REPLICATION STUDY

Wh-Dependency Processing in a Naturalistic Exposure Context: Sensitivity to Abstract Syntactic Structure in High-Working-Memory L2 Speakers

Robyn Berghoff

Department of General Linguistics, Stellenbosch University, Stellenbosch, South Africa
Corresponding author. E-mail: berghoff@sun.ac.za

Abstract

This study replicates Felser and Roberts (2007), which used a cross-modal picture priming task to examine indirect-object dependency processing in classroom L2 learners. The replication focuses on early L2 learners with extensive naturalistic L2 exposure (n = 22)—an understudied group in the literature—and investigates whether these learners, in contrast to those in the original study, reactivate the moved element at its original position in the sentence. Bayesian multilevel regression is used to analyze the data. The results suggest that higher-working-memory participants did reactivate the moved element at its structural origin. By extending previous research to an understudied group, the study contributes to our knowledge regarding sensitivity to abstract syntactic structure in L2 processing.

Introduction

A key point of interest in second language (L2) acquisition research relates to the circumstances under which first language (L1) and L2 processing converge. L1–L2 differences have repeatedly been observed in the processing of syntactically complex constructions, such as those involving movement dependencies (e.g., Berghoff, 2020; Felser & Roberts, 2007; Marinis et al., 2005). These differences have been attributed to a variety of factors, including reduced sensitivity to abstract syntactic structure in L2 compared to L1 speakers and reduced L2 compared to L1 processing automaticity. At the same time, certain theoretical accounts (Clahsen & Felser 2006a, 2006b, 2018; Ullman, 2001) and empirical findings (Pliatsikas & Marinis, 2013; Pliatsikas et al., 2017) suggest that L1–L2 processing convergence may be more likely among (early) L2 learners with naturalistic exposure to the L2. The ability to draw conclusions in this area is currently limited by the dearth of studies that have examined such learners and these studies' focus on only one type of movement dependency, namely long-distance wh-dependencies. The present study extends this body of literature by exploring the

processing of indirect-object dependencies among L2 learners drawn from a context in which naturalistic L2 exposure is extensive and typically begins at an early age. To this end, the study replicates Felser and Roberts (2007), which examined indirect-object dependency processing among classroom learners in a foreign-language context.

Background

The focus of this study is on sentences such as (1), taken from Roberts et al. (2007, p. 185), where *the peacock* originates structurally after the direct object *the nice birthday present* but appears earlier in the sentence.

(1) John saw the peacock to which the small penguin gave the nice birthday present __ in the garden last weekend.

Processing a sentence such as (1) poses a challenge because after the moved element (i.e., the filler) *the peacock* is encountered, it must be retained in short-term memory until it can be linked to the element that licenses it, in a process termed "filler integration." A number of online studies of filler-gap dependency processing have observed a tendency, among L1 speakers, to reactivate the filler at clause boundaries and at its original position (Chow & Zhou, 2019; Fernandez et al., 2018; Gibson & Warren, 2004; Nicol, 1993; Nicol & Swinney, 1989; although see Roberts et al., 2007 and Miller, 2014 for contrary findings). This processing pattern is in line with a Chomskyan account of movement (Chomsky, 1986), in which the filler moves through these positions on its way to its surface structure destination, leaving behind a silent copy of itself—a "trace"—at each position.

Roberts et al. (2007) used a cross-modal picture priming task to investigate whether L1 speakers—children aged 5 to 7 years and adults—reactivated the moved element at its original position, termed the "gap" position. While listening to sentences such as (1), participants were shown pictures that were either identical or unrelated to the entity denoted by the filler at either the gap position or a control position 500 milliseconds earlier in the sentence. They then had to decide whether the depicted entity was alive or not alive, with their reaction times (RTs) to this decision serving as the dependent variable. Both adults and children with relatively high working-memory capacity, as measured in adults by a reading span task and in children by a listening span task, showed reduced RTs to identical versus unrelated targets at the gap position. RTs to identical targets were also lower at this position than at the control position. This finding suggests that the moved element was reactivated at the gap position, thus facilitating responses in the decision task. Participants with relatively lower working-memory scores, however, showed no difference in RTs to identical versus unrelated targets (adults) or a disadvantage for identical versus unrelated targets (children) at the gap position.

