Group level and individual differences in second language sentence processing

Group level and individual differences in second language sentence processing

Ian Cunnings
University of Reading

Variability at the group and individual level can inform understanding in second language research. For example, examining by-group and individual differences can play an important role in teasing apart competing theoretical accounts of first and second language processing. In this paper, I review existing research examining variability in second language sentence processing. Focusing on relative clause attachment, filler-gap dependency resolution and subject-verb agreement, I examine these three phenomena as case studies for how examining variability can inform key debates in second language processing research. I review variability at the group and individual level in both cognitive and linguistic abilities, language experience and proficiency, and in the linguistic environment. I also discuss methodological issues in generalizing findings across studies and in using psycholinguistic tasks to examine individual variation in language processing, which pose important challenges that need to be addressed if the field is to move towards an individual differences perspective of second language processing. Although the review focuses on three linguistic phenomena in second language sentence processing, the issues discussed are relevant to the examination of variability in bi-/multilingual language acquisition and processing more broadly.

Keywords: Second language processing, sentence processing, individual differences, bilingualism, multilingualism

1. Introduction

Understanding variability, at the group or individual level, can inform our understanding of the factors that influence language comprehension. There is increasing interest in how individual differences in particular influence language acquisition and processing across the language sciences (Kidd et al., 2018). That studying individual differences can inform understanding of second language

https://doi.org/10.1075/lab.24093.cun | Published online: 14 October 2025
Linguistic Approaches to Bilingualism ISSN 1879-9264 | E-ISSN 1879-9272

processing has however not always been taken for granted. In an earlier review of individual differences in second language sentence processing, Roberts (2012) noted that the majority of research at the time controlled individual variation in an attempt to keep study participant groups homogeneous, rather than to investigate individual differences as an object of study. However, as noted by Roberts, studying individual differences can contribute to our understanding of second language processing.

There are varied theoretical reasons for examining variability in second language processing. For example, examining individual differences can inform key debate about the similarities and differences between first (L1) and second (L2) language processing (Clahsen & Felser, 2006, 2018; Cunnings, 2017; Hopp, 2018; McDonald, 2006). In this review, I focus on how variability can inform our understanding of L2 sentence processing, as an example of how variability can inform research in bi-/multilingualism more broadly.

1.1 Variability in L2 sentence processing

The similarities and differences between L1 and L2 sentence processing have been widely debated. McDonald (2006) claimed L1 and L2 processing are similar, but that L2 processing places additional strain on capacity-limited cognitive resources. Hopp (2018) also argued for continuity between L1 and L2 processing and predicted L2 difficulty is related to acquisition of the L2 lexicon. Alternatively, the Shallow Structure Hypothesis (SSH; Clahsen & Felser, 2006, 2018) predicts L2 speakers should have difficulty utilising syntactic structure during processing, while Cunnings (2017) claimed that L1/L2 differences need not necessarily implicate shallow L2 parsing, but instead may result from how L1 and L2 speakers are susceptible to memory-based interference during sentence processing. Research in L2 sentence processing can also inform understanding of L2 acquisition, where debate has questioned whether late L2 learners can acquire (certain types of) linguistic properties not instantiated in their L1 (Hawkins & Chan, 1997; Schwartz & Sprouse, 1996).

Studying variability, and in particular individual differences, can inform these debates. Examining individual differences constitutes a key test of any claim that predicts certain L2 properties may not be processed in a nativelike way even at high levels of L2 proficiency and can elucidate the extent to which predicted L1/L2 differences represent particular developmental stages as opposed to purported end-states. For example, assessing how proficiency influences the extent to which L2 speakers adopt shallow processing (Clahsen & Felser, 2006, 2018) or are susceptible to memory-based interference (Cunnings, 2017), provides a key test case of whether predicted L1/L2 differences in processing persist into higher levels of

L2 proficiency. Testing precisely how individual variation influences L2 processing also offers ways to tease apart such theories. For example, finding that variation in L2 processing mostly influences processing of L2 syntax would be one way to test the SSH. Assessing L2 speakers of varying proficiency is also key to assessing claims about L1 transfer (Hawkins & Chan, 1997; Schwartz & Sprouse, 1996). Furthermore, examining individual differences is crucial to teasing apart accounts which predict continuity between L1 and L2 processing. For example, assessing the extent to which variability in L2 sentence processing is related to individual differences in lexical access is key to testing the claim that L2 processing difficulty is related to the L2 lexicon (Hopp, 2018).

Examining variability is also important for approaches that call to abandon comparisons between L1 and L2 processing as a primary research question (e.g., Hopp, 2022). This approach examines interlanguage processing without comparison to a monolingual standard and instead focuses on how bi-/multilingualism influences L1/L2/L-n processing by nature of a language user knowing multiple languages. Research in this vein moves questions of transfer away from only asking about how particular groups of L2 speakers may process an L2 based on their L1, to examine the bidirectional nature of cross-linguistic influence. Assessing how individual differences in, for example, exposure to an L1/L2/L-n influences L1/L2/L-n processing, are important in assessing the extent to which cross-linguistic influence affects bi-/multilingual language processing.

Individual differences can thus inform debates across different subfields of L2 processing research. While I do not intend to argue for or against a particular approach to L2 research, one that either does or does not include an L1/L2 comparison, whichever approach is adopted based on the research questions at hand, it is clear that studying variability has an important role to play in advancing our understanding of L2 processing.

1.2 Sources of variability in L2 sentence processing

Variability in L2 processing can be examined at the group or individual level, in part dependent on the nature of the source of variability. Some factors, like L1 background, where different groups of L2 speakers with different L1 backgrounds may be tested to examine the role of L1 transfer in L2 processing, are most obviously operationalised as between-groups differences. Other sources of variability differ at the level of the individual. Though such variables constitute continuous predictors, L2 researchers often turn them into categorical variables for analysis (Plonsky, 2013). For example, although proficiency is usually assessed on a continuum, participants may be categorised into lower and higher proficiency groups for analysis.

L2 proficiency has long been studied, partially given the importance of examining L2 processing at initial- and end-states, where different theories make competing predictions (Hawkins & Chan, 1997; Schwartz & Sprouse, 1996). Proficiency is also important in examining the developmental trajectory of L2 processing more generally. Proficiency is typically taken to index an L2 speaker's general language comprehension and production abilities, as assessed via different methods, including self-report, standardised assessments and cloze tests (see Tremblay, 2011). Vocabulary measures, such as LexTale (Lemhöfer & Broersma, 2012), are also used as proxy measures of proficiency, though the strength of the correlation between vocabulary and more general proficiency measures should not be taken for granted (Puig-Mayenco et al., 2023).

A variety of factors that reflect linguistic experience and the linguistic environment may also influence L2 processing. These include age of L2 acquisition and the amount and type of L2 exposure. L2 exposure can be analysed as either a group-level or individual difference. For example, L2 speakers could be divided into groups of participants with and without immersive L2 exposure, or alternatively immersion could be assessed on a continuum based on how much time an L2 speaker has lived in an immersive environment. These variables are typically assessed via self-report and play an important role in determining L2 proficiency (Tremblay, 2011), and thus presumably may (directly or indirectly) affect L2 processing.

