). We also manipulated whether the distractor appears linearly before (7c/d)

https://doi.org/10.1017/S0142716423000498 Published online by Cambridge University Press

or after (7a/b) the target. A comprehension question with two answer options probing the actor of the critical predicate followed all conditions. The two answer options were always the target and distractor used in (7a/c). We used this design as including the implausible agent ("the car") as an answer option in conditions (7b/d) would allow participants to answer the question without attending to the sentence. We acknowledge however that this means that the questions would be easier in conditions (7b/d), as participants could answer by simply selecting the NP appearing in the sentence. We were more interested however in the relative differences between (7a/b) and (7c/d), and differences in accuracy of (7a/c) between L1/L2 participants, which obviate this issue.

We expected lower accuracy when the distractor was plausible (7a/c) than implausible (7b/d). This effect should be larger between (7a/b), where a long-distance dependency must be established between the target and retrieval probe, with the distractor linearly intervening between them (Van Dyke & McElree, 2011), compared to (7c/d), when the distance between the target and verb is short. What is crucial is whether L2 speakers show decreased comprehension accuracy when the distractor is plausible (7a/c). Such L1/L2 differences may not arise in (7b/d), where the distractor is an implausible actor. Also, L2 speakers' increased susceptibility to interference may be observed only in long-distance dependency conditions if L1/L2 differences are modulated by the strength of interference.

Participants

The same participants as in Experiment 1 participated in Experiment 2. The participants completed Experiment 2 immediately after Experiment 1.

Materials

Experiment 2 consisted of 24 sets of experimental sentences like (7a–d) and 72 fillers. Each sentence had a comprehension question with two answer options.

Procedure

We administered the comprehension question task in PCIbex Farm. Before each sentence was presented, participants saw an underline masking the whole sentence. Pressing the space bar replaced the underline with the sentence. When participants pressed the space bar again, the sentence disappeared, and a comprehension question with two answer options appeared. Half of the time the correct answer was on the left side and the other half on the right side. The task began with four practice trials.

Preregistered data analysis

We fit a logistic regression to comprehension accuracy rates using generalized linear mixed models. Independent variables were group (L1/L2), dependency (long-distance/short-distance), distractor (plausible/implausible), and their interactions. We fit models in the same manner as in Experiment 1.

https://doi.org/10.1017/S0142716423000498 Published online by Cambridge University Press

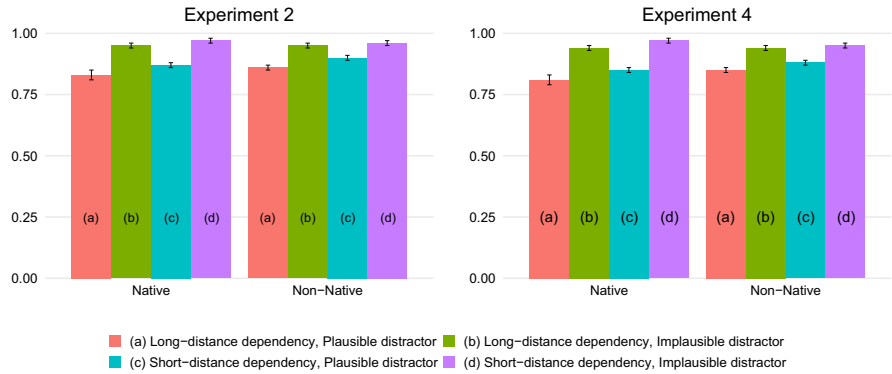

Figure 2. Comprehension accuracy for Experiments 2/4.

Results

The average comprehension accuracy of fillers was 92% (range 75–100) for L1 and L2 participants. Comprehension accuracy of the experimental materials is illustrated in Figure 2, and a summary of the statistical analysis is provided in Table 2.

The model showed a significant main effect of distractor, with lower comprehension accuracy in plausible-distractor than implausible-distractor conditions. There was a numerical tendency for lower accuracy in the long-distance conditions, but the main effect of interference type was not significant. There was also no significant main effect of group, nor any significant interactions.

Discussion

The results showed lower accuracy in plausible-distractor than implausible-distractor conditions, potentially suggesting similarity-based interference, although it may be due to questions being easier to answer in these conditions, as discussed in our description of (7a–d). Accuracy rates were numerically lower in long-distance than short-distance conditions, but this effect was not significant. Importantly, L2 speakers did not show increased interference compared to L1 speakers.

Experiments 1/2 suggested similar interference effects between L1 and L2 speakers. Experiment 3 further examined interference in an online task but also explored the indecisive evidence of interference observed in Experiment 1. Recall that Experiment 1 showed interference effects only at the sentence-final region but not at the critical/spillover regions. We discussed that this inconclusive evidence may be due to the distractor's status as a core argument, which Parker and An (2018) claimed resists interference. Also, the observed interference pattern was partially inconsistent with Lewis and Vasishth (2005) but compatible with the Engelmann et al. (2019) model, which predicts that inhibitory interference in grammatical sentences disappears when the distractor is salient. Experiment 3 explores these issues using a research design where the distractor appears as a less salient, oblique argument inside a prepositional phrase.

https://doi.org/10.1017/S0142716423000498 Published online by Cambridge University Press

Table 2. Statistical analyses for Experiments 2/4

	Experiment 2				Experiment 4			
	Estimate	SE	*z*	*p*	Estimate	SE	*z*	*p*
Intercept	2.887	0.13	21.84	<.001	2.771	0.15	18.90	<.001
Group	0.056	0.09	0.62	.532	0.060	0.08	0.72	.469
Dependency	0.166	0.09	1.94	.053	0.201	0.08	2.48	.013
Distractor	0.683	0.07	9.58	<.001	0.671	0.07	9.22	<.001
Group:Dependency	0.009	0.06	0.14	.886	−0.041	0.06	−0.67	.505
Group:Distractor	−0.094	0.09	−1.07	.283	−0.105	0.07	−1.61	.108
Dependency:Distractor	0.002	0.06	0.03	.980	0.044	0.06	0.73	.467
Group:Dependency:Distractor	0.016	0.07	0.24	.809	−0.021	0.08	−0.27	.790

Experiment 3

Experiment 3 examined interference using experimental sentences as below.

