White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/205509/

Version: Published Version

Article:
Al-Maani, Alaa, Sloggett, Shayne, Grillo, Nino et al. (2024) Testing for Proficiency Effects and Crosslinguistic Influence in L2 Processing: Filler-Gap Dependencies in L2 English by Jordanian-Arabic and Mandarin Speakers. Studies in Second Language Acquisition. ISSN: 1470-1545

https://doi.org/10.1017/S027226312400007X

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Studies in Second Language Acquisition (2024), 1–17
doi:10.1017/S027226312400007X

REPLICATION STUDY

Testing for proficiency effects and crosslinguistic influence in L2 processing: filler-gap dependencies in L2 English by Jordanian-Arabic and Mandarin speakers

Alaa Al-Maani[1], Shayne Sloggett[2], Nino Grillo[2] and Heather Marsden[2]

[1]Al al-Bayt University
[2]University of York
Corresponding author: Heather Marsden; Email: heather.marsden@york.ac.uk

(Received 21 December 2022; Revised 24 October 2023; Accepted 10 November 2023)

Abstract

This study expands on previous research into filler-gap dependency processing in second language (L2) English, by means of a replication of Canales's (2012) self-paced reading study. Canales, among others, found that advanced L2-English speakers exhibited the same processing behavior that Stowe (1986) found for native English processing: On encountering a filler, they posited gaps in licensed positions and avoided positing gaps in grammatically unlicensed island positions. However, the previous L2 studies focused on advanced-level L2 proficiency and did not test specifically for first language (L1) influence. The present study compares two groups of intermediate-level L2-English speakers with contrasting non-wh-movement L1s, Jordanian Arabic and Mandarin, to investigate the effects of L1 influence and individual differences in proficiency. Our results provide evidence that at intermediate level, too, L2 filler-gap processing adheres to grammatical constraints. L1 did not affect this behavior, but proficiency effects emerged, with larger licensed filled-gap effects at higher proficiency.

Introduction

Research on native language (L1) processing of filler-gap dependencies has provided considerable evidence that, on encountering a "filler" such as a *wh*-pronoun at the start of an embedded question, comprehenders actively search for the subsequent associated "gap." For example, using self-paced reading (SPR), Stowe (1986) showed that when native English speakers read sentences containing a filler-gap dependency such as (1a; *who* is the filler), the reading time for the pronoun *us* is slower than for the same pronoun in a sentence with no filler-gap dependency (1b).

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S027226312400007X Published online by Cambridge University Press

1. a. My brother wanted to know who Ruth will bring us home to ___ at Christmas.
   b. My brother wanted to know if Ruth will bring us home to Mom at Christmas.

This behavior is taken to show that, on encountering a filler, the parser posits a gap at each subsequent potential gap site until the correct gap site is identified (the active filler strategy, Frazier & Clifton, 1989).[1] The object of *bring* in (1a) is a potential gap site, so when the parser's expectation of a gap is not met due to the presence of *us*, a reading slowdown occurs.[2] Stowe additionally showed that native speakers' expectations about potential gap sites are constrained by syntactic structure. Gaps cannot occur in so-called syntactic islands, such as complex noun phrases and relative clauses (Ross, 1967). Stowe found that, for complex noun phrases such as the bracketed phrase in (2a–b), there was no difference between the filler-gap (2a) and no-gap (2b) conditions in the speed of processing the first word of the prepositional complement (*Greg's*), despite *about* being a potential a gap licensor in nonisland structures. This suggests that, even though the filler (*what*) indicates a subsequent gap, no attempt is made to fill that gap within the complex noun phrase island, as predicted if the syntactic structure guides parsing.

2. a. The teacher asked what [the silly story about Greg's older brother] was supposed to mean __.
   b. The teacher asked if [the silly story about Greg's older brother] was supposed to mean anything.

A key question in second language (L2) acquisition research concerns the extent to which L2 processing proceeds in the same way as L1 processing. The shallow structure hypothesis (Clahsen & Felser, 2006; 2018; Felser, 2019) proposes that syntax, in particular, may be a source of L1–L2 difference, with real-time computation of syntactic structure potentially being delayed, due to the cognitive demands of L2 processing, and consequently not incorporated into the moment-by-moment parse. Investigation of filler-gap structures offers insight into L2 syntax processing. A number of L2 studies that used the same paradigm as Stowe (1986) found both filled-gap effects and respect of island constraints in advanced L2 English speakers with a range of different L1s (Aldwayan et al., 2010; Canales, 2012; Covey et al., 2022; Johnson et al., 2016). But investigations of other filler-gap dependency structures have not always found L1-like processing. For example, with more complex long-distance filler-gap dependencies, Marinis et al. (2005) found that processing by advanced L2 speakers with several different L1s diverged from that of native speakers in a way that suggested insensitivity to gaps and hence underuse of syntactic structure. Similar results were found by Pliatsikas (2010) in advanced L1-Greek L2-English speakers either with no experience of immersion in an English-speaking environment, or with an average of five years immersion, but Pliatsikas & Marinis (2013) found that a group with an average immersion duration of nine years exhibited native-like processing, suggesting that extensive L2 exposure could lead to real-time integration of syntactic structure in long-distance dependency processing. However, a different long-distance dependency

[1] Alternative approaches exist, which look at dependency formation in terms of direct association between a displaced element and its subcategorizer (Pickering & Barry, 1991). The distinction between gap filling and direct association approaches, while important, is orthogonal to the present work.

[2] The embedded subject *Ruth* is also a potential gap site. Stowe did not find a slowdown here and suggested this was due to its immediate proximity to the filler. Other studies have examined this further using different designs (e.g., Lee, 2004). The present paper focuses only on the object filled-gap position hereafter.

https://doi.org/10.1017/S027226312400007X Published online by Cambridge University Press

study by Berghoff (2022) found that, even among those with extensive L2 exposure, native-like processing was not guaranteed and was subject to individual variation. Berghoff also found a limited effect of proficiency suggestive of a relationship between increased proficiency and increased gap sensitivity, though Berghoff noted that the overall high proficiency of the participants could have meant that clear proficiency effects would not be detectable. An eyetracking study of wh-island processing by Boxell & Felser (2017) found that L2 speakers temporarily violated island constraints by attempting to posit a gap within complex subject islands. Kim et al. (2015) also found island violation during L2-English processing, but only in L1-Korean speakers, and not L1-Spanish, which the authors attributed to L1 influence, due to the absence of island effects in Korean, in contrast to Spanish and English.