In addition to linguistic variables, another potential source of individual variation in L2 processing is cognitive capacity. In her 2012 review, Roberts noted that research on individual differences in cognitive capacity on L2 processing had focused on working memory and, as evidenced below, this remains mostly the case today. Most research on individual differences in working memory in L2 processing has assumed a capacity-based approach (Daneman & Carpenter, 1980), where individual differences are conceptualised as pertaining to the amount of information an individual can keep in memory at one time. Within this framework, L2 processing is assumed to tax working memory capacity more than L1 processing, leading to increased L2 processing difficulty (for discussion, see Cunnings, 2017). Individual differences in memory capacity are typically measured by complex span tasks, such as reading span or digit span, where participants temporarily hold information in memory whilst completing a secondary task (Mathy et al., 2018). In L2 processing, the L2 reading span task (Harrington & Sawyer, 1992) is typically used. Participants with higher span scores are assumed to have higher working memory capacities which, from this perspective, presumably allows them to free up additional cognitive resources for L2 processing.

Research in the L1 literature has examined how executive function/cognitive control influences sentence processing (Kan et al., 2013; Novick et al., 2014). These studies have examined how cognitive ability may influence sentence processing not in terms of memory capacity, but in terms of how readers manage information conflict by flexibly switching attention. Examining individual differences from this perspective has to date received less attention in the L2 processing literature, but I include relevant studies using this approach in the review below.

Note that for each source of individual variation there are both different ways in which it could be assessed (for proficiency, via self-report, standardised assessments, cloze tests etc.) and how it could be analysed (as a continuous vs. categorical variable). These methodological and analytical choices make it difficult to generalise across studies, especially when conflicting results are reported. I discuss this issue in more detail in the General Discussion.

1.3 This review

Against this background, this review provides an overview of variability in L2 sentence processing. I focus on three phenomena in comprehension, namely relative clause attachment, filler-gap dependencies and subject-verb agreement. These phenomena are chosen as they exemplify key aspects of sentence comprehension that any account of L2 processing must address, namely ambiguity resolution, linguistic dependency resolution and, in the case of subject-verb agreement, sensitivity to violations of morphosyntax. These three phenomena also constitute case studies of how early research led to claims about L1/L2 differences, that have subsequently become more nuanced following examination of individual differences. It is beyond the scope of this review to provide a comprehensive overview of individual differences in all aspects of L2 processing, but it aims to illustrate how individual differences can inform understanding of L2 processing.

2. Relative clause attachment

Cross-linguistic variation in relative clause (RC) attachment has long been examined. In (1a), the RC 'who was reading a letter' could modify either 'the secretary' (NP1 attachment) or 'the professor' (NP2 attachment). Some languages, like English, favour NP2 attachment and others, like Spanish, favour NP1 attachment (Carreiras & Clifton, 1993). RCs thus provide a test case of how general processing preferences and learner specific properties influence L2 processing. Studies examining RC processing have examined either globally ambiguous sentences like (1a), or sentences disambiguated to NP1 or NP2 attachment like (1b/c).

(1a) We saw the secretary of the professor who was reading a letter.

(1b) We saw the secretary of the professors who was reading a letter.

(1c) We saw the secretaries of the professor who was reading a letter.

Early studies suggested L1 but not L2 speakers exhibit online attachment preferences for sentences like (1b/c) (Felser et al., 2003; Papadopoulou & Clahsen, 2003). Of particular note in both these studies was that L2 speakers exhibited this pattern irrespective of whether their L1 had similar or different attachment preferences to the L2, suggesting L1 transfer did not influence L2 processing.

Although these results suggest the L1 does not influence L2 RC processing, some studies have reported that L2 attachment preferences influence L1 processing (Dussias, 2003; Dussias & Sagarra, 2007). Dussias and Sagarra examined attachment preferences in Spanish, which has an NP1 preference. In by-group comparisons, they reported an online NP1 preference, as expected in Spanish, for Spanish L2 English speakers with limited exposure to L2 English. However, a group of Spanish L2 English speakers with extensive naturalistic L2 English exposure exhibited an online NP2 preference, as found in English. Length of L2 English immersion also correlated with a larger online NP2 attachment preference in Spanish, in an individual differences analysis. These results suggest the L2 can affect L1 RC attachment preferences, but the amount of L2 exposure is crucial in explaining this cross-linguistic influence. This complicates interpretation of the lack of L1 transfer effects on L2 attachment resolution and leaves open the possibility that any potential L1 transfer effects on L2 attachment resolution may be modulated by L1 exposure. How differing amounts of L1 exposure may influence L2 attachment preferences has, however, not been examined.

In the L1 literature, research has examined how individual differences in working memory capacity influence RC attachment, with some reporting a stronger NP2 preference in readers with higher reading span scores (Cotter & Ferreira, 2024; Kim & Christianson, 2013; Swets et al., 2007), though conflicting results have been found (Traxler, 2007). How to interpret these results is contested. From a capacity-based view in which high-span readers can hold more information in memory, high-span readers should be more likely to consider NP1 attachment. However, more studies have reported the opposite (Cotter & Ferreira, 2024; Kim & Christianson, 2013; Swets et al., 2007). This has been argued to result from a chunking strategy whereby low-span readers are more likely to 'chunk' NP1 and NP2 into a single unit, with NP1 as the phrasal head (Kim & Christianson, 2013; Swets et al., 2007).

Working memory capacity has also been examined in L2 RC ambiguity resolution. Hopp (2014) tested individual differences in reading span in L1 English

speakers and German L2 English speakers. In an offline task, L2 speakers with lower reading span scores were more likely to consider NP1 attachment. Though the L1 speakers did not show this effect, Hopp noted this might be due to the small L1 sample. Cheng et al. (2021) also reported an offline task with L1 and L2 English speakers and found a stronger NP2 preference as reading spans increased for both groups (i.e., lower span readers were more likely to consider NP1 attachment). These offline results are similar to L1 studies (Swets et al., 2007) and suggest, at least offline, L1 and L2 speakers with lower reading span scores are more likely to consider NP1 attachment, which might suggest a chunking strategy (Kim & Christianson, 2013; Swets et al., 2007).

In contrast to these offline results, inconsistent findings have been reported in online measures of RC ambiguity resolution (Cheng et al., 2021; Hopp, 2014; Kim and Christianson, 2017). In addition to an offline task, Hopp (2014) also conducted an eye-tracking study. For sentences like (2), L1 speakers had longer reading times when the reflexive disambiguated to NP1 rather than NP2 attachment, while the L2 speakers, as a group, did not show any significant differences.

(2) The student had liked the secretary of the professor who had almost killed himself in the office.

Given the lack of group level effects for L2 speakers, Hopp also examined individual differences in the L2 group. Neither individual differences in reading span scores nor L2 proficiency, as measured by the Oxford Placement Test (Allan, 1992), correlated with online L2 processing however. Individual differences in L2 lexical automaticity, as assessed in a lexical decision task, did however interact with L2 reading times. L2 speakers with high levels of lexical automaticity exhibited an NP2 attachment preference online, similar to L1 speakers, while L2 speakers with lower levels of lexical automaticity showed no clear effects.

Cheng et al. (2021) tested how individual differences in reading span, lexical automaticity, and for L2er's proficiency, influence L1 and L2 attachment preferences online. In their eye-tracking experiment, reading times were longer when sentences were disambiguated to NP1 attachment, replicating the usual NP2 preference in English, in both L1 and L2 speakers. The L1 group however showed the effect earlier in the eye-movement record than L2 speakers (in first-pass times), and although both groups showed the effect in total reading times, the effect was bigger for L1 speakers. Like Hopp (2014), individual differences in reading span scores and, for L2 speakers, proficiency (English placement test scores) did not correlate with reading times. Individual differences in lexical automaticity did however influence the NP2 preference in total reading times, where disambiguation to NP1 attachment led to increasingly longer reading times in readers with

lower levels of lexical automaticity.[1] Cheng et al. interpreted this finding as indicating that individuals with higher levels of lexical automaticity were able to overcome reanalysis following the (dispreferred) NP1 disambiguation quicker than individuals with lower levels of lexical automaticity. Note this effect did not interact with group, suggesting individual differences in lexical automaticity influence L1 and L2 sentence processing.