(8a) *Plausible target, Plausible distractor*
The thief that was near the robber in the bank calmly stole the diamond last night.

(8b) *Plausible target, Implausible distractor*
The thief that was near the locker in the bank calmly stole the diamond last night.

(8c) *Implausible target, Plausible distractor*
The table that was near the robber in the bank calmly stole the diamond last night.

(8d) *Implausible target, Implausible distractor*
The table that was near the locker in the bank calmly stole the diamond last night.

Region

The thief/that was near the robber/in the bank/calmly stole/the diamond/last night.

(8a–d) are identical to the sentences used in Experiment 1, except that the distractor is embedded within a prepositional phrase.

L1 speakers should show implausibility effects at "stole," with longer reading times in implausible-target conditions. The Lewis and Vasishth (2005) model predicts inhibitory interference in (8a) relative to (8b) and facilitatory interference in (8c) relative to (8d). The Engelmann et al. (2019) model makes the same predictions for (8a–d) because the distractor is now inside a prepositional phrase, a less salient position. Also, Experiment 3 may show clearer or stronger interference effects compared to Experiment 1, given that the distractor is now an oblique argument (Parker & An, 2018). L2 speakers are also expected to show implausibility effects at

https://doi.org/10.1017/S0142716423000498 Published online by Cambridge University Press

"stole" (Fujita & Cunnings, 2022), and if they are more susceptible to interference than L1 speakers, they should show larger interference effects (Cunnings, 2017).

Participants

Ninety-six L1 English speakers (mean age 24, range 18–40) and 96 L1 German, L2 English speakers (mean age 27, range 18–50; mean age of onset to English 9, range 5–13), none of whom took part in Experiments 1/2, participated in Experiment 3.[4] We recruited the participants via Prolific. L2 participants completed the QPT after the experiment. Their mean score was 51 (range 30–60).

Materials

Experiment 3 contained 24 sets of experimental materials like (8a–d) and 72 fillers. A yes/no question followed all sentences.

Procedure and preregistered data analysis

The procedure and preregistered data analysis are the same as Experiment 1.

Results

The mean comprehension accuracy of experimental and filler questions was 93% (range 75–99) for L1 and L2 participants. See Figure 3 for reading times and Table 3 for a summary of the statistical analysis.

Critical and spillover regions

Analysis showed a significant main effect of target and a significant region by target interaction, with nested contrasts indicating implausibility effects only at the spillover region (critical region: Estimate = –0.007, SE = 0.01, t = –0.91, p = .363; spillover region: Estimate = –0.074, SE = 0.01, t = –6.85, p < .001). There was also a significant main effect of distractor due to longer reading times in implausible-distractor than plausible-distractor conditions. Across critical/spillover regions, numerically, L1 speakers showed slightly clearer distractor effects in implausible than plausible conditions, while distractor effects appear consistent in both plausible and implausible sentences for L2 speakers. However, the group by target by distractor interaction was not significant nor was any other fixed effects of theoretical interest.

Discussion

Consistent with Experiment 1, implausibility effects were observed at the spillover region, demonstrating retrieval of the target at the verb. Crucially, we observed interference at the critical/spillover regions with longer reading times in implausible-distractor than plausible-distractor conditions. Also, this distractor effect did not significantly interact with target plausibility. This finding is

https://doi.org/10.1017/S0142716423000498 Published online by Cambridge University Press

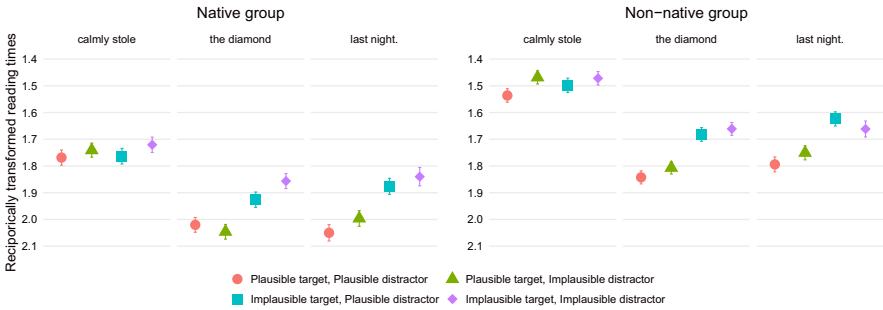

Figure 3. Reciprocally transformed reading times at the critical, spillover, and sentence-final regions in Experiment 1. The y axes are flipped for illustrative purposes.

