

Please share your stories about how Open Access to this article benefits you.

Proficiency and working memory based
explanations for nonnative speakers'
sensitivity to agreement in sentence
processing

by C. Coughlin and A. Tremblay

2013

This is the published version of the article, made available with the permission of the publisher. The original published version can be found at the link below.

Caitlin E. Coughlin, Annie Tremblay. (2013). Proficiency and working memory based explanations for nonnative speakers' sensitivity to agreement in sentence processing. *Applied Psycholinguistics* 34(3):615-646.

Published version: <http://dx.doi.org/10.1017/S0142716411000890>

Terms of Use: <http://www2.ku.edu/~scholar/docs/license.shtml>

Proficiency and working memory based explanations for nonnative speakers' sensitivity to agreement in sentence processing

CAITLIN E. COUGHLIN and ANNIE TREMBLAY
University of Illinois at Urbana–Champaign

Received: August 14, 2010 Accepted for publication: June 16, 2011

ADDRESS FOR CORRESPONDENCE

Annie Tremblay, University of Illinois at Urbana–Champaign, 2090 Foreign Language Building, 707 S. Mathews Avenue, Urbana, IL 61801. E-mail: atrembla@illinois.edu

ABSTRACT

This study examines the roles of proficiency and working memory (WM) capacity in second-/foreign-language (L2) learners' processing of agreement morphology. It investigates the processing of grammatical and ungrammatical short- and long-distance number agreement dependencies by native English speakers at two proficiencies in French, and the relationship between their proficiency and WM capacity in French and their sensitivity to agreement violations. Native English speakers at mid- and high proficiencies in French and native French speakers completed an acceptability judgment task, a self-paced reading task, and a WM task in French, and the English speakers also completed a WM task in English. The results showed that whereas all participants performed at ceiling on the acceptability judgment tasks, only the high-level L2 learners and native speakers showed some sensitivity to number agreement violations. For L2 learners, this sensitivity did not vary as a function of the length of the agreement dependency. The results also indicated that L2 learners tended to be more sensitive to agreement violations as their WM memory capacity in French increased. The implications of these results for theories of L2 morphological processing are discussed.

Agreement morphology such as plural marking (e.g., *the cats*) is known to be difficult for second-/foreign-language (L2) learners, who commonly make agreement errors in production (e.g., Lardiere, 2006; White, Valenzuela, & Kozłowska-MacGregor, 2004) and do not process agreement as efficiently as native speakers do (e.g., Jiang, 2004, 2007; Keating, 2009). Even at high proficiencies, attainment in this domain is highly variable, especially when L2 learning began after the onset of puberty (e.g., Birdsong & Molis, 2001; Johnson & Newport, 1989). This can be so even if the morphological feature to be learned is instantiated in the native language (e.g., Hopp, 2010). Focusing on the processing of agreement morphology, three types of accounts have sought to explain L2 learners' variable sensitivity to agreement: storage accounts (e.g., Ullman, 2001, 2004), which

propose that L2 learners fail to decompose morphologically complex words; structural accounts (e.g., Clahsen & Felser, 2006; Jiang, 2004, 2007; Keating, 2009; Sato & Felser, 2010), which claim that L2 learners lack or cannot deploy the structure necessary to process agreement dependencies; and computational accounts (e.g., Hopp, 2010; McDonald, 2000, 2006), which attribute L2 learners' difficulty with agreement dependencies to reduced processing efficiency due to L2 learners' limited decoding abilities and working memory (WM) capacity. In this study, we review each of these accounts, with particular attention to the role of WM capacity in L2 processing, and we assess their predictive power for understanding inter-L2-learner variability in the processing of agreement morphology. We will present the results of an (offline) acceptability judgment task and an (online) self-paced reading task examining the effects of proficiency and WM capacity on the processing of short- and long-distance agreement dependencies between object clitics and their antecedents in French. We will then discuss the implications of our findings for theories of L2 morphological processing.

THEORIES OF L2 MORPHOLOGICAL PROCESSING

Ullman (2001, 2004; see also Pinker, 1999; Pinker & Ullman, 2002) proposed that native speakers' lexical and grammatical processing depend on two different memory systems: respectively, declarative memory and procedural memory. Declarative memory is used for the learning and knowledge of facts and events, and it can be recollected consciously; for language, it is the memory system that allows native speakers to associate meanings to word forms. In contrast, procedural memory is used for the learning and control of skills and habits, and it cannot be recollected consciously; for language, it is the memory system that enables native speakers to compute grammatical structures (e.g., phonology, morphology, syntax, and semantics). To explain L2 learners' poorer ability to compute grammatical structures (compared to native speakers) but their relatively intact ability to learn lexical words, Ullman (2001) proposed that L2 learners increasingly rely on declarative memory as their age of first exposure to the target language increases, and that their reliance on declarative memory should decrease as their exposure to and use of the target language increases (Ullman, 2004). This does not imply that multimorphemic words in the lexicon are not analyzed but that the lexical analyses are not procedural computations: they are instead derived from pattern recognition or from rules learned during pedagogical instruction. Thus, according to Ullman's model, the processing of grammatical structure, including agreement morphology, is qualitatively different for native and nonnative speakers, at least nonnative speakers at lower levels of proficiency.¹

One prediction that the declarative-procedural (DP) model makes is that L2 learners should process regularly and irregularly inflected words somewhat similarly. That is, they should store morphologically simple and complex words in declarative memory as single units, whether or not morphological, nonprocedural generalizations have taken place within the lexicon. In the psycholinguistic research on this topic, three main methods have been used to determine how L2 learners (and native speakers) process regularly and irregularly inflected words. The first one examines whether the processing of inflected words is subject to

frequency effects; because frequency effects have been associated with lexical storage (i.e., the more frequently a word is encountered, the stronger the memory traces), the DP model predicts that L2 learners should show token frequency effects for both regularly and irregularly inflected words, whereas native speakers should show such effects only for irregularly inflected words. The second method aims to determine whether regularly inflected words prime their stem; if inflected words and their stem have separate lexical entries, as would be predicted by the DP model, then L2 learners' processing of stems should not be primed by their inflected form, unlike that of native speakers. Finally, the third method examines the electrophysiological brain responses evoked by regular words that were irregularly inflected and irregular words that were regularly inflected; the DP model predicts that both word forms should elicit similar electrophysiological responses for L2 learners but different electrophysiological responses for native speakers.

Using these methods, the existing L2 research on this topic has revealed mixed findings. On the one hand, in a multiple-choice task where participants chose the correctly inflected verb form, Birdsong and Flege (2001) found that advanced Korean and Spanish L2 learners of English show stronger frequency effects for irregular verbs (e.g., *swam*) than for regular ones (e.g., *helped*). Similarly, in an event-related potential (ERP) study, Hahne, Mueller, and Clahsen (2006) report that both native German speakers and Russian L2 learners of German showed a P600 effect (typically associated with grammatical violations) for regular verbs that were incorrectly inflected (e.g., *getanzen* "danced"; correct form: *getanzt*) and an N400 effect (typically associated with lexical/semantic violations) for irregular verbs that were regularly inflected (e.g., *gelauft* "run"; correct form: *gelaufen*), consistent with a dual route for processing regularly and irregularly inflected words and unlike what the DP model would predict for L2 learners.

In contrast, in a masked priming study, Silva and Clahsen (2008) found that regularly inflected verbs (e.g., *prayed*) do not prime their stem (e.g., *pray*) for advanced Chinese, German, and Japanese L2 learners of English (unlike native speakers). In line with these findings, using lexical decision and masked priming experiments, Neubauer and Clahsen (2009) found that advanced Polish L2 learners of German showed frequency effects for both regularly (e.g., *gespielt* "played") and irregularly (e.g., *gelaufen* "run") inflected words, whereas native German speakers showed frequency effects only for irregularly inflected ones. Similarly, Bowden, Gelfand, Sanz, and Ullman (2010) found frequency effects for both regular (e.g., *pescó* "(I) fish") and irregular (e.g., *pienso* "I think") present-tense verbs in the production latencies of intermediate to advanced English L2 learners of Spanish, but not for regular verbs in the production latencies of native Spanish speakers. These results suggest that L2 learners store morphologically complex words as single units in declarative memory, consistent with the predictions of the DP model. One possibility, which is also consistent with the DP model, is that L2 learners rely more on declarative memory in earlier stages of development. The findings of Osterhout, McLaughlin, Pitkamen, Frenck-Mesitire, and Molinaro (2006) and Osterhout et al.'s (2008) ERP study are consistent with this hypothesis: for novice English L2 learners of French, morphological anomalies that yield an N400 effect shortly after the onset of instruction eventually yield a P600 effect after several months of instruction (for a proposed development, see Steinhauer, White, & Drury, 2009).