Taken together, this body of research reports mixed findings for filler-gap processing, with a potential role for the type of structure, the L1, and individual differences in L2 experience. However, it is notable that the previous research focuses on highly proficient L2 speakers. This means that little is known about the development of L2 syntactic processing (as observed by Felser (2019) in relation to island constraints). Moreover, few studies have systematically tested for L1 influence during L2 processing. The present study addresses these gaps. We report on an investigation of filler-gap and island processing by intermediate-level L2-English speakers whose L1s are Jordanian Arabic or Mandarin. Although the existing findings, with the exception of Kim et al. (2015), do not suggest L1 influence during L2 processing, it could be that the advanced-proficiency participants in previous studies were beyond the stage where influence could be detectable (as Berghoff (2022) suggested for proficiency effects). Thus, our research questions ask whether between-L1 differences with respect to filler-gap structures lead to differential L2-English processing in intermediate-level speakers and whether individual differences in proficiency affect filler-gap processing. The background and motivations for our study are presented in the following two sections. Subsequently, we detail our experiment (a replication of Canales (2012)) and results, followed by discussion.

Wh-questions in English, Jordanian Arabic and Mandarin

English, Jordanian Arabic, and Mandarin differ with regard to the derivation of wh-questions. Broadly, English wh-questions involve wh-movement, whereas Jordanian Arabic and Mandarin wh-questions do not. In the widely adopted formal grammatical account of wh-movement (Chomsky, 1973), the wh-phrase moves to the beginning of the clause from the position where the relevant argument was generated, leaving a "trace" (*t*), as in (3).

3. My brother wanted to know **who**$_i$ Ruth will bring us home to t_i at Christmas.

Within this approach, the wh-island effect in (2a) arises due to a universal constraint on wh-movement that prohibits a constituent in a syntactic island from moving outside that structure, as illustrated through the contrast in (4a–b), where the island is a relative clause.

4. a. *Who$_i$ did the singer [that bothered t_i] criticize the pianist?
   b. Who$_i$ did the singer [that bothered Peter] criticize t_i?

https://doi.org/10.1017/S027226312400007X Published online by Cambridge University Press

An alternative view is that island effects reflect processing constraints rather than syntactic constraints. Under this view, processing of an island structure is costly in terms of cognitive resources, and this leads to inability to resolve filler-gap dependencies within such structures (e.g., Hofmeister & Sag, 2010; Kluender & Kutas, 1993). However, as argued by Omaki & Schulz (2011), the processing account of islands still assumes computation of complex abstract representations (even if those representations are not the wh-movement structures outlined above). Thus, on either account, if L2 speakers posit gaps in filler-gap structures outside islands but do not do so within an island, this would suggest that they have built a relevant abstract representation. The current study does not aim to differentiate between grammatical versus processing-based accounts of islands.

Turning to Mandarin and Jordanian Arabic, neither has gaps in wh-questions. In Mandarin, wh-words remain in situ, as in (5) where the wh-object *shenme* "what" occurs in the canonical postverbal object position (Huang et al., 2009, p. 262).

5. Zhangsan xiang-zhidao Lisi mai-le shenme
   Zhangsan wonder Lisi buy-ASP what
   "Zhangsan wonders what Lisi bought."

Further, wh-phrase arguments are not subject to island effects. Thus *shenme* in (6) can be interpreted outside the relative clause that contains it (Huang et al., 2009, p. 266):

6. Ni zui xihuan [mai shenme de] ren?
   you most like buy what COMP person
   "What$_i$ do you like the person [that bought t$_i$]?"

In Jordanian Arabic, a number of wh-question structures are possible, including wh-in-situ (7a) and the form in (7b) in which the wh-object *mi:n* "who" occurs before the subject and verb, but a co-indexed resumptive pronoun occupies the object position.

7. a. um-i saʔlat il-binit gabalat mi:n
      mother-my asked the-girl met who
      "My mother asked who the girl met."

   b. um-i saʔlat mi:n$_i$ illi il-binit gabalat-uh$_i$
      mother-my asked who that the-girl met-him
      "My mother asked who the girl met."

Al-Daher (2016) argues that none of the Jordanian Arabic question forms involves wh-movement. (See Aoun et al. (2009) for nonmovement accounts of standard and Lebanese Arabic.) The first author of the present paper, who is a native speaker of Jordanian Arabic, reports that wh-in-situ is mainly limited to echo questions, whereas the form in (7b) is widely used as a typical wh-interrogative.[3] As in other languages with resumptive pronouns, island effects are absent, as illustrated in (8) where *mi:n* "who" is

[3]We are not aware of any published information on the distribution of different wh-question forms in Jordanian Arabic. The limited distribution of wh-in-situ has been observed for Najdi Arabic (Albaty, 2013) and Lebanese Arabic (Aoun et al., 2009), both of which are closely related to Jordanian Arabic.

https://doi.org/10.1017/S027226312400007X Published online by Cambridge University Press

co-indexed with a resumptive pronoun inside a relative clause (cf. the ungrammaticality of the corresponding English form (4a)).

8. Mi:n_i [illi azʕaʒ-uh_i] il-mutrib illi intaqad ʕazef li-piano
   Who$_i$ that bothered-him$_i$ the-singer that criticised player the-piano
   'Who is it that the singer who criticised the pianist bothered?'

Thus, the typical wh-question structure in Jordanian Arabic differs from Mandarin, but the two languages also differ from English in that neither exhibits wh-movement or island effects. These crosslinguistic differences are built into the predictions for our experiment, as detailed in the next section.

Motivation and predictions

The present research builds on four previous L2 studies that, following Stowe's (1986) seminal study, used the filled-gap paradigm in the absence of syntactic islands (e.g., 1a–b) and the presence of syntactic islands (e.g., 2a–b) to investigate grammatical processing in L2 English: Aldwayan et al. (2010) with L1-Najdi Arabic speakers; Canales (2012), L1-Spanish; Johnson et al. (2016), L1-Korean; and Covey et al. (2022), L1-Mandarin. The first three used SPR and found that, like L1-English speakers, L2 speakers slowed down at the filled-gap position in filler-gap sentences with no syntactic island (1a) but exhibited no slowdown within syntactic islands (2a). Covey et al. used electroencephalography and also found effects that differed between filled-gap positions in filler-gap sentences and filled-gap positions within syntactic islands. Notably, the L2-English group exhibited a P600 effect (indicative of difficulty with syntactic integration) at licit filled-gap sites and none within islands, whereas an L1-English group exhibited an N400 effect (indicative of semantic anomaly detection). Covey et al. argued that, although the event-related potenial effects were qualitatively different between the two groups, they nonetheless testified to sensitivity to island constraints on filler-gap processing in the L2 group as well as the L1. Thus, all of these studies' findings suggest that advanced L2 speakers actively search for a gap on encountering a filler and that the search is constrained by syntactic structure, such that positing a gap in an illicit position is not attempted. In short, the findings suggest full use of syntactic structure during L2 processing.