Felser and Roberts (2007) used the same task and materials employed by Roberts et al. (2007) to investigate the L2 processing of indirect-object dependencies among L1 Greek speakers. These participants had first been exposed to the L2 between the ages of 6 and 11 in a classroom setting and did not consider themselves bilingual; further, they had been living in the United Kingdom for an average of 2.9 years at the time of testing (Felser & Roberts, 2007, p. 18). The results for the L2 speakers differed from those obtained for the L1 speakers in Roberts et al. (2007) in two respects: first, there were no working memory effects on processing behavior among the L2 speakers; and second, they showed an advantage for identical versus unrelated targets at both sentence

positions. The latter result suggests that instead of selectively reactivating the moved element at the gap position, the participants may have actively maintained the moved element in working memory during processing, which facilitated their responses at both the gap and the control locations. Importantly, as indirect-object dependencies are formed in essentially the same way in Greek and English, the results do not suggest a transfer of L1 processing strategies to the L2.

Miller (2014, 2015) investigated whether the reduced automaticity of L2 compared to L1 processing might inhibit trace reactivation in L2 speakers. From this perspective, delays in L2 lexical access lead to a delay in the construction of syntactic representations, and it is this delay in L2 processing that precludes the observation of a trace reactivation effect. As such, this account predicts that if experimental stimuli are designed in such a way that L2 lexical access is facilitated, L2 speakers will show sensitivity to movement traces during real-time processing.

In Miller (2014, 2015), fillers were denoted by L1–L2 cognates (e.g., English–French *gorilla–gorille*), with the rationale that the facilitative effect of cognates on lexical access (see e.g., Costa et al., 2000) would mitigate the potential confounding effects of reduced L2 processing automaticity. In line with this prediction, Miller's (2014) intermediate L2 learners showed RT patterns consistent with trace reactivation at the gap position. Miller (2015) obtained similar results with indirect-object cleft sentences in which the filler crossed a clause boundary, where a subset of learners showed evidence of filler reactivation at both the clause boundary and the gap position.

Miller's (2014, 2015) findings are suggestive of a role for processing automaticity in facilitating the construction of fully specified syntactic representations. In turn, they predict that the construction of such representations should also be more likely given the presence of individual characteristics associated with greater processing automaticity. One such characteristic is L2 exposure, which has been proposed to exert a practice effect on the L2 system, leading L2 processing to become more proceduralized (e.g., Ullman, 2001). Indeed, a few studies have observed differences in L2 processing across L2 learners with classroom L2 exposure and naturalistic L2 exposure (e.g., Dussias & Sagarra, 2007). Regarding the processing of movement dependencies specifically, Pliatsikas and Marinis (2013; see also Pliatsikas et al., 2017), in their study of long-distance wh-dependency processing, observed trace reactivation at the clause boundary among L2 learners with naturalistic L2 exposure (an average of 9 years), but not among L2 learners whose exposure was limited to the classroom. Some accounts of L2 processing—for example, the Shallow Structure Hypothesis (Clahsen & Felser, 2006a, 2006b, 2018)—additionally attribute a central role to age of L2 acquisition (AoA) in increasing sensitivity to morphosyntactic information during L2 processing. There is variation in the literature regarding the timing of the so-called sensitive period for grammar, with some studies reporting an offset at around age six (Long, 1990) and others only at the end of adolescence (Hartshorne et al., 2018; Johnson & Newport, 1989). Here, too, though, type of exposure is crucial: Research has established that AoA is less relevant for L2 outcomes in instructed L2 settings in which L2 exposure is limited (Muñoz, 2006).

This article reports on a close replication of Felser and Roberts (2007) conducted in South Africa with L1 Afrikaans–L2 English speakers with AoAs ranging from 1–14 (mean 5.3 years). We refer to these as "early" L2 learners because the maximum AoA still falls within the upper bound of the proposed sensitive period for grammar. While South Africa has 11 official languages, English is a prominent societal language (Posel & Zeller, 2016). Exposure to English often commences before it is formally introduced as a school subject and is not limited to the classroom context, with studies indicating that Ln speakers of English use this language extensively with both family and friends

(Berghoff, 2021; Coetzee van Rooy, 2013). At the same time, however, L2 English speakers are not immersed in the L2 in South Africa, and the L1 is typically maintained alongside English (Berghoff, 2021; Coetzee van Rooy, 2012; Posel et al., 2020). The consequences of such societally multilingual settings for language processing remain poorly understood. This study aims to extend our knowledge in this domain by investigating whether L2 learners of this background show evidence of trace reactivation at the gap position during indirect-object dependency processing.

Method

Participants

The study's participants were 22 L1 Afrikaans–L2 English speakers[1] (mean age 20.75 years, standard deviation [SD] 1.06 years, range 19–23 years) who were students at a university in South Africa. All had normal or corrected-to-normal vision. The study was approved by the university's research ethics committee (project number 0382) and informed consent was obtained from all participants prior to the beginning of the experiment. Participants received course credit for their participation.

