Working memory and L2 sentence processing

Book or Report Section

Accepted Version

Chapter 27

Cunnings, Ian ORCID logoORCID: https://orcid.org/0000-0002-5318-0186 (2022) Working memory and L2 sentence processing. In: Schwieter, John W. and Wen, Zhisheng (eds.) The Cambridge Handbook of Working Memory and Language. Cambridge University Press, pp. 593-612. ISBN 9781108955638 doi: https://doi.org/10.1017/9781108955638.033 Available at https://centaur.reading.ac.uk/106198/

To link to this article DOI: http://dx.doi.org/10.1017/9781108955638.033

Publisher: Cambridge University Press

Working memory and L2 sentence processing

Ian Cunnings

University of Reading

Address for correspondence:

Dr Ian Cunnings

School of Psychology and Clinical Language Sciences

The University of Reading

Whiteknights Campus

Reading

RG6 7BE

Abstract

Working memory based limitations have increasingly been proposed as a way of explaining differences between native (L1) and non-native (L2) sentence processing. However, while there has been increasing interest in the role that working memory may play in L2 sentence processing, different approaches to L2 processing rely on different conceptualisations of the role that working memory plays in sentence processing. These different conceptualisations lead to different predictions about both the source of L1/L2 differences in sentence processing, and how individual differences in L2 processing should be measured. In this chapter, I critically evaluate different models of working memory during L2 sentence comprehension, review existing studies that have examined how working memory influences L2 sentence processing, and discuss the importance of considering how individual differences in working memory and sentence processing can be measured.

Keywords: second language processing; working memory; capacity-based theory; similarity-based interference.

1 Introduction

Understanding either a native (L1) or non-native (L2) language in real-time relies on the ability to keep track of who did what to whom during sentence and discourse comprehension. That real-time language processing relies on working memory, broadly defined as the ability to manipulate a limited amount of task-relevant information at one time (Baddeley, 2007; Cowan, 2017), is not contested, and it is clear that storing and accessing information from memory is crucial for successful comprehension (Jäger et al., 2017; Lewis et al., 2006; Vasishth et al., 2019). As such, it is not surprising that L2 researchers are interested in the role that working memory plays in determining outcomes in L2 acquisition and processing (e.g. Linck et al., 2014; Shin, 2020). However, different conceptualizations of working memory make different predictions about how working memory should influence language comprehension. In this chapter, I discuss how these different conceptualizations make different predictions about how working memory may influence L2 processing. I begin below by outlining different approaches to working memory during sentence processing, before discussing existing research that has examined how working memory may influence L2 processing. I end by discussing the importance of considering how task-based differences may influence conclusions that can be drawn about working memory and individual differences in L2 processing.

2 Capacity-based and interference-based approaches to working memory during sentence processing

A key distinction in different accounts of working memory is the extent to which there is hypothesised to be a distinct component dedicated to working memory. According to multi-component theories (e.g. Baddeley, 2007), there is a capacity-limited working memory component distinct from long-term memory. Other accounts assume there is no dedicated working memory, and instead describe working memory as the activated part of long-term memory (e.g. Cowan, 2017). Other accounts describe working memory in terms of the ability to allocate attention to task-relevant information (e.g. Engle, 2002). These different approaches make different predictions about the role of working memory during L2 sentence processing. Below I discuss two different approaches in turn.

2.1 Capacity-based approaches to working memory during L2 processing

A dominant approach to working memory during L2 processing is the capacity-based view, originally proposed in the L1 processing literature (Daneman & Carpenter, 1980) and extended to L2 acquisition by Harrington and Sawyer (1992). Such accounts, which are broadly compatible with multi-component views, assume that there is a dedicated working component which has a capacity that differs between individuals, which in turn leads to individual differences in language comprehension. Applied to the L2 context, capacity-based approaches predict that L2 comprehension is slower and more effortful than L1 comprehension, taxing the capacity-limited working memory component to a greater extent in L2 as compared to L1 processing. Such accounts would predict that L2 sentence processing difficulty may be neutralised for L2 learners with high enough L2 memory capacity, or alternatively that L1 processing may become error-prone if taxed in such a way as to mimic the capacity-limitations of L2 processing (McDonald, 2006).

A contested issue in capacity-based approaches to sentence processing is how capacity is defined. Capacity could be described in terms of the number of words and/or phrases that an individual can maintain in working memory at one time or alternatively, especially in the case of syntactic ambiguity resolution, capacity-limitations may constrain how many interpretations of an ambiguous input a reader can consider in parallel at one time, or how many different sources of information a reader may be able to consult at one time to guide sentence processing (Daneman & Carpenter, 1980; Just & Carpenter, 1992; Just et al., 1996). Just and Carpenter

(1992) claimed that individual differences in memory capacity influence sentence processing such that high capacity readers are able to consult discourse and pragmatic information to guide syntactic ambiguity resolution and consider multiple parses of an ambiguous input, while low capacity readers may focus on the syntactically simplest parse and have difficulty utilising discourse-based information to guide ambiguity resolution.