Across two self-paced reading studies, Kim and Christianson (2017) examined how individual differences in reading span influence RC ambiguity resolution in Korean L2 English speakers in both their L1 and L2. High span readers exhibited longer reading times for globally ambiguous relative clauses in both their L1 and L2. Kim and Christianson interpreted this as indicating that higher span readers were more likely to consider both interpretations of the ambiguous relative clause during processing. While these results suggest online effects of reading span scores in L1 and L2 processing, note that Cheng et al. (2021) also tested globally ambiguous sentences but did not find significant effects of individual differences in reading span during processing.

2.1 Relative clause attachment: Summary

In summary, to date there has been no evidence of L1 transfer influencing L2 RC ambiguity resolution. Though earlier studies reported a lack of online attachment preferences for L2 speakers (Felser et al., 2003; Papadopoulou & Clahsen, 2003), other studies reported similar online attachment preferences in L1 and L2 speakers, especially when individual differences are taken into account (Cheng et al., 2021; Hopp, 2014). While Hopp and Cheng et al. reported that individual differences in lexical automaticity influence RC processing, the nature of these effects differed across studies. It is difficult to draw direct comparisons though as the L2 group as a whole in Hopp's study did not show attachment preferences online, while the L2 speakers in Cheng et al.'s study did.

In offline tasks, L1 and L2 attachment resolution is influenced by reading span, with higher span readers showing stronger NP2 attachment preferences (Cheng et al., 2021; Cotter & Ferreira, 2024; Hopp, 2014; Swets et al., 2007). Online studies however have not consistently shown effects of individual differences in reading span on L2 RC attachment during processing (compare Cheng et al., 2021; Hopp, 2014; Kim & Christianson, 2017).

---

1. This was further modulated by whether the RC appeared in subject vs. object position (see Cheng et al., 2021).

3. Filler-gap dependencies

In filler-gap dependencies, two non-adjacent constituents need to be integrated for successful comprehension. Gap-filling is an 'active' process in L1 and L2 comprehension (Cunnings, 2017). Consider (3), adapted from Felser et al. (2012). Here, the dependency is between the filler ('magazine'/'shampoo') and the preposition 'about', but there is an earlier potential gap at 'read'. Although either noun is plausible at 'about', at 'read' 'shampoo' is implausible. If readers fill gaps actively, longer reading times should be observed at 'read' for the implausible filler ('shampoo') compared to the plausible one ('magazine') (Traxler & Pickering, 1996). (4) exemplifies the filled-gap diagnostic (Stowe, 1986). Again, the filler-gap dependency is between the filler ('magazine') and preposition ('about'), but if a dependency is formed actively at 'read', reading times for the following word should be longer when it is an overt direct object ('articles' compared to 'quickly'). L1 and L2 speakers are sensitive to both diagnostics, indicating active gap-filling (for review, see Cunnings, 2017).

(3) Everyone liked the magazine/shampoo that the hairdresser read extensively about before going to the salon.

(4) Everyone liked the magazine that the hairdresser read quickly/articles about before going to the salon.

Variability in L2 gap-filling has also been investigated. Al-Maani et al. (2024) examined filled-gap effects in self-paced reading and reported that the size of the filled-gap effect was influenced by individual differences in proficiency (placement test scores), with higher proficiency L2 speakers showing larger filled-gap effects. A number of studies have examined individual differences in memory capacity. Johnson et al. (2016) tested L1 English speakers and Korean L2 English speakers in self-paced reading with a filled-gap diagnostic and assessed individual differences via a composite measure of memory span. Both groups showed filled-gap effects during reading. The composite span score did not predict individual differences in reading times for L1 speakers, but for L2 speakers higher span scores led to smaller filled-gap effects. Johnson et al. argued that L2 speakers with larger memory spans overcome the filled-gap effect more easily.

Other studies have reported null effects of individual differences in memory capacity. In self-paced reading, Juffs (2005) tested different types of filler-gap dependencies, but neither individual differences in reading span nor word span scores correlated with L2 speakers' reading times. Studies investigating filler-gap dependencies using cross-modal priming paradigms have also reported conflicting results (Berghoff, 2023; Dekydtspotter & Miller, 2013; Felser & Roberts, 2007; Miller, 2014, 2015).

In addition to behavioural studies, individual differences in brain responses during L2 gap-filling have also been examined using event-related potentials (ERPs). Dallas et al. (2013) manipulated the plausibility of filler-gap dependencies in an ERP study testing L1 English speakers and Chinese L2 English speakers. L1 English speakers exhibited N400 effects for implausible compared to plausible filler-gap dependencies. While the L2 group as a whole did not show this effect, the L2 data correlated with individual differences in proficiency, as measured by a composite vocabulary assessment score, with higher proficiency L2 speakers showing the predicted N400. A composite score of working memory span however did not correlate with L2 speakers' N400 responses.

Dong et al. (2023) investigated gap-filling using ERPs in L1 English speakers and Chinese L2 English speakers using the filled-gap diagnostic. While the L1 speakers exhibited a P600 effect, the L2 speakers did not, but they did show a smaller, prefrontal-central positivity. This however, was not influenced by individual differences in proficiency test or reading span scores. Covey et al. (2024) also tested filled-gap effects with Chinese L2 English speakers and reported P600 effects. Furthermore, the size of this P600 effect was modulated by individual differences, with larger effects for L2 speakers with greater attentional control, as assessed using a number Stroop task.

Syntactic restrictions known as island constraints limit when filler-gap dependencies can be formed (Ross, 1967). Consider (5), from Felser et al. (2012). Here, although there is a potential dependency between the filler ('magazine'/'shampoo') and the verb 'read', this is ruled out because the verb appears inside a relative clause. If island constraints restrict gap-filling, the plausibility effect observed at 'read' in (3) should not occur in (5) (Traxler & Pickering, 1996). Both L1 and L2 speakers show plausibility effects in non-island environments like (3) but not sentences like (5) (Felser et al., 2012; Omaki & Schulz, 2011), suggesting dependency formation is guided by island constraints.

(5) Everyone liked the magazine/shampoo that the hairdresser who read extensively bought before going to the salon.

L1 transfer in L2 processing of island constraints has also been examined. Kim et al. (2015) tested sentences like (3/5) in self-paced reading in two L2 groups, with L1 Spanish or L1 Korean. Spanish is a wh-movement language and behaves like English with respect to the relevant island constraint, while Korean is a wh-in-situ language. Although the Spanish L2 English speakers behaved like L1 English speakers with evidence of plausibility effects in non-island but not island environments, the Korean L2 English speakers had longer reading times for implausible than plausible sentences in both non-island and island environments, suggesting temporary violation of the island constraint. Whilst this suggests difficulty

in applying island constraints for L2 speakers from a wh-in-situ L1 background, other studies using a variety of diagnostics have found L2 speakers from different L1 backgrounds, including wh-movement and wh-in-situ L1s, obey island constraints during processing (Aldwayan et al., 2010; Al-Maani et al., 2024; Covey et al., 2024; Johnson et al., 2016).

Individual differences in L2 sensitivity to island constraints have also been examined. Some studies discussed above examining individual differences in filler-gap dependencies also tested dependency formation in island environments (Al-Maani et al., 2024; Covey et al., 2024; Johnson et al., 2016). In each case, the L2 speakers did not create filler-gap dependencies inside islands, irrespective of individual differences in proficiency (Al-Maani et al., 2024), memory span (Johnson et al., 2016) or attentional control (Covey et al., 2024).