A different approach to L2 learners' processing of agreement morphology is Clahsen and Felser's (2006a) shallow structure hypothesis (SSH), according to which late L2 learners compute shallower and less detailed syntactic structures, and rely more on lexical, semantic, and pragmatic information than native speakers. SSH claims that the processing of grammatical structure is qualitatively different for native and nonnative speakers, regardless of proficiency in the target language. This hypothesis was proposed to explain reduced or absent structure-based effects in L2 learners' processing of temporary syntactic ambiguities and filler-gap dependencies (e.g., Felser & Roberts, 2007; Felser, Roberts, Marinis, & Gross, 2003; Marinis, Roberts, Felser, & Clahsen, 2005; Papadopoulou & Clahsen, 2003). Whereas Clahsen and Felser (2006a) originally claimed that highly proficient L2 learners may process agreement morphology in a similar manner as native speakers (unlike syntax), in more recent studies, Clahsen and colleagues suggested that L2 learners may be less affected by internal morphological structure than native speakers (for a discussion, see Clahsen, Felser, Neubauer, Sato, & Silva, 2010). Clahsen and Felser (2006b) also suggest that L2 learners may have difficulty computing agreement dependencies, with nativelike processing being restricted to local domains such as "morphosyntactic agreement between closely adjacent constituents" (p. 111). This then makes the prediction that although L2 learners may be able to decompose morphologically complex words, they should have difficulty computing agreement when it is not local.

Keating (2009) tested this prediction in an eye-tracking study on gender agreement with beginner, intermediate, and advanced English L2 learners of Spanish. He found that advanced L2 learners, but not beginner or intermediate L2 learners, were sensitive to gender agreement violations, and, unlike native speakers, they showed this sensitivity only when the agreement dependency was between two adjacent words (e.g., **una fiesta pequeño* "a small-masc party-fem"). Keating concluded that his results support SSH, but he admitted that linear and structural distances were confounded in his study. Using self-paced reading experiments, Sagarra (2007) and Sagarra and Herschensohn (2010) similarly found that intermediate English L2 learners of Spanish, but not beginner L2 learners, were sensitive to gender and number violations that occurred between two adjacent words (e.g., **el prototipo famosa* "the prototype-masc famous-fem"), whereas Jiang (2004, 2007) reported that, unlike native speakers, intermediate to advanced Chinese L2 learners of English were not sensitive to number agreement violations in sentences where the two words in the agreement dependency were not adjacent (e.g., **The bridge to the island were about ten miles away*). The distance between two words in an agreement dependency may thus contribute to L2 learners' lack of sensitivity to number agreement violations.

In contrast, Foote (2011) found that both advanced English L2 learners of Spanish and native Spanish speakers slowed down when parsing gender and number agreement violations, even when the two words in the agreement dependency were not adjacent (e.g., **El pollo del taco está rica pero picante* "The chicken-masc of the taco is tasty-fem but spicy"). Her findings perhaps revealed proficiency differences between her group of L2 learners and those in Keating (2009), suggesting that nativelike attainment is not impossible, even in longer agreement dependencies. Other behavioral studies have indeed shown that high-proficiency L2 learners

can perform similarly to native speakers in experiments that require them to use agreement morphology in sentence processing (e.g., Havik, Roberts, Van Hout, Schreuder, & Haverkort, 2009; Hopp, 2006, 2010), and recent ERP studies have even reported that very advanced L2 learners can detect agreement violations in a qualitatively similar way (i.e., eliciting a LAN effect and a P600 effect) as native speakers (Ojima, Nakata, & Kakigi, 2005; Rossi, Gugler, Friederici, & Hahne, 2006); the ease with which L2 learners detect such violations, however, is at least partially dependent on whether the native language has a similar agreement rule (e.g., Foucart & Frenck-Mestre, 2011; Tokowicz & MacWhinney, 2005).

The results of Jiang (2004) and Keating (2009) suggest that linear distance between two elements in an agreement dependency can potentially increase the difficulty with which L2 learners process such dependencies. Such a difficulty should arise if long-distance agreement dependencies impose a higher WM load on L2 learners than short-distance ones. We might thus expect a relationship between L2 learners' ability to process such dependencies and their WM capacity.

A third type of account that predicts such a relationship is McDonald's (2006) cognitive processing account. McDonald (2006) proposed that L2 learners' variable use of and sensitivity to grammatical structures is due to difficulties in basic cognitive processes that depend in part on grammatical knowledge (i.e., these processes should improve with increasing proficiency) but that are not grammar specific: "(1) low L2 [working-]memory capacity, (2) poor L2 decoding ability, and (3) slow L2 processing speed" (p. 382). Previous research has indeed shown that learners tend to have a lower WM capacity in the L2 than in the native language (e.g., Harrington & Sawyer, 1992; Osaka & Osaka, 1992; Service, Simola, Mesanheimo, & Maury, 2002; Van den Noort, Bosch, & Hugdahl, 2006), suggesting that potential limitations in proficiency may have a negative outcome on L2 learners' performance in WM tasks. Furthermore, L2 learners have been found to have more difficulty identifying words in noise (e.g., Mayo, Florentine, & Buus, 1997; Meador, Flege, & MacKay, 2000; Van Wijngaarden, Steeneken, & Houtgast, 2002) and to have slower processing times than native speakers (for discussion, see Frenck-Mestre, 2002). McDonald (2006) proposed that these cognitive processes interact with grammatical processing such that L2 learners' performance on corresponding cognitive tasks should correlate with their performance on grammatical tasks, and the impact of these cognitive processes should increase as the complexity of grammatical structure increases. Given the research that shows strong ties between L2 WM capacity and L2 comprehension (e.g., Altekin & Erçetin, 2010; Harrington & Sawyer, 1992; Miyake & Friedman, 1998; Service et al., 2002; Walter, 2004), she predicted L2 grammatical performance to be more strongly dependent on WM capacity in the L2 than on WM capacity in the native language.

To test these predictions, McDonald (2006) conducted a first set of experiments with L2 learners of English from various language backgrounds and native English speakers. The participants' grammatical performance was assessed using an auditory grammaticality judgment task that included agreement violations (regular past tense, third person singular, regular plural, present progressive, irregular past tense, irregular plural). WM capacity was assessed by asking the participants to report a list of English words they had heard in order of sizes (i.e., from smallest to largest; for details, see Montgomery, 2000). Decoding ability was assessed with

a gating task in which the participants identified words after hearing segments of the words. Processing speed was assessed by asking the participants to detect words in sentences. Note that L2 learners' performance on all the cognitive tasks depended in part on their proficiency in English, at least their knowledge of words and phonological processes. The L2 learners' results on each of the tasks were significantly different from those of the native speakers: L2 learners evidenced poorer grammaticality judgments, smaller WM capacity, poorer decoding abilities (i.e., more gates necessary), and slower word detection times than native speakers. Significant correlations (in the predicted directions) were found between L2 learners' judgments (accuracy and/or response times) and their WM capacity and decoding ability. Because the word detection task involved the recognition of words in complete sentences, McDonald (2006) suggested that it may have involved too much syntactic knowledge to be a good measure of processing speed alone.

In a second set of experiments, McDonald (2006) asked groups of native English speakers to complete the same grammaticality judgment task under processing loads related to either WM capacity (low or high digit loads), decoding ability (listening through white noise), or processing speed (eliciting responses under time pressure or presenting compressed speech). The participants also completed the same WM and decoding (i.e., gating) tasks as in the first set of experiments. The results showed poorer performance under a load (in the second set of experiments) than under no load (in the first set of experiments) for all the groups except for the group under a low memory load. For the grammaticality judgment task, significant correlations were found between the performances of (a) the participants under a memory load and their WM capacity, (b) the participants under a decoding load and their decoding ability (as measured by the gating task), and (c) the participants under time pressure and their response times in the grammaticality judgment task. The performances of the native speakers who had a high memory load or who listened to speech through white noise correlated most with L2 learners' performance.

The findings of this study, specifically those for native speakers, suggest relationships between the above cognitive processes (i.e., WM capacity, decoding ability, and processing speed) and grammatical accuracy. For L2 learners, because their performance on the cognitive tasks (at least those testing for decoding ability and processing speed) should have been correlated with their general proficiency in English (including lexical and phonological knowledge), it is unclear whether it was these cognitive abilities rather than their general proficiency that were significant predictors of grammatical accuracy. Decoding ability and processing speed, in particular, may be strongly tied to L2 learners' automaticity in the target language, which increases with proficiency (for discussion, see Segalowitz & Hulstijn, 2005). Although WM capacity should improve with increasing proficiency (e.g., Service et al., 2002; Van den Noort et al., 2006), it is likely to be less dependent on proficiency than decoding ability and processing speed, as native speakers show a great deal of variability in their WM capacity (e.g., Baddeley & Wilson, 1985; Daneman & Carpenter, 1980; Roberts & Gibson, 2002; Salt-house, 1990; Turner & Engle, 1989; Waters & Caplan, 1996a, 1996b). Hence, by testing for both general proficiency and WM capacity and by disentangling them, researchers may capture a great deal of variation among L2 learners in their

sensitivity to agreement morphology. Different models of WM capacity make different predictions for L2 learners' processing of agreement morphology, however. Let us consider some of these models before we discuss the role of WM capacity in L2 morphosyntactic processing.