Language background information was obtained using the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007). Participants' English proficiency was assessed using a C-test consisting of three short texts, each of which contained 20 incomplete words with the first half of their letters provided. The participants' scores were comparable to those obtained from a sample of 53 L1 English speakers who were students at the same university (mean 76.92%, SD 11.6%).[2]

One participant who indicated their age of first exposure to English as 0 years was removed from further analyses. The characteristics of the remaining participants are summarized in Table 1.

Working memory was assessed using a computerized reading span task (Stone & Towse, 2015; von Bastian et al., 2013). In the task, participants were presented with a set

Table 1. Participant characteristics

	Range	Mean	SD
Age of L2 acquisition (years)	1–14	5.27	3.01
Length of L2 exposure (years)	6–21	15.5	3.52
C-test score (%)	36.67–91.67	71.15	13.67
Current average global L2 exposure (%)[a]	2[b]–70	42.88	16.29
L2 speaking ability (self-rated)[c]	3–10	7.40	1.84
L2 spoken comprehension ability (self-rated)	3–10	8.44	1.76
L2 reading ability (self-rated)	2–10	8.11	1.97

[a]The relevant question in the LEAP-Q here is "Please list what percentage of the time you are *currently* and *on average* exposed to each language" (italics in original).
[b]The lowest values for L2 Exposure and the three self-rated variables all come from one participant. This participant obtained a C-test score of 80%, suggesting that their self-ratings were not reliable. The value of 2 for L2 Exposure is also implausible, given the study's context. Due to the small sample, however, we did not wish to exclude this participant's data.
[c]Self-ratings are on a scale of 0 to 10, with 0 indicating "none" and 10 indicating "perfect."

---

[1]This sample size is identical to the final sample size used in Felser and Roberts (2007). As a reviewer points out, the sample is relatively small, which can cause issues in frequentist analysis due to low statistical power. Bayesian techniques like those adopted in this article have been argued to be better suited to analyzing small-sample data (e.g., Baldwin & Fellingham, 2013).
[2]Felser and Roberts (2007) used the Oxford Placement Test (OPT) to assess their participants' English proficiency. We used a C-test instead due to concerns about potential ceiling effects among our participants.

of sentences and had to judge each sentence as either "makes sense" or "does not make sense." Each sentence was followed by a number that had to be remembered until the end of the set of sentences, at which point the participant had to provide all of the numbers they had seen in that set in order of appearance. The number of sentences in a set ranged from two to five, and scoring was done based on the proportion of numbers the participant recalled correctly. The data from one participant who scored 0 on this task was removed from further analyses. The mean proportion correct of the remaining participants was 53.97% (SD 13.9%, range 26–76%).

Materials

The task involved 20 experimental sentences, which were identical to those used in Roberts et al. (2007) and Felser and Roberts (2007). As in these studies, the task also contained 60 filler sentences similar in length to the experimental sentences, 12 of which were similar in structure to the experimental sentences, but where the visual target was displayed at a position other than the two critical test points.

The 80 sentences were recorded by a female L1 English speaker using Audacity (Audacity Team, 2019). All but two of the target pictures were obtained from Snodgrass and Vanderwart's (1980) dataset.[3] Each experimental sentence was paired with a visual target that was either identical to the referent of the indirect object or unrelated.

In each experimental sentence, the visual target (identical or unrelated) appeared at one of two critical points: the offset of the direct object noun phrase (i.e., the gap position) or a pregap control position 500 milliseconds prior to this offset. This yielded four experimental conditions, illustrated in (2) (Felser & Roberts, 2007, p. 20). It is noted that, like English, Afrikaans is also a wh-movement language in which the indirect object canonically follows the direct object (de Stadler, 1995).

(2) Fred chased the squirrel to which the nice monkey explained …
　　 a. Identical, gap position:
　　　　 …the game's difficult rules [SQUIRREL] in the class last Wednesday.
　　 b. Identical, pregap position:
　　　　 …the game's [SQUIRREL] difficult rules in the class last Wednesday.
　　 c. Unrelated, gap position:
　　　　 …the game's difficult rules [TOOTHBRUSH] in the class last Wednesday.
　　 d. Unrelated, pregap position:
　　　　 …the game's [TOOTHBRUSH] difficult rules in the class last Wednesday.