The filler-gap dependencies discussed thus far are triggered by overt lexical items. In other cases, linguistic theory posits gaps licensed by structural positions (Chomsky, 1995). Consider (6), from Marinis et al. (2005). In (6a/b), the filler-gap dependency is between the filler ('the nurse') and the verb 'angered'. In (6a), linguistic theory also posits an earlier gap at the spec-CP position at the clause headed by 'that'. No such intermediate gap is posited in (6b). Marinis et al. (2005) reported shorter reading times at the verb in sentences like (6a) compared to (6b), suggesting the intermediate gap facilitated processing, for L1 but not L2 speakers.

(6a) The nurse who the doctor argued that the rude patient had angered is refusing to work late.

(6b) The nurse who the doctor's argument about the rude patient had angered is refusing to work late.

In a by-group comparison, Pliatsikas and Marinis (2013) however reported facilitation for sentences like (6a) compared to (6b) for Greek L2 English speakers with extensive immersion in an English-speaking environment but not Greek L2 English speakers with only classroom English exposure. Berghoff (2022) did not report facilitation for a group of Afrikaans L2 English speakers who also had naturalistic L2 exposure however. Note that the L2 speakers in Berghoff's study were in a multilingual environment (South Africa) compared to the English dominant environment (UK) in Pliatsikas and Marinis' study. This may mean that the participants in Pliatsikas and Marinis had a larger amount of immersive L2 exposure. These findings suggest the importance of L2 exposure when examining L2 sentence processing, but further examination of how different types of linguistic environments influence gap-filling are required.

3.3 Filler-gap dependencies: Summary

Existing research using varied methods indicates that L2 speakers posit filler-gap dependencies actively and utilise island constraints to restrict when dependencies are formed. While some results suggest L1 influence on L2 gap-filling (Kim et al., 2015), research as a whole suggest L2 speakers process filler-gap dependencies similarly to L1 speakers. Whether L2 speakers posit certain syntactically licensed gaps has been debated (Berghoff, 2023; Marinis et al., 2005; Pliatsikas & Marinis, 2013), but the results of Pliatsikas and Marinis (2013) suggest this is possible given sufficient, naturalistic exposure.

Individual differences in proficiency may influence filler-gap dependency resolution (Al-Maani et al., 2024; Dallas et al., 2013), though this has not been found in all studies. Effects of memory span on L2 filler-gap dependency resolution have been similarly inconsistent, though the results to date suggest that individual differences in memory span and attentional control do not influence the violability of island constraints in either L1 or L2 processing (Covey et al., 2024; Johnson et al., 2016).

4. Subject-verb agreement

Subject-verb agreement has received attention in L2 processing for various reasons. L2 sensitivity to agreement features either instantiated or missing in a learner's L1 provides a test-case of competing accounts of L1 transfer (Hawkins & Chan, 1997; Schwartz & Sprouse, 1996). This is often tested in violation paradigms. For example, longer reading times at 'were' in ungrammatical sentences like (7b) compared to grammatical sentences like (7a) can be taken as evidence of online sensitivity to agreement. Studies have also tested sentences like (7c), where a so-called 'distractor' intervenes between the verb and its grammatical controller. This distractor causes interference, such that the processing difficulty associated with reading ungrammatical sentences is ameliorated when the distractor matches the properties of the verb (Wagers et al., 2009). Although the precise characterisation of this effect is contested, it is typically explained in terms of how information is encoded and/or retrieved from memory during processing (for discussion, see Yadav et al., 2023).

(7a) The boys quite unsurprisingly were late for class.

(7b) The boy quite unsurprisingly were late for class.

(7c) The boy near the girl(s) quite unsurprisingly were late for class.

Conflicting results have been reported in behavioural studies with regard to sensitivity to agreement violations in L2 speakers whose L1 lacks agreement (Cheng, Rothman, et al., 2022; Jiang 2004). Lim & Christianson (2015) investigated individual differences in sensitivity to number agreement in Korean L2 English speakers, whose L1 lacks subject-verb agreement. They tested sentences such as "The teacher(s) who instructed the student(s) were very strict" and found that L1 speakers showed longer reading times for ungrammatical than grammatical sentences at the critical verb and spillover region ('were' and 'very'), while L2 speakers exhibited this effect during spillover only. L2 reading times were also influenced by individual differences in proficiency, as assessed via a cloze test, such that higher proficiency L2 speakers showed larger grammaticality effects than lower proficiency L2 speakers. Both L1 and L2 speakers reading times were also influenced by the distractor, with faster reading times for ungrammatical sentences if the distractor matched the number of the verb.

Proficiency effects have also been reported for L2 speakers whose L1 has agreement. In a self-paced study in Spanish, Jegerski (2016) reported grammaticality effects for both L1 and L2 Spanish speakers, though the grammaticality effect was larger for near-native as compared to advanced L2 speakers, when grouped based on proficiency test scores. Alaskar & Cunnings (2025) reported grammaticality effects in a self-paced reading study for both L1 English speakers and Arabic L2 English speakers. This effect was however larger for the L1 group. Similar to Lim and Christianson (2015), individual differences in proficiency (placement test scores) influenced processing, such that higher proficiency L2 speakers only showed the predicted longer reading times for ungrammatical sentences.

Individual differences in L2 memory capacity have also been examined. Rattanasak et al. (2024) reported grammaticality effects in both L1 and L2 speakers, with larger grammaticality effects for L2 speakers with higher reading span scores. Foote (2011) also tested whether reading span scores influence subject-verb agreement in English L2 Spanish speakers. While the L2 speakers were sensitive to agreement violations, the size of this effect was not significantly influenced by individual differences in reading span. Note that Rattansak et al. tested L2 speakers whose L1 (Thai) lacks agreement, while Foote tested L2 speakers whose L1 also has subject-verb number agreement. Although this might suggest memory span interacts with L1 background, further research is needed to systematically assess this possibility.

Conflicting findings have also been reported in ERP studies, especially for L2 English speakers whose L1 lacks agreement (Armstrong et al., 2018; Bian et al., 2021; Chen et al., 2007; Cheng, Cunnings, et al., 2022; Ojima et al., 2005). Bian et al. (2021) and Cheng, Cunnings et al. (2022) reported P600 effects to agreement violations that were smaller for L2 speakers than L1 speakers. L2 P600

effects were however not significantly influenced by either individual differences in L2 proficiency (via a composite proficiency score in Armstrong et al., 2018, and a placement test in Cheng, Cunnings, et al., 2022) or memory span (Armstrong et al., 2018, via a composite span score). Note two of the three studies that reported L2 P600 effects (Armstrong et al., 2018; Cheng, Cunnings, et al., 2022) tested L2 speakers in an immersion setting, while two studies that did not report L2 P600s (Chen et al., 2007; Ojima et al., 2005) tested non-immersed L2 speakers. Whilst this may suggest a role for immersion, future direct comparisons are required here.

P600 effects, though smaller for L2 than L1 speakers, have also been reported for Spanish L2 English speakers, whose L1 has agreement (Tanner et al. 2012). Tanner et al. (2014) examined individual differences in Spanish L2 English speakers. At the group level, L2 speakers showed a biphasic pattern, with violations eliciting an N400 followed by P600. However, analysis of individual brain responses indicated some L2 speakers elicited N400s and others P600. Individual differences in L2 proficiency test scores correlated with overall response magnitude, but not with whether L2 speakers showed N400 or P600 effects. Individual differences in age of arrival in an English-speaking environment, and motivation to speak, both correlated with P600 effects, with larger effects for L2 speakers with earlier immersion and a higher motivation to speak. While the pattern of results indicating N400s for some participants and P600 for others may indicate different processing strategies across different L2 speakers, note that biphasic patterns have also been reported in L1 studies (Tanner & Van Hell, 2014), suggesting this is not a specific property of L2 processing.