Table 3. Statistical analysis for Experiment 3 (critical/spillover regions; "calmly stole/the diamond")

	Estimate	SE	*t*	*p*
Intercept	1.738	0.03	52.33	<.001
Group	−0.117	0.03	−3.86	<.001
Region	0.117	0.01	7.96	<.001
Target	−0.040	0.01	−5.38	<.001
Distractor	−0.016	0.01	−2.98	.003
Group:Region	0.010	0.01	1.23	.220
Group:Target	−0.002	0.01	−0.28	.776
Group:Distractor	−0.002	0.01	−0.34	.730
Region:Target	−0.033	0.01	−5.41	<.001
Region:Distractor	0.004	0.00	0.97	.330
Target:Distractor	−0.003	0.01	−0.59	.553
Group:Region:Target	−0.001	0.01	−0.17	.868
Group:Region:Distractor	0.000	0.00	0.12	.907
Group:Target:Distractor	0.011	0.01	1.95	.051
Region:Target:Distractor	−0.007	0.00	−1.45	.148
Group:Region:Target:Distractor	0.003	0.00	0.71	.477

compatible with Experiment 1, which observed similar distractor effects at the sentence-final region. The absence of the target by distractor interaction is inconsistent with the Lewis and Vasishth and Engelmann et al. models, however. Also, while there were some descriptive L1/L2 differences, there was no significant evidence that L2 speakers are more susceptible to interference than L1 speakers. This finding, we argue, is consistent with Experiments 1/2 and suggests that L1 and L2 speakers similarly weight retrieval cues during subject-verb dependency formation.

https://doi.org/10.1017/S0142716423000498 Published online by Cambridge University Press

In summary, Experiments 1/3 indicate implausibility and interference effects during sentence processing. Also, across Experiments 1–3, we did not find increased interference in L2 memory retrieval. Experiment 1 examined interference during sentence processing, and we conducted Experiment 3 to explore if similar results obtain with the distractor in a syntactically different position. Following these experiments, we conducted Experiment 4 to replicate the findings of Experiment 2, using a similar, offline research design but with the distractor occurring in an object position (Experiment 4) rather than a subject position (Experiment 2) in the long-distance conditions.

Experiment 4

Experiment 4 investigated interference using a research design comparable to Experiment 2. (9a–d) are akin to the sentences used in (7a–d) but contain a distractor as an oblique argument in (9a/b).

(9a) *Long-distance dependency, Plausible distractor*
The girl that was near the boy played with friends.

(9b) *Long-distance dependency, Implausible distractor*
The girl that was near the car played with friends.

(9c) *Short-distance dependency, Plausible distractor*
The boy was near the girl that played with friends.

(9d) *Short-distance dependency, Implausible distractor*
The car was near the girl that played with friends.

Question: Who played with friends? (The girl/The boy)

The predictions are the same as Experiment 2. That is, lower comprehension accuracy is predicted in (9a/c) than (9b/d), with increased differences for (9a/b). Also, if L2 speakers are more susceptible to interference than L1 speakers, they should show decreased accuracy in (9a/c).

Participants

Those who participated in Experiment 3 completed Experiment 4. We conducted the offline experiment after the online experiment.

Materials

The materials consisted of 24 sets of experimental sentences like (9a–d) and 72 fillers. A comprehension question with two answer options followed each sentence.

Procedure and preregistered data analysis

The procedure and preregistered data analysis are the same as Experiment 2.

https://doi.org/10.1017/S0142716423000498 Published online by Cambridge University Press

Results

Average comprehension accuracy of fillers was 92% (range 75–100) for both L1 and L2 participants. Comprehension accuracy for experimental materials is shown in Figure 2, and a summary of the statistical analysis is provided in Table 2.

There was a significant main effect of dependency, with lower comprehension accuracy in long-distance than short-distance conditions. There was also a significant main effect of distractor due to lower comprehension accuracy in plausible-distractor than implausible-distractor conditions. Crucially, there were no significant interactions.

Discussion

The results showed lower comprehension accuracy in plausible-distractor than implausible-distractor conditions. As discussed for Experiment 2, although this effect suggests interference, it may be due to the nature of the question presented across conditions. There was also evidence of reduced comprehension accuracy in long-distance conditions, suggesting greater interference effects when comprehenders establish a long-distance dependency and the distractor intervenes between the target and retrieval probe (Van Dyke & McElree, 2011). Importantly, we did not find that L2 speakers were more susceptible to interference than L1 speakers.

Below, we report Experiments 5/6, which tested semantic interference but used speeded judgment tasks. Experiments 5/6 had two aims. One was to replicate interference effects observed in Experiments 1/3 using a different task. Experiments 5/6 thus tested materials akin to Experiment 1/3. Recall that while we observed clear interference effects in Experiment 3, Experiment 1 showed interference effects only at the sentence-final region, and we discussed that this might be due to the distractor's argument status (Parker & An, 2018). Speeded judgment tasks may be useful for testing this claim, because they have proved to robustly observe interference effects (Fujita & Cunnings, 2022; González Alonso et al., 2021; Schlueter et al., 2018; Wagers et al., 2009). Thus, if Experiment 5 does not show interference, it can provide cogent evidence that the distractor's argument status prevents interference. The other aim was to replicate the finding that L1/L2 memory retrieval is analogously susceptible to interference.

Experiments 5/6

Experiments 5/6 investigated semantic interference using sentences identical to Experiments 1/3, but excluding the original sentence-final region ("last night"), as in (10/11).

Experiment 5

(10a) *Plausible target, Plausible distractor*
The thief that the robber was near in the bank calmly stole the diamond.

https://doi.org/10.1017/S0142716423000498 Published online by Cambridge University Press

(10b) *Plausible target, Implausible distractor*
The thief that the locker was near in the bank calmly stole the diamond.

(10c) *Implausible target, Plausible distractor*
The table that the robber was near in the bank calmly stole the diamond.

(10d) *Implausible target, Implausible distractor*
The table that the locker was near in the bank calmly stole the diamond.

Experiment 6

(11a) *Plausible target, Plausible distractor*
The thief that was near the robber in the bank calmly stole the diamond.

(11b) *Plausible target, Implausible distractor*
The thief that was near the locker in the bank calmly stole the diamond.