The experimental items were divided across four presentation lists, so that each participant saw only one version of each experimental sentence. The 20 experimental items in each list were combined with the 60 fillers and pseudorandomized.

Procedure

The cross-modal picture priming task was designed and administered in PsychoPy (Peirce et al., 2019). Participants performed the task on a laptop with a 15-inch

---

[3]The two exceptions were "hippopotamus" and "panda." For these, black-and-white line drawings in the style of the Snodgrass and Vanderwart (1980) pictures were used.

screen (resolution: 1366 × 768). At the beginning of the session, the experiment administrator told the participant to listen carefully to the prerecorded sentences, which were presented over headphones, and watch the screen for a picture of an animal or an object that would be displayed at an undetermined point during the sentence. They were instructed further that when a picture appeared, they had to decide as quickly as possible whether the animal/object was alive or not alive and indicate their choice by pressing either the green ("yes") or red ("no") key on the keyboard. As in Felser and Roberts (2007), the task also included 38 comprehension questions, which were distributed across the experiment and auditorily presented. The experiment was preceded by a short practice round to allow participants to familiarize themselves with the procedure. The task included four self-timed breaks and on average took around 30 minutes to complete. After the completion of the experiment, participants completed the working memory task, the LEAP-Q, and the C-test.

Analysis

RTs were analyzed using Bayesian regression. A key advantage of the Bayesian approach (see e.g., Norouzian et al., 2019) is that it allows for the strength of evidence both for and against the null hypothesis to be evaluated. In contrast, the conventional null hypothesis significance testing approach does not provide evidence in favor of the null hypothesis, as the failure to obtain a significant effect may be due to, for example, a lack of statistical power, rather than the nonexistence of the effect. Another benefit offered by this approach is the ability to specify, by means of so-called priors, the expected direction and magnitude of an effect based on extant research findings or expert opinion. Here, we use Felser and Roberts's (2007) results as a basis for the specification of informative priors for the effects of target type, sentence position, and their interaction. Details of the prior specification are provided in Appendix A. Because Felser and Roberts (2007) report no effect of working memory, we use noninformative priors for this term; noninformative priors were also used for the standard deviations. For robustness purposes, we also reran the models using informative priors based on the results of Roberts et al. (2007). The Bayes factors remain robust. These results are available upon request.

All models were fit with four chains, each of which contained 10,000 samples following a warmup of 2,000 samples. For each model parameter, we report the parameter estimate b; the 95% credible interval, or the range within which b can be taken to fall with 95% certainty; and the evidence ratio *P(b).* Following Jeffreys (1998), we consider an evidence ratio of 0.3 or smaller as substantial evidence for the absence of an effect and an evidence ratio of 3 or greater as substantial evidence for the presence thereof.

Results
Accuracy

Accuracy scores were 80.13% (SD 7.4%, range 67.6–94.6%) on the end-of-trial comprehension questions and 96.3% (SD 3.7%, range 84.2–100%) on the aliveness decision task. These results are comparable to those of Felser and Roberts (2007) and Roberts et al. (2007).

Table 2. Mean RTs (SD) to visual targets per condition

	Control position	Trace position
Unrelated target	882 (257)	867 (196)
Identical target	838 (222)	815 (194)

Table 3. Model results: RTs to visual targets

	Estimate	Est. Error	CI L	CI U
Intercept	**−0.2**	0.04	−0.28	−0.12
Target Type	**−0.04**	0.03	−0.10	0.02
Position	−0.02	0.04	−0.09	0.06
Working Memory Score	**−0.03**	0.04	−0.11	0.06
Target Type × Position	−0.04	0.06	−0.17	0.09
Working Memory Score × Target Type	**0.04**	0.03	−0.01	0.09
Working Memory Score × Position	0.02	0.03	−0.04	0.07
Working Memory Score × Target Type × Position	**−0.05**	0.05	−0.14	0.04

Note: Estimate = parameter estimate; Est. Error = standard error; CI L = lower end of the 95% credible interval; CI U = upper end of the 95% credible interval. Parameter estimates in bold are effects that are reliably present (Bayes factor ≥ 3).

Reaction Times

In line with previous studies (Felser & Roberts, 2007; Roberts et al., 2007), only trials in which the aliveness decision was correct were analyzed, which led to the removal of 3.7% of the data. No RTs on this task exceeded 2,000 milliseconds, nor were there any individual outliers exceeding two SDs from each participant's mean per condition; thus, no additional data points were omitted.

Table 2 provides the means and SDs of the participants' RTs per condition. As is evident, RTs to identical targets were shorter than those for unrelated targets at both the control and trace position, but the advantage for identical targets is slightly larger at the trace position (52 vs. 44 ms).