WM CAPACITY AND ITS ROLE IN L2 MORPHOSYNTACTIC PROCESSING

The first model of WM, put forward by Baddeley and Hitch (1974), consists of a phonological loop, a visuospatial sketchpad, and a central executive component responsible for attention control. The first two systems were proposed to hold speech-based and visuospatial information in temporary stores, whereas the third system was responsible for allocating attention and controlling the information in these stores. More recently, Baddeley (2000) proposed a fourth component, the episodic buffer, which represents an interface between the three WM systems and long-term memory (for a discussion, see Baddeley, 2007).

Whereas much research on auditory language comprehension has focused on the role of the phonological loop (e.g., Gathercole & Baddeley, 1993), in the area of sentence processing, it is the ability of speakers to process sentences while holding information in memory that has received more attention. Two types of WM theories have been proposed: the single-resource theory, according to which humans have a set of verbal processing resources dedicated to all verbal tasks (e.g., Just & Carpenter, 1992; King & Just, 1991; MacDonald, Just, & Carpenter, 1992; Miyake, Carpenter, & Just, 1994); and a separate sentence-interpretation resource theory, according to which part of verbal WM is dedicated to "interpretive" processes, such as assigning syntactic structure to a sentence and using it to understand its meaning, and part of it is dedicated to "postinterpretive" processes, such as using the meaning of the sentence to perform other operations (e.g., Caplan & Waters, 1995, 1996; Waters & Caplan, 1996a, 1996b). Caplan and Waters (1999) propose that although both interpretive and postinterpretive processes degrade as WM capacity decreases or as WM load increases, interpretive processes are not differentially affected by the size of WM capacity and of WM load. They reviewed several studies showing that speakers with lower reading spans or under a larger WM load do not show a greater effect of syntactic complexity than speakers with higher reading spans or under a smaller WM load; by contrast, speakers with lower reading spans or under a large WM load have more difficulty processing sentences that contain more propositions than speakers with higher reading spans or under a small WM load, suggesting that postinterpretive processes are differentially affected by WM capacity (for a discussion, see Caplan & Waters, 1999).

WM capacity in sentence processing studies is typically measured with reading span tasks, which assess the ability of speakers to process sentences while holding words or digits in memory. One of the well-established reading span tasks is that of Daneman and Carpenter (1980), in which speakers read aloud sentences and recall the last word of the set of sentences they read, with each set containing from two to six sentences. The participants' reading span score is the total number of words that were correctly recalled. Waters and Caplan (1996a) identify a number of limitations with this task, including the possibility that the participants may

process sentences only at a superficial level when reading them aloud and that there might be a trade-off between their focus on the sentence meanings and their word recall. The test may thus reflect the storage component of WM more than the processing component. Instead, Waters and Caplan (1996a) developed a reading span task in which the participants judged the acceptability of sentences based on thematic role assignments (from the animacy requirements of verbs), and then recalled the last word of sets of sentences (for other versions requiring the participants to focus on the meaning or acceptability of sentences, see Tirre & Pena, 1992; Turner & Engle, 1989). The WM score is a composite score that reflects both the processing component (i.e., accuracy and reaction times when judging the acceptability of sentences) and the storage component (i.e., accuracy of the word recall) of WM, with the latter being computed as a function of set sizes (i.e., the highest set size to yield an accurate word recall), thus better capturing the upper limit of the storage component. In L2 studies, however, the dependent variables associated with the processing component of L2 reading-span tests are likely to be correlated with L2 proficiency. Solutions to this problem include to test for WM capacity in the native language (assuming that WM is language independent; e.g., Osaka & Osaka, 1992) or to use only the L2 word recall scores as a measure of WM capacity, as it is less likely to covary with proficiency.

A number of studies have investigated the role of WM capacity in L2 sentence processing (for a review of studies that focus on L2 learning in general, see Ardila, 2003; Williams, 2011). Some of these morphosyntactic studies suggest that reading span is not a determining factor in L2 learners' ability to process complex syntactic structures. For example, Juffs (2004, 2005) report that the ability of Chinese, Japanese, and Spanish L2 learners of English to process sentences that contain temporary syntactic ambiguities (e.g., *After the children cleaned the house looked very neat and tidy*) or *wh*-movements (e.g., *Who did the woman suggest the manager liked at the office?*) is not differentially affected by WM capacity in the native or target language, as measured by Harrington and Sawyer's (1992) reading-span test, which is based on Daneman and Carpenter's (1980) test. In a cross-modal priming study, Felser and Roberts (2007) found that high-span native English speakers, but not low span ones (as assessed by Harrington & Sawyer's [1992] reading-span test), showed evidence of antecedent reactivation in indirect object (experimental) gap positions but not in pregap (control) positions of *wh*-sentences, indicating that they posited a syntactic gap in indirect object position; by contrast, Greek L2 learners of English showed evidence of antecedent reactivation in both experimental and control positions, and their performance was not differentially affected by WM capacity.

In contrast, studies on the processing of agreement morphology suggest that reading span can be good predictor of L2 learners' ability to process agreement dependencies. Sagarra (2007) examined the relationship between low-proficiency Spanish L2 learners' WM capacity and their sensitivity to gender agreement violations. Because L2 learners' WM capacity has been suggested to be language independent (Osaka & Osaka, 1992), and because WM capacity in the target language can be correlated with L2 proficiency, Sagarra (2007) measured WM capacity in L2 learners' native language, English. The L2 learners completed a reading-span test adapted from Waters and Caplan (1996a), with the scores

including both the L2 learners' word recall and accuracy rates. She found that the L2 learners with higher WM scores processed sentences with gender agreement violations more accurately than L2 learners with lower WM scores. Similarly, Sagarra and Herschensohn (2010) found that the WM scores of their intermediate Spanish L2 learners (assessed in the same manner as in Sagarra, 2007) correlated positively with these L2 learners' reading times at the critical region of sentences with gender agreement violations and with their grammaticality judgments on the same sentences, suggesting that these L2 learners were more sensitive to gender agreement violations as their WM capacity increased.

Havik et al. (2009) also report WM effects in their study on the processing of subject–object relative clause ambiguities by German L2 learners of English. Their study focused on sentences where agreement on the verb disambiguated between the two different syntactic structures (e.g., *Daar is de machinist die de conducteurs heeft/hebben bevrijd uit het brandende treinstel* "That is the engine driver who has saved the guards/who the guards have saved from the burning train-carriage"). They manipulated the length of the agreement dependency, such that an adverbial phrase intervened between the relative clause subject and verb (e.g., *Daar is de machinist die de conducteurs na het ongeluk met de trein heeft/hebben bevrijd uit het brandende treinstel* "That is the engine driver who after the accident with the train has saved the guards/who the guards after the accident with the train have saved from the burning train-carriage"). Their results revealed that only the high-span L2 learners (as measured by Dutch and German versions of Daneman & Carpenter's [1980] reading-span task) had more difficulty in the processing of object relative clauses than in the processing of subject ones, and they did so only when the agreement dependency was short. This suggests that WM capacity influenced their use of agreement morphology for the processing of syntactic structure, but that the long agreement dependencies they tested might have been too taxing for WM capacity to be a good predictor of L2 learners' performance. Similar results were found for the low-span native speakers, whereas the high-span native speakers showed an effect of relative clause type in both the sentences containing short and long agreement dependencies.

By contrast, Foote (2011) did not find a relationship between her advanced Spanish L2 learners' sensitivity to gender and number agreement violations as a function of length of the agreement dependency and L2 learners' WM capacity in Spanish (as assessed by a reading-span test adapted from Waters & Caplan, 1996a). To compute sensitivity to agreement as a function of distance, she first subtracted the participants' reaction times on grammatical items from their reaction times on ungrammatical items separately for the two conditions where the agreeing words were and were not adjacent. The latencies that these subtractions yielded in the nonadjacent condition were then subtracted from the corresponding latencies in the adjacent condition. The WM scores were the total number of words that each participant had recalled in the task. This absence of relationship, however, could be due to the way in which she computed the effect of grammaticality as a function of distance or to the way in which she computed WM scores.²

In summary, the findings of Havik et al. (2009), McDonald (2006), Sagarra (2007), and Sagarra and Herschensohn (2010) indicate that WM capacity can be a good predictor of L2 learners' ability to process agreement morphology. The

present study further investigates the roles WM capacity, as well as the role of proficiency, in the processing of number agreement dependencies as a function of the length of the agreement dependency. It adds to the previous studies on the processing of agreement morphology in that (a) it focuses on a different target language, French; (b) unlike Keating (2009) and Foote (2011), it carefully controls the structure of the sentences in which the agreement dependency is short versus long; and (c) it examines the relationship between L2 learners' processing of agreement morphology and their WM capacity in both the native and target languages, as well as the relationship between their WM capacity in both languages.