(11c) *Implausible target, Plausible distractor*
The table that was near the robber in the bank calmly stole the diamond.

(11d) *Implausible target, Implausible distractor*
The table that was near the locker in the bank calmly stole the diamond.

L1 participants should judge (10a/b) and (11a/b) to be plausible and (10c/d) and (11c/d) to be implausible. Regarding interference, for implausible conditions, the cue-based model predicts that participants should misjudge (10c) and (11c) to be plausible more often than (10d) and (11d), due to facilitatory interference. For plausible sentences, the Lewis and Vasishth model does not predict interference in judgment accuracy in either experiment. The Engelmann et al. model, however, may predict reduced accuracy when the distractor is a subject, if subjecthood substantially increases saliency, and consequently, the distractor is misretrieved. This would predict lower accuracy in (10b) than (10a) because the distractor is in a salient subject position, but not in (11a/b), when the distractor is an object and thus less salient. Alternatively, if interference effects are modulated by the distractor's argument status (Parker & An, 2018), they should be absent in Experiment 5, when the distractor is a core argument, but not in Experiment 6, when it appears in an oblique position. However, both Experiments 1/3 suggested increased reading times when the distractor was implausible in both plausible and implausible sentences. If this interference pattern is robust in our materials, participants may misjudge both (10b) and (11b) to be implausible more often than (10a) and (11a).

Based on Experiments 1–4, we predicted that L2 speakers should judge sentence plausibility correctly. Crucially, if L2 speakers are more susceptible to interference than L1 speakers, they should show larger interference effects than L1 speakers (Cunnings, 2017).

https://doi.org/10.1017/S0142716423000498 Published online by Cambridge University Press

Participants

One hundred ninety-two participants, recruited from Prolific, completed Experiment 5 (96 L1 English participants: mean age 24, range 18–45; 96 L1 German, L2 English participants: mean age 29, range 18–50), and 192 participants completed Experiment 6 (96 L1 English participants: mean age 24, range 18–48; 96 L1 German, L2 English participants: mean age 27, range 18–50). [5] Participants did not complete any of the other experiments reported here. L2 participants' mean age of onset to English was 8 for both Experiments 5 (range 5–14) and 6 (range 5–15). L2 participants completed the QPT after the judgment tasks. The mean score was 52 for both Experiments 5 (range 33–59) and 6 (range 34–60).

Materials

Experiments 5/6 contained 24 sets of experimental sentences like (10a–d) and (11a–d) and 60 fillers. Experimental sentences were identical to those used in Experiments 1/3 except that they did not have the sentence-final region. Half of the fillers were plausible and the other implausible. To ensure participants paid attention, a yes-no comprehension question, which did not probe the critical dependency, followed eight experimental sentences and 20 fillers.

Procedure

We administered the speeded judgment tasks in PCIbex Farm. At the beginning of each trial, participants saw a cross. When participants pressed the space bar, the cross was replaced by a sentence presented word by word at the center of the screen. The pacing was 400 milliseconds. The last phrase was followed by a question mark where participants judged whether the sentence was plausible or implausible. Participants had to answer within 2000 milliseconds, and feedback was provided only if participants missed this timeout. For trials which had a comprehension question, this appeared after participants made their speeded judgment.

Preregistered data analysis

We conducted data analysis for Experiments 5/6 separately and fit mixed effect logistic regressions to judgment responses in R. The models contained the same fixed and random effects as in Experiments 1/3, excluding the region variables. We treated "plausible" responses in plausible-target conditions and "implausible" responses in implausible-target conditions as correct (correct response = 1, incorrect response = 0). Trials where participants failed to make a judgment within the timeout accounted for less than 1% of the data and were treated as missing data.

Results

For Experiment 5, the mean accuracy of comprehension questions of experimental materials was 90% (range 75–100), and for Experiment 6, 91% (range 75–100).

https://doi.org/10.1017/S0142716423000498 Published online by Cambridge University Press

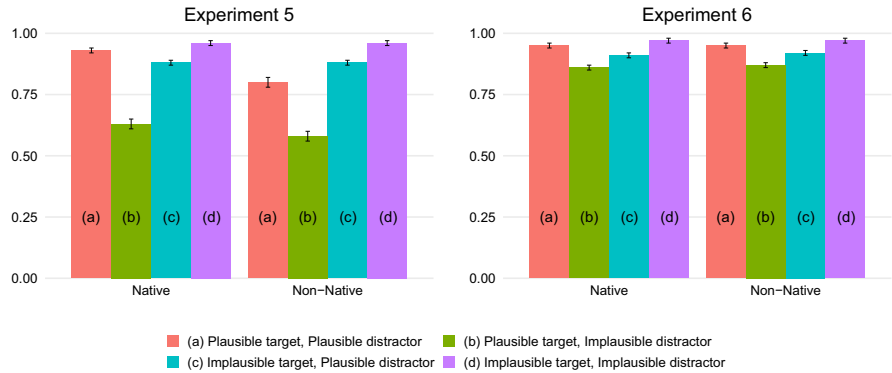

Figure 4. Judgment accuracy rates for Experiments 5/6.

Figure 4 illustrates accuracy rates in each experiment, while Table 4 contains a summary of the statistical analysis.