Log-transformed RTs were analyzed using a Bayesian linear mixed regression model fit with the brms package (version 2.16.3, Bürkner, 2017) in the R environment for statistical computing (version 4.1.2, R Core Team, 2021). The model included Position (Control or Trace, sum contrast coded as –0.5 and 0.5), Target Type (Unrelated or Identical, sum contrast coded as –0.5 and 0.5), and Working Memory Score (scaled and centered around the mean) as fixed effects, as well as the interaction between Working Memory Score, Target Type, and Position. Model comparisons indicated that adding C-test score, L2 Exposure, and AoA did not improve the fit of the model; therefore, no additional predictors were included. The random effects structure included random intercepts for participants and items and by-participants and by-items random slopes for Position, Target Type, and their interaction. Model results are provided in Table 3. Bayes factors indicating the extent of support for the existence of an effect in the direction specified in the model output were calculated using the "hypothesis" function from the brms package; in each case, the Bayes factor indicates the ratio of the hypothesis (e.g., $b > 0$) to its complement (e.g., $b < 0$; see Winter & Bürkner, 2021). The estimates of robust effects (Bayes factor ≥ 3) are indicated in bold.

There was a reliable effect of Target Type, which indicates that RTs were faster for identical versus unrelated targets ($P(b < 0) = 10.39$). There was also a reliable effect of Working Memory Score, such that participants with higher working memory had lower

Table 4. Mean RTs (SDs) to visual targets in low-span and high-span participants

	Low-span participants		High-span participants	
	Control position	Trace position	Control position	Trace position
Unrelated target	904 (237)	879 (190)	857 (273)	864 (238)
Identical target	821 (203)	816 (194)	843 (232)	816 (192)

Table 5. Model results for low-span and high-span participants

	Low-span participants				High-span participants			
	Est.	Est. Error	CI L	CI U	Est.	Est. Error	CI L	CI U
Intercept	**−0.17**	0.05	−0.28	−0.07	**0.21**	0.07	−0.35	−0.08
Target Type	**−0.09**	0.04	−0.16	−0.01	−0.01	0.05	−0.10	0.08
Position	−0.02	0.05	−0.12	0.07	−0.01	0.04	−0.10	0.07
Target Type × **Position**	0.00	0.08	−0.17	0.15	**−0.06**	0.09	−0.22	0.12

RTs overall ($P(b < 0) = 3.13$). In addition, there were reliable interactions between Working Memory Score and Target Type ($P(b > 0) = 15.1$) and between Working Memory Score, Target Type, and Position ($P(b < 0) = 6.6$). The former effect indicates that participants with higher working-memory scores showed less of an RT advantage for identical compared to unrelated target pictures; the latter effect indicates that participants with higher working-memory scores showed a larger RT advantage for identical pictures at the gap compared to the control position.

Given the interactions between Working Memory Score and the factors of interest, we split participants into two groups based on the median working memory score (55.7%). This yielded two groups of 10 participants each. Importantly, these groups did not differ significantly in terms of either AoA or L2 exposure ($ps > .2$). Table 4 shows the mean RTs (SDs) per working memory group in the four conditions.

We then analyzed RTs in the low-span and high-span participants separately, again using Bayesian linear mixed regression models with Position and Target Type as fixed effects and the same maximal random effects structure reported in the preceding text. As in the main analysis, informative priors based on Felser and Roberts's (2007) results were used for the effects of Target Type, Position, and their interaction. Model results are provided in Table 5, with the estimates of effects that are reliably present marked in bold. Figures 1 and 2 illustrate the posterior distributions of the model parameters (i.e., estimates of the distributions that take the new data into account) for the low- and high-span groups, respectively.