Using a different task, Kim et al. (2015) found contrasting L2-English processing patterns in L1-Spanish speakers compared with L1-Korean speakers. The former group appeared to posit gaps only at licit (nonisland) gap sites and not within islands, while the latter appeared to posit gaps in both nonisland *and* island structures. Both groups exhibited knowledge of island constraints in an offline task, leading the authors to interpret the online island violation by the L1-Korean group as reduced ability to recognize the island structure in real time due to influence from Korean, where wh-phrases are interpretable within corresponding structures (similarly to Mandarin (6)). The contrast between these results and Johnson et al.'s (2016), who found no island violation in L1-Korean participants, could relate to proficiency or immersion. While no direct comparison of English proficiency can be made across the two studies, Kim et al.'s participants had had, on average, 3.6 years' immersion in the United States while Johnson et al.'s had had around 6. If lower proficiency or immersion played a role, then investigation of lower proficiency speakers, as in the current study, could lead to detection of L1-influenced processing.

https://doi.org/10.1017/S027226312400007X Published online by Cambridge University Press

As outlined above, the two L1s in the present study are both non-wh-movement languages. Following Kim et al., this could lead to L1-influence whereby both groups attempt to fill gaps in islands, as well as in nonisland structures. However, we predict that such behavior will be more evident in L1-Mandarin speakers than L1-Jordanian Arabic because of the difference between the two languages in wh-object question form: in Mandarin, wh-phrases occur in situ, whereas the wh-phrase is typically clause-initial in Jordanian with a resumptive pronoun in object position (7b). In the Jordanian Arabic structure, a dependency exists between the wh-phrase and the resumptive that is superficially similar to English filler-gap dependencies. Processing research on filler–resumptive dependencies in Hebrew has yielded surprisal effects akin to filled-gap effects (Keshev & Meltzer-Asscher, 2017).[4] This suggests that processing of the clause-initial wh-word triggers a search for the associated resumptive pronoun, similarly to an active gap search in English. Influence from such a processing strategy could accelerate development of Jordanian Arabic speakers' L2 English filler-gap processing relative to Mandarin speakers.[5] Thus, Jordanian Arabic speakers may differentiate between nonisland and island structures, attempting to fill a gap in the former but not in the latter, whereas Mandarin speakers may treat the two structures the same and attempt to fill gaps within both.

Turning to L2 proficiency, it could be the case that at lower proficiency, a speaker who comprehends a filler-gap dependency sentence after reading it may nonetheless be delayed in their real-time processing of the dependency, in the same way that Hopp (2015) found integration of morphosyntactic information to be delayed in lower proficiency compared to higher proficiency speakers during L2 German processing. If filler-gap dependency processing is delayed, lower proficiency speakers may exhibit no slowdown, or a smaller slowdown, at licensed filled-gap sites. This would lead to no differentiation between nonisland and island structures at lower proficiency, with no filled-gap effect in either.

Bringing all of the above together, we predict that the intermediate-level L2 speakers in the present study will demonstrate both a filled-gap effect and sensitivity to islands (like previous studies' advanced L2 speakers) but that these effects will be modulated by L1 and proficiency. Specifically, the filled-gap effect will be larger in the Jordanian Arabic group than in the Mandarin group, relative to the absence of such an effect in islands, and it will be larger in participants with higher proficiency than lower proficiency. We articulate our hypotheses in relation to the experiment design in the next section, where we also set out how we operationalize detection of a filled-gap effect.

Method

Participants

Eighty L2 English speakers participated in the SPR experiment, 40 L1-Jordanian Arabic, and 40 L1-Mandarin. They were all university students: the Jordanian speakers on an English-medium programme in Jordan, the Mandarin speakers in the UK. All had begun learning English at primary school in Jordan or China. Prior to completing the experiment, a larger group of participants (60 Jordanian, 45 Mandarin) completed a 40-point multiple choice proficiency test (Quick Placement Test Part

[4]To our knowledge, there is no processing research on filler-resumptive dependencies in Jordanian Arabic.

[5]Some studies of relative clause processing in Mandarin suggest an active gap strategy, whereby detection of a gap triggers a search for its filler (Ng, 2008). However, transfer of such a strategy to the L2 would not directly facilitate filler-gap processing.

https://doi.org/10.1017/S027226312400007X Published online by Cambridge University Press

1, Oxford University Press et al., 2001) and a brief background questionnaire, using the online survey tool Qualtrics. Those who attained a proficiency score corresponding to the B2 (upper intermediate) range of the Common European Framework of Reference for Languages went on to complete the SPR task. Table 1 summarizes each group's age details and proficiency scores. Although there was no significant difference in proficiency between the two groups, ($t = -0.047$, $p = 0.96$), we acknowledge that the English language exposure differs between the two groups: none of the Jordanian speakers had lived in an English-speaking country, though they did experience regular immersion in English through teaching, assessment, and academic-related communication on their English-medium university programme. The Mandarin speakers, on the other hand, had all lived in the UK from between seven months to five years (mean: 28 months, *SD*: 12.74). We note that previous L2 studies (outlined in the introduction) found effects of length of immersion on long-distance dependency processing only in those with considerably longer immersion than our L1-Mandarin participants had. In the "Data Analysis" section, we probe the relationship between proficiency and length of UK residence further for the Mandarin group.

No native English group was included because processing of filler-gap dependencies by this population has been extensively examined previously by Stowe (1986) and the subsequent L2 studies, yielding findings that unambiguously converge on the pattern outlined in the introduction.

Materials

Our materials replicate those used by Canales (2012), who in turn partially replicated Aldwayan et al.'s (2010) adaptation of Stowe (1986). An SPR task was created, with two sets of critical sentences corresponding to two sub-experiments: the filled-gap subexperiment (9a–b) and the wh-island subexperiment (10a–b, where the island is a relative clause). The sentence pairs in both sub-experiments comprised a gap condition (9a, 10a) containing sentences with an embedded wh-question, and a No Gap condition containing sentences with an if-clause (9b, 10b).

9. Filled-gap subexperiment
   a. My cousin wondered who David will put Liz near ___ at the wedding.
   b. My cousin wondered if David will put Liz near Jack at the wedding.