Applied to L2 processing, such accounts would predict that capacity-limited L2 learners may maintain fewer words/phrases in working memory at one time than L1 speakers, which presumably may impact on L2 learners ability to process sentences with non-adjacent constituents, as in the case of different linguistic dependencies. If the ability to take different sources of information, such as discourse context, into consideration during processing is capacity-limited (Just & Carpenter, 1992), capacity-based approaches would also predict that L2 learners may have difficulty utilising discourse context to guide sentence processing. However, L2 learners have been shown to be sensitive to how discourse-level information guides syntactic ambiguity resolution (Pan et al., 2015; Pan & Felser, 2011; Pozzan & Trueswell, 2016). Indeed, the work of Pan and colleagues has suggested that for at least certain types of syntactic ambiguity, L2 learners may be *more* sensitive to discourse-related biases than L1 speakers. For example, Pan et al. (2011) examined temporary syntactic ambiguities in which a prepositional phrase modified either a preceding verb (e.g. Bill glanced at the customer *with strong suspicion*) or a preceding noun (e.g. Bill glanced at the customer *with ripped jeans*). The prior discourse created interpretive bias for either verb or noun modification. L2 learners' reading times of the critical sentences were faster for verb-modifying sentences in verb-biasing contexts, and faster for noun-modifying sentences in noun-biasing contexts. Although L1 readers demonstrated sensitivity to this discourse bias in an offline comprehension task (as did the L2 group), they did not show sensitivity to it during online reading. These findings would be unexpected under a view in which purportedly capacity-limited L2 readers have difficulty taking multiple sources of information into account during real-time sentence processing.

2.2 Interference-based approaches to working memory during L2 processing

In contrast to capacity-based models, other accounts do not posit a separate working memory component, but instead distinguish between items in memory and those in the focus of attention. Working memory, in this case, is characterised as the ability to allocate attention to bring task-relevant information in and out of the focus of attention (e.g. Engle, 2002). Such accounts focus on how memory representations are encoded, stored and retrieved during the completion of cognitive tasks. Cue-based parsing (Lewis et al., 2006, 2006; Lewis & Vasishth, 2005; McElree et al., 2003; Van Dyke & Johns, 2012) provides a framework for how memory representations are encoded and retrieved from memory during sentence processing.

To illustrate how memory encoding and retrieval influences language processing, consider anaphora resolution from the perspective of cue-based parsing. In (1), representations of each word and phrase in the sentence will be encoded as each are encountered during incremental processing. Reaching the pronoun 'him' will initiate a memory retrieval operation to find an antecedent for the pronoun. In cue-based parsing, memory retrieval is achieved by matching a set of retrieval cues against all items in memory in parallel, with the item that provides the best match being retrieved. A pronoun such as 'him', as in (1a), will initiate retrieval using a set of cues including, but not limited to, a cue for the current discourse topic (e.g. [+TOPIC]), and a masculine antecedent (e.g. [+MASC]), along with cues related to syntactic constraints on anaphora resolution. As memory retrieval involves matching cues against all items in memory in parallel, items that partially match a set of retrieval cues may be accidently retrieved causing interference. In (1a), the retrieval cues as described above uniquely identify 'the boy' as the antecedent to be retrieved, but in (1b) 'the man' provides a partial match as it is [+MASC]. Cue-based parsing predicts that partially matching items in memory may sometimes be retrieved, causing *interference* during processing. As interference is based on the similarity between a set of retrieval cues and the items in memory, this is called similarity-based interference. Cue-based parsing thus predicts that how information is retrieved from memory, especially in terms of similarity-based interference, is a key determinant of successful language comprehension (e.g. Lewis et al., 2006; Van Dyke & Johns, 2012; Vasishth et al., 2019). Note that although I focus here on memory encoding and retrieval operations during the processing of different linguistic dependencies, such as in anaphora resolution, in typical instantiations of cue-based parsing (e.g. Lewis & Vasishth, 2005), memory encoding and retrieval operations occur at each word in the sentence.

(1a) The boy said that the lady spoke to him last night.

(1b) The boy said that the man spoke to him last night.

From this perspective of working memory during sentence processing, L1/L2 differences can be explained in terms of how L2 learners encode and retrieve information from memory. Cunnings (2017) proposed that L2 learners may be more susceptible to similarity-based interference as a result of how they weight memory retrieval cues. Specifically, L2 learners may weight discourse-based cues more heavily than L1 speakers.

Note that there are different instantiations of cue-based parsing (e.g. Parker et al., 2017; Vasishth et al., 2019). Simplifying somewhat, some recent debate has centred around whether particular findings index differences in the speed verses accuracy of a retrieval operation during sentence processing as a result of interference. It is beyond the scope of this chapter to discuss these issues here. Suffice to say that a key difference between capacity-based and interference-based approaches is that capacity-based views predict that a key determinant of successful sentence processing is the *quantity* or number of types of information in working memory at one time, while the interference-based account focuses instead on the *content* of information in memory. Another possibility is that both capacity and interference influence processing, although one of the two factors may still play a larger role in determining outcomes in L2 acquisition and processing.

2.3 Teasing apart capacity-based and interference-based approaches

It can sometimes be difficult to tease apart capacity-based and interference-based accounts of sentence processing as capacity and interference are often confounded. Consider a so-called filler-gap dependency as exemplified in (2), where the displaced noun 'the book' needs to be interpreted as the direct object of the verb 'read' for comprehension to be successful. Both capacity-based and interference-based approaches would predict that the short dependency in (2a) should be easier than the longer dependency in (2b). The capacity-based view would explain this in terms of the amount of information that needs to be maintained at one time being larger in (2b) than (2a). However, as dependencies become longer, the number of constituents that could potentially lead to interference will also naturally increase, in line with interference-based accounts. For example, the inclusion of another 'readable' noun ('the newspaper') in (2b) should lead to increased interference in comparison to (2a). Such interference would not be predicted in (2c), where the dependency is equally as long as (2b) in terms of the number of words or phrases, but importantly the additional noun ('the coffee') is not plausibly 'readable'. How the capacity-based view would predict differences between sentences like (2b) and (2c) is less clear, as the amount of information that needs to be maintained in both cases is similar, while it naturally follows that (2b) may cause difficulty in comparison to (2c) in the interference-based view, as the content of the sentences differ.