Experiment 5

There was a significant main effect of target, with lower accuracy in plausible-target than implausible-target conditions. This main effect of target significantly interacted with group because L2 participants showed significantly lower accuracy than L1 participants in plausible-target conditions (Estimate = –0.428, SE = 0.11, z = –3.77, p < .001) but not in implausible-target conditions (Estimate = 0.058, SE = 0.14, z value = 0.43, p value = .667). There was also a significant main effect of distractor due to lower accuracy in implausible-distractor than plausible-distractor conditions and a significant interaction between group and distractor. Nested contrasts indicated lower accuracy for L2 than L1 participants in plausible-distractor conditions (Estimate = –0.331, SE = 0.10, z = –3.20, p = .001) but not in implausible-distractor conditions (Estimate = –0.038, SE = 0.11, z = –0.34, p = .734). These effects of distractor are however difficult to interpret, as there was also a significant target by distractor interaction. A follow-up analysis testing simple effects of distractor indicated lower accuracy in the implausible-distractor than plausible-distractor condition when the target was plausible (Estimate = –1.013, SE = 0.09, z = –11.37, p < .001), and lower accuracy in the plausible-distractor than implausible-distractor condition when the target was implausible (Estimate = 0.649, SE = 0.13, z = 5.07, p < .001). Importantly, the interaction between group, target, and distractor was not significant.

Experiment 6

The model showed a significant main effect of target, which indicates lower accuracy in plausible-target than implausible-target conditions. There was also a significant target by distractor interaction, and a follow-up analysis examining simple effects of distractor showed the same interference pattern as in Experiment 5. That is, accuracy was significantly lower in the implausible-distractor than plausible-distractor condition for plausible sentences (Estimate = –0.643, SE = 0.12,

https://doi.org/10.1017/S0142716423000498 Published online by Cambridge University Press

Table 4. Statistical analysis for Experiments 5/6

	Experiment 5				Experiment 6			
	Estimate	SE	*z*	*p*	Estimate	SE	*z*	*p*
Intercept	2.444	0.11	21.61	<.001	3.205	0.14	23.36	<.001
Group	−0.187	0.09	−2.04	.042	0.095	0.09	1.02	.309
Target	0.853	0.13	6.54	<.001	0.279	0.12	2.26	.024
Distractor	−0.191	0.07	−2.65	.008	0.008	0.07	0.11	.916
Group:Target	0.242	0.09	2.73	.006	0.021	0.08	0.25	.804
Group:Distractor	0.146	0.06	2.37	.018	0.017	0.07	0.23	.815
Target:Distractor	0.828	0.08	10.87	<.001	0.641	0.09	6.76	<.001
Group:Target:Distractor	−0.108	0.06	−1.70	.089	−0.012	0.07	−0.17	.864

$z = -5.49$, $p < .001$) and significantly lower in the plausible-distractor than implausible-distractor condition for implausible sentences (Estimate = 0.628, SE = 0.12, $z = 5.38$, $p < .001$). As in Experiment 5, the interaction between group, target, and distractor was not significant.

Discussion

Both Experiments 5/6 suggested interference from the implausible distractor in plausible sentences and from the plausible distractor in implausible sentences. These interference patterns are unexpected from the cue-based model but consistent with Experiments 1/3. Regarding L1/L2 differences, although there were some interactions with group in Experiment 5, neither Experiment 5 nor 6 revealed a significant group by target by distractor interaction.

Descriptively, interference effects appear to be larger in Experiment 5 than Experiment 6 (see Figure 4). This is especially true in plausible-target conditions, where the difference between plausible- and implausible-distractor conditions is numerically larger in Experiment 5 than Experiment 6. However, an additional non-preregistered analysis, which examined accuracy rates from the two experiments with an additional sum-coded fixed effect of experiment (Experiment 5/Experiment 6), did not show a significant distractor by experiment interaction (Estimate = 0.070, SE = 0.04, $z = 1.59$, $p = .113$), nor a significant three-way interaction between target, distractor and experiment (Estimate = −0.062, SE = 0.04, $z = -1.42$, $p = .156$). Note also that this descriptive difference could be due to differences in the felicity of the sentence-initial relative clause. In condition (10b), which contains an implausible distractor in a plausible sentence ("The thief that the locker was near . . . "), judgments may have been influenced by the fact that it is pragmatically awkward for an inanimate object to be the subject of a relative clause. The descriptively lower judgments for sentences like (10b) compared to (10a) could thus in part be due to this pragmatic infelicity. Note that we also observed clear interference effects in (11a/b), where the sentence-initial relative clause is not pragmatically awkward in the inanimate condition ("The thief that was

https://doi.org/10.1017/S0142716423000498 Published online by Cambridge University Press

near the locker . . . "). As such, while we cannot rule out pragmatic infelicity influencing the results of Experiment 5, the fact that we still observed interference effects in Experiment 6 we argue indicates that our results cannot be reduced to this issue. For these reasons, we maintain that we observed interference in both Experiments 5 and 6, but do not draw strong conclusions here about potential cross-experiment differences. Below, we discuss these findings along with those observed in Experiments 1–4.

General discussion

The present study examined interference in L1 and L2 subject-verb dependency formation. Our six preregistered experiments showed largely consistent findings. Experiments 1/3 demonstrated implausibility effects during sentence processing, suggesting that participants retrieved the target after encountering the retrieval probe. These experiments also suggested interference effects in both plausible and implausible sentences, which are partially but not fully consistent with the cue-based model (Engelmann et al., 2019; Lewis & Vasishth, 2005). Although interference effects observed in Experiment 1 were present only at the sentence-final region, Experiment 5, which aimed to replicate the findings of Experiment 1, showed clear interference effects in a speeded judgment task. Experiments 3/6 also observed similar interference patterns. Experiment 4 indicated lower comprehension accuracy in long-distance than short-distance conditions. Experiment 2 showed a similar trend as well, although the effect was not significant. These findings suggest increased retrieval difficulty when a dependency between two elements is long and a distractor is interpolated between the encoding and retrieval of the target (Van Dyke & McElree, 2011). Across Experiments 1–6 and in three different tasks, we did not find that L2 speakers were more susceptible to semantic interference than L1 speakers. Also, there was no evidence that the distractor's argument status and saliency influenced memory retrieval. Below, we discuss the implications of these results.