4.1 Subject verb agreement: Summary

While earlier studies suggested insensitivity to subject-verb agreement violations during L2 processing (Chen et al., 2007; Jiang, 2004; Ojima et al., 2005), more recent results indicate L2 sensitivity to such violations (Alaskar & Cunnings, 2025; Armstrong et al., 2018; Cheng, Cunnings, et al., 2022; Cheng, Rothman, et al., 2022; Lim & Christianson, 2015). Though not found in all studies, sensitivity to agreement violations may be influenced by L2 proficiency (Alaskar & Cunnings, 2025; Lim & Christianson, 2015), and other factors related to language experience, such as age of arrival in an immersion environment (Tanner et al., 2014).

There have been inconsistent results with regards to how memory capacity influences L2 sensitivity to subject-verb agreement violations (Armstrong et al., 2018; Foote, 2011; Rattanasak et al., 2024). Distractors interfere in subject-verb agreement in both L1 and L2 speakers (Lim & Christianson, 2015; Tanner et al., 2012), suggesting memory-based interference influences L1 and L2 processing (for discussion, see Alaskar & Cunnings, 2025).

5. General discussion

Across the studies reviewed here, the results have illustrated how individual differences influence L2 processing of different phenomena, though it should be noted that study findings have not always been consistent in terms of precisely *which factors* influence L2 processing. Below I discuss the theoretical implications of these findings, before discussing some methodological issues that limit the conclusions that can currently be drawn about individual differences in L2 processing. Though I focus on the three phenomena reviewed above, I also touch upon results from other phenomena from the wider literature on individual differences in L2 sentence processing where appropriate.

5.1 Theoretical implications of variability in L2 sentence processing

A common finding across this review has been that whilst earlier studies reported L1/L2 differences, subsequent research suggested these may be ameliorated. In relative clause attachment, whilst early studies suggested L2 speakers do not show structure-based attachment preferences during processing (Felser et al., 2003; Papadopoulou & Clahsen, 2003), subsequent research indicated this is possible, at least for L2 speakers with sufficiently automatised lexical access (Hopp, 2014). In filler-gap dependencies, though the results of Marinis et al. (2005) suggested L2 speakers do not posit structurally-mediated gaps, the results of Pliatsikas and Marinis (2013) indicated this is possible, given sufficient L2 immersion. For subject-verb agreement, though earlier studies suggested L2 insensitivity to agreement violations (Chen et al., 2007; Jiang, 2004; Ojima et al., 2005), subsequent work suggested otherwise (Armstrong et al., 2018; Cheng, Cunnings, et al., 2022), with sensitivity being influenced by proficiency (Alaskar & Cunnings, 2025; Jegerski, 2016; Lim & Christianson, 2015). Research in other types of agreement, such as noun-adjective agreement, has reported similar findings (Alemán Bañón et al., 2018; Gabriele et al., 2013, 2021; Keating, 2009; Sagarra & Herschensohn, 2010).

Results such as these are compatible with accounts which predict continuity between L1 and L2 processing (e.g., Hopp, 2022). Though these results do not necessarily rule out the possibility that L2 speakers utilise shallow processing (Clahsen & Felser, 2006, 2018), or are more susceptible to memory-based interference (Cunnings, 2017), at certain stages of development, these findings do not support strong versions of such accounts which would predict L1/L2 differences even at high levels of L2 proficiency or after many years of naturalistic L2 exposure. Of course, from this review of three psycholinguistic phenomena, it is not possible to conclude that L1/L2 differences do not persist in other phenomena,

but the results surveyed here do suggest caution in drawing strong conclusions about L1/L2 differences until study findings are replicated across different L2 populations in different linguistic environments.

L1 transfer has played an important role in theory development in L2 research and its influence has been widely debated (Clahsen & Felser, 2006; Hawkins & Chan, 1997; Hopp, 2022; Schwartz & Sprouse, 1996). In studies of relative clause attachment, there was no clear evidence of L1 influence on L2 processing (Felser et al., 2003; Papadopoulou & Clahsen, 2003). For filler-gap dependencies, though effects consistent with L1 transfer have been reported (Kim et al., 2015), the results largely suggest similar gap-filling processes across L2 speakers from different L1 backgrounds (Aldwayan et al., 2010; Felser et al., 2012; Omaki & Schulz, 2011; Williams et al., 2001). For subject-verb agreement, though some early results suggested L2 insensitivity to agreement violations, which might be taken as evidence of indicating difficulty with L2 morphosyntactic features especially when they are not instantiated in the L1 (Chen et al., 2007; Jiang, 2004; Ojima et al., 2005), subsequent work has indicated sensitivity to agreement violations, even in L2 speakers whose L1 lacks agreement (Cheng, Cunnings, et al., 2022; Lim & Christianson, 2015). These results do not support strong claims about L2 speakers not being able to fully acquire L2 features that are not instantiated in the L1 (Hawkins & Chan, 1997; see also studies on L2 processing of noun-adjective agreement for similar conclusions, e.g., Alemán Bañón et al., 2018; Gabriele et al., 2013, 2021). While the lack of transfer effects in some studies might seem inconsistent with claims that transfer influences L2 acquisition (Schwartz & Sprouse, 1996), note that not all studies tested for transfer effects at stages of acquisition where its influence is predicted to be largest (i.e., the initial stages of L2 acquisition).

Some of the reviewed literature on filler-gap dependency resolution in particular highlights how examining individual differences in L2 processing can inform wider debate beyond the L1/L2 processing literature. Recall that some studies on filler-gap dependencies reported that individual differences in memory span (Johnson et al., 2016) and attentional control (Covey et al., 2024) influence L2 filler-gap dependency resolution. Importantly, although these studies found evidence of individual differences influencing L2 gap formation, neither study found that these individual differences influenced whether or not L2 speakers violated island constraints that restrict the formation of such dependencies. These findings can inform theoretical debate in the linguistics literature regarding the characterisation of island constraints, where some argue that island constraints are syntactic in nature, while others have argued for a resource-based account in which island effects result from processing difficulty (compare Hofmeister & Sag, 2010; Sprouse et al., 2012). The fact that individual differences in memory span and attentional control did not influence island sensitivity, even in L2 speakers

where processing burden should be highest, was taken by these authors as indicating support for syntactic rather than processing-based accounts of islands. These studies thus provide a novel example of how L2 research can inform key theoretical issues in the wider linguistics literature.

5.2 Which sources of variability influence L2 sentence processing?

I now turn to discussion of which sources of variability influence L2 sentence processing. Note that in discussing which sources of variability influence L2 sentence processing here, I do not intend to imply that such effects assume an L1/L2 comparison. While individual differences in L2 processing can inform this debate, they can also contribute to L2 research that abandons the L1/L2 comparison as the primary research question (see Hopp, 2022).