(2a) John saw the book that the man read happily.

(2b) John saw the book that the man with the newspaper read happily.

(2c) John saw the book that the man with the coffee read happily.

The capacity-based and interference-based approaches make different predictions about how individual differences should influence working memory during sentence processing. According to the capacity-based view, capacity-based limitations in L2 sentence processing should restrict the amount or types of information that an L2 reader can consult at one time. A key issue from this perspective is how memory capacity is assessed. This is typically assessed through administration of complex span tasks, where participants must remember a list of items whilst completing a secondary task (see e.g. Conway et al., 2005; Mathy et al., 2018). In L2 sentence processing, this is most often operationalised in terms of performance on variants of the reading span task, in which participants read a series of sentences and must remember the final word of each sentence for later recall (e.g. Daneman & Carpenter, 1980; Harrington & Sawyer, 1992). The number of sentences, and thus the number of words to recall, usually varies from two to five, and participants often also have to engage in a secondary task, such as grammaticality or plausibility judgements of the presented sentences. The number of words recalled is typically taken as the dependent measure, with a higher recall rate being taken as evidence of a larger working memory capacity. Note that the task is administered in different variants, for example it is sometimes tested in the L1 or L2 (or both), and is also not always scored in the same way, which may impact potential correlations between reading span scores and other variables (e.g. Shin, 2020). I return to this issue in Section 4.

Alternatively, according to the interference-based account, L2 sentence processing is predicted to be dependent on how L2 learners encode and retrieve information from memory during processing. One way in which individual differences may be explained in this case is in terms of individual differences in the quality of representations in memory. Van Dyke et al. (2014), for example, who studied individual differences in L1 sentence processing, argued that individuals with larger receptive vocabulary knowledge encode higher quality representations in memory that are less susceptible to interference. This proposal is in ways similar to some claims in the L2 processing literature, where Hopp's (2018) Lexical Bottleneck Hypothesis predicts that successful L2 sentence processing is reliant on the quality of L2 lexical representations. Another source of individual variation from the perspective of cue-based parsing comes from how individuals weight different cues during memory retrieval (Vasishth et al., 2019). As noted above, Cunnings (2017) argued L2 learners may weight discourse-based retrieval cues more heavily during processing than L1 speakers. This hypothesis would predict that a primary determinant of L2 processing ability is the amount and type of linguistic experience an L2 learner has in the L2. Both robust lexical representations, and the appropriate weighting of different memory retrieval cues, require adequate linguistic experience of the relevant L2. This importance of linguistic experience in determining individual differences in L2 sentence processing parallels debate in the L2 acquisition literature regarding the roles of language proficiency and L1 transfer in L2 acquisition. L2 acquisition, from the perspective of cue-based parsing, involves learning to weight retrieval cues in an appropriate manner, while transfer can be described in terms of how an L2 learners' L1 may influence how different L2 learner groups weight memory retrieval cues during L2 processing. More generally, accounting for individual differences in L2 sentence processing in this way can help provide greater integration between research on real-time L2 sentence processing and L2 acquisition.

3 Memory capacity and interference in L2 sentence processing

I now turn to existing research that bears on how memory capacity and interference may influence L2 sentence processing. For memory capacity, I focus on existing research that has examined the interaction between scores on the reading span task and different aspects of sentence processing, before discussing research on anaphora resolution as a case study on how interference influences L2 sentence processing.

3.1 Reading span and relative clause attachment in L2 sentence processing

Studies examining relative clauses in L2 processing have tested sentences like (3). (3a) is ambiguous, as the relative clause ('who injured himself') can modify either the first noun phrase (NP; 'the brother'), called high attachment, or the second ('the man'), called low attachment. A large literature has examined attachment preferences, and while results have been mixed, L1 English speakers typically prefer low attachment, while preferences in other languages differ (e.g. Carreiras & Clifton, 1993; Cuetos & Mitchell, 1988; Hemforth et al., 2015). Whether L2 speakers demonstrate L1-like attachment preferences has been widely debated (e.g. Felser et al., 2003; Hopp, 2014; Papadopoulou & Clahsen, 2003; Witzel et al., 2012).

(3a) James saw the brother of the man who hurt himself yesterday afternoon.

(3b) James saw the sister of the man who hurt himself yesterday afternoon.

(3c) James saw the brother of the lady who hurt himself yesterday afternoon.

Individual differences in L2 attachment resolution have been examined in both offline comprehension tasks, and online reading time experiments. In an offline task testing ambiguous sentences, Hopp (2014) reported a negative correlation between reading span, tested in the L2, and high attachment in L2 learners of English, with higher L2 span readers preferring low attachment. This finding has recently been replicated by Cheng et al. (2021), who found this pattern in both L1 and L2 English readers (with the reading span task again being administered in English). Similar findings have also been reported in some previous L1 studies (e.g. Swets et al., 2007). However, the finding that lower span readers chose high attachment more frequently than higher span readers is perhaps unexpected from a capacity-based view, which most obviously would predict that low span readers should prefer the linearly closer NP. This would predict a low attachment preference in low-span readers, i.e. the opposite pattern to what has been observed. Swets et al. argued that these results may reflect how lower and higher span readers chunk sentences into prosodic units, with different chunking strategies leading to different NPs in the relative clause becoming more salient. Whatever the source of this effect, the fact that similar findings have been observed in both L1 and L2 comprehension suggests a similar underlying cause in both L1 and L2 processing.