Interference in L1 and L2 comprehension

Our study indicated that L1 and L2 comprehension is influenced by distractors. This finding is partially compatible with studies showing interference in L1 and L2 comprehension across various dependencies (e.g., Fujita & Cunnings, 2022; Lago & Felser, 2018; Tanner et al., 2012). However, we found no consistent evidence of significant L1/L2 differences. In Experiment 5 (a speeded judgment task), we did observe significant group by distractor and group by target interactions, with L2 speakers having lower accuracy in plausible-distractor and plausible-target conditions, respectively. Descriptively, L2 speakers seem to have lower accuracy in the plausible-target/plausible-distractor condition, but why L2 speakers should have difficulty with this condition is not clear, given that one might expect this condition to be easy as there are two plausible actors for the critical verb. Additionally, these distractor and target effects are difficult to interpret on their own, given we also observed a significant target by distractor interaction that importantly did not interact with group. We also did not find any significant interactions with group in Experiment 6. As such, we contend that across

https://doi.org/10.1017/S0142716423000498 Published online by Cambridge University Press

experiments, L1 and L2 speakers were influenced by the distractor in a largely similar way.

Our finding across experiments that L2 speakers were not significantly more influenced by distractors than L1 speakers is not consistent with the hypothesis that L1 and L2 speakers weight retrieval cues differently (Cunnings, 2017), at least in subject-verb dependencies and in proficient L2 speakers. Instead, our results suggest L1 and L2 speakers weight retrieval cues in a similar way. Also, our experiments manipulated multiple properties of the distractor, such as its argument status and saliency. Given that our results showed similar interference between L1 and L2 speakers, irrespective of these factors, they suggest that the argument status and saliency of a distractor may not influence memory retrieval in L1 and L2 subject-verb dependency resolution.

While the absence of L1/L2 differences in dependency formation is consistent with some studies (Fujita & Cunnings, 2022; Tanner et al., 2012), it is inconsistent with others (Felser et al., 2009; Felser & Cunnings, 2012). One possible account of these incompatible results is that L1/L2 differences in interference are dependent on discourse information. That is, in previous studies which observed similar interference patterns between L1 and L2 speakers (Fujita & Cunnings, 2022; Lago & Felser, 2018; Tanner et al., 2012), the distractor was in a non-salient position in the discourse, typically appearing as an oblique object similar to the distractor in our Experiments 3 and 6. However, L1/L2 differences have been observed in studies that included a distractor in a salient discourse topic position (Felser et al., 2009; Felser & Cunnings, 2012). For example, in Felser et al.'s (2009) study on reflexive resolution, the distractor appeared in a discourse salient position, as the matrix subject of the critical sentence, as in (4). Note that in our experiments, although the distractor was in a relatively salient subject position in Experiments 1, 2, and 5, it was never the discourse topic. Given these differences, discourse topichood may be a key source of L1/L2 differences in memory retrieval. As studies examining how discourse topics affect L2 memory retrieval are scarce, we leave this issue open for future research.

Our observations across self-paced reading and speeded judgment experiments that participants were influenced by the distractor, when it was both a subject and oblique object, might be taken to suggest a role for recency in dependency resolution, given that the distractor was always linearly closer to the retrieval probe than the retrieval target. A preference for the linearly closer distractor could potentially be taken as an indication of shallow L2 processing (for discussion, see Fujita & Cunnings, 2022). However, we did not find any consistent L1/L2 differences across studies suggesting that L2 speakers were more influenced by the linearly closer distractor than L1 speakers. Thus, if our results are taken to indicate a role for linear proximity in dependency formation, they do not support this being an L2-only strategy, and as such, we contend our results do not provide support for shallow L2 processing (Clahsen & Felser, 2006, 2018).

We acknowledge that the lack of consistent L1/L2 differences across our experiments could be due to the lack of power to observe the relevant interactions (Brysbaert, 2021). We note however that our sample sizes are larger than previous studies that have observed L1/L2 differences (e.g., Felser et al., 2009; Felser & Cunnings, 2012). While further research with larger samples is thus required to

https://doi.org/10.1017/S0142716423000498 Published online by Cambridge University Press

elucidate the nature of interference effects in L1 and L2 processing, we reiterate that the present study, despite reporting six experiments, did not find evidence for L1/L2 differences.

In summary, our results did not reveal L1/L2 differences in cue weighting during sentence processing (cf. Cunnings, 2017), and instead suggest L1 and (proficient) L2 speakers resolve subject-verb dependencies in a similar way.

Interference in dependency formation

Experiment 3 showed interference, with longer reading times in implausible-distractor than plausible-distractor conditions, irrespective of target plausibility. Experiment 1 showed a similar interference pattern, but only at the sentence-final region. The disputable nature of this evidence might pertain to the distractor's argument status (Parker & An, 2018). As discussed in the Introduction, Parker and An claimed that distractors that are core arguments nullify interference effects. They argued that such arguments are encoded in memory with distinct representations that can be easily rejected as a retrieval candidate. However, Experiment 5 tested sentences similar to those used in Experiment 1 in a speeded judgment task and showed clear interference effects compatible with those observed at the sentence-final region in Experiment 1. Thus, the combined findings from Experiments 1/5 suggest that core argument distractors do not nullify interference in both L1 and L2 comprehension.[6]

The interference effects observed in implausible sentences (i.e., reduced implausibility effects in Experiments 1/3 and misperception of plausibility in Experiments 5/6) are compatible with previous results reported by Cunnings and Sturt (2018) and Fujita and Cunnings (2022), who observed facilitatory semantic interference in filler-gap dependencies. This finding also aligns with the cue-based model (Engelmann et al., 2019; Lewis & Vasishth, 2005), which considers thematic relations as retrieval cues (Smith & Vasishth, 2020). However, our experiments also suggested distractor effects in plausible (target match) sentences, with longer reading times and lower speeded judgment accuracy when the distractor was an implausible NP. This finding is inconsistent with previous studies which showed no interference effects in target-match sentences (e.g., Wagers et al., 2009). The results are also challenging for the Lewis and Vasishth (2005) model, which predicts inhibitory interference (i.e., the opposite pattern) in plausible sentences.