Table 5 indicates that for the low-span participants, the only reliable effect was an RT advantage for identical compared to unrelated targets ($P(b < 0) = 69.95$). The data are inconclusive regarding a potential advantage for identical targets at the gap relative to the control position (Target Type × Position: $P(b < 0) = 1$). For the high-span participants, the only effect that is reliably present is the Target Type × Position interaction ($P(b < 0) = 3.1$).
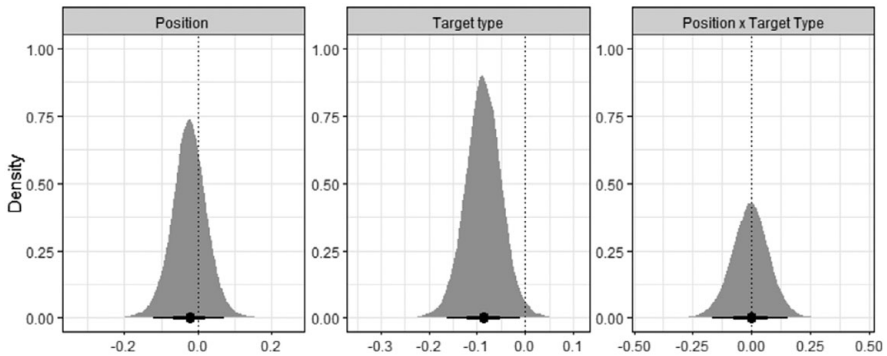
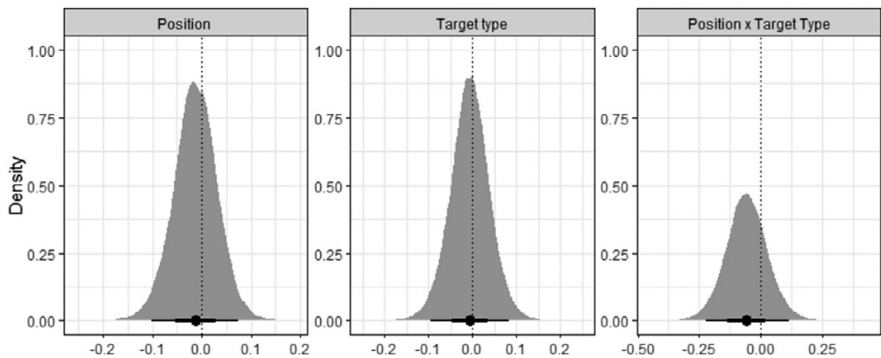**Figure 1.** Posterior distributions: Low-span group.

Figure 2. Posterior distributions: High-span group.

Discussion and Conclusion

This article's aim was to extend previous research on indirect-object dependency resolution to a group of L2 speakers that is understudied in the L2 processing literature, namely early L2 acquirers with extensive (though nonimmersive) naturalistic L2 exposure. This focus was motivated by accounts of L2 processing that predict greater processing automaticity among L2 learners of this profile (e.g., Clahsen & Felser, 2006a, 2006b, 2018; Ullman, 2001), as well as previous studies that have found increased sensitivity to abstract syntactic structure among learners in naturalistic exposure environments (e.g., Pliatsikas & Marinis, 2013). We conducted a close replication of Felser and Roberts (2007). In contrast to these authors, but like Roberts et al. (2007), we observed a working memory effect on our participants' response patterns. Follow-up analyses indicated that while low-working-memory participants responded more quickly to identical targets at both the gap and the earlier control position, high-working-memory participants' RTs to identical targets were lower at the gap than the control position.

The low-working-memory participants' processing pattern, which mirrors that of Felser and Roberts's (2007) participants, would be consistent with a strategy in which the filler was actively maintained in working memory, leading to lower RTs at both test

positions. However, we note a caveat here, which arises due to the affordances of the Bayesian analysis: Specifically, the data do not provide evidence that the low-working-memory group did *not* show a position-specific RT advantage to identical targets; the Bayes factor was inconclusive at 1. We therefore cannot comment further on whether trace reactivation occurred among this group.

There does, however, appear to be a difference in processing pattern between our low-span L2 group and the low-span L1 group in Roberts et al. (2007), who did not show an advantage for identical targets at either position. This difference suggests that even among individuals who share relatively lower working-memory capacity, L1 and L2 processing of movement dependencies may differ. The divergence here may be attributable to different allocations of cognitive resources during processing: For example, Williams (2006) found that L2 speakers with relatively low working-memory capacity, unlike L1 speakers, seemed not to process input incrementally when they also had to perform a memory task, suggesting that the L2 speakers had directed their cognitive resources toward the memory task.

Our high-span group's processing pattern is compatible with a strategy in which the filler is selectively reactivated at the gap position. This finding is in line with the proposal that when a filler is encountered, the parser predicts an upcoming syntactic gap, and retrieval of the filler from memory is triggered when such a gap is reached (e.g., Frazier, 1987). In this respect, our high-working-memory participants showed the same processing pattern as the high-working-memory L1 groups (both adults and children) in Roberts et al. (2007). In turn, this finding aligns with the results of Miller (2014, 2015), in that it shows that L2 learners can make use of abstract syntactic structure during real-time processing. In Miller's (2014, 2015) studies, however, it was a task characteristic, specifically the use of cognates as visual targets, that seemed to facilitate sensitivity to the gap. The present results, like those of Pliatsikas and Marinis (2013), provide an indication that this sensitivity can arise in the absence of targeted attempts to elicit it. Considering, however, that our low- and high-span groups did not differ in terms of AoA or L2 exposure, we cannot say that either of these factors is decisive in engendering trace reactivation. Ultimately, working memory capacity seemed to be the deciding factor in this regard.

