RESEARCH PAPER

Working memory and L2 gender agreement processing in structural distance conditions

Ezequiel M. Durand-López

Abstract Previous studies examining the role of executive function in L2 processing show that working memory (WM) facilitates the processing of agreement in local domains in adult second language (L2) learners. Furthermore, other studies explored whether L2 learners can establish the agreement operation across phrases (i.e., structural distance) and whether WM intervenes in the said linguistic computation. However, these studies have often included both syntactic and linear distance in their stimuli, making it impossible to discern whether WM effects emerge from physical or syntactic reasons. The present study assesses how verbal WM updating and L2 proficiency modulate syntactic processing. Beginner and advanced adult English L2 learners of Spanish and Spanish monolinguals completed a verbal WM updating task, and a self-paced reading task containing Spanish sentences with gender agreement and disagreement within and across phrases. Results show that Spanish monolinguals exhibited sensitivity to gender agreement violations in local domains and in structural distance conditions, while beginner L2 learners were not sensitive to violations in either condition. Advanced learners, on the other hand, detected violations in local domains, and their verbal WM updating spans were associated with sensitivity to

violations across phrases. Taken together, the findings suggest that (a) morphosyntactically complex structures consume cognitive resources in great number, and (b) L2 processing is qualitatively similar albeit quantitatively different from native processing, thus providing evidence that late bilinguals may process the L2 in a native-like manner.

Keywords Structural distance · Working memory · Gender agreement · L2 processing · L2 Spanish

Human beings make use of a wide range of cognitive resources to perform daily tasks. The executive function administers this pool of resources, which includes working memory (WM), the cognitive system(s) responsible for the control, regulation, and active maintenance of information of many kinds in the face of distracting information (Conway et al., 2007). WM has been found to be a reliable predictor of performance on multiple language-unrelated tasks (for a review, see Chai et al., 2018). More recently, WM has also been found to play an important role in language acquisition and processing, especially in the L2, since comprehending and producing speech in the L2 is perceived as more cognitively taxing than L1 processing mainly due to cross-linguistic automatization differences (Vejnović & Zdravković, 2010). Over the years, it has become more evident that WM regulates different aspects of L2 processing, such as

E. M. Durand-López (✉)
College of Charleston, 9 Liberty Street, Charleston, SC 29401, USA
e-mail: durande@cofc.edu

reading (e.g., Walter, 2004) and listening comprehension (e.g., Fay & Buchweitz, 2014; Miki, 2012), writing (e.g., Mavrou, 2018), and spontaneous and planned speaking (e.g., Guará-Tavares, 2013).

Additionally, certain L2 processing models propose WM as a factor explaining quantitative processing differences between native speakers and L2 learners (McDonald, 2006). On the other hand, other computational accounts posit that L2 processing is qualitatively different from L1 computation regardless of speakers' cognitive resources, as L2 learners exhibit trouble processing hierarchical structures at the syntactic level (Clahsen & Felser, 2006a, 2006b, 2018). Empirical studies have found that L2 learners with higher WM spans effectively process morphosyntactic operations in local domains (e.g., Sagarra & Herschensohn, 2010), while other studies also claim that L2 learners can establish complex syntactic relations even outside of local domains (e.g., Alemán Bañón et al., 2014). However, multiple methodological issues from studies investigating structural distance (also known as syntactic distance; i.e., morphosyntactic operations across constituents) prevent their results from being conclusive on whether L2 learners can establish agreement between syntactically distant words. One such issue is the inclusion of both linear and structural distance in the stimuli, complicating interpretation. In the following section, I will provide an overview of L2 processing models and studies on the association between WM and L2 (morpho)syntactic processing, and I will discuss methodological implications of previous studies assessing the computation of structural distance in late bilinguals.

Literature review

L2 processing accounts

Currently, researchers disagree on whether L2 learners can (i.e., accessibility accounts) or cannot (i.e., deficit accounts) achieve native-like morphological and morphosyntactic processing in the L2. Both types of accounts explain morphosyntactic acquisition and processing from a representational and computational point of view. Deficit, computational accounts such as the Shallow Structure Hypothesis (Clahsen & Felser, 2006a, 2018) argue that L2 processing is qualitatively different from that of native speakers, and that they differ in the morphosyntactic computation they perform during language comprehension. Specifically, L2 learners rely more on lexical and semantic cues rather than on inflectional morphology to process the target language. This reasoning suggests why L2 learners, regardless of their proficiency level, exhibit difficulty computing morphosyntactic information. Importantly, the Shallow Structure Hypothesis proposes that L2 learners can indeed achieve native-like morphosyntactic processing. However, they also argue that L2 learners may process nonlocal dependencies (i.e., agreement across constituents rather than in the same constituent) differently than native speakers (Clahsen & Felser, 2006b, p. 565).

In contrast to deficit accounts, accessibility theories posit that L2 learners can acquire both representation and computation that are qualitatively like that of native speakers of the target language. These theories, however, acknowledge that achieving an L2 grammar comparable to that of native speakers is a daunting task. Computational accounts within accessibility theories posit that L2 learners can achieve complex structural L2 processing, as non-native and native grammars and processors are identical. These accounts have tried to explain the fact that L2 learners have difficulty computing incoming morphosyntactic information by means of both linguistic and cognitive explanations. Hopp's (2007, 2010) Fundamental Identity Hypothesis states that, although L1 and L2 grammars are qualitatively comparable, L2 processing is much less efficient than L1 processing. Differences between L1 and L2 processing may be found, but these variances are not due to a critical period or to L2 learners' ability to integrate multiple sources of knowledge during processing, but to L1 transfer, weaker decoding abilities or shortage of computational resources. McDonald's (2006) account also explains differences between L1 and L2 processing in terms of cognitive resources. This model claims that bilinguals use the same systems to process both the L1 and the L2 (i.e., L1 and L2 processing are qualitatively similar). However, these processes may be quantitatively different, since the system is not as effective when processing the L2 mostly because (a) slower processing, weaker decoding abilities, and low WM capacity lead to a lack of sensitivity to morphosyntactic violations, (b) greater cognitive demands that arise during L2 processing hinder access to WM and other attentional resources, which are necessary for the

computation of incoming morphosyntactic information.

Working memory and L2 morphological and morphosyntactic processing

Baddeley and Hitch (1974) proposed the first WM model, which described WM as a limited-capacity multicomponent system comprised of two temporary store systems and a central executive system. In a more recent version of the model, Baddeley (2000) added another component to the WM model, the episodic buffer, which stores and integrates information from a myriad of cognitive sources, including that of other components of WM. Other influential WM theories were put forward after the multiple-resource model of WM. Single-resource models claim that individuals possess a set of verbal processing resources for all verbal tasks (e.g., Just & Carpenter, 1992). Both processing and storage depend on a common pool of resources within a limited cognitive ability whose capacity differs among individuals. On the other hand, the multiple-resource models posit that storage and processing function independently (Baddeley, 2007). Finally, domain-general models do not necessarily distinguish between storage and processing, as they argue that WM is the active part of long-term memory and is not conceived as a cascade of cognitive processes (Cowan, 1998; Unsworth et al., 2009). For comprehensive reviews about the different WM models, the reader is referred to Cowan (2016).

Researchers have been particularly interested in exploring whether L2 learners use this cognitive capacity while processing word structure in real time in the L2. Studies addressing the effect of WM on the processing of word structure within words suggest that L2 morphological processing does not significantly tap into the executive function (e.g., Rizaoğlu & Gürel, 2020). However, a large body of research has found that WM affects L2 morphosyntactic processing (Arnold, 2019; Sagarra, 2007; Sagarra & Herschensohn, 2010; Sagarra & LaBrozzi, 2018). In general, these studies have investigated the processing of morphosyntactic operations in local domains. Other studies have also assessed syntactic and physical factors, and their results will be discussed in the following section.