Less consistent results have been observed in online studies. Kim and Christianson (2017) tested ambiguous relative clauses in Korean L2 learners of English in a self-paced reading experiment. Participants were tested on relative clauses in both their L1 (Korean) and L2 (English), and completed a reading span task in their L1. Kim and Christian found longer reading times for individuals with higher L1 reading spans, a pattern that was found in both the readers' L1 and L2. They interpreted this as indicating that high-span readers are more likely to consider both possible attachment sites in parallel, leading to competition. This would be consistent with claims that individuals with higher memory capacity are able to consider multiple interpretations of an ambiguous sentence at one time (Just & Carpenter, 1992). Note that as Kim and Christianson only tested ambiguous sentences in their study, and did not compare reading times of ambiguous sentences to those disambiguated to either low or high attachment, as in (3b) and (3c) respectively, this interpretation of their results is difficult to distinguish from one in which participants with higher L1 reading span scores were simply slower, more careful readers overall. Cheng et al. (2021) did not report any significant effects of L2 reading span in an eye-tracking study that tested both ambiguous sentences like (3a), and sentences disambiguated to either low or high attachment as in (3b/c) respectively. Hopp (2014) also did not report any significant effects of L2 reading span in an eye-tracking study of L2 readers testing sentences disambiguated to either low or high attachment.

In sum, while reading span appears to be correlated with attachment preferences in offline tasks, whether reading span influences online processing of relative clauses during L2 comprehension is less clear. Note that cross-study comparisons are complicated in this case, as while Cheng et al. (2021) and Hopp (2014) tested reading span scores in the L2, Kim and Christianson (2017) tested L1 reading spans.

3.2 *Reading span and filler-gap dependencies in L2 sentence processing*

Individual differences in filler-gap dependency resolution has also been investigated. Dallas et al. (2013) examined filler-gap dependencies as in (4), which manipulated whether a non-adjacent noun was a plausible (‘which player’) or implausible (‘which football’) direct object of the verb ‘threatened’, using event-related brain potentials (ERPs). Implausible sentences were expected to yield the N400 ERP component (Kutas & Hillyard, 1980). L1 English speakers displayed the expected N400 effect, as did higher proficiency but not lower proficiency Chinese L2 learners of English. Participants completed three memory span tasks (alphabet span, subtract two span and English reading span), but a composite memory span score did not predict ERP effects in the L2 group.

(4) The umpire asked which player/football the coach threatened before the game.

Other studies examining filler-gap dependencies using various different tasks have also not found significant effects of reading span. Felser and Roberts (2007) did not find any significant effects of L2 reading span in a cross-modal priming paradigm that investigated filler-gap dependency resolution during L2 listening comprehension, while Miller (2014) did not find any significant effects of L2 reading span in a reading comprehension task that also involved picture verification. Juffs (2005) examined filler-gap dependencies in a self-paced reading experiment but did not find any significant effects of either (L1 or L2) reading span or word span.

In contrast to these null effects, Dussias and Piñar (2010) found that only L2 learners with high L2 reading spans behaved like L1 speakers in their self-paced reading experiment. However, Dussias and Piñar tested sentences like (5), where the displaced filler 'who' must be interpreted as the subject of 'killed' for successful comprehension, but it may initially be interpreted as the direct object of 'know' during incremental processing. As such, these results might be related to reanalysis following temporary syntactic ambiguity, rather than effects related to filler-gap dependency resolution per se.

(5) Who did the policeman know killed the pedestrian?

While these results might indicate a specific effect of L2 reading span during reanalysis, given uncertain results in other aspects of ambiguity resolution, as discussed in Section 3.1, further research is required here to assess the replicability of this finding. In sum, although Dussias and Piñar reported an interaction between L2 reading span and (a particular aspect of) filler-gap dependency resolution, other studies have not consistently found correlations between span scores and different sentence processing measures indexing L2 processing of filler-gap dependencies.

3.3 Reading span and morphosyntactic agreement in L2 sentence processing

A large literature has examined L2 processing of morphosyntactic agreement (for review, see Cunnings, 2017). It is beyond the scope of this chapter to provide a comprehensive overview, and instead I focus on studies that have examined how the length of agreement dependencies interact with reading span scores. Studies have examined how length influences the processing of L2 agreement based on the hypothesis that longer dependencies presumably impose greater demands on working memory than shorter ones.

In an eye-tracking during reading study, Keating (2010) examined noun-adjective gender agreement in L1 Spanish speakers and English L2 Spanish learners. He compared reading times for sentences such as (6), in which gender marking on an adjective (‘abierta’/’abierto’ and ‘rosado’/’rosada’) must agree with the sentence subject (‘la tienda’ and ‘el vestido’). Sentences contained either grammatical or ungrammatical gender marking on the adjective, and length was also manipulated by including both short (6a) and long (6b) conditions. The long condition contained an additional noun (‘la muchacha’) intervening between the adjective and sentence subject.

(6a) La tienda está abierta/∗abierto los sábados y domingos por la tarde.

‘The store-FEM is open-FEM/∗open-MASC Saturdays and Sundays in the afternoon.’

(6b) El vestido de la muchacha es rosado/∗rosada y tiene lunares blancos.