One possible retrieval-based account of our reading time data in plausible sentences relates to the distractor's saliency (Engelmann et al., 2019). Engelmann et al. proposed that as a distractor's saliency increases, it is more likely to get retrieved than the target, counteracting inhibitory interference that would typically be predicted by the cue-based model. In our experiments, for sentences containing a sufficiently salient distractor, this would predict longer reading times in implausible than plausible-distractor conditions, in both implausible and plausible-target conditions. In Experiment 1, the distractor was in a subject position and thus expected to be salient. The main effect of distractor observed in this experiment thus would be compatible with the Engelmann et al. proposal. Regarding Experiment 3, the distractor was an oblique argument inside a prepositional phrase, which is unlikely to be salient. Given this, the Engelmann et al. model would predict

https://doi.org/10.1017/S0142716423000498 Published online by Cambridge University Press

inhibitory interference in plausible sentences here, like the Lewis and Vasishth model. However, we still observed a significant main effect of distractor, with longer reading times when the distractor was implausible, irrespective of target plausibility. It is not clear that either the Lewis and Vasishth or Engelmann et al. model can account for this finding.[7]

Note however that although the results in Experiment 3 showed only a significant main effect of distractor, this distractor effect seems larger for implausible sentences, at least for L1 participants (see Figure 3). L2 participants however showed numerically longer reading times for the implausible-distractor than plausible-distractor conditions in both plausible and implausible sentences. This descriptive observation would be compatible with the Engelmann et al. model for L1 participants. We do not draw any strong conclusion here however, given that the three-way interaction between group, target, and distractor was not significant. Also, our speeded judgment data in Experiments 5/6 showed clear distractor effects in plausible sentences which are difficult to reconcile with the Engelmann et al. model. Thus, whether the Engelmann et al. model can account for our data is unclear.

Finally, we note again that the distractor effects observed in our study, especially in plausible (target match) sentences, may indicate that distractor recency affected our results. How recency influences memory retrieval has been considered in other dependencies, such as pronoun resolution (Cunnings et al., 2014). However, recency-based accounts do not seem to hold in general, as numerous previous studies on subject-verb agreement have reported the mismatch asymmetry, with no significant effects observed in target-match sentences (e.g., Wagers et al., 2009). How recency effects, especially in target-match sentences, may vary across dependencies is thus an avenue for future research.

Conclusion

We investigated semantic interference in L1 and L2 subject-verb dependency formation, manipulating the distractor's argument status and saliency. Our main aim was to test whether L2 speakers are more susceptible to interference than L1 speakers. Our experiments did not provide evidence that L2ers are more susceptible to interference than L1ers. These findings suggest that L1 and L2 speakers weight retrieval cues analogously during subject-verb dependency formation. Also, the observed interference patterns were not fully consistent with the cue-based model and the typical finding of interference effects in language comprehension (the target-mismatch asymmetry). We suggested that retrieval-based accounts may be compatible with our results if they consider distractor saliency or the role of recency in dependency resolution.

Replication package. Data, materials, and analysis code are available at https://osf.io/f7uwe/.

Funding statement. This study was funded 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.

Competing interests. The authors declare no competing interests.

Notes

1 Following the preregistration, we continued participant recruitment until we had 96 L1 speakers and 96 L2 speakers whose comprehension accuracy was higher than 75%, as an index that they paid attention. Also, L2 participants needed to score at least 30 out of 60 on the QPT to be included in the data analysis. Eight L1 participants and four L2 participants were excluded due to these criteria.
2 The QPT contains 60 questions designed to gauge English proficiency. Achieving a score of 30 indicates an upper-intermediate level of English proficiency (B1 level according to the CEFR).
3 In response to a reviewer's recommendation, we examined the potential impact of L2 participants' English proficiency on the results across Experiments 1–6. However, we did not observe any significant evidence of proficiency influencing interference effects, and as such do not discuss this further.
4 As in Experiment 1, we continued participant recruitment until we had 96 L1 participants and 96 L2 participants whose comprehension accuracy was at least 75%. L2 participants needed to score at least 30 out of 60 on the QPT. Accordingly, we excluded 11 L1 participants and 7 L2 participants.
5 Some of our experimental materials contained a comprehension question after participants made a judgment as an index of their attention, and we continued participant recruitment until we had 96 L1 and 96 L2 participants who answered at least 75% of the comprehension questions correctly in each experiment. Accordingly, we excluded data from seven L1 participants and eight L2 participants for Experiment 5 and five L1 participants and eight L2 participants for Experiment 6.
6 As noted by a reviewer, another difference between our study and Parker and An is that while we used the copula predicate "was near" across items (e.g., "The thief that was near the robber"), Parker and An used lexical verbs (e.g., "The waitress who sat near the girl"). As such, while the predicate in Parker and An has a semantic structure akin to "sat(the waitress), near(the waitress, the girl)", our sentences arguably include a single predicate "be-near(the thief, the robber)". Whilst we acknowledge this difference across studies, if we take this analysis, then the distractor is always a core argument in our experiments, irrespective of whether it appears in a subject or oblique object position. The fact that we observed interference across experiments would thus still be incompatible with Parker and An's claim that core arguments resist interference.
7 In the L1 literature, another class of models, named representational theories (Hammerly et al., 2019), could potentially explain this pattern. However, these theories are formulated to account for interference from morpho-syntactic features such as number, and it is not clear how they can be extended to the types of semantic features manipulated in this study. For this reason, we do not discuss these models further here.