Working memory and structural distance

Studies exploring WM effects on L2 morphosyntactic processing can be divided in two: those that include distance in their stimuli, and those that do not. Distance between two words engaged in an agreement operation can be linear (i.e., interfering words between the two agreeing units) or syntactic (i.e., syntactic relations are established across syntactic constituents). Overall, those studies that did not include distance in the stimuli report no associations between WM and syntactic processing (Felser & Roberts, 2007; Juffs, 2004, 2005, 2006).

On the other hand, L2 syntactic processing studies exploring distance found WM effects. Those including linear distance in their stimuli report that, within L2 speakers, only highly proficient learners with high WM spans can compute syntactic relations with linear distance (Coughlin & Tremblay, 2013; Keating, 2010; Reichle & Coughlin, 2013), while their native speaker peers do not show an effect of WM while processing linear distance. These findings suggest that L2 processing might be quantitatively rather than qualitatively different from L1 processing. Also, they suggest that L2 processing is less automatized than L1 processing, as learners need a greater amount of cognitive resources to keep linguistic information in memory while processing interfering words before reaching to the agreeing word.

In contrast to linear distance studies, those exploring structural distance effects are rare. Thus, it is not known whether the processing of agreement established across constituents that are physically close to one another is achievable for L2 learners. These studies have explored structural distance with behavioral and neurocognitive methodologies but did not include WM tasks. Thus, the role of WM in the processing of agreement in more complex syntactic (not physical) conditions is also unclear.

Regarding behavioral studies, Keating (2009) researched structural distance effects through the computation of gender agreement in Spanish monolinguals and beginning, intermediate and advanced English L2 learners of Spanish using the eye-tracking methodology. There were three conditions: within-phrase agreement without linear distance (e.g., *Un trabajo aburrida es ideal para alguien que no tolera el estrés*), structural distance (i.e., agreement between the DP and the VP) with medium linear distance (e.g.,

Un trabajo es bastante mala cuando no ofrece vacaciones o días libres), and structural distance (agreement between the DP and a subordinate clause) with long linear distance (e.g., *Una biblioteca no tiene computadoras cuando es pequeño y no falta dinero*). The results showed that while native Spanish speakers were sensitive to the gender agreement violations in all conditions, advanced L2 learners of Spanish were only sensitive in the DP condition, and both beginning and intermediate L2 learners were not sensitive in any condition. This study advances our understanding of the processing of structural distance, although it is not exempt from methodological pitfalls. To start, the two conditions engaged in structural distance also have linear distance. This fact is particularly problematic in the third condition, such as *Un refresco tiene muy buen sabor cuando está fría y no caliente*, where the elements engaged in the N-A disagreement relation, apart from being in different constituents, are too far apart. Second, the number of participants per group for this study was low, as only 12 advanced, 14 intermediate and 18 beginning L2 learners participated in the online task. To conclude, although Keating's (2009) study is informative on the effects of distance on online processing of gender agreement, its results should be taken warily.

Two other studies explored structural distance through gender and number agreement with neurocognitive methodologies solely in advanced L2 learners. Dowens et al. (2010) investigated whether English-speaking advanced L2 learners of Spanish were able to gain native-like processing of features that were present and absent in their L1 (i.e., number and gender, respectively) using event-related potentials (ERPs). The sentences contained gender and number violations within (i.e., D-N relations) and across (i.e., N-A relations) phrases (e.g., *El/*La/*Los suelo está plano/*a/*s; The_{masc}/*The_{fem}/*The_{plural} floor_{masc} is flat_{masc}/*_{fem}/*_{plural}*). For the within-phrase condition, both types of violations elicited a P600 effect in native speakers, while advanced learners also showed a P600 effect, but this effect was greater for number violations compared to gender violations. For the structural distance condition, the differences between native speakers and L2 learners were more pronounced. Native speakers showed a P600 pattern similar to the within-phrase condition for both gender and number, while learners showed a lack of a left negativity effect, but they also showed a P600 effect with greater

amplitudes for the number than for the gender condition. This study was the first one to disentangle structural distance from linear distance. Nevertheless, the authors mention WM as a main factor modulating structural distance effects but failed to assess WM capacity.

One last study on structural distance effects, that of Alemán Bañón et al. (2014), investigated the said phenomenon in the processing of sentences with gender discord relations in Spanish monolingual speakers and advanced English L2 learners of Spanish using ERPs. The conditions relevant to the present review included: a) N-A gender concord and discord within a DP (e.g., *El cerebro es un órgano muy complejo/*a; The brain is a very complex_{masc}/*_{fem} organ*) and b) N-A gender concord and discord across constituents (e.g., *El cuadro es auténtico/*a; The_{masc} painting_{masc} is authentic_{masc}/_{fem}*). The results showed robust P600 effects for gender violations across constituents. Furthermore, native speakers and advanced L2 learners showed very similar results, as within-constituent concord yielded more positive waveforms than the across-constituent version in both gender concord and discord sentences. The authors interpreted these results as advanced L2 learners being able to establish agreement outside of a single constituent, and native-like processing being attainable in adult L2 acquisition. Also, the researchers concluded that morphosyntactic processing in advanced L2 learners is not confined to local domains.

This study managed to disentangle syntactic and lineal distance effects, but it could be improved in several ways. First, the authors did not assess WM capacity, which leaves the question of whether cognitive abilities aid syntactic processing unanswered. Second, the structures used as stimuli in the study, albeit located in different phrases, are not syntactically complex, which could have brought ceiling effects. Specifically, in sentences with structural distance such as *El cuadro es auténtico...*, there is a small clause that raises to the specifier of the sentence. Its syntactic complexity is shallow; the speaker does not need to process the meaning of the verb, since *ser* is stative in Spanish, making the sentence a simple structure to process. Third, the raising structure is a common, prototypical construction with a copula, which does not present a cognitive challenge for the comprehender. Non-canonical albeit grammatical raising structures, in turn, may be more

suitable for this type of study. Fourth, the head of the specifier (*el* in *el cuadro*) has overt gender, which activates a priori the morphosyntactic information such as grammatical gender related to the noun it modifies, and therefore, to the adjective modifying the noun. This technical concern could be detrimental for the sensitivity to the N-A agreement relation since a functional item with overt gender is participating in the agreement relation as well. Finally, the study only compares Spanish monolinguals and advanced L2 learners of Spanish, thus not tracking how L2 learners' syntactic processing behavior might change as their further advance in their L2 acquisition process.

In sum, the few studies that have explored structural distance have two aspects in common. First, they included experimental sentences either in which agreement across phrases was not isolated and therefore mixed with linear distance, or in which the difficulty of the sentences was too low to test learners' capacity to process structural distance. Second, though these studies claim that structural distance is a particularly taxing phenomenon, none of them included cognitive measures such as WM to test their assumptions. Testing whether this cognitive ability is also involved in the processing of agreement across phrases without linear distance could shed light on the issues.