‘The dress-MASC of the girl-FEM is pink-MASC/∗pink-FEM and has white polka dots.’

L1 readers had longer reading times for ungrammatical sentences irrespective of length, while L2 readers as a group showed this grammaticality effect for short dependencies only. Reading span scores, assessed in the participants’ L1, however positively correlated with the grammaticality effect in L2 readers only, with higher span learners showing larger grammaticality effects in both short and long conditions.

Two other studies are also relevant here. Coughlin and Tremblay (2013) used self-paced reading to examine sensitivity to number agreement violations during the processing of French clitics in L1 and L2 readers. L2 participants completed a reading span task in both their L1 (English) and L2 (French). Using a length manipulation in a similar vein to Keating (2010), Coughlin and Tremblay found that high proficiency, but not intermediate proficiency, L2 learners had longer reading times for ungrammatical than grammatical clitics in short and long conditions. This grammaticality effect was numerically larger in the short condition however, and there was a tendency for L2, but not L1, reading span scores to interact with grammaticality effects, though the relevant correlations were only marginally significant. Foote (2011) also reported a self-paced reading experiment that tested noun-adjective gender agreement and subject-verb number agreement in L1 and L2 Spanish speakers, again comparing short and long dependencies. Both groups showed grammaticality effects in short and long conditions, with differences between grammatical and ungrammatical sentences being smaller in the long conditions for both gender and number agreement. L2 reading span scores however did not significantly correlate with the size of these grammaticality effects.

In sum, there have been mixed results in terms of whether reading span scores predict L2 sensitivity to agreement violations. L2 learners do however more consistently show grammaticality effects in shorter rather than longer conditions across the studies discussed above. This might be taken as evidence of a capacity-based disadvantage in L2 readers. However, in the studies above, the long conditions always include an additional intervening element that may interfere in the agreement dependency. In (6b), for example, the intervening noun (‘la muchacha’) matches the gender of the adjective in the ungrammatical condition. As such, it is difficult to tease apart a capacity-based explanation from an interference-based one based on these results. One way to tease these issues apart would be to systematically manipulate the agreement properties of the intervening element in the long dependencies, as cue-based parsing would predict that the properties of this constituent should influence the extent to which readers are susceptible to interference (for discussion, see Cunnings, 2017). Further research is required here to tease apart a capacity-based and interference-based account of these L1/L2 differences.

3.4 Interference in L2 sentence processing

The results in the preceding sections highlight inconsistencies in whether reading span scores interact with L2 sentence processing, and some results are confounded between capacity-based and interference-based accounts. I now turn to further discussion of how interference can provide an explanation for some findings in the L2 processing literature. For reasons of space, I discuss here briefly two relevant studies on the processing of reflexives. For in-depth review, see Cunnings (2017).

Felser et al. (2009) examined the processing of reflexives in an eye-tracking during reading study with L1 English speakers and L2 learners with L1 Japanese. They tested sentences as in (7). In (7), the only grammatical antecedent for the reflexive ‘himself’ is ‘Richard’, while the gender of an ungrammatical potential antecedent has also been manipulated. In (7a), this ungrammatical antecedent (‘John’) matches the gender of the reflexive, while in (7b), it mismatches (‘Jane’).

(7a) John noticed that Richard had cut himself with a very sharp knife.

(7b) Jane noticed that Richard had cut himself with a very sharp knife.

Although L2 learners demonstrated native-like understanding of reflexives in an offline task, Felser et al. found longer reading times at the reflexive for L2, but not L1, readers when the ungrammatical antecedent matched the gender of the reflexive. This finding might be unexpected from a capacity-based perspective, given the ungrammatical antecedent is linearly more distant from the reflexive than the grammatical antecedent, but it follows naturally from an interference-based account. From this perspective, reflexive resolution involves retrieving an antecedent from memory. Retrieval is achieved by matching a set of retrieval cues against all items in memory. Retrieval cues will include structural cues that guide retrieval to the grammatical antecedent, but importantly also morphosyntactic agreement features, in this case [+MASC]. The longer reading times for L2 learners in (7a) may thus index interference from the discourse prominent, gender-matching sentence subject.

Felser and Cunnings (2012) reported L2 difficulty of a slightly different nature in their eye-tracking study of reflexives in L1 English speakers and German L2 English learners. They tested sentences similar to (7) but manipulated gender congruency between the reflexive and both the grammatical and ungrammatical antecedent. When readers first encountered the reflexive, L1 readers had longer reading times when the grammatical antecedent mismatched the gender of the reflexive, while L2 readers showed longer reading times when the ungrammatical antecedent mismatched the reflexive's gender. L2 learners demonstrated native-like understanding of reflexives in an offline task however, and their reading times were influenced by the gender of the grammatical antecedent, like L1 readers, in reading times for regions of text after the reflexive (i.e. in spillover processing). In sum, although the precise pattern of results in Felser and Cunnings (2012) and Felser et al. (2009) differ, they both suggest L2 difficulty during processing as a result of L2 learners' temporary consideration of discourse-prominent, but ungrammatical, antecedents during reflexive resolution.

Note that whether anaphora resolution in L1 processing is impervious to interference is debated (Dillon et al., 2013; Jäger et al., 2020), and interference clearly influences L1 processing more generally (for review, see Jäger et al., 2017). These L2 results are however compatible with the claim that a source of L1/L2 differences during processing is how the different populations weight different cues to memory retrieval (Cunnings, 2017).