Proficiency effects were found in online measures in at least some studies on filler-gap dependencies (Al-Maani et al., 2024; Dallas et al., 2013) and subject-verb agreement (Alaskar & Cunnings, 2025; Jegerski, 2016; Lim & Christianson, 2015; Tanner et al., 2014). It is thus likely uncontroversial to say that proficiency influences L2 processing. Nevertheless, proficiency was not found to influence RC attachment during online processing (Cheng et al., 2021; Hopp, 2014), and null effects of proficiency were also reported in some studies of filler-gap dependencies and agreement (Armstrong et al., 2018; Cheng, Cunnings, et al., 2022; Dong et al., 2023). Drawing conclusions is difficult here however, given inconsistent findings could be a result of the varied range in proficiencies tested across studies, or the different proficiency measures used. It could however be that proficiency influences certain linguistic phenomena more than others. For example, proficiency may have a greater influence on processing of obligatory arguments, as in filler-gap dependencies and subject-verb agreement, rather than adjuncts like relative clauses, where NP1 and NP2 attachment preferences are biases rather than being deterministic. The extent to which the processing of certain linguistic phenomena is particularly affected by L2 proficiency is an avenue for future research.

Linguistic exposure has also been reported as influencing L2 processing, though again inconsistent results have been found. How L2 exposure may influence relative clauses during L2 processing was not examined in the literature reviewed here, though effects of L2 exposure on L1 relative clause attachment have been reported (Dussias & Sagarra, 2007). That relative clause attachment is influenced by exposure is perhaps a result of the fact that attachment biases reflect preferences. However, from this perspective, it is still puzzling why lifelong exposure to an L1 does not result in L1 transfer of attachment preferences to the L2. How relative amounts of L1 and L2 exposure may influence L2 attachment preferences is an avenue for future research.

Evidence of immersion playing a role in filler-gap dependency resolution comes from Pliatsikas and Marinis (2013), who reported effects consistent with L2 speakers positing structurally-mediated gaps in immersed but not non-immersed environments. In subject-verb agreement, Tanner et al. (2014) reported larger P600 effects to grammatical violations for L2 speakers with earlier naturalistic exposure to the L2. Though these studies suggest a role of immersion, further research is required here to draw strong conclusions about how the linguistic environment may influence different aspects of L2 processing.

Turning to working memory, the most consistent evidence comes from offline tasks on relative clause attachment, where a higher rate of NP1 attachment was found for lower capacity readers (Cheng et al., 2021; Cotter & Ferreira, 2024; Hopp, 2014; Swets et al., 2007). Inconsistent results were reported in terms of memory capacity influencing online L2 processing of relative clauses (Cheng et al., 2021; Hopp, 2014; Kim & Christianson, 2017). Note that working-memory-based effects in offline and online tasks have had different interpretations. In offline tasks, a stronger NP1 preference in lower span speakers is argued to result from a 'chunking' strategy in which NP1 and NP2 form a single unit, with NP1 as the head (Kim & Christianson, 2013; Swets et al., 2007). For online processing, it has been claimed that only higher span readers are able to maintain the two interpretations of a globally ambiguous relative clause in memory at one time (Kim & Christianson, 2017). It might be that these two effects, a chunking strategy and consideration of different interpretations, conflict with each other. Future research is required to tease these issues apart.

There is also conflicting evidence with regards to individual differences in working memory for filler-gap dependencies (Dallas et al., 2013; Dong et al., 2023; Johnson et al., 2016; Juffs, 2005) and subject-verb agreement (Armstrong et al., 2018; Foote, 2011; Rattanasak et al., 2024). Conflicting results have also been found in studies on various other phenomena (for adjective-noun agreement, compare Foote, 2011; Keating, 2010; Sagarra & Herschensohn, 2010; for subject-object ambiguities, compare Brothers et al., 2021; Hopp, 2015; Juffs, 2004; for subject/object filler-gap dependencies, see Havik et al., 2009; Hopp et al., 2024).

A number of factors contribute to these inconsistent findings. One issue relates to the conceptualisation of how working memory influences sentence processing. In the L1 literature, some approaches have eschewed the notion of memory capacity, in terms of the amount of information held in memory at one time, influencing comprehension, and instead focus on how memory interference, described in terms of the content and quality of representations in memory, influences processing (Van Dyke & Johns, 2012). Although it is beyond the scope of this review to go into detail, these different conceptualisations make different predictions about which types of sentences, and when during processing, work-

ing memory demands should influence processing. Future systematic investigation of the roles of memory capacity and memory interference in L2 processing are required to tease these issues apart (for discussion, see Cunnings 2017, 2022a, 2022b).

Another important consideration in this regard is whether individual differences in cognitive ability are best considered in terms of memory capacity and/or interference, or in terms of executive function/cognitive control. To date, only a handful of studies have examined how executive function influences L2 processing (e.g.,Brothers et al., 2021; Covey et al., 2024; Hopp et al., 2024), and future research is required here to tease these issues apart.

Importantly, systematic examination of any claims about how presumably domain general cognitive abilities, such as memory capacity, memory interference and/or executive function, influence sentence processing requires assessment of these different potential sources of variation in different populations of speakers. Testing these issues in L2 processing provides a way of assessing these claims that is not possible by examining L1 processing alone. For example, given the increased variation in proficiency seen in L2 as compared to L1 adult speakers, examining how individual variation in, for example, executive function might interact with individual variation in proficiency, provides a way of assessing claims about how executive function influences language processing in a way that is not possible by testing typical adult L1 speakers alone. Further research is required to elucidate how individual differences in bi-/multilingual sentence processing can inform wider theoretical debate on the nature of how memory capacity, memory interference and/or executive function influence sentence processing.

5.3 Methodological considerations

Methodological issues also make it difficult to draw conclusions across studies. Considering memory capacity for example, although span tasks have been used in various studies, the precise span tasks used, their scoring methods and, especially for reading span, whether the task was conducted in the L1 and/or L2, varied across the studies reviewed here. Variation in task administration and scoring contribute to difficulty in drawing conclusions across studies (see Juffs & Harrington, 2011; Leeser & Sunderman, 2016). In tandem with careful theoretical consideration of what precisely a given working memory task is believed to assess (memory capacity, cognitive control etc.), standardisation of the administration and scoring of working memory tasks is important in addressing this issue. Note that this issue also applies to other measures, such as proficiency tests and language background questionnaires, which have also varied across the studies discussed here. It is also

important not to ignore that the administration and analysis of psycholinguistic tasks is not entirely standardised, which also contributes to difficulty in making cross-study comparisons.

A final methodological challenge, specific to examining variability at the individual level, relates to how well cognitive tasks measure individual differences (Hedge et al., 2018; Parsons et al., 2019). Note that a "task" here is intended to be construed broadly, to include both individual differences measures (e.g., reading span tasks) and psycholinguistic experiments used to examine sentence processing. For a task to *reliably* measure individual differences, it must be able to meaningfully capture sufficient individual variation to systematically rank individuals along a continuum. A typical way to assess whether a task reliably measures individual differences is to calculate test-retest reliability. That is, if a task consistently measures individual differences in a particular construct, we would expect participant performance to be similar across multiple instances of completing the same task, leading to a high test-retest correlation. When participants only complete a task once, as is typical in psycholinguistic experiments, another way to assess this reliability of individual differences is by calculating split-half reliability (see e.g., Parsons et al., 2019). That is, participant responses in a task are split and the two halves are correlated. A high correlation would suggest a task that is consistently measuring individual differences. A correlation of .7 or above is typically considered desirable and, although not a strict cutoff, low reliability is detrimental to statistical inference (Parsons et al., 2019). To date, whether psycholinguistic tasks reliably measure individual differences has received little attention (Cunnings & Fujita, 2020; James et al., 2018).