The study

Experimental studies on morphosyntactic processing show that proficient L2 learners with more cognitive resources such as WM compute L2 morphosyntax, both in local domains and in structural distance conditions, similarly to native speakers, suggesting that L1 and L2 mechanisms are qualitatively similar. However, previous studies' methodological concerns as well as unexplored questions reveal the necessity for a new study examining WM effects on the processing of structural distance in the L1 and L2. First, some studies have mixed linear and structural distance in the stimuli presented (e.g., Keating, 2009). The results of studies with hybrid stimuli should be taken with caution, as it is impossible to disentangle linear distance from structural distance effects. Only one study, to my knowledge, has managed to isolate structural distance (Alemán Bañón et al., 2014). However, the structure chosen might not have been

syntactically complex enough for structural distance effects to emerge. Second, studies on L2 processing of structural distance only included L2 learners with high proficiency. Examining the role of L2 proficiency in the computation of morphosyntactic operations in complex syntactic conditions will track how learners change their syntactic processing behaviors as their further advance in the acquisition of their L2. Finally, previous studies draw a clear picture that a high WM capacity facilitates L2 processing of morphosyntax in local domains or with linear distance, but it is not clear whether learners also make use of WM to process syntax with no linear distance. This question can inform how non-linguistic cognitive abilities such as WM aid language processing.

To address the methodological and theoretical matters mentioned above, the present study investigates how WM and L2 proficiency modulate the processing of gender agreement in local domains and across phrases in adult beginner and advanced English L2 learners of Spanish and in Spanish monolinguals. The research questions of the study are as follows:

RQ1: Is the processing of gender agreement across phrases more taxing than in local domains for English L2 learners of Spanish and Spanish monolinguals? Does proficiency modulate how L2 learners process structural distance?

Based on previous studies on syntactic processing, I predict that both Spanish natives and L2 learners will have more difficulty processing gender agreement if the morphosyntactic operation is established in a more syntactically complex setting, namely across constituents. Also, based on L2 morphosyntactic processing studies, only advanced learners will be sensitive to gender agreement violations within and across phrases.

RQ2: Does WM capacity affect the processing of structural distance in beginner and advanced English L2 learners of Spanish?

I anticipate that structural distance will be more taxing for L2 learners regardless of their proficiency level than for native speakers and, as a result, WM facilitation effects will emerge only for L2 learners, in line with the findings of previous studies.

The findings of the study inform L2 processing models by providing evidence about (a) whether L2 learners exhibit similar processing mechanisms than native speakers of the target language, (b) whether the executive function is associated with the degree of

sensitivity to morphosyntactic violations, (c) whether L2 learners can establish morphosyntactic operations outside of local domains, and (d) whether proficiency intervenes in the attention paid to morphosyntactic cues to effectively process incoming linguistic information in the L2.

Method

Participants

Seventy-four participants took part in this study: 48 were English late L2 learners of Spanish (28 beginners and 20 advanced learners) and 26 were Spanish monolinguals. All were students in an American (L2 learners) or Argentinian (Spanish monolinguals) university, and they were between 18 and 31 years of age ($M = 21.68$, $SD = 3.87$). To be included in the study, Spanish monolinguals could not have spent more than two months in a country whose primary language was other than Spanish, while L2 learners (a) had to have lexical representations of all experimental nouns and adjectives, which was measured through a translation recognition task, (b) had to achieve a minimum of 80% accuracy on a gender assignment task, and (c) had to correctly identify as correct or incorrect at least 75% of the sentences in each condition on a grammaticality judgement task. Finally, beginner L2 learners were enrolled in a 3rd or 4th-semester Spanish language class, whereas advanced learners were graduate students in a Spanish literature program.

Materials and procedure

L2 learners performed the following tasks in a 75-min session: a language background questionnaire, a Spanish proficiency test, a self-paced reading task, a WM task, a translation recognition task, a gender assignment task, and a grammaticality judgement task. Spanish monolinguals, on the other hand, completed the same tasks apart from the Spanish proficiency test, the translation recognition task, the gender assignment task, and the grammaticality judgement task.

Screening tasks

The screening tasks were the language background questionnaire, the Spanish proficiency test, the translation recognition task, the gender assignment task, and the grammaticality judgement task.

Language background questionnaire L2 learners and Spanish monolinguals first completed the *Language Experience and Proficiency Questionnaire* (LEAPQ; Marian & Kaushanskaya, 2007). According to the information retrieved from the LEAPQ, all L2 learners began acquiring Spanish after the age of 11 ($M = 13.27$, $SD = 2.26$), they were raised in the United States and were not exposed to the Spanish language at home. On the other hand, the Spanish monolingual group was raised in Buenos Aires, Argentina, and did not study any L2 besides a mandatory English class in middle school and high school. Finally, monolinguals' self-perceived ratings for L2 proficiency were very low ($M = 2.36$, $SD = 1.31$; possible range: 1–10).

Spanish proficiency test L2 learners' proficiency in Spanish was measured through an adapted version of the grammar section of the *Diploma de Español como Lengua Extranjera*. Participants read sentences and a text in Spanish and filled in the blanks with one of four options. Each correct answer received one point. Participants achieving a score under 30 points were considered beginners, while those scoring above were advanced learners, following common cut-off points in the literature (e.g., Montrul & Slabakova, 2003).

Translation recognition task L2 learners performed a translation recognition task containing all experimental nouns and adjectives engaged in the gender agreement operation in the self-paced reading task. Participants matched the Spanish words located in a column on the left with their English translations that were randomized in a column on the right. To be included in the study, participants had to score perfectly (i.e., no errors) on this task. This decision was made to ensure that longer latencies in the self-paced reading task were not due to participants not knowing the meaning of the experimental words.

Grammaticality judgement task This task was administered with the aim of ensuring that the L2 learner groups had the knowledge of the gender agreement operation in the L2 both in local domains and across constituents. Learners classified sentences as correct or incorrect, and identified the source of the

error in the case of incorrect sentences by highlighting the incorrect word(s). A total of 32 sentences were generated following four conditions: structural distance with gender concord, structural distance with gender discord, local agreement with gender concord, and local agreement with gender discord. The sentences were similar in structure to the ones contained in the self-paced reading task. To be included in the study, L2 learners had to correctly identify as correct or incorrect at least six out of eight (75%) sentences in each condition.

Gender assignment task Finally, to measure L2 learners' knowledge of the grammatical gender of the experimental items, participants completed a gender assignment task. They were given a list of all the critical nouns present in the self-paced reading task and chose the correct determiner from the options *el* ('the-masc-sing') and *la* ('the-fem-sing'). Participants had to achieve a minimum of 80% to be included in the study.

Language processing and WM tasks

Self-paced reading task

Participants performed a non-cumulative self-paced reading task. Participants sat in front of a computer screen and read sentences in Spanish at their own pace, and they responded to yes/no comprehension questions about the sentences read. The task was designed using the E-Prime 3.0 software. Each trial began with a fixation cross in the center of the screen. Upon pressing the spacebar, a series of dashes in place of the letters for each word in the sentence to be read appeared. Participants read the sentence verbatim, and to advance to the next word, they pressed the spacebar button. The software recorded the reading times (RTs) in ms from when a word was first displayed until the participant pressed the spacebar. After pressing the spacebar for the last word in each sentence, participants responded to a yes/no comprehension question, with no time limit. At the beginning of the task, participants were provided with three practice trials that consisted of three sentences with their corresponding yes/no questions.

Participants read 128 sentences. Half of the total number of sentences (64 sentences) were experimental, while the rest served as fillers and had a different syntactic structure. The 64 experimental sentences followed four conditions. Half of the experimental items (32 sentences) had a noun-adjective gender agreement operation within the direct object (1 and 2), whereas the other half of the experimental sentences had the same agreement operation but between the head of the direct object and a secondary predicate (3 and 4). Half of the sentences for each of the syntactic conditions mentioned above were engaged in gender agreement (1 and 3), while the other half were engaged in gender agreement violations (2 and 4). Thus, participants saw a total of sixteen sentences per condition.