This study focuses specifically on L2 learners' sensitivity to number agreement violations in third-person direct object clitics that are close to or distant from their antecedents. Third-person direct object clitics in French include *le* "him/it-masc.," *la* "her/it-fem.," and *les* "them." They agree in gender (when singular) and number with their antecedent. Unlike in English, these pronouns occur in preverbal position (*Marie le mangera*, "Marie it will-eat") and they cannot be stressed (**Marie LE mangera*; *Marie le mangera LUI*, "Mary it will-eat IT"). L2 learners' sensitivity to number agreement is examined in sentences where the object clitic agrees or does not agree with its left-dislocated antecedent (e.g., *Ce fruit/Ces fruits Marie le (*les)/les (*le) mangera pour sa collation avant l'entretien*, "This fruit/These fruits Marie it (*them)/them (*it) will-eat for her snack before the interview"), and where the object clitic is close to or distant from its left-dislocated antecedent (e.g., *Ce fruit/Ces fruits avant l'entretien Marie le (*les)/les (*le) mangera pour sa collation*, "This fruit/These fruits before the interview Marie it (*them)/them (*it) will-eat for her snack").³ Left-dislocated phrases are common in spoken French, and they make it possible to test for the participants' sensitivity to short and long number agreement dependencies without introducing a context prior to the critical sentences.

Because English pronouns also agree in number with their antecedents, the goal of the present study is not to determine whether nonnative speakers can learn to use morphosyntactic information not instantiated in the native language, but rather to investigate the effects of proficiency and WM capacity on L2 learners' sensitivity to number agreement violations as a function of distance between object clitics and their antecedents. Because the intervening material between the clitic and its antecedent adds to the propositional content of the sentence processed immediately before the clitic, both the single-resource theory and the separate sentence-interpretation-resource theory of WM capacity would predict that L2 learners' sensitivity to agreement violations in short- and long-distance dependencies will be differentially affected by their WM capacity. This would also be consistent with the findings of the previous studies on L2 learners' processing of agreement morphology (e.g., Havik et al., 2009; McDonald, 2006; Sagarra, 2007; Sagarra & Herschensohn, 2010). Furthermore, we predict that less proficient L2 learners will be less sensitive to agreement violations than more proficient L2 learners, and L2 learners will be less sensitive to agreement violations when the clitic is distant from its antecedent than when it is close to it.

The participants completed a region by region self-paced reading task that tested these predictions. They also completed an acceptability judgment task that assessed whether they had explicit knowledge of number agreement between

Table 1. Mean (standard deviation) L2 learners' biographical information

	Cloze (/45)	AFE	YrsInstr	MthsRes	%Use
Mid L2 (<i>n</i> = 26)	19.3 (4.8)	12.2 (3.0)	8.3 (2.5)	1.2 (1.6)	15.7 (12.6)
High L2 (<i>n</i> = 26)	30.2 (3.5)	10.5 (4.1)	8.8 (2.9)	9.2 (12.0)	16.1 (14.5)

Note: L2, second language; AFE, age of first exposure to french; YrsInstr, number of years of instruction in/on French; MthsRes, months of residence in a french-speaking environment; %Use, percentage of weekly use of French.

object clitics and left-dislocated antecedents. Finally, all the participants completed a reading-span task in French, and the L2 learners also completed a reading-span task in English, both of which were adapted from Waters and Caplan (1996a).

METHOD

Participants

Fifty-two adult native English speakers who learned French as an L2 (experimental group; age: 19–35; 39 females, 12 males) and 16 native French speakers (control group; age: 20–34; 12 females, 4 males) participated in this study. All participants had normal or corrected to normal vision, and were undergraduate or graduate students at the University of Illinois. In return for their participation, the L2 learners received financial compensation or course credit, and the native French speakers received financial compensation.

The native French speakers did not speak languages other than French before puberty. They all had functional knowledge of English, as they lived and were tested in the United States. The native English speakers did not speak languages other than English before puberty. At the time of the testing, they had completed at least four semesters of French instruction, but many of them were in at least one advanced (300-level or above) French class. The L2 learners' proficiency in French was assessed with the help of a cloze test. Such tests are commonly used as proficiency measures in L2 research, because they correlate highly with standardized proficiency tests (e.g., Bachman, 1985; Fotos, 1991). The validity, reliability, and discriminability of the particular cloze test used in this study were established independently (Tremblay, 2011; Tremblay & Garrison, 2010).⁴ Based on their cloze test scores, the L2 learners were evenly divided into two proficiency groups: mid (*n* = 26) and high (*n* = 26).

The participants completed a short language background questionnaire that collected relevant biographical information. Among other things, the L2 learners were asked their age of first exposure to French, the number of years of French instruction they had received, the number of months they had spent in a French-speaking environment, and their percent weekly use of French. The L2 learners' cloze test scores and their biographical information are summarized in Table 1.

Table 2. *Experimental conditions*

Distance	Clitic Number	1	2	3	4
Short	Singular	Ce fruit *Ces fruits	Marie le mangera Marie le mangera	pour sa collation pour sa collation	avant l'entretien. avant l'entretien.
	Plural	Ces fruits *Ce fruit	Marie les mangera Marie les mangera	pour sa collation pour sa collation	avant l'entretien. avant l'entretien.
Long	Singular	Ce fruit *Ces fruits	avant l'entretien avant l'entretien	Marie le mangera Marie le mangera	pour sa collation. pour sa collation.
	Plural	Ces fruits *Ce fruit	avant l'entretien avant l'entretien	Marie les mangera Marie les mangera	pour sa collation. pour sa collation.

A linear regression was performed on L2 learners' cloze test scores, with age of first exposure to French, years of instruction on French, months of residence in a French-speaking environment, and percent weekly use of French as independent variables. The regression indicated that months of residence in a French-speaking environment was a reliable predictor of cloze test scores ($r = .489, p < .004$). This further validates the use of the cloze test as proficiency measure.

Materials

The participants completed a region by region self-paced reading task. The experiment included 154 sentences, 48 of which were experimental, 96 of which were distracter sentences, and 10 of which were practice sentences. The experimental stimuli were 48 unrelated sentences containing a left-dislocated noun phrase and an object clitic (*le, la, or les*). Three variables were manipulated: the distance between the antecedent and the clitic (i.e., short vs. long), the number of the object clitic (i.e., singular vs. plural), and the grammaticality of the object clitic (i.e., agreeing vs. not agreeing in number with its antecedent). The distance and grammaticality manipulations aimed to examine L2 learners' sensitivity to agreement violations as a function of the distance between the object clitic and its antecedent, and the number manipulation was included to determine whether L2 learners would be sensitive to both missing and superfluous plural marking on the object clitic. The eight conditions resulting from crossing the three variables ($k = 6$) are illustrated in Table 2. The clitic number and grammaticality variables were counterbalanced in four lists in a Latin square design such that no list would contain more than one version of the same sentence. A complete list of the experimental items is provided in Appendix A.

The antecedents of the clitics were all nouns that were likely to be known by the L2 learners. These nouns were preceded by a demonstrative adjective (*ce* "this-masc/that-masc," *cette* "this-fem/that-fem," or *ces* "these/those") in order to avoid the definite articles *le* "the-masc-sing," *la* "the-fem-sing," and *les* "the-plur," which could have potentially cued the participants to the grammaticality or ungrammaticality of the object clitics. Half of the antecedent nouns were masculine and half were feminine so that the participants' general sensitivity to number agreement (irrespective of gender) would be assessed. The subjects were all proper nouns, and the gender of the subject was the opposite of the gender

of the clitic to avoid potential confusability between the subject and the object. The subject-clitic-verb region contained verbs that were frequent and likely to be known by L2 learners. All the verbs were consonant-initial, and thus none of the singular clitics had their vowel elided (*l'*).

The experimental items were interspersed with 96 distracter items. All sentences were presented in four regions and had similar lengths, containing 10–15 words. Some of the distracter sentences included subject clitics with right-dislocated antecedents so that the experimental sentences would not stand out. Half of the distracters were grammatical and half were ungrammatical. Ungrammatical distracters included gender and person violations between subject clitics and their right-dislocated antecedents, subject-verb person and number agreement violations, and “phonological” violations (e.g., *à le* “at the” instead of *au* “at-the”; *de le* “of the” instead of *du* “of-the”; *le* “the” + vowel-initial word rather than *l'*).