This discussion is not intended to be an exhaustive overview of how interference may influence L2 anaphora resolution (see Cunnings, 2017). However, it illustrates how interference may influence L2 processing, and highlights how consideration of the content of information in memory, in this case potential antecedents for an anaphor, rather than merely the amount of information in memory, can help inform our understanding of L2 processing. That L2 processing may be influenced by the content, rather than mere amount, of information in memory, is predicted by interference-based accounts that focus on memory encoding and retrieval, rather than capacity, during sentence processing. Although the studies discussed above may suggest a role of interference during L2 processing, further research is clearly required here to disentangle potential roles of memory capacity and interference in L2 anaphora resolution more broadly. Whether capacity or interference can explain L1/L2 differences in other linguistic dependencies also requires further systematic examination.

4 Quantifying individual differences in L2 sentence processing

As discussed above, there have been inconsistent results in terms of individual differences in reading span and L2 sentence processing. At least some of this inconsistency may be related to the sample sizes tested in existing research, as these have varied across studies. Individual differences research requires large samples, and further research with large L2 samples is required here. Another difficulty in assessing the (in)consistency of these results is that the reading span task has not been administered and scored in a systematic way across studies.

Additionally, the extent to which there is indeed consistent individual variation in the psycholinguistic tasks used to examine L2 sentence processing has not been systematically examined. Below, I discuss the importance of considering these two issues when assessing the influence of working memory in L2 sentence processing.

4.1 Measuring and characterising individual differences in L2 reading span

Inconsistency in the administration and scoring of reading span tasks makes it difficult to systematically compare results across L2 studies, as the different findings may at least in part be due to different ways in which the reading span task is administered and scored. Shin (2020) conducted a meta-analysis of the effects of reading span on L2 reading comprehension, and although she did not examine L2 sentence processing, her results are illustrative of this issue. Shin found that differences in how the task was administered, for example in which language (the L1 or L2) the task was tested, whether words needed to be recalled in the order in which they were remembered, whether there was a secondary task, the length of the sentences used, and also the scoring method used, all influenced the size of any potential correlation between reading span and L2 reading comprehension scores. Shin's results indicate the need for standardisation in the administration and scoring of the reading span test when assessing the role that reading span may play in L2 comprehension. It is less clear how such issues may influence L2 sentence processing, as they have not been systematically examined in L2 processing research, but it is likely that at least some of the inconsistences in the literature may be due to different ways in which the reading span task has been administered and scored across studies. For example, the language in which the reading span task was administered was not consistent in the studies discussed in Sections 3.1-3.3. Some studies tested L1 reading span, some L2 reading span, and some both. The studies also did not use consistent scoring methods.

These differences make it difficult to draw conclusions about how reading span may influence L2 sentence processing across studies.

Additionally, how to interpret interactions between reading span scores and sentence processing is contested. Although reading span scores are often interpreted as indexing memory capacity, whether the reading span task measures individual differences in capacity rather than individual differences in language experience has been debated (Kidd et al., 2018; MacDonald & Christiansen, 2002). MacDonald and Christiansen (2002), for example, argued that the reading span task is a test of language processing ability with its own set of task demands, rather than an index of processing capacity, and Farmer et al. (2017) recently found that linguistic experience influenced reading span scores in L1 speakers. The extent to which L2 linguistic experience influences L2 scores in the reading span task has also been debated (Juffs & Harrington, 2011). Juffs and Harrington suggested that one way to alleviate this issue is to create a composite reading span score based on L2 learners' performance on reading span tasks administered in both their L1 and L2. Additionally, a number of researchers (e.g. Waters & Caplan, 2003) have advocated calculation of composite span scores based on performance across several complex span tasks, including those that do not rely so heavily on language such as operation span (Turner & Engle, 1989), rather than assessing memory capacity using a single task. The rationale being that a composite score will provide a better estimate of an individual's memory capacity that is not as confounded with performance in a single task. Further research is required here to tease apart the extent to these issues can explain the mixed results related to correlations between reading span and L2 sentence processing, and whether purported differences in L2 memory capacity can be explained in terms of individual differences in L2 linguistic experience.

It is also important to note that differences in how psycholinguistic tasks are administered may also influence results. Recall that Kim and Christianson (2017) found an interaction between reading span and L2 processing of ambiguous relative clauses, whereas Cheng et al. (2021) did not. In addition to the fact that Kim and Christianson tested L1 reading span while Cheng et al. tested L2 reading span, another difference between the two studies is that Kim and Christian used self-paced reading whereas Cheng et al. adopted eye-tracking during reading. Self-paced reading, which typically does not allow rereading of earlier parts of a sentence, may place different demands on working memory than eye-tracking during reading, where rereading is possible. Note also that while both experiments required participants to answer comprehension questions about the sentences they read, only Kim and Christianson's comprehension task probed interpretation of the relative clause ambiguity itself. These task-related differences make it difficult to draw direct comparisons between the two studies, and further research is required to examine how such task demands may influence individual differences in L2 sentence processing.

4.2 Measurement reliability and individual differences

A final issue to consider when assessing individual differences, be it in reading span or some other measure, is whether widely used cognitive tasks systematically measure individual differences to begin with (Hedge et al., 2018; Parsons et al., 2019). For an experimental task to measure individual differences, there must be systematic individual variation in the task to begin with, and to meaningful correlate two measures, such as reading span scores and a measure of sentence processing, there must be systematic individual variation in *both* tasks. The systematicity with which a task measures individual differences is assessed by measurement reliability statistics, such as split-half reliability (for discussion, see Parsons et al., 2019). In split-half reliability, each participant's data from a task is randomly split in half, and the two halves correlated. A correlation of .7 or above is considered consistent enough for a task to be used as a measure of individual differences (Nunnally, 1978). While this is not a strict cut-off, low measurement reliability is detrimental to statistical inference (Parsons et al., 2019).