(1) La madre vendió su cuadro caro en la feria.

The mother sold her painting$_{masc}$ *expensive*$_{masc}$ at the fair
"The mother sold her expensive painting at the fair."

(2) *La madre vendió su cuadro *cara en la feria.*

The mother sold her painting$_{masc}$ **expensive*$_{fem}$ at the fair
"The mother sold her expensive painting at the fair."

(3) *La madre vendió caro su cuadro en la feria.*

The mother sold expensive$_{masc}$ her painting$_{masc}$ at the fair
"The mother sold her expensive painting at the fair."

(4) La madre vendió *cara su cuadro en la feria.

The mother sold *expensive$_{fem}$ her painting$_{masc}$ at the fair
"The mother sold her expensive painting at the fair."

All experimental and filler sentences were nine words long. Experimental sentences with local agreement followed the same syntactic structure: DP (e.g., *La madre* in 1), regular, transitive verb in the past tense and perfective aspect (e.g., *vendió* in 1), DP (e.g., *su cuadro caro* in 1), and finally a PP (e.g., *en la feria* in 1). Sentences with structural distance had a similar structure, but the adjective moved from the direct object to a secondary predicate between the verb and the direct object and, therefore, generated a new constituent (e.g., *caro* in 3). To control for gender markedness, half of the sentences had a feminine

noun, whereas the other half had a masculine noun (see Appendix for the complete list of sentences). Given that RTs are modulated by syllable length (i.e., the longer the word to read, the longer the RTs), all nouns, verbs, adjectives, and the determiners and prepositions at the other critical regions (i.e., N + 1 and N + 2 positions) were disyllabic. Moreover, determiners in the direct object were ungendered (e.g., *su*), as pre-activation of phi features can potentially inhibit gender discord effects. Experimental nouns across conditions were also matched for word frequency. A t-test for independent samples showed that the word frequency of experimental nouns did not differ significantly across conditions (structural distance and local agreement), $t(30) = 0.26$, $p = 0.79$.

Items were split into four lists, so that no participants would see more than one version of an experimental item. Within lists, sentences were pseudorandomized to avoid having two sentences of the same condition appear successively. The computer recorded the RTs for each word in the sentence and the accuracy for the comprehension questions. For the latter dependent variable, one point was assigned to correct answers and zero points to incorrect ones. Statistical analyses were performed only on sentences with correct responses to comprehension questions.

Verbal working memory updating task

Participants took an adapted version of the Math span task, a language-unrelated tests that measures verbal WM updating capacity (Shahnazari-Dorcheh & Roshan, 2012). Each group performed this task in their L1 (English for the L2 speaker group and Spanish for the monolingual group), as verbal memory spans can be confounded with L2 proficiency (e.g., Juffs & Harrington, 2011). Participants saw a series of basic calculations (e.g., 6–2 = ?) on a computer screen, and they said the result of these operations aloud in their L1 and remembered the second digit of each set of calculations (in the case of 6–2, 2 is the number to be remembered). The task started with two operations per set and went up to six calculations. At the end of each set, participants recalled the second digits in order. There was a total of 15 sets. The time between operations was 2.5 s. Participants were given 1 point per correct answer in the processing section and 1 point per each second digit recalled in the correct order

in the storage section. The task specifically measures WM updating because participants need to constantly update the type of operation in question (i.e., subtraction or addition) along with the numbers in each operation. Finally, math ability effects cannot be discarded, but it is important to consider that all operations were subtraction and addition with one-digit numbers, and that all participants were college students. The combination of these two factors reduces potential math ability effects to a minimum.

Results

Descriptive statistics and data modeling for inferential statistics

Table 1 shows the mean accuracy scores for the comprehension questions. Both RT and accuracy data were analyzed through mixed-effects linear and logistic regressions in R (R Core Team, 2020). The models were fitted using the *lme4* package (Bates et al., 2015). After fitting, the estimated marginal means (EMM), contrasts and p-values were calculated using the package *emmeans* (Lenth, 2020). All contrasts were adjusted for multiple comparisons using the Tukey's HSD test, as implemented in the latter package. All models were run with the maximal random effect structure as recommended in Barr et al. (2013). The models were then incrementally simplified until they converged or were not a singular fit. The specific random structure for the models is presented with each specific model. All models were compared against a null model (a variance component model with the same random structure as the model of interest). Alpha was set at 0.05 for all analyses. Only models that were significantly different from the null model were analyzed.

Reading times

First, all sentences whose comprehension questions resulted in incorrect responses were not included in the statistical analyses, which affected 8.65% of the dataset (4.75% for Spanish monolinguals, 6.56% for advanced L2 learners, and 14.04% for beginners). After this procedure, RTs faster than 200 ms and slower than 3000 ms were discarded. This process affected 0.01% of the previously trimmed dataset for

Table 1 Mean accuracy for the comprehension questions by group (standard deviation in parentheses)

Agreement	Local domains		Structural distance	
	Agreement	Disagreement	Agreement	Disagreement
Beginners	0.88 (0.02)	0.84 (0.03)	0.92 (0.01)	0.88 (0.02)
Advanced	0.93 (0.02)	0.93 (0.02)	0.96 (0.01)	0.97 (0.01)
Natives	0.96 (0.01)	0.95 (0.01)	0.98 (0.005)	0.98 (0.008)

Spanish monolinguals, 0.02% for advanced L2 learners and 0.02% for beginners. RTs for the N–1 position (the word prior to the critical item), the N position (the critical word), the N + 1 position (the word immediately after the critical item) and the N + 2 position (two items after the critical word) were analyzed for each group via a linear mixed-effects model with agreement (agreement and violation) and syntactic relation (structural distance and within-constituent agreement), and all the interactions as fixed effects. It is important to highlight that across conditions, the positions mentioned in the sentence entailed different word categories; however, the N position consistently marked the conclusion of the agreement operation. For instance, in the local domain condition, the N position represented the adjective (e.g., *fría* in *tomó su sopa fría*), whereas in the structural distance condition, it corresponded to the noun (e.g., *sopa* in *tomó fría su sopa*). In both scenarios, irrespective of the word category (adjective or noun), the N position signaled whether the relationship was one of agreement or disagreement.

Monolinguals

For the N-1 position, the model had random intercepts by participant and item and random slopes of syntactic structure by participant and item, and it was significantly different from the null model, $\chi^2(7) = 16.44$, $p = 0.021$. There was no main effect of agreement, $t(1502) = 0.22$, $p = 0.82$, although there was a main effect of syntactic relation, $t(10) = 3.06$, $p = 0.01$ (Estimate = 3.21, SE = 14.46), as this group took significantly more time to process the structural distance condition compared to the local agreement condition. This is not surprising, as the word for the structural distance condition had two syllables, while the word for the local agreement condition only had one. To further explore whether this difference was due to syntactic structure *per* se or to word length, a

paired t-test comparing the sum of the RTs for the N-1 and N-2 positions between the two conditions (e.g., structural distance and local agreement) was run, as the two conditions presented the same number of syllables (i.e., three) considering the combination of N-1 and N-2 positions. There were no significant differences in RTs between the conditions, $t(25) = 0.10$, $p = 0.92$, which suggests that the significant difference in RTs between the two conditions at the N-1 position was due exclusively to word length.