10. Wh-island subexperiment
   a. The director questioned who the singer [that bothered Becky last season] criticized ___ after the concert.
   b. The director questioned if the singer [that bothered Becky last season] criticized the pianist after the concert.

Twenty sentence pairs were created for each subexperiment. In the filled-gap subexperiment, the critical region was a three-letter name after the embedded verb (*Liz*,

Table 1. Participants' age and proficiency task scores, by group

	Age		Proficiency scores (/40)	
	Mean (*SD*)	Range	Mean (*SD*)	Range
Jordanian Arabic (n=40)	31.6 (4.48)	21–40	32.4 (2.47)	30–38
Mandarin (n=40)	25.6 (4.43)	18–37	32.5 (2.23)	30–37

https://doi.org/10.1017/S027226312400007X Published online by Cambridge University Press

in 9). In the wh-island experiment, the critical region was a five-letter name after the verb inside the relative clause (*Becky*, in 10). Sentences were divided across two presentation lists so that no participant read the same sentence in both conditions. Within each list, the 40 critical sentences were randomly combined with 80 fillers, which were designed to distract participants' attention from the critical sentences using a range of structures that matched the critical sentences in complexity and length.

Following the 40 critical trials and twenty of the fillers, a comprehension question was presented, such as (11), which followed (9a):

11. Did the sentence suggest that Liz will attend the wedding?

Half of the comprehension questions required a *Yes* answer and half *No*. Our use of yes-no questions deviates from Canales (and the other L2 replications), who used a fill-the-blank task, whereby, after each SPR trial, the whole sentence appeared with one word missing, and participants selected from two options to fill the blank with the word they had just read. We used yes-no questions on the grounds that they stimulate reading for meaning whereas the fill-the-blank task could rely on more superficial memorization of words. A further deviation was to include comprehension questions after only 20 of the fillers instead of all of them, in order to decrease the task-taking burden, in light of our participants' less advanced proficiency.[6]

The experiment was built using *Linger* Software (http://tedlab.mit.edu/~dr/Linger/). Masked noncumulative word-by-word presentation was used. Participants pressed the space bar on a computer keyboard to reveal the next word. The time taken for each button press was measured. Participants completed the experiment on a laptop, in a quiet room, in the presence of a researcher.

Experimental hypotheses

We tested the hypotheses in (12–13), expressed in terms of the size of the filled-gap effect, which is defined immediately below.

12. Hypothesis 1, L1 influence:
    Jordanian Arabic speakers will demonstrate a larger filled-gap effect in the filled-gap subexperiment relative to the wh-island subexperiment than Mandarin speakers.

13. Hypothesis 2, proficiency:
    Higher proficiency L2 speakers will demonstrate a larger filled-gap effect in the filled-gap subexperiment relative to the wh-island subexperiment than lower proficiency L2 speakers.

We operationalise the size of the filled-gap effect as the difference in reading times at the critical and spillover (i.e., post-critical) words between the gap and no gap conditions in the filled-gap subexperiment (9a–b) relative to the wh-island subexperiment (10a–b). In other words, we include experiment as an interaction term within our statistical modeling. This is a departure from the previous studies (Canales, 2012; Aldwayan et al. 2010; and others) which analyzed the two subexperiments separately. However, under the logic of null hypothesis significance testing, separate analyses do not statistically address the

[6]The materials are available in the IRIS L2 repository: https://www.iris-database.org/details/EmLpY-5yKlU.

https://doi.org/10.1017/S027226312400007X Published online by Cambridge University Press

question of whether a null effect in the wh-island subexperiment (i.e., the predicted absence of a filled-gap effect) can be construed as meaningfully different behavior from a filled-gap effect in the filled-gap subexperiment. We contend that the interaction of experiment (filled-gap versus wh-island) and clause type (gap versus no gap) is crucial for establishing whether processing differs between island and nonisland gap structures.[7]

Data processing and analysis

Before analyzing the reading times, we conducted two preliminary investigations relating to accuracy on the comprehension questions. The two groups' mean proportions of comprehension question accuracy were similar, at 0.77 (SD, 0.42) for the Jordanian Arabic group and 0.8 (SD, 0.4) for the Mandarin group.

The first investigation concerned the L1 Mandarin group's length of immersion in the UK. Unsurprisingly, this was strongly correlated with proficiency ($r = 0.69$, $p<0.001$). However, subsequent analysis of comprehension question accuracy showed that proficiency was a better predictor of accuracy than years of immersion. This was determined by fitting three separate mixed-effects logistic regression models to comprehension accuracy, with fixed effects as in (14) (and random effects for participants and items):[8]

14. a. Model 1: proficiency only
    b. Model 2: immersion only
    c. Model 3: proficiency + immersion

We compared Models 1 and 2 to Model 3 using the analysis of variance (ANOVA) function in R. This showed that including both immersion and proficiency (Model 3) provided no significant improvement in the model fit relative to including only proficiency (Model 1; $p = 0.95$) but relative to only immersion (Model 3 versus Model 2), the fit was marginally improved ($p = 0.05$). Model 1 showed a robust effect of proficiency ($\beta = 0.126$, $p <0.01$) on question accuracy, while Model 2 showed no effect of immersion ($\beta = 0.014$, $p = 0.1$). Model 3 ($\beta = 0.129$, $p = 0.044$) reflected the effect of proficiency found in Model 1 (given that the comparison of Model 3 with Model 1 showed no significant improvement of fit). These results suggest that inclusion of proficiency in our subsequent analyses will capture any effects that could arise from our L1-Mandarin participants' immersion duration; though we acknowledge that we cannot discount the contribution of immersion to the L1 Mandarin group's proficiency and processing development.

The second preliminary investigation concerned the effect of proficiency score on comprehension accuracy between groups. A mixed-effects logistic regression model was fitted to the comprehension scores. The fixed effects were proficiency score and L1, with random effects for participants and items. The L1 factor was sum-coded (Mandarin = −1, Jordanian Arabic = 1), and proficiency scores were centered around the means. The results (Table 2) indicate a significant main effect of proficiency, with

[7]We note the reasonable concern that across the two subexperiments, the critical word is in a different position, and is itself a different word. If the focus of the experiment were only island versus nonisland environments, this inconsistency would constitute a real problem. However, because the crucial effect is an interaction of clause type (gap versus no gap) and experiment (island versus nonisland), any effects that can be solely attributed to lexical differences in the island/nonisland manipulation should not affect our estimation of the interactive term.