Our observation of a working memory effect bears on another important question in SLA, namely whether individual cognitive differences are equally relevant to L2 outcomes across early and late L2 learners. Theories of SLA and L2 processing in which AoA plays a central role (e.g., Clahsen & Felser, 2006a, 2006b, 2018) typically do not discuss the potential effects of individual differences on early learners' L2 attainment, with the implicit assumption being that an early start to learning and sufficient exposure should together ensure acquisition success. However, some studies have observed effects of, for example, language aptitude on L2 outcomes among early learners (Abrahamsson & Hyltenstam, 2008; Granena, 2014). Our results align with these findings and highlight the complex, multifactorial nature of early L2 acquisition (cf. Granena, 2014). Future research might aim to shed additional light on the interplay between environmental and individual-level variables among early L2 learners, particularly with respect to the parsing of complex syntactic structures.

References

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, *30*, 481–509. https://doi.org/10.1017/S027226310808073X

Audacity Team. (2019). Audacity: Free audio editor and recorder. https://audacityteam.org/

Baldwin, S. A., & Fellingham, G. W. (2013). Bayesian methods for the analysis of small sample multilevel data with a complex variance structure. *Psychological Methods*, *18*, 151.

Berghoff, R. (2020). L2 processing of filler-gap dependencies: Attenuated effects of naturalistic L2 exposure in a multilingual setting. *Second Language Research*, *25*, 026765832094575. https://doi.org/10.1177/0267658320945757

Berghoff, R. (2021). The role of English in South African multilinguals' linguistic repertoires: A cluster-analytic study. *Journal of Multilingual and Multicultural Development*, *12*, 1–15. https://doi.org/10.1080/01434632.2021.1941066

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*. https://doi.org/10.18637/jss.v080.i01

Chomsky, N. (1986). *Barriers*. MIT Press.

Chow, W.-Y., & Zhou, Y. (2019). Eye-tracking evidence for active gap-filling regardless of dependency length. *Quarterly Journal of Experimental Psychology*, *72*, 1297–1307. https://doi.org/10.1177/1747021818804988

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

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

Clahsen, H., & Felser, C. (2018). Some notes on the Shallow Structure Hypothesis. *Studies in Second Language Acquisition*, *24*, 1–14. https://doi.org/10.1017/S0272263117000250

Coetzee-Van Rooy, S. (2012). Flourishing functional multilingualism: Evidence from language repertoires in the Vaal Triangle region. *International Journal of the Sociology of Language*, *2012*, 87–119. https://doi.org/10.1515/ijsl-2012-0060

Coetzee-Van Rooy, S. (2013). Afrikaans in contact with English: Endangered language or case of exceptional bilingualism? *International Journal of the Sociology of Language*, *2013*, 179–207.

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1283–1296. https://doi.org/10.1037/0278-7393.26.5.1283

de Stadler, L. (1995). The indirect object in Afrikaans. *South African Journal of Linguistics*, *13*, 26–38. https://doi.org/10.1080/10118063.1995.9723972

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish–English bilinguals. *Bilingualism: Language and Cognition*, *10*, 101. https://doi.org/10.1017/S1366728906002847

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

Fernandez, L., Höhle, B., Brock, J., & Nickels, L. (2018). Investigating auditory processing of syntactic gaps with L2 speakers using pupillometry. *Second Language Research*, *34*, 201–227. https://doi.org/10.1177/0267658317722386

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

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

Granena, G. (2014). Language aptitude and long-term achievement in early childhood L2 learners. *Applied Linguistics*, *35*, 483–503. https://doi.org/10.1093/applin/amu013

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, *177*, 263–277. https://doi.org/10.1016/j.cognition.2018.04.007

Jeffreys, H. (1998). *Theory and probability*. Oxford: Clarendon Press.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*, 60–99. https://doi.org/10.1016/0010-0285(89)90003-0

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, *12*, 251–285. https://doi.org/10.1017/s0272263100009165

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech Language and Hearing Research*, *50*, 940. https://doi.org/10.1044/1092-4388(2007/067)

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language processing. *Studies in Second Language Acquisition*, *27*, 483. https://doi.org/10.1017/S0272263105050035

Miller, A. K. (2014). Accessing and maintaining referents in L2 processing of wh-dependencies. *Linguistic Approaches to Bilingualism*, *4*, 167–191. https://doi.org/10.1075/lab.4.2.02mil

Miller, A. K. (2015). Intermediate traces and intermediate learners. *Studies in Second Language Acquisition*, *37*, 487–516. https://doi.org/10.1017/S0272263114000588

Muñoz, C. (2006). *Age and the rate of foreign language learning*. Multilingual Matters.