The model for the N position that successfully converged had random intercepts for participant and item and no random slopes; it was significantly different from the null model, $\chi^2(7) = 60.11$, $p < 0.01$. While monolinguals were significantly slower in the violation condition than in the agreement condition, $t(1525) = 4.26$, $p < 0.001$, there were no main effects of syntactic relation (Estimate = 24.281, SE = 15.99; $t(1526) = 1.51$; $p = 0.12$). Pairwise comparisons averaging over the factor of syntactic relation, however, revealed that monolinguals were significantly slower, $t(1525) = 3.02$, $p = 0.003$, while reading violations (EMM = 621.59, SE = 48.06) than agreement (EMM = 552.52, SE = 48.09) in the control condition as well as in the structural distance condition (*violation*: EMM = 644.82, SE = 47.95; *agreement*: EMM = 577.86, SE = 47.90; $t(1525) = 3.00$, $p = 0.003$).

With respect to the N + 1 position, the model that successfully converged had random intercepts for participant and items and random slopes of agreement by participant; it was significantly different from the null model, $\chi^2(7) = 21.45$, $p < 0.01$. There was a main effect of agreement, $t(23.63) = 2.47$, $p = 0.021$, whereby monolinguals were significantly slower in the violation condition (EMM = 534.9, SE = 27.1) than in the agreement condition (EMM = 501.6, SE = 30.3) when averaging over syntactic relation. Posthoc comparisons revealed that, for the control syntactic condition, monolinguals took significantly

more time to read violations, $t(76) = 2.82$, $p < 0.001$ (EMM = 553.20, SE = 28.45) than agreement relations (EMM = 502.06, SE = 31.55). However, there were no significant differences between violation and agreement for the structural distance condition (*violation*: EMM = 516.61, SE = 28.34; *agreement*: EMM = 501.28, SE = 31.42; $t(70) = 0.86$, $p = 0.39$).

Finally, the model that successfully converged for the N + 2 position included random intercepts by participant and item and was significantly different from the null model, $\chi^2(7) = 23.67$, $p < 0.01$. The model revealed no significant main effect of agreement, $t(1537) = 0.90$, $p = 0.36$, or syntactic relation, $t(1539) = 0.70$, $p = 0.48$. Pairwise comparisons revealed no significant differences between violation and agreement neither for the structural distance condition (*violation*: EMM = 455.19, SE = 26.40; *agreement*: EMM = 461.91, SE = 26.39; $t(1537) = 0.49$, $p = 0.62$) nor for the control condition (*violation*: EMM = 477.52, SE = 26.50; *agreement*: EMM = 453.30, SE = 26.49; $t(1537) = 1.73$, $p = 0.08$).

Advanced L2 learners

The model for the N–1 position had random intercepts by participant and item and random slopes of syntactic structure by participant. There was no main effect of agreement, $t(1115) = 1.19$, $p = 0.23$ (Estimate = 18.24, SE = 15.21), or syntactic relation, $t(19) = 0.99$, $p = 0.33$ (Estimate = 26.38, SE = 26.38).

For the N position, the model for advanced L2 learners had random intercepts by participant and item. There were no main effects of agreement (Estimate = 4.53, SE = 15.68, $t(1142) = 0.28$, $p = 0.77$) or syntactic relation (Estimate = 8.47, SE = 15.69, $t(1145) = 0.54$, $p = 0.59$). Post-hoc comparisons averaging over the syntactic relation factor showed that advanced learners were not significantly slower at reading violations than agreement relations neither for the control condition (*violation*: EMM = 516.27, SE = 38.00; *agreement*: EMM = 516.51, SE = 37.91; $t(1142) = 0.011$, $p = 0.99$) nor for the structural distance condition (*violation*: EMM = 503.50, SE = 37.80; *agreement*: EMM = 512.34, SE = 37.81; $t(1142) = 0.40$, $p = 0.68$).

For the N + 1 position, a model with random intercepts for participant and item, and random slopes

of agreement by participant was marginally significantly different from the null model, $\chi^2(7) = 27.02$, $p < 0.01$. The model revealed no significant main effect of agreement (Estimate = 20.40, SE = 14.57; $t(19) = 1.40$, $p = 0.17$). Post-hoc comparisons showed that advanced learners were not significantly slower at reading violations than agreement relations neither for the control syntactic condition (*violation*: EMM = 462.51, SE = 23.93; *agreement*: EMM = 445.59, SE = 23.93; $t(51) = 0.90$, $p = 0.36$) nor for the structural distance condition (*violation*: EMM = 429.47, SE = 23.73; *agreement*: EMM = 405.58, SE = 20.16; $t(47) = 1.30$, $p = 0.19$).

Finally, the model that successfully converged for the N + 2 position had random intercepts by participant and item; it was significantly different from the null model, $\chi^2(7) = 28.82$, $p < 0.01$. While advanced learners were significantly slower in the violation condition than in the agreement condition overall, $t(1145) = 3.17$, $p < 0.001$, there were no main effect of syntactic relation (Estimate = 11.58, SE = 7.20; $t(1148) = 1.60$; $p = 0.10$). Pairwise comparisons averaging over the factor of syntactic relation, however, revealed that the advanced learner group was significantly slower while reading violations (EMM = 407.26, SE = 17.25) than agreement (EMM = 375.45, SE = 17.22) in the control condition, $t(1145) = 3.08$, $p = 0.002$, but there was no significant difference for the structural distance condition (*violation*: EMM = 386.73, SE = 17.13; *agreement*: EMM = 372.82, SE = 17.15; $t(1144) = 1.38$, $p = 0.16$).

Beginning L2 learners

The model for the N-1 position had random intercepts by participant and item and random slopes of syntactic structure by participant. There was no main effect of agreement, $t(1115) = 1.19$, $p = 0.23$, or syntactic relation, $t(19) = 0.99$, $p = 0.33$ (Estimate = 3.21, SE = 14.46).

The model for beginning L2 learners, for the N position, had a random effect structure of random intercepts by participants and items. No significant main effects of agreement (Estimate = 6.82, SE = 21.49; $t(1401) = 0.31$, $p = 0.751$) or syntactic relation (Estimate = 24.87, SE = 21.57, $t(1406) = 1.15$, $p = 0.25$) were found.

Regarding the N + 1 position, the model had random intercepts for participants and items. There was no significant main effect of agreement (Estimate = 11.33, SE = 13.80; $t(1420) = 0.82$, $p = 0.41$) or syntactic relation (Estimate = 6.74.87, SE = 13.81, $t(1413) = 0.48$, $p = 0.62$). Posthoc tests revealed that beginners were not significantly slower at reading violations than agreement relations neither for the control syntactic condition (*violation*: EMM = 483.25, SE = 21.59; *agreement:* EMM = 497.98, SE = 21.44; $t(1420) = 0.74$, $p = 0.45$) nor for the structural distance condition (*violation*: EMM = 479.90, SE = 21.41; *agreement:* EMM = 487.84, SE = 21.26; $t(1421) = 0.41$, $p = 0.68$).

Finally, the model that successfully converged for the N + 2 position had random intercepts by participant and item. The model revealed no significant main effect of agreement, $t(1412) = 0.81$, $p = 0.41$, or syntactic relation, $t(1416) = 1.05$, $p = 0.29$. Pairwise comparisons revealed no significant differences between violation and agreement neither for the structural distance condition (*violation*: EMM = 428.34, SE = 19.97; *agreement:* EMM = 406.96, SE = 19.85; $t(1413) = 1.37$, $p = 0.17$) nor for the control condition (*violation*: EMM = 404.31, SE = 20.08; *agreement:* EMM = 407.53, SE = 20.02; $t(1412) = 0.20$, $p = 0.83$). Figure 1 shows a summary of the raw RTs word by word for the different groups and conditions.