A true or false question followed each sentence in the self-paced reading task. This increased the likelihood that the participants would read the sentences for meaning rather than for form (for a similar study that instead used end of sentence form-focused questions, see Coughlin & Tremblay, 2011). For the experimental items, the comprehension questions included both the clitic and the verb used in the sentence, and they focused on the meaning of the adverbial portion of the sentence rather than on the clitic or its antecedent. For ungrammatical sentences, the clitic used in the comprehension question appeared in the same form (*le, la, les*) as in the sentence. To ensure that the participants would not ignore direct objects throughout the experiment, the distracter items included true or false questions that focused on the direct object in the sentence.

To determine whether L2 learners had explicit knowledge of number agreement in object clitics, the participants also completed an acceptability judgment task. Thirty-two of the experimental sentences from the self-paced reading task (four in each condition) and 36 of the distracter items were used in this task. The four different lists from the reading task were used to create four corresponding lists for the acceptability judgment task.

Of particular interest is the relationship between the participants' WM capacity and their processing of agreement dependencies as a function of distance in the self-paced reading task. In order to examine this relationship, reading-span tasks were created in both French and English using the design of Waters and Caplan (1996a). By having the L2 learners complete the WM task in both languages, we can examine the relationship between WM capacity in the native language and in the L2, determine whether WM capacity in the L2 is correlated with L2 proficiency, and establish whether the processing of agreement dependencies can be predicted by WM capacity in either the L1 or the L2. These two reading-span tasks were designed to test both the processing and storage components of WM capacity.

Each WM task contained 56 unrelated sentences. They included 28 semantically acceptable and 28 semantically unacceptable sentences split equally into four different groups that differed in sentence structure: cleft-subject (e.g., *It was the movie that terrified the child because it showed a monster*), cleft-object (e.g., *It was the secretary that the boutique delighted yet it was closed*), subject-object (e.g., *The senator that the strike angered became the leader*), and object-subject (e.g., *The millionaire favored the new law that frustrated the mayor*). For each sentence type, half of the sentences contained embedded verbs that required an

animate object and half contained embedded verbs that required an inanimate object. The semantically unacceptable sentences were created by inverting the subject and the object of the verb (e.g., cleft–subject: *It was the kid that appealed to the water because it was hot*; cleft–object: *It was the radio that the child scared because it was too loud*; subject–object: *The performance that the audience frightened left the theater*; object–subject: *The fence surrounded the carpenter that pleased the garden*). Other than for the embedded verbs, which had specific animacy requirements, the sentences in the French and English versions of the task contained different lexical items in order to avoid possible practice effects between the two tasks. If a given sentence was acceptable in one language, its corresponding sentence (which had the same embedded verb but otherwise different lexical items) was unacceptable in the other language. None of the sentence-final words (that would be recalled) were used more than once across the two languages. The number of words in each sentence and the length of the phrase-final word were controlled across the French and English versions of the test: both the French and English sentences contained on average 10.7 words ($SD = 5$), 43 of which had nouns and 13 of which had adjectives as their last word (in both languages), and these words were on average 5.4 letters long ($SD = 1.7$ in English, $SD = 1.3$ in French). The two tasks were therefore very similar.

Procedures

The participants completed the tasks in a booth in a quiet room. All the experiments were run with E-Prime (Psychology Software Tools).

The testing session began with the self-paced reading experiment. The sentences were presented one region at a time following a noncumulative moving-window procedure. A region by region presentation was favored over a word by word presentation, because the former is likely to impose a smaller processing load than the latter (at least for L2 learners), and it results in a more natural task because object clitics do not occur in isolation in French. At the beginning of each trial, a fixation cross appeared on the left side of the screen. The participants pressed the space bar to advance to the next region. They were instructed to read at a normal pace, but to press the space bar as soon as they finished reading one region to move on to the next. Each sentence in the experiment ended with a true or false comprehension question. The experiment began with 10 practice sentences, none of which contained agreement violations involving object clitics and their left-dislocated antecedents. The participants received feedback on the accuracy of their responses to the comprehension question in the practice session, but not in the main session. The order of test items in the main session was randomized across participants. The reading times at the four regions and the accuracy rates on the true or false comprehension questions were recorded.

The WM task in French was administered after the self-paced reading task. The participants read sets of two, three, four, or five sentences region by region, with each sentence containing six regions. The regions were presented using a noncumulative moving window procedure, and the participants pressed the space bar to advance to the next region. After each sentence, the participants were asked if the meaning of the sentence they had just read was acceptable, and after each

set, they were prompted to recall the last word of the sentences in that set in the order in which the sentences had appeared. A brief practice session preceded the main task. No feedback was given in the practice session or main experiment for either the participants' judgment accuracy or their word recall accuracy. The participants' accuracy rates and reaction times on the acceptability judgments and their word recall responses were recorded. This task was followed by a short break during which the participants filled out the language background questionnaire.

The acceptability judgment task was the next experiment administered. The participants were asked to read sentences on the screen and decide if they were or were not grammatically acceptable (yes, no). If they judged that a sentence was not acceptable, they were then asked to rewrite the sentence, making one correction that would render the sentence acceptable. There was no time limit imposed on their responses, and the entire sentence remained on the screen until the participant either indicated that it was acceptable or gave their corrected version of the sentence. There was a practice session of 10 sentences before the main experiment, none of which contained agreement violations involving object clitics and their left-dislocated antecedents. The participants received feedback from the researcher on the accuracy of their responses in the practice session, but not in the main session.

After the above tasks, the L2 learners completed the cloze test in French and the WM task in English. The procedures for the WM task in English were identical to the procedures for the WM task in French. The entire testing session took approximately two hours for L2 learners and 1 hr for native speakers.

Data analysis and predictions

For the acceptability judgment task, responses were considered accurate if the participants accepted grammatical sentences and corrected the number agreement violation on either the clitic or the antecedent after rejecting ungrammatical sentences. Grammatical sentences that were judged inaccurate for other (e.g., stylistic) reasons were not considered as errors, as long as the agreement dependency was not changed. Because all the groups obtained very high accuracy rates on this task, statistical analyses were not conducted.

For the self-paced reading task, the dependent variable was the participants' reading times. Reading times were included in the analyses only if the participants correctly answered the true or false comprehension question that followed each sentence. This resulted in the exclusion of 18.8% of L2 learners' data and 13.5% of native speakers' data. Reading times per region that were longer than 5000 ms were excluded from the analyses. This resulted in the exclusion of another 0.8% of L2 learners' data and 0.3% of native speakers' data. Linear mixed models with grammaticality (grammatical, ungrammatical), clitic number (singular, plural), and region (critical, spillover) as fixed variables, and with participant and test item as random variables, were then conducted on the participants' reading times in the critical regions (i.e., the second region in the short-distance conditions and the third region in the long-distance conditions) and in the postcritical regions to capture spill-over effects. Because the critical regions and lexical items were not the same in the short- and long-distance conditions, separate linear mixed models

were performed on these two sentence types. The L2 learners and native speakers' results were also analyzed separately given the unequal number of participants in each group. For L2 learners, proficiency (mid, high) also entered the linear mixed models as additional fixed variable. Subsequent linear mixed models were conducted whenever a variable interacted with grammaticality or with proficiency.

The participants' WM capacity was computed following Waters and Caplan's (1996a) analysis: the accuracy rates and response times on the acceptability judgments, and the largest number of words recalled in the correct order for three out of four sets (with an additional 0.5 word if two of the four sets of the immediately following span size were correctly recalled) were each transformed into Z scores (first for all the participants, and then separately for L2 learners and native speakers) so that they would be on a comparable scale and could be averaged together.⁵ Correlational analyses were first conducted between L2 learners' WM capacity in French and their proficiency scores as established by the cloze test, and then between L2 learners' WM scores in French and their WM scores in English.⁶ Regression analyses were then performed on the participants' reading times, with WM scores in French and English as independent variables.

If the participants are sensitive to number agreement violations in object clitics, they should show slower reading times for the subject-clitic-verb region when the clitic is ungrammatical than when it is grammatical, and the effect of grammaticality should be larger for higher proficiency L2 learners than for lower proficiency ones. If the length of the agreement dependency increases WM load, the L2 learners should be more sensitive to number agreement violations in the short-distance conditions than in the long-distance ones, and the participants' sensitivity to agreement violations should vary as a function of their WM capacity.