Nicol, J. L. (1993). Reconsidering reactivation. In G. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing* (pp. 321–350). Erlbaum.

Nicol, J. L., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, *18*, 5–19.

Norouzian, R., Miranda, M. D. E., & Plonsky, L. (2019). A Bayesian approach to measuring evidence in L2 research: An empirical investigation. *The Modern Language Journal*, *25*, 1. https://doi.org/10.1111/modl.12543

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*, 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pliatsikas, C., Johnstone, T., & Marinis, T. (2017). An fMRI study on the processing of long-distance wh-movement in a second language. *Glossa: A Journal of General Linguistics*, *2*, 1052. https://doi.org/10.5334/gjgl.95

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

Posel, D., & Zeller, J. (2016). Language shift or increased bilingualism in South Africa: Evidence from census data. *Journal of Multilingual and Multicultural Development*, *37*, 357–370. https://doi.org/10.1080/01434632.2015.1072206

Posel, D., Hunter, M., & Rudwick, S. (2020). Revisiting the prevalence of English: Language use outside the home in South Africa. *Journal of Multilingual and Multicultural Development*, *218*, 1–13. https://doi.org/10.1080/01434632.2020.1778707

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/

Roberts, L., Marinis, T., Felser, C., & Clahsen, H. (2007). Antecedent priming at trace positions in children's sentence processing. *Journal of Psycholinguistic Research*, *36*, 175–188. https://doi.org/10.1007/s10936-006-9038-3

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174.

Stone, J. M., & Towse, J. N. (2015). A working memory test battery: Java-based collection of seven working memory tasks. *Journal of Open Research Software*, *3*, 92. https://doi.org/10.5334/jors.br

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, *4*, 105–122. https://doi.org/10.1017/s1366728901000220

von Bastian, Claudia C., Locher, A., & Ruflin, M. (2013). Tatool: A Java-based open-source programming framework for psychological studies. *Behavior Research Methods*, *45*, 108–115. https://doi.org/10.3758/s13428-012-0224-y

Williams, J. N. (2006). Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition*, *9*, 71–88. https://doi.org/10.1017/S1366728905002385

Winter, B., & Bürkner, P.-C. (2021). Poisson regression for linguists: A tutorial introduction to modelling count data with brms. *Language and Linguistics Compass*, *15*, e12439.

Appendix A

Informative priors for the effects of Target Type, Position, and their interaction were based on the reaction times in Felser and Roberts (2007). Each prior was normally distributed with a mean equal to the result obtained in Felser and Roberts (2007) and a standard deviation equal to the mean. The prior distributions are visualized in Figure A1.
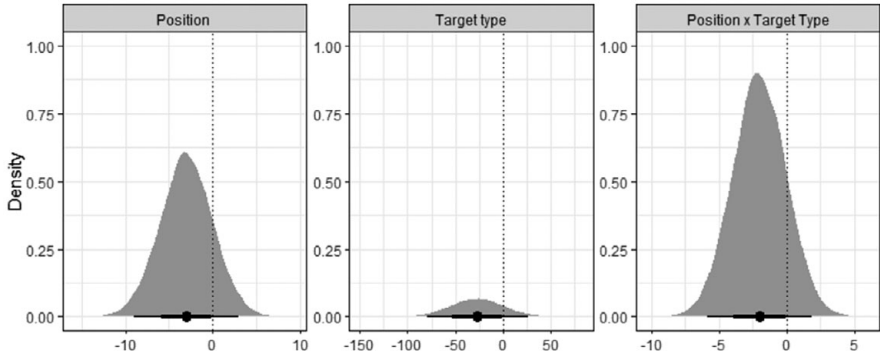

Figure A1. Prior distributions.

Cite this article: Berghoff, R. (2023). Wh-Dependency Processing in a Naturalistic Exposure Context: Sensitivity to Abstract Syntactic Structure in High-Working-Memory L2 Speakers. *Studies in Second Language Acquisition,* **45,** 586–598. https://doi.org/10.1017/S0272263122000304