In Shin's (2020) meta-analysis of reading span and L2 reading comprehension, measurement reliability was only reported in a minority of the studies surveyed. Although it is difficult to quantify how often reliability measures are reported for studies examining the role of reading span or other individual differences in L2 sentence processing, as the relevant meta-analysis has not been conducted, reliability measures are not routinely reported in L2 sentence processing research (Cunnings & Fujita, 2020).

Cunnings and Fujita examined individual differences in a self-paced reading study that tested temporarily ambiguous and unambiguous sentences as in (8a) and (8b) respectively. Longer reading times are expected at 'played' in (8a), as 'the cat' may initially be interpreted as being conjoined with 'the dog' as the direct object of 'washed' when it is in fact the subject of 'played'. (8b) is an unambiguous control in which the temporary ambiguity does not arise due to use of the conjunction 'while' instead of 'and'.

(8a) Ken washed the dog and the cat in the garden played with a ball.

(8b) Ken washed the dog while the cat in the garden played with a ball.

Although whole group analyses indicated the expected differences in reading times between (8a) and (8b) in L1 and L2 readers, and split-half reliabilities for measures of overall participant reading speed were high (> .9), split-half reliabilities for the crucial ambiguity effect (the ambiguous/unambiguous difference) were low (< .2). These results indicate little systematic individual variation in syntactic ambiguity resolution in the sample tested to begin with, making the question of which factors may module individual differences in L2 sentence processing, be it memory capacity, interference or another factor, a moot point in this study.

As such, future investigations assessing the role of individual differences in L2 sentence processing, be it in terms of reading span scores or other individual differences measures, must consider whether the tasks used actually measure systematic individual differences to begin with (for further discussion, see Cunnings & Fujita, 2020).

5 Conclusion

In this chapter, I have discussed different accounts of working memory during L2 sentence processing. Crucial to this debate is the conceptualisation of working memory itself, and while the capacity-based view may have been dominant in previous L2 processing research, consideration of memory interference can also inform our understanding of the similarities and differences between L1 and L2 sentence processing. It is also important to consider differences in how tasks are administered across studies. This includes both the tasks used to measure memory capacity but also the psycholinguistic tasks used to examine sentence processing, as task-based differences can potentially lead to non-trivial differences in the pattern of results obtained. The extent to which commonly used psycholinguistic tasks measure systematic individual differences in L2 processing also needs to be considered. Further research that systematically examines these issues is required to further tease apart the roles of memory capacity and interference during L2 sentence processing.

References

Baddeley, A. D. (2007). *Working memory, thought, and action*. Oxford University Press.

Carreiras, M., & Clifton, C. (1993). Relative Clause Interpretation Preferences in Spanish and English. *Language and Speech*, *36*, 353–372. https://doi.org/10.1177/002383099303600401

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

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*, 769–786. https://doi.org/10.3758/BF03196772

Coughlin, C. E., & Tremblay, A. (2013). Proficiency and working memory based explanations for nonnative speakers' sensitivity to agreement in sentence processing. *Applied Psycholinguistics*, *34*, 615–646. https://doi.org/10.1017/S0142716411000890

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review*, *24*, 1158–1170. https://doi.org/10.3758/s13423-016-1191-6

Cuetos, F., & Mitchell, D. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition*, *30*, 73–105. https://doi.org/10.1016/0010-0277(88)90004-2

Cunnings, I. (2017). Parsing and Working Memory in Bilingual Sentence Processing. *Bilingualism: Language and Cognition*, *20*, 659–678. https://doi.org/10.1017/S1366728916000675

Cunnings, I., & Fujita, H. (2020). Quantifying individual differences in native and nonnative sentence processing. *Applied Psycholinguistics*, 1–21. https://doi.org/10.1017/S0142716420000648

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*, 766–799. https://doi.org/10.1111/lang.12026

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language, 69*, 85–103. https://doi.org/10.1016/j.jml.2013.04.003

Dussias, P. E., & Piñar, P. (2010). Effects of reading span and plausibility in the reanalysis of wh-gaps by Chinese-English second language speakers. *Second Language Research, 26*, 443–472. https://doi.org/10.1177/0267658310373326

Engle, R. W. (2002). Working Memory Capacity as Executive Attention. *Current Directions in Psychological Science, 11*, 19–23. https://doi.org/10.1111/1467-8721.00160

Farmer, T. A., Fine, A. B., Misyak, J. B., & Christiansen, M. H. (2017). Reading Span Task Performance, Linguistic Experience, and the Processing of Unexpected Syntactic Events. *Quarterly Journal of Experimental Psychology, 70*, 413–433. https://doi.org/10.1080/17470218.2015.1131310

Felser, C., & Cunnings, I. (2012). Processing reflexives in a second language: The timing of structural and discourse-level constraints. *Applied Psycholinguistics, 33*, 571–603. https://doi.org/10.1017/S0142716411000488

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

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*, 453–489. https://doi.org/10.1017/S0142716403000237

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

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

Harrington, M., & Sawyer, M. (1992). L2 Working Memory Capacity and L2 Reading Skill. *Studies in Second Language Acquisition*, *14*, 25–38. https://doi.org/10.1017/S0272263100010457