RESULTS

Table 3 presents the participants' accuracy rates on the acceptability judgment task. As can be seen from these results, the L2 learners and native speakers performed at ceiling on all the conditions. These results indicate that all the participants had

Table 3. *Participants' mean (SE) percentage accuracy on acceptability judgment task*

Distance	Clitic		Mid L2	High L2	Natives
	Number	Grammaticality			
Short	Singular	Grammatical	98.1 (0.01)	99.0 (0.01)	100 (0)
		Ungrammatical	98.1 (0.01)	99.0 (0.01)	96.5 (0.02)
	Plural	Grammatical	99.1 (0.01)	100 (0)	98.3 (0.02)
		Ungrammatical	99.0 (0.01)	100 (0)	96.4 (0.02)
Long	Singular	Grammatical	98.1 (0.01)	99.0 (0.01)	100 (0)
		Ungrammatical	98.1 (0.01)	96.2 (0.02)	98.2 (0.02)
	Plural	Grammatical	98.0 (0.01)	100 (0)	100 (0)
		Ungrammatical	97.1 (0.02)	97.1 (0.02)	100 (0)

Note: L2, second language.

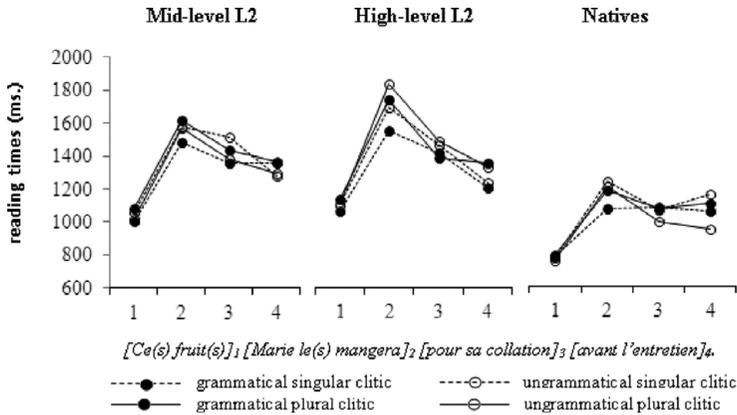


Figure 1. Participants' reading times in the short-distance condition.

Table 4. Mean (SE) reading times in critical and spillover regions for the short-distance condition

Clitic Number	Grammaticality	Critical Spillover					
		Mid L2		High L2		Natives	
Singular	Grammatical	1482 (94)	1358 (82)	1568 (98)	1434 (95)	1074 (93)	1083 (90)
	Ungrammatical	1578 (96)	1517 (111)	1708 (125)	1627 (95)	1272 (120)	1067 (73)
Plural	Grammatical	1615 (122)	1437 (86)	1756 (138)	1401 (74)	1188 (137)	1073 (110)
	Ungrammatical	1567 (116)	1384 (104)	1852 (135)	1505 (85)	1212 (143)	998 (72)

Note: L2, second language.

explicit knowledge of number agreement dependencies between object clitics and their left-dislocated antecedents.

Figure 1 shows the participants' reading times in the short-distance conditions, and Table 4 presents the participants' reading times and standard errors in the critical (i.e., second) and spillover (i.e., third) regions of the short-distance conditions. As can be seen from these results, all the groups seemed to slow down when reading ungrammatical singular clitics (compared to grammatical ones), and only the high-level L2 learners slowed down when reading ungrammatical plural clitics (compared to grammatical ones).

Mixed linear models on L2 learners' reading times in the second and third regions, with grammaticality, clitic number, region, and proficiency as fixed

variables, and with participant and item as random variables, revealed significant effects of grammaticality, $F(1, 1927) = 4.593, p < .032$, and region, $F(1, 1925) = 48.958, p < .001$, as well as significant two-way interactions between clitic number and region, $F(1, 1925) = 5.588, p < .018$, and between region and proficiency, $F(1, 1925) = 5.720, p < .017$. No other effect reached significance: number, $F(1, 1927) = 1.745, p < .187$; Clitic Number \times Proficiency, $F(1, 1927) = 1.671, p < .196$; Grammaticality \times Clitic Number \times Proficiency, $F(1, 1927) = 1.981, p < .159$; all other F s < 1 . Given the significant interaction involving proficiency, subsequent linear mixed models were conducted separately on the mid- and high-level L2 learners. For the mid-level L2 learners, these analyses yielded only a significant effect of region, $F(1, 958) = 9.773, p < .002$; Grammaticality \times Clitic Number, $F(1, 959) = 2.398, p < .122$; Clitic Number \times Region, $F(1, 958) = 1.336, p < .248$; all other F s < 1 . For the high-level L2 learners, they yielded a significant effect of grammaticality, $F(1, 968) = 4.620, p < .032$, and a significant effect of region, $F(1, 967) = 48.127, p < .001$, as well as a significant two-way interaction between number and region, $F(1, 967) = 4.991, p < .026$. None of the other effects were significant: number, $F(1, 968) = 3.724, p < .054$; all other F s < 1 . These results suggest that the high-level L2 learners contributed the most to the effect of grammaticality found in the overall analysis.

Similar mixed level models on native speakers' reading times in the second and third regions, with grammaticality, clitic number, and region as fixed variables, and with participant and item as random variables, revealed only a significant effect of region, $F(1, 627) = 12.766, p < .001$, and a marginally significant interaction between grammaticality and region, $F(1, 627) = 3.718, p < .054$. No other results reached significance: Grammaticality \times Clitic Number, $F(1, 627) = 1.314, p < .252$; all other F s < 1 . Given the interaction involving grammaticality, subsequent linear mixed models were conducted separately on the critical and spillover regions. These analyses did not yield any significant results: critical region: grammaticality, $F(1, 305) = 2.539, p < .112$; spillover region: grammaticality, $F(1, 307) = 1.119, p < .291$. These results thus indicate that the numerical tendency for native speakers to slow down at ungrammatical clitics is not reliable.

Figure 2 shows the participants' reading times in the long-distance conditions, and Table 5 presents the participants' reading times and standard errors in the critical (i.e., third) and spillover (i.e., fourth) regions of the long-distance conditions. As can be seen from these results, whereas the midlevel L2 learners do not show an effect of grammaticality, the high-level L2 learners seemed to slow down when parsing ungrammatical plural clitics (compared to grammatical ones), and the native speakers slowed down when parsing ungrammatical singular and plural clitics (compared to grammatical ones).

Mixed linear models on L2 learners' reading times in the third and fourth regions, with grammaticality, clitic number, region, and proficiency as fixed variables, and with participant and item as random variables, revealed a marginally significant effect of grammaticality, $F(1, 1992) = 3.721, p < .054$, and significant effects of clitic number, $F(1, 1992) = 4.418, p < .036$, and region, $F(1, 1991) = 17.477, p < .001$, as well as a significant three-way interaction between

Table 5. Mean (SE) reading times in critical and spillover regions for all long-distance condition

Clitic Number	Grammaticality	Critical Spillover					
		Mid L2		High L2		Natives	
Singular	Grammatical	1551 (122)	1420 (94)	1577 (95)	1454 (73)	1082 (93)	1223 (118)
	Ungrammatical	1492 (108)	1523 (112)	1627 (117)	1541 (89)	1222 (131)	1078 (75)
Plural	Grammatical	1676 (130)	1505 (105)	1591 (93)	1554 (87)	1031 (69)	1125 (84)
	Ungrammatical	1648 (142)	1586 (118)	1736 (121)	1462 (72)	1179 (131)	1206 (136)

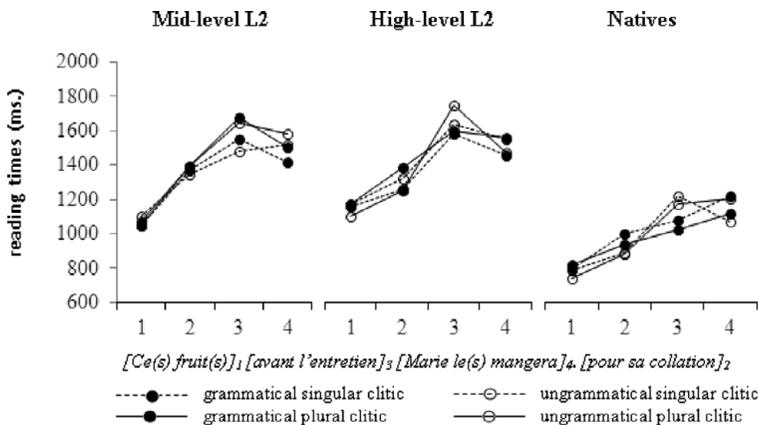


Figure 2. Participants' reading times in the long-distance condition.