References

Armstrong, A., Bulkes, N., & Tanner, D. (2018). Quantificational cues modulate the processing of English subject-verb agreement by native Chinese speakers: An ERP study. *Studies in Second Language Acquisition*, **40**(4), 731–754. https://doi.org/10.1017/S0272263118000013

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

Brysbaert, M. (2021). Power considerations in bilingualism research: Time to step up our game. *Bilingualism: Language and Cognition*, **24**(5), 813–818. https://doi.org/10.1017/S1366728920000437

Cai, L., Shu, H., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject–verb agreement in L2 learning. *Bilingualism: Language and Cognition*, **10**(2), 161–174. https://doi.org/10.1017/S136672890700291X

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

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

Clahsen, H., & Felser, C. (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. *Bilingualism: Language and Cognition*, **20**(4), 659–678. https://doi.org/10.1017/S1366728916000675

Cunnings, I., & Felser, C. (2013). The role of working memory in the processing of reflexives. *Language and Cognitive Processes*, **28**(1–2), 188–219. https://doi.org/10.1080/01690965.2010.548391

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

Cunnings, I., & Fujita, H. (2023). Similarity-based interference and relative clauses in second language processing. *Second Language Research*, **39**(2), 539–563. https://doi.org/10.1177/02676583211063534

Cunnings, I., Patterson, C., & Felser, C. (2014). Variable binding and coreference in sentence comprehension: Evidence from eye movements. *Journal of Memory and Language*, **71**(1), 39–56. https://doi.org/10.1016/j.jml.2013.10.001

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

Deniz, N. D. (2022). Processing syntactic and semantic information in the L2: Evidence for differential cue-weighting in the L1 and L2. *Bilingualism: Language and Cognition*, 1–13. https://doi.org/10.1017/S1366728921001140

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). 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., 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

Fujita, H. (2021a). *An R package for creating experimental files in IbexFarm.* [Computer software]. https://doi.org/10.17605/OSF.IO/7RVX6

Fujita, H. (2021b). 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. (2023a). Online revision process in clause-boundary garden-path sentences. *Memory & Cognition.* https://doi.org/10.3758/s13421-023-01444-0

Fujita, H. (2023b). Predictive structure building in language comprehension: A large sample study on incremental licensing and parallelism. *Cognitive Processing*, **24**(2), 301–311. https://doi.org/10.1007/s10339-023-01130-8

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

Fujita, H., & Cunnings, I. (2022). Interference and filler-gap dependency formation in native and non-native language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **48**(5), 702–716. https://doi.org/10.1037/xlm0001134

Fujita, H., & Cunnings, I. (2023). Interference in quantifier float and subject-verb agreement. *Language, Cognition and Neuroscience*, **38**(7), 1001–1019. https://doi.org/10.1080/23273798.2023.2189738

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. (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

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

Ito, A., Corley, M., & Pickering, M. (2018). A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension. *Bilingualism: Language and Cognition*, **21**(2), 251–264. https://doi.org/10.1017/S1366728917000050

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

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, 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

Kush, D. (2013). *Respecting relations: Memory access and antecedent retrieval in incremental sentence processing* [PhD Thesis, University of Maryland]. http://hdl.handle.net/1903/14589

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. (2022). Semantic attraction in sentence comprehension. *Cognitive Science*, **46**(2), e13086. https://doi.org/10.1111/cogs.13086

Lee, E.-K. R., & Phillips, C. (2023). Why non-native speakers sometimes outperform native speakers in agreement processing. *Bilingualism: Language and Cognition*, **26**(1), 152–164. https://doi.org/10.1017/S1366728922000414

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

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, **27**(1), 53–78. https://doi.org/10.1017/S0272263105050035

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

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

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in LI 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., & An, A. (2018). Not all phrases are equally attractive: Experimental evidence for selective agreement attraction effects. *Frontiers in Psychology*, **9**, 1566. https://doi.org/10.3389/fpsyg.2018.01566

Patterson, C., Trompelt, H., & Felser, C. (2014). The online application of binding condition B in native and non-native pronoun resolution. *Frontiers in Psychology*, **5**. https://www.frontiersin.org/article/10.3389/fpsyg.2014.00147

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

Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics*, **32**(2), 299–331. https://doi.org/10.1017/S0142716410000421

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

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

Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition*, **16**(2), 367–382. Cambridge Core. https://doi.org/10.1017/S1366728912000302

Tanner, D., Nicol, J., Herschensohn, J., & Osterhout, L. (2012). Electrophysiological markers of interference and structural facilitation in native and nonnative agreement processing. In *Proceedings of the 36th Annual Boston University Conference on Language Development*, 594–606.

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., & 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

Wagers, M., 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

Zehr, J., & Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX)* [Computer software]. https://doi.org/10.17605/OSF.IO/MD832

Cite this article: Fujita, H. & Cunnings, I. (2024). Subject-verb dependency formation and semantic interference in native and non-native language comprehension. *Applied Psycholinguistics* **45**, 81–109. https://doi.org/10.1017/S0142716423000498