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*, 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Hemforth, B., Fernandez, S., Clifton, C., Frazier, L., Konieczny, L., & Walter, M. (2015). Relative clause attachment in German, English, Spanish and French: Effects of position and length. *Lingua*, *166*, 43–64. https://doi.org/10.1016/j.lingua.2015.08.010

Hopp, H. (2014). Working Memory Effects in the L2 Processing of Ambiguous Relative Clauses. *Language Acquisition*, *21*, 250–278. https://doi.org/10.1080/10489223.2014.892943

Hopp, H. (2018). The Bilingual Mental Lexicon in L2 Sentence Processing. *Second Language*, *17*, 5–27. https://doi.org/10.11431/secondlanguage.17.0_5

Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, *94*, 316–339. https://doi.org/10.1016/j.jml.2017.01.004

Jäger, L. A., Mertzen, D., Van Dyke, J. A., & Vasishth, S. (2020). Interference patterns in subject-verb agreement and reflexives revisited: A large-sample study. *Journal of Memory and Language*, *111*, 104063. https://doi.org/10.1016/j.jml.2019.104063

Juffs, A. (2005). The influence of first language on the processing of wh-movement in English as a second language. *Second Language Research*, *21*, 121–151. https://doi.org/10.1191/0267658305sr255oa

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, *44*, 137–166. https://doi.org/10.1017/S0261444810000509

Just, Marcel A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149. https://doi.org/10.1037/0033-295X.99.1.122

Just, Marcel Adam, Carpenter, P. A., & Keller, T. A. (1996). The capacity theory of comprehension: New frontiers of evidence and arguments. *Psychological Review*, *103*, 773–780. https://doi.org/10.1037/0033-295X.103.4.773

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. https://doi.org/10.1075/lald.53.05kea

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual Differences in Language Acquisition and Processing. *Trends in Cognitive Sciences*, *22*, 154–169. https://doi.org/10.1016/j.tics.2017.11.006

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*, 365–388. https://doi.org/10.1177/0267658315623322

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–205. https://doi.org/10.1126/science.7350657

Lewis, R. L., & Vasishth, S. (2005). An Activation-Based Model of Sentence Processing as Skilled Memory Retrieval. *Cognitive Science*, *29*, 375–419. https://doi.org/10.1207/s15516709cog0000_25

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*, 447–454. https://doi.org/10.1016/j.tics.2006.08.007

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, *21*, 861–883. https://doi.org/10.3758/s13423-013-0565-2

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*, 35–54. https://doi.org/10.1037/0033-295X.109.1.35

Mathy, F., Chekaf, M., & Cowan, N. (2018). Simple and Complex Working Memory Tasks Allow Similar Benefits of Information Compression. *Journal of Cognition*, *1*, 31. https://doi.org/10.5334/joc.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*, 381–401. https://doi.org/10.1016/j.jml.2006.06.006

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, *48*, 67–91. https://doi.org/10.1016/S0749-596X(02)00515-6

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

Nunnally, J. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

Pan, H.-Y., & Felser, C. (2011). Referential context effects in L2 ambiguity resolution: Evidence from self-paced reading. *Lingua*, *121*, 221–236. https://doi.org/10.1016/j.lingua.2010.08.003

Pan, H.-Y., Schimke, S., & Felser, C. (2015). Referential context effects in non-native relative clause ambiguity resolution. *International Journal of Bilingualism*, *19*, 298–313. https://doi.org/10.1177/1367006913515769

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*, 501–528. https://doi.org/10.1017/S0272263103000214

Parker, D., Shvartsman, M., & Van Dyke, J. A. (2017). The cue-based based retrieval theory of sentence comprehension: New findings and new challenges. In L. Escobar, V. Torrens, & T. Parodi (Eds.), *Language Processing and Disorders* (pp. 121–144). Cambridge Scholars Publishing.

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*, 378–395. https://doi.org/10.1177/2515245919879695

Pozzan, L., & Trueswell, J. C. (2016). Second language processing and revision of garden-path sentences: A visual word study. *Bilingualism: Language and Cognition, 19*, 636–643. https://doi.org/10.1017/S1366728915000838

Shin, J. (2020). A meta-analysis of the relationship between working memory and second language reading comprehension: Does task type matter? *Applied Psycholinguistics*, *41*, 873–900. https://doi.org/10.1017/S0142716420000272

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,* 64–81. https://doi.org/10.1037/0096-3445.136.1.64

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*, 127–154. https://doi.org/10.1016/0749-596X(89)90040-5

Van Dyke, J. A., & Johns, C. L. (2012). Memory Interference as a Determinant of Language Comprehension: Interference in Comprehension. *Language and Linguistics Compass*, *6*, 193–211. https://doi.org/10.1002/lnc3.330

Van Dyke, J. A., Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition*, *131,* 373–403. https://doi.org/10.1016/j.cognition.2014.01.007

Vasishth, S., Nicenboim, B., Engelmann, F., & Burchert, F. (2019). Computational Models of Retrieval Processes in Sentence Processing. *Trends in Cognitive Sciences*, *23,* 968–982. https://doi.org/10.1016/j.tics.2019.09.003

Waters, G. S., & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, & Computers*, *35*, 550–564. https://doi.org/10.3758/BF03195534

Witzel, J., Witzel, N., & Nicol, J. (2012). Deeper than shallow: Evidence for structure-based parsing biases in second-language sentence processing. *Applied Psycholinguistics*, *33*, 419–456. https://doi.org/10.1017/S0142716411000427