grammaticality, region, and proficiency, $F(1, 1991) = 5.224, p < .022$. No other effect reached significance: Clitic Number \times Proficiency, $F(1, 1991) = 1.295, p < .255$; Region \times Proficiency, $F(1, 1991) = 1.113, p < .291$; Grammaticality \times Clitic Number \times Region, $F(1, 1991) = 2.609, p < .106$; all other $F_s < 1$. Given the significant interaction involving grammaticality and proficiency, subsequent linear mixed models were conducted separately on the mid- and high-level L2 learners. For these comparisons, the alpha level was adjusted to 0.025. For the midlevel L2 learners, these analyses yielded a significant effect of number, $F(1, 977) = 4.599, p < .032$, and region, $F(1, 976) = 4.300, p < .038$. No other effects were significant: Grammaticality \times Region, $F(1, 976) = 3.254, p < .072$; all other $F_s < 1$. For the high-level L2 learners, they yielded a significant effect of grammaticality, $F(1, 1016) = 3.967, p < .047$, and a significant effect of

region, $F(1, 1015) = 15.687, p < .001$. None of the other effects reached significance, $F(1, 1015) = 3.739, p < .053$; Grammaticality \times Region, $F(1, 1015) = 1.967, p < .161$; all other F s < 1 . These results suggest again that the high-level L2 learners contributed the most to the effect of grammaticality found in the overall analysis. They also indicate that the midlevel L2 learners were particularly affected by the clitic number, indicating that they do not ignore number marking on the clitic.

Similar mixed linear models on native speakers' reading times in the third and fourth regions, with grammaticality, clitic number, and region as fixed variables, and with participant and item as random variables, revealed only a significant two-way interaction between grammaticality and region, $F(1, 651) = 5.551, p < .019$. No other effects reached significance, Grammaticality \times Clitic Number, $F(1, 651) = 3.093, p < .079$; Grammaticality \times Clitic Number \times Region, $F(1, 651) = 2.512, p < .113$; all other F s < 1 . Subsequent linear mixed models were therefore conducted separately on the critical and spillover regions. These analyses yielded a significant effect of grammaticality for the critical region, $F(1, 319) = 6.551, p < .011$, and a significant interaction between grammaticality and clitic number for the spillover region, $F(1, 317) = 5.317, p < .022$. No other effect reached significance in the critical region (all F s < 1) or in the spillover region: grammaticality, $F(1, 319) = 1.091, p < .297$; all other F s < 1 . Given the significant interaction involving grammaticality in the spillover region, a third linear mixed model was conducted on that region separately for sentences containing singular and plural clitics. For these comparisons, the alpha level was adjusted to .0125. These analyses revealed a significant effect of grammaticality only for sentences containing a singular clitic, $F(1, 151) = 6.244, p < .014$. Notice, however, that this effect of grammaticality is in the opposite direction to that in the critical region. Hence, unlike the results for the short-distance conditions, those for the long-distance conditions indicate that native speakers slowed down in the critical region when they parsed ungrammatical object clitics, and for singular clitics, this is followed by a trade-off in the spillover region.

Together, the results of the self-paced reading experiment indicate that the high-level L2 learners were somewhat more sensitive to agreement violations than the midlevel L2 learners. Numerically, the effect of grammaticality appeared stronger in the short-distance conditions than in the long-distance ones, but our use of different sentences in the two conditions preclude the use of statistical analyses to establish whether the two distance conditions differed in the effect of grammaticality they elicited. The results also show that L2 learners' sensitivity to agreement violations was manifested in both the critical and spillover regions. By contrast, the native speakers' sensitivity to agreement violations varied as a function of region and distance, with the effect of grammaticality reaching significance only in the critical region of the long-distance condition. On the one hand, we suspect that the sentences in the short-distance condition might have been too easy to parse for native speakers to show a significant effect of grammaticality; on the other hand, in the long-distance conditions, the fact that native speakers show an effect of ungrammaticality only in the critical region (rather than in both the critical and spillover regions) suggest that they integrate morphological information more rapidly than L2 learners.

Table 6. *Participants' mean (SE) scores on the WM task in French*

	Mid L2	High L2	Natives
Accuracy rates (acceptability judgments)	-0.66 (0.10)	0.20 (0.10)	0.74 (0.13)
Response times (acceptability judgments)	0.00 (0.17)	0.02 (0.19)	-0.05 (0.21)
Words recalled	-0.24 (0.16)	-0.26 (0.17)	0.81 (0.26)
Overall WM scores	-0.30 (0.08)	-0.01 (0.10)	0.50 (0.15)

Note: WM, working memory; L2, second language.

One question that arises from the previous results is whether individual variability in WM capacity could explain L2 learners' sensitivity to number agreement violations in the short- and long-distance conditions. To answer this question, the participants' reading span was examined. Table 6 presents the L2 learners' and native speakers' Z scores on each component of WM task in French (i.e., accuracy rates on the acceptability judgments, response times on the acceptability judgments, and words recalled), as well as their overall WM scores (i.e., the three components averaged). This first set of Z scores was computed over all the participants so that L2 learners and native speakers could be compared.

As can be seen in Table 6, native speakers have much higher WM scores than L2 learners, and as a group, the high-level L2 learners have higher WM scores than the midlevel L2 learners, suggesting that the WM scores may be correlated with proficiency. Correlational analyses between the participants' Z scores on each WM component and their cloze test scores, and between their overall WM scores and their cloze test scores, were therefore performed. They revealed a moderate correlation between L2 learners' accuracy Z scores and the cloze test scores ($r = .66, p < .001$), as well as a weak correlation between L2 learners' overall WM scores and their cloze test scores ($r = .28, p < .05$). One would indeed expect that L2 learners' ability to determine whether or not a sentence is semantically acceptable in French is at least partially dependent on their proficiency in French. This component of reading span is thus responsible for the weak correlation between L2 learners' overall WM scores and their cloze test scores.

For the L2 learners, we also examined the individual WM score components across the French and English versions to investigate possible relationships. To do so, L2 learners' Z scores in French were recalculated without those of native speakers, and correlational analyses between the English and French scores were performed. Table 7 presents the L2 learners' Z scores in French and English, and Table 8 presents the results of the correlational analyses.

Numerically, the results in Table 7 suggest that the WM capacity of the midlevel L2 learners in English is lower than that of the high-level L2 learners in English. However, a one-way analysis of variance on the overall WM scores in English does not reveal a significant difference between the two groups ($F < 1$). The

Table 7. L2 Learners' mean (SE) scores on the WM task in French and English

	Mid L2		High L2	
	French	English	French	English
Accuracy rates				
(acceptability judgments)	-0.45 (0.11)	-0.17 (0.14)	0.45 (0.11)	0.18 (0.14)
Response times				
(acceptability judgments)	-0.01 (0.16)	-0.06 (0.17)	0.01 (0.19)	0.06 (0.17)
Words recalled	0.01 (0.19)	0.08 (0.18)	-0.01 (0.20)	-0.08 (0.22)
Overall WM scores	-0.15 (0.11)	-0.05 (0.10)	0.15 (0.11)	0.05 (0.13)

Note: WM, working memory; L2, second language.

Table 8. Correlations between second language learners' working memory scores in French and English

French	English			
	Accuracy Rates	Response Times	Words Recalled	Overall
Accuracy rates	.26	.24	.10	—
Response times	.25	.68***	-.06	—
Words recalled	.14	-.08	.38**	—
Overall	—	—	—	.51***

** $p < .01$. *** $p < .001$.

results in Table 8 indicate that the L2 learners' response-time and word-recall Z scores in French correlate with those in English, as do their overall WM scores. By contrast, their accuracy Z scores on the WM task in French do not correlate with their accuracy Z scores on the WM task in English. This is indeed what we should find given the relationship between L2 learners' accuracy rates in the WM task in French and their proficiency in French.

Given the significant correlation between L2 learners' reading span and their proficiency in French, in order to examine the relationship between WM capacity and sensitivity to agreement violations, we used L2 learners' word-recall scores in French (which are not correlated with proficiency) and their overall WM scores in English. For each distance and number condition, we subtracted L2 learners' reading times in the critical regions of grammatical items from their reading times in the critical regions of ungrammatical items. We then performed linear regressions on these latencies, with WM in French (from the word-recall Z scores) and WM in English (from the overall Z scores) as independent variables. These analyses revealed that WM in French is a marginally significant predictor of the

grammaticality effect that L2 learners displayed on sentences with plural clitics in the short-distance condition ($r = .254, p < .083$) and on sentences with singular clitics in the long-distance condition ($r = .267, p < .068$). Although the observed relationship is somewhat weak, these results suggest that L2 learners with higher WM capacity in French tend to be more sensitive to number agreement violations in French than L2 learners with lower WM capacity in French. Notably, however, this relationship does not seem to increase as a function of the distance between the clitic and its antecedent. In contrast, WM in English was not found to be a significant predictor of the grammaticality effect that L2 learners displayed on any of the sentence types. We will return to the significance of these results in the discussion section.