[8]All mixed-effects models were run using the lme4 package (Bates et al., 2015) in the R statistical environment (R Core Team, 2022).

https://doi.org/10.1017/S027226312400007X Published online by Cambridge University Press

Table 2. Mixed-effects logistic regression model results for comprehension question accuracy

Coefficient names	Estimate	SE	*p*
L1: Jordanian Arabic	–0.059	0.074	0.429
Proficiency	0.164	0.033	<0.001
L1 × Proficiency	0.039	0.033	0.243

the positive coefficient (β = 0.16) confirming that increased proficiency predicts increased comprehension accuracy.

Given this result and our goal of investigating individual differences in proficiency, we decided *not* to exclude any reading time data on the basis of comprehension question accuracy—contra the procedures in Canales (2012) and other previous studies. Exclusion of data due to lower comprehension scores could lead to failure to detect an effect of proficiency. Instead, proficiency is included as a factor in our reading times analyses.

For these analyses, we excluded reading times falling outside of the range of 100 ms and 2,500 ms, which resulted in removal of 3.5% of the Jordanian data and 0.67% of the Mandarin data. We fitted linear mixed-effects models to the log-transformed reading times for the critical and spillover words, with fixed effects and interactions of experiment (filled-gap, wh-island; sum-coded as –1, 1), clause type (no gap, gap; –1, 1), L1 (Jordanian Arabic, Mandarin; –1, 1), and proficiency score (centered). Maximal random-effects models failed to converge, so we iteratively excluded the random effects associated with the least amount of variance until convergence was achieved. The models that converged included random intercepts for participants and items, with experiment, clause type, and their interaction as by-participant random slopes in the critical region model. For the spillover region, the model that converged additionally included random by-item slopes for clause type and proficiency. Follow-up nested models were used to shed light on significant interactions in the omnibus models. The lmerTest package (Kuznetsova et al., 2017) was used to calculate *p*-values for fixed effects via Satterthwaite approximation.

Results

The mean raw reading times by segment for the filled-gap subexperiment and the wh-island subexperiment are presented in Figures 1 and 2, respectively.

Figure 1 shows that, in the filled-gap subexperiment, both groups had descriptively longer reading times at the critical region (*Liz*) and the spillover region (*near*) in the gap condition than in the no gap condition. However, in the wh-island subexperiment in Figure 2, reading times at the critical region (*Becky*) and the spillover region (*last*) appear similar in the gap and no gap conditions, for both groups.

The results of the mixed-effects models are given in Table 3.

The results of particular importance for our goals are the interactions that include experiment with clause type. If, as found in previous research, the L2 speakers posit gaps in licensed positions and avoid positing gaps in illicit positions then in the filled-gap experiment but not the wh-island experiment, reading times should be longer in gap clauses than no gap clauses. Such behavior is confirmed by the significant two-way interactions of experiment with clause type at both the critical and spillover regions ($\beta = -0.011$, $p = .008$; $\beta = -0.009$, $p = .010$), in conjunction with examination of the descriptive results in Figures 1 and 2. However, to test our hypotheses about L1

https://doi.org/10.1017/S027226312400007X Published online by Cambridge University Press

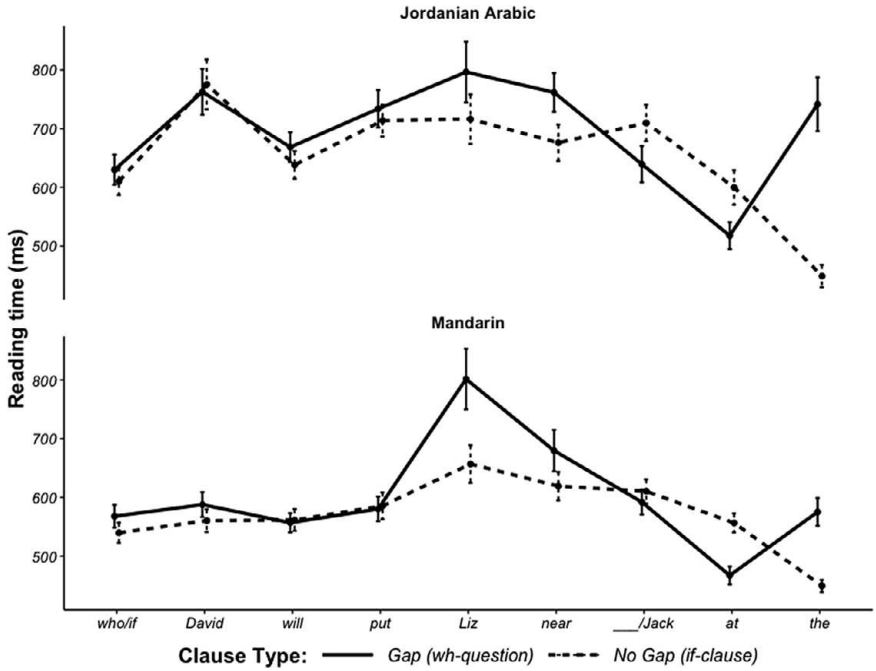

Figure 1. Mean raw readings times with standard errors, by group, in the filled-gap subexperiment, for sentences such as (9a–b).

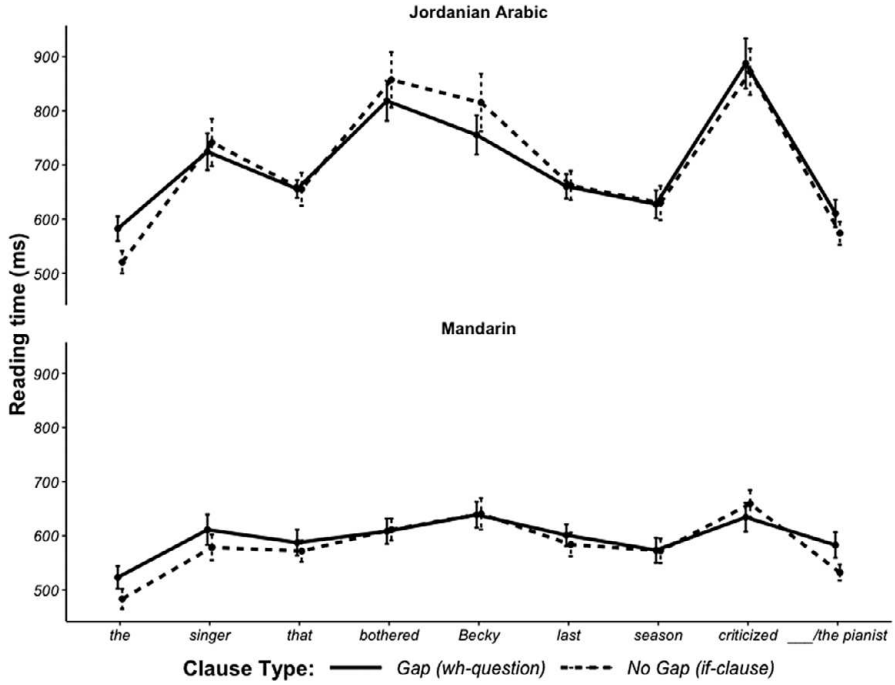

Figure 2. Mean raw readings times with standard errors, by group, in the wh-island subexperiment, for sentences such as (10a–b).