Consider Cunnings and Fujita (2020), who assessed reliability of self-paced reading times as a measure of individual differences in ambiguity resolution in L1 and L2 speakers. They compared reading times of temporarily ambiguous garden path sentences ("Ken washed the dog and the cat in the garden played with a ball") to unambiguous controls ("Ken washed the dog while the cat in the garden played with a ball"). While split-half reliabilities for individual differences in *overall* reading speed, averaged across conditions, were reliable ($r > .9$ for both groups), split-half reliabilities for the garden-path effect (the difference between experimental conditions) were low ($< .2$). These results highlight how it cannot be taken for granted that psycholinguistic tasks constitute reliable measures of individual differences in L2 processing (see also Hui & Wu, 2024).

Low measurement reliability can be addressed in part by running more powerful studies (see Parsons et al., 2019). However, it is difficult to draw robust inferences about individual differences when trial-level noise is larger than individual variation (Rouder et al., 2023). An important first step in addressing this issue would be for researchers to report the reliability of their experimental tasks.

Across the literature reviewed here however, this was rarely if ever reported. It might also be that new tasks need to be developed that can systematically measure individual differences in L2 sentence processing. Whatever the answer, addressing this methodological issue is crucial for the field to advance an individual differences perspective on second language processing.

6. Conclusions

In her 2012 review, Roberts predicted that future research would benefit from an individual differences approach to L2 sentence processing. It is clear from the research reviewed here that the field has embraced this approach. I have argued that current findings indicate the importance of proficiency, and the amount and type of linguistic exposure, but future research is required to examine how these and other linguistic factors influence L2 processing. The role that individual differences in cognitive capacity may play in explaining individual differences I have argued is currently less certain, but this is partly due to different perspectives on how individual differences in memory capacity, memory interference and/or cognitive control are best characterised.

At the same time, there are important methodological challenges that need to be addressed. These issues go beyond the phenomena studied here, extend beyond language comprehension to production, and are equally important when examining individual differences in adult and child populations that go beyond the L2 groups reviewed here. Through addressing these challenges, future research should help refine our understanding of how individual differences influence not only L2 processing, but language acquisition and processing in bi-/multilingual populations more broadly.

Funding

Open Access publication of this article was funded through a Transformative Agreement with University of Reading.

Acknowledgements

I would like to thank the editor, Leah Roberts and two anonymous reviewers for their comments on previous drafts of this paper.

References

Alaskar, S., & Cunnings, I. (2025). Agreement and reflexives in non-native sentence processing. *Bilingualism: Language and Cognition*, *28*, 897–910.

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.), *Language Acquisition and Language Disorders* (Vol. *53*, pp. 65–86). John Benjamins Publishing Company.

Alemán Bañón, J., Fiorentino, R., & Gabriele, A. (2018). Using event-related potentials to track morphosyntactic development in second language learners: The processing of number and gender agreement in Spanish. *PLOS ONE*, *13*(7), e0200791.

Allan, D. (Ed.). (1992). *Oxford Placement Test. Test 1: Test pack: grammar test; listening test* (7th impr). Oxford University Press.

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*, *46*(2), 564–580.

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

Berghoff, R. (2022). L2 processing of filler-gap dependencies: Attenuated effects of naturalistic L2 exposure in a multilingual setting. *Second Language Research*, *38*(2), 373–393.

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*(2), 586–598.

Bian, J., Zhang, H., & Sun, C. (2021). An ERP study on attraction effects in advanced L2 learners. *Frontiers in Psychology*, *12*, 616804.

Brothers, T., Hoversten, L. J., & Traxler, M. J. (2021). Bilinguals on the garden-path: Individual differences in syntactic ambiguity resolution. *Bilingualism: Language and Cognition*, *24*(4), 612–627.

Carreiras, M., & Clifton, C. (1993). Relative clause interpretation preferences in Spanish and English. *Language and Speech*, *36*(4), 353–372.

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

Cheng, Y., Cunnings, I., Miller, D., & Rothman, J. (2022). Double number marking matters for both L1 and L2 processing of nonlocal agreement similarly: An ERP investigation. *Studies in Second Language Acquisition*, *44*(5), 1309–1329.

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

Cheng, Y., Rothman, J., & Cunnings, I. (2022). Determiner-number specification and non-local agreement computation in L1 and L2 processing. *Journal of Psycholinguistic Research*, *51*(4), 847–863.

Chomsky, N. (1995). *The Minimalist Program* (4th print). MIT Press.

Clahsen, H., & Felser, C. (2006). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, *27*(1), 107–126.

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

Cotter, B.T., & Ferreira, F. (2024). The relationship between working memory capacity, bilingualism, and ambiguous relative clause attachment. *Memory & Cognition*, *52*(7), 1530–1547.

Covey, L., Fiorentino, R., & Gabriele, A. (2024). Island sensitivity in L2 learners: Evidence from acceptability judgments and event-related potentials. *Second Language Research*, *40*(1), 19–50.

Cunnings, I. (2017). Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition*, *20*(4), 659–678.

Cunnings, I. (2022a). Interference-based and capacity-based approaches to working memory in second language sentence processing. In T. Leal, E. Shimanskaya, & C.A. Isabelli (Eds.), *Language Acquisition and Language Disorders* (Vol. *67*, pp. 243–264). John Benjamins Publishing Company.

Cunnings, I. (2022b). Working memory and L2 sentence processing. In J.W. Schwieter & Z. (Edward) Wen (Eds.), *The Cambridge Handbook of Working Memory and Language* (1st ed., pp. 593–612). Cambridge University Press.

Cunnings, I., & Fujita, H. (2020). Quantifying individual differences in native and nonnative sentence processing. *Applied Psycholinguistics*, 1–21.

Dallas, A., DeDe, G., & Nicol, J. (2013). An event-related potential (ERP) investigation of filler-gap processing in native and second language speakers: ERP study of L2 filler-gap processing. *Language Learning*, *63*(4), 766–799.

Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466.

Dekydtspotter, L., & Miller, A.K. (2013). Inhibitive and facilitative priming induced by traces in the processing of *wh*-dependencies in a second language. *Second Language Research*, *29*(3), 345–372.

Dong, Z.R., Han, C., Hestvik, A., & Hermon, G. (2023). L2 processing of filled gaps: Non-native brain activity not modulated by proficiency and working memory. *Linguistic Approaches to Bilingualism*, *13*(6), 767–800.

Dussias, P.E. (2003). Syntactic ambiguity resolution in L2 learners: Some Effects of Bilinguality on L1 and L2 Processing Strategies. *Studies in Second Language Acquisition*, *25*(4), 529–557.

Dussias, P.E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish–English bilinguals. *Bilingualism: Language and Cognition*, *10*(01), 101.

Felser, C., Cunnings, I., Batterham, C., & Clahsen, H. (2012). The timing of island effects in nonnative sentence processing. *Studies in Second Language Acquisition*, *34*(1), 67–98.

Felser, C., & Roberts, L. (2007). Processing wh-dependencies in a second language: A cross-modal priming study. *Second Language Research*, *23*(1), 9–36.

Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, *24*(3), 453–489.

Foote, R. (2011). Integrated knowledge of agreement in early and late English–Spanish bilinguals. *Applied Psycholinguistics*, *32*(1), 187–220.

Gabriele, A., Alemán Bañón, J., Hoffman, L., Covey, L., Rossomondo, A., & Fiorentino, R. (2021). Examining variability in the processing of agreement in novice learners: Evidence from event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(7), 1106–1140.

Gabriele, A., Fiorentino, R., & Bañón, J.A. (2013). Examining second language development using event-related potentials: A cross-sectional study on the processing of gender and number agreement. *Linguistic Approaches to Bilingualism*, *3*(2), 213–232.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition*, *14*(1), 25–38.