WM and gender agreement processing

Correlational analyses were conducted between WM scores and the magnitude of the grammaticality effect for each group and position to determine whether there was a relationship between sensitivity to gender discord and WM. To determine the magnitude of the grammaticality effect, the procedure used by Waters and Caplan (1996) and by Keating (2010) was used. RTs for the agreement conditions were subtracted from RTs for discord conditions and then divided by RTs on agreement conditions for each subject and locality conditions.

For the Spanish monolingual group, no correlations were significant (N position, within constituents: $r = 0.03$, $p = 0.912$; N position, structural distance: $r = 0.12$, $p = 0.551$; N + 1 position, within constituents: $r = 0.09$, $p = 0.845$; N + 1 position, structural distance: $r = 0.17$, $p = 0.233$; N + 2 position,

within constituents: $r = 0.18$, $p = 0.201$; N + 2 position, structural distance: $r = 0.04$, $p = 901$). For the advanced L2 learner group, there was a strong, positive correlation between WM and the magnitude of grammaticality effect for the structural distance condition at the N + 2 position ($r = 0.67$, $p < 0.01$), while correlations for the other regions were non-significant (N position, within constituents: $r = 0.08$, $p = 0.74$; N position, structural distance: $r = 0.11$, $p = 0.454$; N + 1 position, within constituents: $r = 0.17$, $p = 0.184$; N + 1 position, structural distance: $r = 0.21$, $p = 0.114$; N + 2 position, within constituents: $r = 0.11$, $p = 0.498$) For their beginner peers, no correlations proved significant (N position, within constituents: $r = 0.11$, $p = 0.611$; N position, structural distance: $r = 0.10$, $p = 0.504$; N + 1 position, within constituents: $r = 0.13$, $p = 0.638$; N + 1 position, structural distance: $r = 0.06$, $p = 0.901$; N + 2 position, within constituents: $r = 0.14$, $p = 0.431$; N + 2 position, structural distance: $r = 0.20$, $p = 0.119$).

Comprehension accuracy

Accuracy was coded as a binary factor, with correct responses coded as 1 and incorrect ones as 0. Logistic mixed-effect regression models were fit to analyze accuracy data for the three groups. Monolinguals showed an overall accuracy of 95%, while advanced learners had an accuracy of 93%, and their beginner peers responded to comprehension questions accurately 86% of the times.

Summary of results

(a) Spanish monolinguals were sensitive to gender disagreement relations both in local domains and across constituents at the N position. At the poststimulus position, they lost sensitivity to structural distance, although they showed sensitivity to gender discord in local domains.

(b) Advanced L2 learners of Spanish, on the other hand, did not take significantly more time to process gender discord than gender concord within or across constituents at the critical word or poststimulus position. However, they showed delayed sensitivity to gender discord in local domains (N + 2 position). Additionally, those
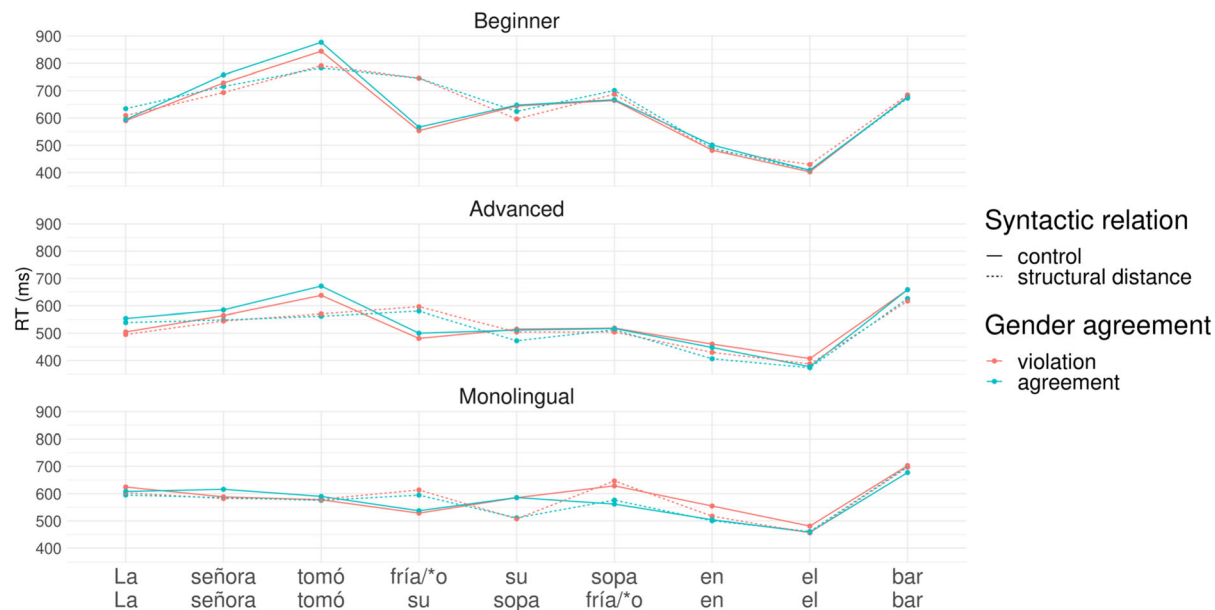
Fig. 1 Raw RTs word by word for the three groups and four conditions

advanced learners with higher WM were also sensitive to gender disagreement in structural distance conditions.

(c) Beginner L2 learners of Spanish did not show sensitivity to gender disagreement at any position.

(d) As for overall comprehension accuracy, all groups had very high overall scores.

Discussion

The study investigated the effects of structural distance, WM, and L2 proficiency on the processing of noun-adjective gender agreement by beginner and advanced English L2 learners of Spanish and Spanish monolinguals. The first research question pertained to whether L2 learners process L2 syntax qualitatively quite similar to how Spanish native speakers process L1 syntax. This question was assessed through the inclusion of gender agreement relations in local domains and across phrases. The second research question was whether WM modulates L1 and L2 syntactic processing. In the view of the fact that the results for each research question complement each other, these will be discussed together in the following paragraphs.

My predictions were threefold: (a) both monolinguals and L2 learners would have more trouble processing structural distance compared to agreement in local domains, (b) within the L2 learner group, only advanced learners would effectively process agreement operations within and across phrases, and (c) a greater WM capacity would facilitate syntactic processing in L2 learners, but not in native speakers. The predictions were partially supported by online data. In line with the interpretation of data by previous self-paced reading studies, I assumed that significantly longer RTs in the N, N + 1 or N + 2 positions for the gender disagreement conditions compared to the gender agreement conditions would indicate that a given group paid attention to morphological cues in the nouns and adjectives in question, thus being sensitive to gender agreement violations.