https://doi.org/10.1017/S027226312400007X Published online by Cambridge University Press

Table 3. Linear mixed-effects model coefficients for log-transformed reading times at the critical and spillover regions

Coefficient names	Critical region			Spillover region		
	Estimate	SE	*p*	Estimate	SE	*p*
L1: Mandarin	−0.021	0.012	.096	−0.021	0.010	.048
Experiment: Wh-island	−0.002	0.005	.593	−0.014	0.004	<.001
ClauseType: Gap	0.010	0.003	.004	0.011	0.003	<.001
Proficiency	0.012	0.005	.025	0.012	0.005	.011
L1 × Experiment	−0.014	0.004	<.001	−0.001	0.003	.716
L1 × ClauseType	0.003	0.003	.314	−0.000	0.003	.871
Experiment × ClauseType	−0.011	0.004	.008	−0.009	0.003	.010
L1 × Proficiency	0.002	0.005	.775	−0.002	0.004	.717
Experiment × Proficiency	−0.009	0.002	<.001	−0.002	0.002	.236
ClauseType × Proficiency	0.005	0.002	<.001	0.003	0.001	.014
L1 × Experiment × ClauseType	−0.002	0.004	.698	0.003	0.003	.299
L1 × Experiment × Proficiency	0.002	0.002	.359	−0.003	0.001	.045
L1 × ClauseType × Proficiency	<−0.001	0.002	.852	0.003	0.001	.035
Experiment × ClauseType × Proficiency	−0.005	0.002	.007	−0.003	0.001	.014
L1 × Experiment × ClauseType × Proficiency	<0.001	0.002	.779	<0.001	0.001	.871

influence and L2 proficiency, we need to examine the three-way interactions of L1 × Experiment × ClauseType and Experiment × ClauseType × Proficiency, and the four-way interaction.

Considering L1 influence first, Hypothesis 1 (12) proposed that Jordanian Arabic speakers would demonstrate a larger filled-gap effect in the filled-gap subexperiment relative to the wh-island subexperiment than Mandarin speakers. This predicts an interaction of L1 with experiment and clause type. However, neither the three-way interaction of these factors nor the four-way interaction that additionally included proficiency were significant at either the critical or the spillover region. Thus, Hypothesis 1 is not supported. Although L1 was involved in other significant effects, since those do not involve clause type and experiment together, they do not provide evidence pertinent to Hypotheses 1. In short, there is no evidence of Mandarin speakers' processing being different from Jordanian Arabic speakers' between island and nonisland gap structures.

Turning to proficiency, Hypothesis 2 (13) predicted that higher proficiency L2 speakers would demonstrate a larger filled-gap effect in the filled-gap subexperiment relative to the wh-island subexperiment than lower proficiency L2 speakers. Here, the interaction of Experiment × ClauseType × Proficiency was significant at the critical and spillover regions ($\beta = -0.005$, $p = .007$; $\beta = -0.003$, $p = .014$). To probe the source of this interaction, we ran follow-up nested linear mixed-effects models: proficiency nested within clause type within experiment, with participants and items as random effects. Table 4 presents the results.

The bottom four rows of Table 4 are informative about the source of the interaction. Notably, the penultimate row shows an effect of proficiency in the gap condition of the filled-gap sub-experiment, with reading times increasing significantly with proficiency at both the critical and spillover regions ($\beta = 0.002$, $p < .001$; $\beta = 0.002$, $p < .001$). There was no such effect in the corresponding no gap condition. In the wh-island subexperiment there was also no effect of proficiency with the gap or no gap conditions in the critical region model, though there were significant effects at the spillover region ($\beta = 0.019$, $p = .033$; $\beta = 0.010$, $p = .049$). This suggests that, while proficiency affected

https://doi.org/10.1017/S027226312400007X Published online by Cambridge University Press

Table 4. Results of nested models

Coefficient names	Critical region			Spillover region		
	Estimate	SE	*p*	Estimate	SE	*p*
Experiment: Wh-island	–0.003	0.004	.414	–0.014	0.003	<.001
Filled-gap / ClauseType	0.022	0.004	<.001	0.019	0.004	<.001
Wh-Island / ClauseType	<–0.001	0.004	.958	0.002	0.004	.620
Filled-gap / NoGap / Proficiency	0.002	0.006	.060	0.007	0.005	.137
Wh-Island / NoGap / Proficiency	0.032	0.006	.754	0.019	0.005	.033
Filled-gap / Gap / Proficiency	0.002	0.006	<.001	0.002	0.005	<.001
Wh-Island / Gap / Proficiency	0.003	0.006	.635	0.010	0.005	.049

response times at the spillover region in the wh-island subexperiment, it did so regardless of the presence or absence of a gap. However, in the filled-gap subexperiment, proficiency only significantly impacted responses if the host structure contained a gap. The plots in Figure 3 provide a visualization of the relative magnitudes of the proficiency effects across the two subexperiments.

Figure 3 shows that, across both the critical and spillover regions, reading times in the filled-gap subexperiment become notably slower in the gap condition relative to no gap as proficiency increases. However, there is no discernible interaction of proficiency and clause type within the wh-island subexperiment (despite the significant effects in the spillover region model). The evidence from Table 4 and Figure 3 together confirms Hypothesis 2: Increased proficiency predicted increased sensitivity to filler-gap dependencies, in the form of slower reading times at the critical filled-gap position in gap clauses in the filled-gap subexperiment relative to the wh-island experiment. We discuss the implications of these findings in the next section.

Discussion

The aim of this replication study was to investigate whether between-L1 differences with respect to filler-gap structures lead to differential L2 English processing and whether individual differences in L2 proficiency modulate intermediate-level L2 filler-gap processing. We discuss each of these in turn.