Havik, E., Roberts, L., van Hout, R., Schreuder, R., & Haverkort, M. (2009). Processing subject-object ambiguities in the L2: A self-paced reading study with German L2 Learners of Dutch. *Language Learning*, *59*(1), 73–112.

Hawkins, R., & Chan, C.Y. (1997). The partial availability of Universal Grammar in second language acquisition: The 'failed functional features hypothesis'. *Second Language Research*, *13*(3), 187–226.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186.

Hofmeister, P., & Sag, I.A. (2010). Cognitive constraints and island effects. *Language*, *86*(2), 366–415.

Hopp, H. (2014). Working memory effects in the L2 processing of ambiguous relative clauses. *Language Acquisition*, *21*(3), 250–278.

Hopp, H. (2015). Individual differences in the second language processing of object–subject ambiguities. *Applied Psycholinguistics*, *36*(2), 129–173.

Hopp, H. (2018). The bilingual mental lexicon in L2 sentence processing. *Second Language*, *17*, 5–27.

Hopp, H. (2022). Second language sentence processing. *Annual Review of Linguistics*, *8*(1), 235–256.

Hopp, H., Schimke, S., Gastmann, F., Öwerdieck, D., & Poarch, G.J. (2024). Processing to learn noncanonical word orders: Exploring linguistic and cognitive predictors of reanalysis in early L2 sentence comprehension. *Studies in Second Language Acquisition*, *46*(3), 686–709.

Hui, B., & Wu, Z. (2024). Estimating reliability for response-time difference measures: Toward a standardized, model-based approach. *Studies in Second Language Acquisition*, *46*(1), 227–250.

James, A.N., Fraundorf, S.H., Lee, E.-K., & Watson, D.G. (2018). Individual differences in syntactic processing: Is there evidence for reader-text interactions? *Journal of Memory and Language*, *102*, 155–181.

Jegerski, J. (2016). Number attraction effects in near-native Spanish sentence comprehension. *Studies in Second Language Acquisition*, *38*(1), 5–33.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, *25*(4), 603–634.

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*.

Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society*, *102*(2), 199–225.

Juffs, A. (2005). The influence of first language on the processing of wh-movement in English as a second language. *Second Language Research*, *21*(2), 121–151.

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, *44*(2), 137–166.

Kan, I. P., Teubner-Rhodes, S., Drummey, A. B., Nutile, L., Krupa, L., & Novick, J. M. (2013). To adapt or not to adapt: The question of domain-general cognitive control. *Cognition*, *129*(3), 637–651.

Keating, G. D. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning*, *59*(3), 503–535.

Keating, G. D. (2010). The effects of linear distance and working memory on the processing of gender agreement in Spanish. In B. VanPatten & J. Jegerski (Eds.), *Language Acquisition and Language Disorders* (Vol. *53*, pp. 113–134). John Benjamins Publishing Company.

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, *22*(2), 154–169.

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

Kim, J. H., & Christianson, K. (2013). Sentence complexity and working memory effects in ambiguity resolution. *Journal of Psycholinguistic Research*, *42*(5), 393–411.

Kim, J. H., & Christianson, K. (2017). Working memory effects on L1 and L2 processing of ambiguous relative clauses by Korean L2 learners of English. *Second Language Research*, *33*(3), 365–388.

Leeser, M. J., & Sunderman, G. L. (2016). Methodological implications of working memory tasks for L2 processing research. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Bilingual Processing and Acquisition* (Vol. *3*, pp. 89–104). John Benjamins Publishing Company.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343.

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

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, *27*(01).

Mathy, F., Chekaf, M., & Cowan, N. (2018). Simple and complex working memory tasks allow similar benefits of information compression. *Journal of Cognition*, *1*(1), 31.

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

Miller, A. K. (2014). Accessing and maintaining referents in L2 processing of *wh* -dependencies. *Linguistic Approaches to Bilingualism*, *4*(2), 167–191.

Miller, A. K. (2015). Intermediate traces and intermediate learners: Evidence for the use of intermediate structure during sentence processing in second language french. *Studies in Second Language Acquisition*, *37*(3), 487–516.

Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., & Bunting, M. F. (2014). Clearing the garden-path: Improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience*, *29*(2), 186–217.

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience*, *17*(8), 1212–1228.

Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island constraints in second-language sentence processing. *Studies in Second Language Acquisition*, *33*(4), 563–588.

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition*, *25*(4), 501–528.

Parsons, S., Kruijt, A.-W., & Fox, E. (2019). Psychological science needs a standard practice of reporting the reliability of cognitive-behavioral measurements. *Advances in Methods and Practices in Psychological Science*, *2*(4), 378–395.

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

Plonsky, L. (2013). Study quality in SLA: An assessment of designs, analyses, and reporting practices in quantitative L2 Research. *Studies in Second Language Acquisition*, *35*(4), 655–687.

Puig-Mayenco, E., Chaouch-Orozco, A., Liu, H., & Martín-Villena, F. (2023). The LexTALE as a measure of L2 global proficiency: A cautionary tale based on a partial replication of Lemhöfer and Broersma (2012). *Linguistic Approaches to Bilingualism*, *13*(3), 299–314.

Rattanasak, S., Pongpairoj, N., & Christianson, K. (2024). Effects of working memory capacity and distance-based complexity on agreement processing: A crosslinguistic competition account. *Applied Linguistics Review*, *15*(3), 1151–1176.

Roberts, L. (2012). Individual Differences in Second Language Sentence Processing. *Language Learning*, *62*(s2), 172–188.

Ross, J. (1967). *Constraints on variables in syntax*. MIT.

Rouder, J. N., Kumar, A., & Haaf, J. M. (2023). Why many studies of individual differences with inhibition tasks may not localize correlations. *Psychonomic Bulletin & Review*, *30*(6), 2049–2066.

Sagarra, N., & Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua*, *120*(8), 2022–2039.

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, *12*(1), 40–72.

Sprouse, J., Wagers, M., & Phillips, C. (2012). A test of the relation between working-memory capacity and syntactic island effects. *Language*, *88*(1), 82–123.

Stowe, L. A. (1986). Parsing wh-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, *1*(3), 227–245.

Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General*, *136*(1), 64–81.

Tanner, D., Herschensohn, J., & Osterhout, L. (2012). Electrophysiological markers of interference and structural facilitation in native and nonnative agreement processing. In A. Biller, A. Chung, & A. Kimball (Eds.), *Proceedings of the 36th Boston University Conference on Language Development* (pp. 294–606). Cascadilla.

Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition*, *17*(2), 277–293.

Tanner, D., & Van Hell, J.G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia*, *56*, 289–301.

Traxler, M.J. (2007). Working memory contributions to relative clause attachment processing: A hierarchical linear modeling analysis. *Memory & Cognition*, *35*(5), 1107–1121.

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

Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research: "Clozing" the gap. *Studies in Second Language Acquisition*, *33*(3), 339–372.

Van Dyke, J.A., & Johns, C.L. (2012). Memory interference as a determinant of language comprehension: Interference in comprehension. *Language and Linguistics Compass*, *6*(4), 193–211.

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

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

Yadav, H., Smith, G., Reich, S., & Vasishth, S. (2023). Number feature distortion modulates cue-based retrieval in reading. *Journal of Memory and Language*, *129*, 104400.

Address for correspondence

Ian Cunnings
School of Psychology and Clinical Language Sciences
University of Reading
Reading, RG6 7BE
UK

i.cunnings@reading.ac.uk
https://orcid.org/0000-0002-5318-0186

Publication history

Date received: 3 December 2024
Date accepted: 2 October 2025
Published online: 14 October 2025