Spanish monolinguals showed early sensitivity (at the N position) to gender agreement violations in both local domains and structural distance conditions, although sensitivity for the latter condition faded away rapidly. Beginners, on the other hand, were not sensitive to either condition, whereas their advanced learner peers showed delayed sensitivity (at the N + 2 position) to violations in local domains only. In addition, advanced learners' verbal WM updating spans were associated with sensitivity to violations in the syntactically distant condition as well. Taken

together, the findings suggest that native processing relies on grammatical cues to process hierarchical relations across words such as agreement operations even in syntactically complex conditions. On the other hand, beginner learners did not seem to focus on morphosyntactic cues, or at least those absent in their L1 (i.e., gender cues). If we had included only this proficiency group in the study, we might have arrived at a misleading conclusion: L1 processing relies on grammatical information, while L2 processing relies on lexical cues to process incoming linguistic information, which would agree with the Shallow Structure Hypothesis (Clahsen & Felser, 2006a, 2006b, 2018). Nevertheless, crucial data come from the advanced learner group. Highly proficient learners did process hierarchical cues in the L2, as evidenced by sensitivity to gender agreement errors in local domains. This reveals that proficiency increases attention paid to morphological cues during L2 computation, thus resembling L1 processing patterns. Moreover, advanced learners with higher WM were also able to focus on gender cues to disentangle complex syntactic relations. These findings go against the predictions put forward by the Shallow Structure Hypothesis and, in turn, agree with accessibility computational accounts, such as the Fundamental Identity Hypothesis, claiming that proficiency and availability of cognitive resources permit the reach of native-like processing in late bilinguals (Hopp, 2007, 2010; McDonald, 2006). In other words, the results suggest that differences exhibited between L1 and L2 processing are quantitative in nature, rather than qualitative.

Surprisingly, structural distance effects emerged both for L2 learners and native speakers: L2 learners with high proficiency were not sensitive to agreement violations in the structural distance condition overall, while Spanish monolinguals detected gender agreement violations between phrases but their sensitivity to these errors faded away in the following word (N + 1 position). This rapid decay in sensitivity suggests that establishing syntactic operations outside of local domains might be taxing also in the L1. The said difficulty in both groups may not obey to the rate of occurrence of type of agreement (i.e., agreement within or across phrases), as both operations are exceptionally frequent in the Spanish language. Two possible explanations may account for structural distance effects. From a computational perspective, speakers may need to finish processing an entire constituent and its inner operations before passing onto a new constituent. Computational advantages for agreement in local domains would emerge from these operations not exceeding the scope of their locality. Also, speakers may identify morphosyntactic cues upon reading them, and they may need to put them on hold until they complete the processing of the phrase in which they are embedded. Under this view, operations across constituents would be harder to process because speakers may need to wait until they are able to mark the extent of the constituent in which the second agreeing cue is located, even in absence of linear distance. From a generative perspective, the stimuli in all studies investigating structural distance, whether copulas (Alemán Bañón et al., 2014; Dowens et al., 2010) or secondary predicates (the present study), entail a specific type of raising structures, namely small clauses. These constructions raise either to the specifier of the sentence (in copulas) or to an empty position in the VP (in secondary predicates). Under this view, structural distance effects may be due to small clauses raising to previously empty positions in the syntactic tree.

The findings of the present study agree with studies showing that L2 proficiency facilitates the processing of distance by late bilinguals (Coughlin & Tremblay, 2013; Keating, 2010; Reichle & Coughlin, 2013), particularly in the case of structural distance (Alemán Bañón et al., 2014; Dowens et al., 2010). In addition, the present findings complement the latter type of studies by assessing the role of WM in the processing of structural distance, and suggesting that WM updating intervenes in the computation of morphosyntax across constituents in the L2 only in proficient learners. Overall, the relationship between cognitive domains and hierarchical relations seems to be more clear: only proficient learners with high WM seem to reach native-like computation of distance, whether it is physical or structural in nature. The present study also agrees with previous morphosyntactic processing studies in that WM does not facilitate local morphosyntactic error detection in beginners (Sagarra, 2007) but in proficient learners (Arnold, 2019; Coughlin & Tremblay, 2013; Keating, 2010; Reichle & Coughlin, 2013; Sagarra & Herschensohn, 2010; Sagarra & Labrozzi, 2018). The question of why WM modulates online processing in advanced learners and not in the early stages of L2 acquisition remains unclear and should be thoroughly addressed in future

research. A tentative hypothesis could be put forward, however. The early stages of L2 acquisition entail the consumption of many attentional resources (Hasegawa et al., 2002), and their depletion would not allow beginners to focus on detailed morphosyntactic information. As proficiency increases, a larger portion of cognitive resources is available for learners. While processing incoming information in the L2, advanced learners may thus make use of a more available executive function that allows them to focus on relevant information, like morphosyntactic cues and word order to disentangle hierarchical relations. Thus, the findings from the present investigation support L2 processing theories positing that it is possible for L2 learners to attain native-like processing of L2 syntax, and that this reach is mediated by the availability of cognitive resources such as WM and linguistic experience such as L2 proficiency (Hopp, 2007, 2010; McDonald, 2006).

On the other hand, this research contradicts experimental studies that did not find a WM facilitation effect in morphosyntactic processing in local domains by proficient late bilinguals (Felser & Roberts, 2007; Foote, 2011; Juffs, 2004, 2005, 2006). A number of simple explanations for these null results can be listed: WM spans were computed through their storage component only (Felser & Roberts, 2007; Foote, 2011), the WM task were performed in the participants' non-dominant language (Foote, 2011), the linguistic task and the WM task did not share the same modality (i.e., visual or auditory; Felser & Roberts, 2007), or WM was taken as a nominal variable (Juffs, 2004, 2005, 2006). Finally, the findings do not support deficit computational accounts positing that "representations adult L2 learners compute for comprehension are shallower and less detailed than those of native speakers … and rely more on non-structural information in parsing" (Clahsen & Felser, 2006a, p. 3). Instead, proficient L2 learners seem to make full use morphosyntactic information during real-time processing to compute complex hierarchical structures in the L2.

The current study had certain limitations that may have influenced the results. First, the chosen syntactic structure (i.e., secondary predicate modifying a direct object) is not as frequent as other syntactic patterns in Spanish. All participants whose data have been included in the statistical analysis showed that they were familiar with the said Spanish structure, but it is possible that the processing of the structural distance condition may have been hindered by the fact that it is not as frequent as the local condition. Second, the WM test used does not offer reaction time data, which is usually used for two purposes: (a) discarding responses that took too long, as they might not tap directly into WM, and (b) having alternate cognitive processing data across participants. However, the test is useful in L2 studies because it is a language-independent cognitive task, unlike most WM tests previously used in the L2 literature (e.g., reading span task), and has been used in multiple L2 processing studies (e.g., Durand López, 2021; Mavrou & Bustos-López, 2019).

Conclusion

This study investigated whether the processing of structural distance is more taxing than the processing of agreement in local domains for English L2 learners of Spanish of varying proficiency levels, and whether verbal WM updating capacity intervened in the processing of structural distance without linear distance. Based on self-paced reading data, Spanish monolinguals were sensitive to gender agreement violations both in local domains and across phrases, beginners were insensitive to all violations, advanced learners were sensitive to within-phrase errors, and those with higher WM updating spans also detected gender agreement violations in syntactic distant conditions. These findings support computational models within accessibility accounts, such as the Fundamental Identity Hypothesis (Hopp, 2007, 2010; McDonald, 2006), as they indicate that linguistic (i.e., L2 proficiency) and cognitive (i.e., WM updating capacity) aspects allow late bilinguals to attain native-like processing of the L2. Thus, L1 and L2 processing seem to be qualitatively similar, albeit quantitatively different.

Author contributions I am the sole author of this article.

Declarations

Conflict of interest The author declares that they have no conflict of interest.

Ethical approval All procedures performed in this study had been approved by the Office of Research and Regulatory Affairs prior to data collection.