The findings showed that both L1-Jordanian Arabic and L1-Mandarin L2-English speakers exhibited processing behavior akin to native English processing: They slowed down at an object-filled gap position in wh-questions but not at a corresponding position in if-clauses with no gap, and they did not slow down at unlicensed gap sites inside wh-islands. This behavior suggests that intermediate L2-English speakers instigate an active search for a gap on encountering a filler and that the search is guided by real-time building of a detailed syntactic representation, as argued for advanced L2 speakers in previous studies using the same materials design.

With regard to L1 influence, there was no difference between the two groups. Thus, even though L1-Jordanian Arabic has filler-resumptive structures that may be processed similarly to English filler-gap structures, whereas Mandarin has no parallel, there was no evidence of facilitation in the Jordanian Arabic group relative to the Mandarin group. The posited facilitation could have accelerated development of filler-gap processing in Jordanian Arabic speakers, leading to more pronounced slowdowns at licit gap sites or earlier development of sensitivity to island structures in this group, relative to the Mandarin group. Either of these outcomes would have led to an interaction of L1 × Experiment × ClauseType, but no interaction arose.

https://doi.org/10.1017/S027226312400007X Published online by Cambridge University Press

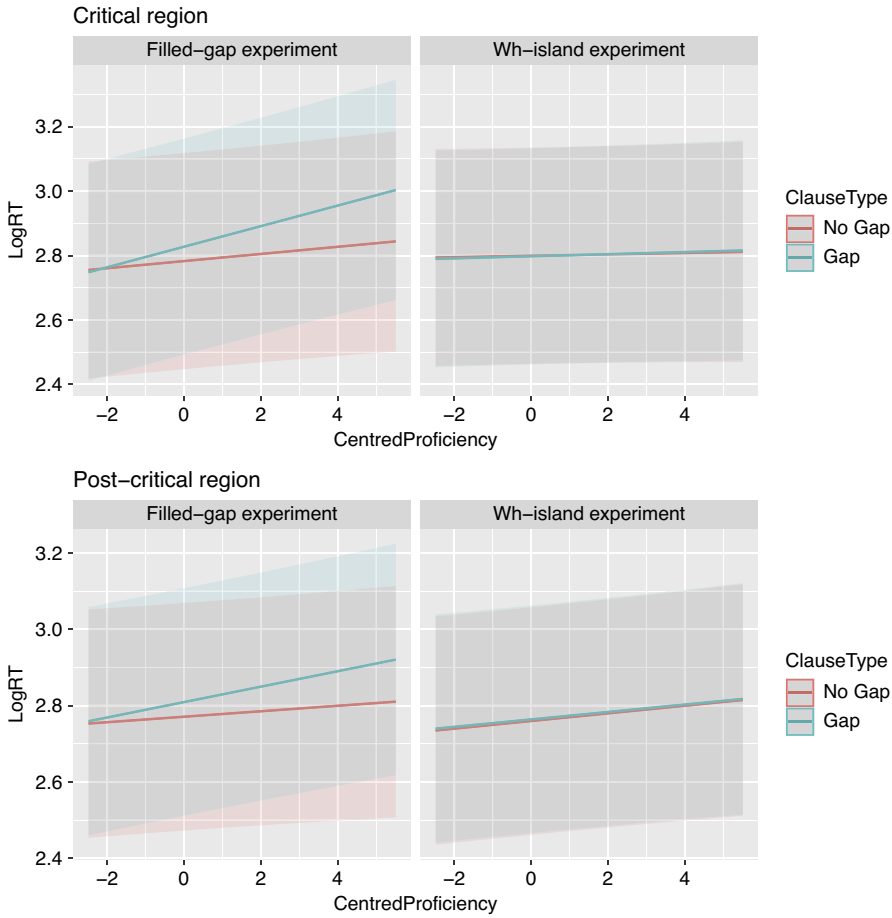

Figure 3. Model-predicted log-response times for the interaction of clause type and experiment as a function of centred participant proficiency at the critical and spillover regions.

The absence of a between-L1 difference in our study could have a number of explanations. First, we acknowledge that immersion experience could play a role. Even though the groups were matched for proficiency, the immersion experience in the L1-Mandarin group could have facilitated their English processing relative to the L1-Jordanian group and consequently obscured a possible between-L1 difference (though previous studies—outlined in the introduction—only found effects of immersion in participants with considerably longer immersion experience than our L1-Mandarin group). Investigation of L1-Mandarin speakers with no English immersion experience could resolve this issue.

Second, considering the difference between our findings and Kim et al.'s (2015), the salience of the islandhood cue could play a role (as noted by Felser (2019)). In our study, this was the relative pronoun *that*, whereas in Kim et al., which also used subject-relative-clause islands, the relative pronoun was *who*. Analysis of written English corpora shows that *that* is increasingly more frequent as a relative clause marker than

https://doi.org/10.1017/S027226312400007X Published online by Cambridge University Press

who (Fajri & Okwar, 2020), which could mean the island structure was easier to identify in our sentences than in Kim et al.'s, though further research is needed to test this suggestion. Alternatively, the grammar of the L1 may make a difference. Korean sentence structure, with rigid verb-final word order, arguably differs more greatly from English than that of Jordanian Arabic or Mandarin, which both make considerable use of subject-verb-object word order. The different tasks used in the current study and Kim et al. could also play a role.

Finally, our finding of no L1 effect is in line with what previous L2 processing research suggests, though few dependency processing studies have directly compared different L1 groups. However, recalling that Boxell & Felser (2017) found island violation in early, but not late, eye-tracking measures (in considerably more complex sentences than those in the current study), it would be worth using eye-tracking for more fine-grained investigation of possible L1 influence.

In terms of proficiency, there was a clear effect whereby the size of the filled-gap effect increased with increasing proficiency. This was due to a less pronounced slow-down by lower proficiency speakers at the filled gap in the gap condition of the filled-gap subexperiment (and not to any attempt to posit gaps in islands). As proposed above, this could be a result of delayed integration of syntactic information. Less proficient speakers may be less efficient in real-time integration of syntactic information so that any reflex of processing a filled gap is diffused over more than just the critical and spillover word, leading to a smaller slowdown at these regions. Within islands, such an effect would obscure any attempt to posit a gap, meaning that it becomes less clear whether the absence of island effects at lower proficiency can be interpreted as evidence of real-time building of a detailed syntactic representation. Nonetheless, the overall adherence in the present data to the target-like processing pattern found in previous studies' advanced L2 speakers, suggests that lower proficiency speakers also process filler-gap dependencies in the same way as in native-language processing, though efficiency of processing increases with proficiency.

Conclusion

The key contributions of this replication study are to investigate L2 filler-gap dependency processing at a lower proficiency than in previous research and to test for L1 influence by comparing speakers of two contrasting L1s. Our findings suggest that intermediate-level L2 speakers process filler-gap dependencies in the same way as advanced speakers and native speakers: on encountering a filler, they posit gaps in licensed positions and avoid positing gaps in illicit positions. There was no evidence of L1 influence, which is consistent with most existing findings, though L1 influence in filler-gap dependency processing merits further research. Regarding proficiency, we found an effect of individual differences: the filled-gap effect increased with increasing proficiency relative to the absence of such an effect in islands, showing that even within a narrowly defined proficiency range, individual differences predict L2 performance.

Acknowledgments. We would like to acknowledge that this research was supported by a PhD scholarship to the first author from Al al-Bayt University (Grant Ref: 4/2/2/6063). Aspects of the research were presented at EuroSLA 29 (Lund, 2019) and EuroSLA 30 (Barcelona [online], 2021). We are grateful to the SSLA reviewers and editor for detailed, helpful feedback.

Competing interest. The authors declare none.

https://doi.org/10.1017/S027226312400007X Published online by Cambridge University Press

Data availability statement. The experiment in this article earned Open Data and Open Materials badges for transparent practices. The materials are available at https://doi.org/10.48316/EmLpY-5yKlU and the data at https://www.iris-database.org/details/kf3XJ-zbaFr.

References

Albaty, Y. A. (2013). Wh-in-situ in Najdi Arabic. *Working Papers of the Linguistics Circle of the University of Victoria*, *23*, 1–14.

Al-Daher, Z. (2016). Pseudo wh-fronting: A diagnosis of wh-constructions in Jordanian Arabic. (Unpublished doctoral dissertation). University of Manitoba. http://hdl.handle.net/1993/31940

Aldwayan, S., Fiorentino, R., & Gabriele, A. (2010). Evidence of syntactic constraints in the processing of wh-movement: A study of Najdi Arabic learners of English. In B. VanPatten & J. Jegerski (Eds.), *Research in second language processing and parsing* (pp. 65–86). John Benjamins. https://doi.org/10.1075/lald.53.03ald

Aoun, J., Benmamoun, E., & Choueiri, L. (2009). *The syntax of Arabic*. Cambridge University Press. https://doi.org/10.1017/CBO9780511691775

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

Berghoff, R. (2022). L2 processing of filler-gap dependencies: Attenuated effects of naturalistic L2 exposure in a multilingual setting. *Second Language Research*, *38*, 373–393. https://doi.org/10.1177/0267658320945757

Boxell, O., & Felser, C. (2017). Sensitivity to parasitic gaps inside subject islands in native and non-native sentence processing. *Bilingualism: Language and Cognition*, *20*(3), 494–511. https://doi.org/10.1017/S1366728915000942

Canales, A. J. (2012). *Online processing of wh-dependencies in English by native speakers of Spanish*. (Unpublished doctoral dissertation) University of Kansas. http://hdl.handle.net/1808/10281

Chomsky, N. (1973). Conditions on transformations. In S. Anderson & P. Kiparsky (Eds.), *A fest-schrift for Morris Halle* (pp. 232–286). Holt, Rinehart, and Winston.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*, 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*, 693–706. https://doi.org/10.1017/S0272263117000250

Covey, L., Fiorentino, R., and Gabriele, A. (2022). Island sensitivity in L2 learners: evidence from event-related potentials and acceptability judgments. *Second Language Research*. Published online August 24, 2022. https://doi.org/10.1177/02676583221116039

Fajri, M. S. A., & Okwar, V. (2020). Exploring a diachronic change in the use of English relative clauses: A corpus-based study and its implication for pedagogy. *SAGE Open*, *10*. https://doi.org/10.1177/2158244020975027

Felser, C. (2019). Structure-sensitive constraints in non-nativesentence processing. *Journal of the European Second Language Association*, *3*, 12–22. https://doi.org/10.22599/jesla.52

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

Hofmeister, P., & Sag, I. A. (2010). Cognitive constraints and island effects. *Language*, *86*, 366–415. https://doi.org/10.1353/lan.0.0223

Hopp, H. (2015). Semantics and morphosyntax in predictive L2 sentence processing. IRAL – International Review of Applied Linguistics in Language Teaching, *53*, 277–306. https://doi.org/10.1515/iral-2015-0014

Huang, C.-T. J., Li, Y.-H. A., & Li, Y. (2009). *The syntax of Chinese*. Cambridge University Press. https://doi.org/10.1017/CBO9781139166935

Johnson, A., Fiorentino, R., & Gabriele, A. (2016). Syntactic constraints and individual differences in native and non-native processing of wh-movement. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/FPSYG.2016.00549/FULL

Keshev, M., & Meltzer-Asscher, A. (2017). Active dependency formation in islands: How grammatical resumption affects sentence processing. *Language*, *93*, 549–568. https://doi.org/10.1353/LAN.2017.0036

https://doi.org/10.1017/S027226312400007X Published online by Cambridge University Press

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

Kluender, R., & Kutas, M. (1993). Subjacency as a processing phenomenon. *Language and Cognitive Processes*, *8*, 573–633. https://doi.org/10.1080/01690969308407588

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*. https://doi.org/10.18637/jss.v082.i13

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

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

Ng, S. (2008). An active gap strategy in the processing of filler-gap dependencies in Chinese. In M.K.M. Chan & H. Kang (Eds.) *Proceedings of the 20th North American Conference on Chinese Linguistics (NACCL-20)*. (pp. 943–957). Ohio State University.

Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island constraints in second-language sentence processing. *Studies in Second Language Acquisition*, *33*, 563–588. http://doi.org/10.1017/S0272263111000313

Oxford University Press, University of Cambridge, & Association of Language Testers in Europe. (2001). *Quick placement test: Paper and pen test*. Oxford University Press.

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

Pliatsikas, C. (2010). *Grammatical processing in second language learners of English*. (Unpublished doctoral dissertation). University of Reading.

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

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

Ross, J. R. (1967). *Constraints on variables in syntax*. (Unpublished doctoral dissertation). MIT.

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

Cite this article: Al-Maani, A., Sloggett, S., Grillo, N., & Marsden, H. (2024). Testing for proficiency effects and crosslinguistic influence in L2 processing: filler-gap dependencies in L2 English by Jordanian-Arabic and Mandarin speakers. *Studies in Second Language Acquisition,* 1–17. https://doi.org/10.1017/S027226312400007X

https://doi.org/10.1017/S027226312400007X Published online by Cambridge University Press