Informed consent All participants read and signed a written informed consent before participating in the study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Appendix

List of sentences by condition

Agreement within phrases	Structural distance
1. La joven tomó su caldo frío/*a en el restaurante	1. La joven tomó frío/*a su caldo en el restaurante
2. El marido dejó su cuarto sucio/*a por la pereza	2. El marido dejó sucio/*a su cuarto por la pereza
3. El hombre vendió su piso caro/*a a su prima	3. El hombre vendió caro/*a su piso a su prima
4. El chico tiene su cuadro listo/*a en la habitación	4. El chico tiene listo/*a su cuadro en la habitación
5. La chica dejó su vaso lleno/*a en la mesa	5. La chica dejó lleno/*a su vaso en la mesa
6. La señora llevó su carro roto/*a a la reunión	6. La señora llevó roto/*a su carro a la reunión
7. El señor hizo su voto corto/*a de la mañana	7. El señor hizo corto/*a su voto de la mañana
8. La mujer compró su gorro caro/*a en la tienda	8. La mujer compró caro/*a su gorro en la tienda
9. La señora tomó su sopa fría/*o en el bar	9. La señora tomó fría/*o su sopa en el bar
10. El marido dejó su casa sucia/*o por la pereza	10. El marido dejó sucia/*o su casa por la pereza
11. El hombre vendió su cama cara/*o a su prima	11. El hombre vendió cara/*o su cama a su prima
12. El chico tiene su fiesta lista/*o en la playa	12. El chico tiene lista/*o su fiesta en la playa
13. La chica dejó su cesta llena/*o en la mesa	13. La chica dejó llena/*o su cesta en la mesa
14. La señora llevó su falda rota/*o a la reunión	14. La señora llevó rota/*o su falda a la reunión
15. El señor hizo su charla corta/*o de la mañana	15. El señor hizo corta/*o su charla de la mañana
16. La mujer compró su prenda cara/*o en la tienda	16. La mujer compró cara/*o su prenda en la tienda

References

Alemán Bañón, J., Fiorentino, R., & Gabriele, A. (2014). Morphosyntactic processing in advanced second language (L2) learners: An event-related potential investigation of the effects of L1–L2 similarity and structural distance. *Second Language Research, 30*(3), 275–306.

Arnold, E. W. (2019). *Using Eye Tracking to Examine Working Memory and Verbal Feature Processing in Spanish* (Publication No. 7472) [Master's thesis, Brigham Young University]. Scholars Archive.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417–423.

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

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *The Psychology of Learning and Motivation* (pp. 47–89). Academic Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Chai, W. J., Abd Hamid, A. I., & Abdullah, J. M. (2018). Working memory from the psychological and neurosciences perspectives: A review. *Frontiers in Psychology, 9*, 401.

Clahsen, H., & Felser, C. (2006a). Grammatical processing in language learners. *Applied Psycholinguistics, 27*(1), 3–42.

Clahsen, H., & Felser, C. (2006b). How native-like is non-native language processing? *Trends in Cognitive Sciences, 10*(12), 564–570.

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

Conway, A. R. A., Jarrold, C., Kane, M. J., Miyake, A., & Towse, J. N. (Eds.). (2007). *Variation in working memory*. Oxford: New York, NY.

Coughlin, C. E., & Tremblay, A. (2013). Proficiency and working memory based explanations for nonnative speakers' sensitivity to agreement in sentence processing. *Applied Psycholinguistics, 34*(3), 615–646.

Cowan, N. (1998). *Attention and memory: An integrated framework*. Oxford University Press.

Cowan, N. (2016). *Working memory capacity* (Classic). Routledge.

Dowens, M. G., Vergara, M., Barber, H. A., & Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience, 22*(8), 1870–1887.

Durand López, E. M. (2021). Morphological processing and individual frequency effects in L1 and L2 Spanish. *Lingua, 257*, 103093.

Fay, A., & Buchweitz, A. (2014). Listening comprehension and individual differences in working memory capacity in beginning L2 learners. *Letrônica, 7*(1), 113–129.

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

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

Guará-Tavares, M. D. G. (2013). Working memory capacity and L2 speech performance in planned and spontaneous conditions: A correlational analysis. *Trabalhos Em Linguística Aplicada, 52*(1), 9–29.

Hasegawa, M., Carpenter, P. A., & Just, M. A. (2002). An fMRI study of bilingual sentence comprehension and workload. *NeuroImage, 15*(3), 647–660.

Hopp, H. (2007). *Ultimate attainment at the interfaces in second language acquisition*. Unpublished Ph.D. Dissertation. Groningen University.

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between nonnative and native speakers. *Lingua, 120*(4), 901–931.

Juffs, A. (2006). Processing reduced relative vs. main verb ambiguity in English as a Second Language: A replication study with working memory. In R. Slabakova, S. Montrul & P. Prevost (Eds.), *Inquiries in linguistic development: In honor of Lydia White* (pp. 213–232). John Benjamins.

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.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122.

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. *Research in Second Language Processing and Parsing, 53*, 113–134.

Lenth, R. (2020). Response-Surface Analysis [R package rsm version 2.10. 3].

Mavrou, I. (2018). Memoria operativa y complejidad cognitiva en tareas de expresión escrita en español como lengua extranjera. *Onomázein Revista De Lingüística, Filología y Traducción De La Pontificia Universidad Católica De Chile, 41*, 125–146.

Mavrou, I., & Bustos-López, F. (2019). The influence of working memory and emotional intelligence on emotional expression in Spanish as a second language. *Revista De Lingüística y Lenguas Aplicadas, 14*, 117–129.

McDonald, J. (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.

Miki, S. (2012). Working memory as a factor affecting L2 listening comprehension sub-skills. *Kumamoto University Departmental Bulletin Paper, 10*, 119–128.

Montrul, S., & Slabakova, R. (2003). Competence similarities between native and near-native speakers: An investigation of the preterite-imperfect contrast in Spanish. *Studies in Second Language Acquisition, 25*(3), 351–398.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reichle, R. V., Tremblay, A., & Coughlin, C. E. (2013). Working-memory capacity effects in the processing of non-adjacent subject-verb agreement: An event-related brain potentials study. In E. Voss, S. J. D. Tai & Z. Li (Eds.), *Selected Proceedings of the 2011 Second Language Research Forum: Converging theory and practice*. Somerville, MA: Cascadilla Proceedings Project.

Rızaoğlu, F., & Gürel, A. (2020). Second language processing of English past tense morphology: The role of working memory. International Review of Applied Linguistics in Language Teaching, 1–29.

Sagarra, N. (2007). Online processing of gender agreement in low proficient English-Spanish late bilinguals. In M. J. Cabrera, J. Camacho, V. Deprez, N. Flores, & L. Sánchez (Eds.), *Current Issues in Linguistic Theory Series* (pp. 240–253). John Benjamins.

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.

Sagarra, N., & LaBrozzi, R. (2018). Benefits of study abroad and working memory on L2 morphosyntactic processing. In C. Sanz & A. Morales-Front (Eds.), *The Routledge Handbook*

of Study Abroad Research and Practice (pp. 149–164). Routledge.

Shahnazari-Dorcheh, M., & Roshan, S. (2012). Developing a non-language related span test for the use in language-specific and cross-language studies. *English Language Teaching, 5*(11), 104–110.

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory, 17*(6), 635–654.

Vejnovic, D., Milin, P., & Zdravkovic, S. (2010). Effects of proficiency and age of language acquisition on working memory performance in bilinguals. *Psihologija, 43*(3), 219–232.

Walter, C. (2004). Transfer of reading comprehension skills to L2 is linked to mental representations of text and to L2 working memory. *Applied linguistics, 25*(3), 315–339.

Waters, G. S., & Caplan, D. (1996). Processing resource capacity and the comprehension of garden path sentences. *Memory and Cognition, 24*(3), 342–355.