Finally, we also examined the relationship between native speakers' reading span (in French) and their sensitivity to agreement violations. We performed similar regressions on the difference between native speakers' reading times in the critical regions of ungrammatical and grammatical sentences, using both their overall WM Z scores and their word-recall Z scores (for the sake of comparison with the L2 learners) as independent variables. None of these analyses reached significance, indicating that WM capacity is not related to native speakers' sensitivity to agreement violations in French.

Let us now turn to a discussion of these findings and their implications for understanding L2 learners' processing of agreement morphology.

GENERAL DISCUSSION

The results of the present study showed that whereas both mid- and high-level L2 learners were able to identify (and correct) number agreement violations in object clitics in an offline acceptability judgment task, only the high-level L2 learners showed sensitivity to such violations in an online self-paced reading task. Of importance, the high-level L2 learners' sensitivity to number agreement violations did not appear to vary (other than numerically) as a function of the distance between the clitic and its antecedent. Furthermore, despite the tendency for these L2 learners to show more sensitivity to agreement violations in the long-distance condition when the clitic is plural than when it is singular, the results did not show a significant interaction between grammaticality and clitic number, suggesting that these L2 learners were able to detect agreement violations, irrespective of clitic number.

These results are in line with those of Foote (2011), who also found that L2 learners were sensitive to agreement violations irrespective of the length of the agreement dependency, but they differ from those of Keating (2009), who report sensitivity to agreement only when the two words in the agreement dependency were adjacent. In his study, Keating manipulated his sentences by increasing both the linear and structural distance between the two words in the agreement dependency. In our study, other than for the intervening adverbial phrase, which increased the linear distance between the clitic and its antecedent, the sentences in the long-distance conditions had identical syntactic structures to those in the short-distance conditions. One possibility is that the structural distance between

two elements in an agreement dependency increases L2 learners' processing load more than does linear distance. Alternatively, the distance between the clitic and its antecedent in the long-distance condition may not have been sufficiently large to affect the high-level L2 learners' sensitivity. L2 research that teases linear and structural distances apart and that manipulates the length of agreement dependencies is necessary to determine which of these explanations is likely to be correct.

One question that arises from these results, then, is why the midlevel L2 learners were able to identify (and correct) number agreement violations in the acceptability judgment task and yet failed to show a significant effect of grammaticality in the self-paced reading task. Because these L2 learners showed an effect of clitic number in their results, it is unlikely that they simply failed to parse the plural marking on the object clitic. A more likely explanation is that they failed to establish an agreement dependency between the clitic and its antecedent, possibly due to two factors: the left-dislocated syntactic structure that was used to express the antecedent, and the preverbal position of the object clitic in French, unlike that of object pronouns in English. The midlevel L2 learners may not have yet developed the processing routines necessary to link dislocated phrases to preverbal object clitics, which might then make it more difficult for them to establish an agreement dependency between these phrases and object clitics.

The results of this study also showed that L2 learners had lower reading-span scores in French than native French speakers, consistent with the findings of previous studies (e.g., Harrington & Sawyer, 1992; Osaka & Osaka, 1992; Service et al., 2002; Van den Noort et al., 2006). L2 learners' acceptability judgments on the WM task in French were found to correlate with their proficiency scores, and their response times and words recalled on the WM task in French were found to correlate with, respectively, their response times and words recalled on the WM task in English. This indicates that WM capacity is at least partially dependent on proficiency, and it covaries with WM capacity in the native language, as has been found in previous work (e.g., Altekin & Erçetin, 2010; Harrington & Sawyer, 1992; Service et al., 2002; Van den Noort et al., 2006). These findings confirm that, as a construct, WM capacity is likely to be language independent, but potential limitations in proficiency can have a negative outcome on L2 learners' performance in such tasks.

WM capacity in French, as computed from the word-recall scores (which did not correlate with proficiency), was also found to be a weak predictor of L2 learners' sensitivity to agreement violations in sentences containing a plural clitic in the short-distance condition and in sentences containing a singular clitic in the long-distance condition. These results are consistent with those of Havik et al. (2009), who found that only high-span L2 learners could use verbal agreement to extract the syntactic structure of relative clauses in Dutch. In the present study, it is unclear why the relationship between reading span and sensitivity to agreement morphology was not found in the four distance and number conditions. Havik et al. (2009) report that high-span L2 learners did not outperform low-span ones when the distance between the subject and the agreeing verb was too

large, potentially because this distance was simply too taxing for the L2 learners. In our study, one might suspect that the sentences with singular clitics in the short-distance condition were the easiest and that those with plural clitics in the long-distance condition were the most difficult, thus perhaps resulting in insufficient variability in L2 learners' sensitivity to agreement violations for a relationship between it and WM capacity to be found. Because clitic number was not found to interact with grammaticality, this hypothesis is of course speculative. Nonetheless, the results suggest that WM capacity can be a factor in determining whether L2 learners show sensitivity to agreement violations in online sentence processing.

Contrary to predictions, WM capacity did not differentially affect L2 learners' performance on sentences in the short- and long-distance conditions. Because the intervening material between the clitic and its antecedent adds to the propositional content of the sentence processed immediately before the clitic, both the single-resource theory and the separate sentence-interpretation-resource theory of WM capacity predicted that L2 learners' sensitivity to agreement violations in short- and long-distance dependencies would be differentially affected by their WM capacity, but this was not the case. As mentioned above, the linear distance imposed between the clitic and its antecedent may not have been sufficiently large to elicit a difference between the short- and long-distance conditions. Further research should further investigate the effect of the size of agreement dependencies on L2 learners' processing of agreement morphology.

The results also showed that despite the significant correlation between L2 learners' WM scores in French and English, WM capacity in English was not a significant predictor of L2 learners' sensitivity to agreement violations. The fact that the word-recall scores in the French task could predict L2 learners' sensitivity to agreement perhaps suggest that a similar process underlies their storing of French words and their detecting agreement violations in French. These findings, together with the absence of clear distance effect in L2 learners' processing of object clitics, are compatible with Ullman's (2001, 2004) DP model. The DP model would predict that L2 learners would store singular and plural object clitics as separate lexical entries, and their establishing an agreement dependency between the clitic and its antecedent would not necessarily be more difficult when the clitic is close to its antecedent than when it is distant from it. Although this model does not make predictions that are specific to WM, it might predict that L2 learners' ability to store words in WM would be related to their ability to process inflected words that are stored as single units in declarative memory. In the case of French clitics, although the plural clitic is formed by adding the plural marker "s" to the masculine singular clitic "le," the phonology of these forms is suppletive rather than agglutinative, with the plural "s" not being pronounced and with the vowel changing from /ə/ or /a/ to /e/. It is therefore possible that L2 learners store these words as separate lexical entries, although whether or not native speakers do so cannot be established on the basis of our results. Further research should try to tease these two scenarios apart.

Even though the present results are compatible with both views of L2 morphological processing, namely, that L2 learners and native speakers process agreement

morphology in a quantitatively or qualitatively different manner, the findings of this study are quite revealing for understanding the factors that regulate L2 learners' processing of agreement morphology. They indicate that L2 learners become more sensitive to agreement morphology in sentence processing as their proficiency in French increases, and they are more likely to show sensitivity to agreement morphology if they have a high WM capacity in the target language. Models that incorporate both proficiency and WM capacity can thus explain a great deal of inter-L2-learner variability in the processing of number agreement. Theories such as McDonald's (2006) cognitive processing account assume that WM capacity is a determining factor in L2 learners' sensitivity to agreement violations. The DP model assumes that declarative memory plays an important role in L2 learners' learning of agreement morphology. By integrating the roles of general proficiency, WM capacity, and declarative memory, these two theories can potentially explain a great deal of inter-L2-learner variability, arguably more than Clahsen and Felser's (2006) SSH.

The main limitation of this study perhaps lies in its use of number agreement violations as a diagnosis for L2 learners' and native speakers' sensitivity to morphological information in sentence processing. This method, which was also used by Foote (2011), Jiang (2004, 2007), Keating (2009), Sagarra (2007), Sagarra and Herschensohn (2010), and can lead participants to alter their normal processing routines either by drawing their attention to the agreement violations or by leading them to develop processing strategies for ignoring these violations. The relationships found between L2 learners' sensitivity to number agreement violations and their WM capacity and proficiency suggest that the task successfully tapped into L2 learners' processing of agreement morphology, but further research should examine their processing of agreement morphology in grammatical contexts where it has interpretive consequences (e.g., in cases where it would influence the syntactic structure assigned to the sentence, as in Havik et al., 2009; Hopp, 2006, 2010).

CONCLUSION

The present study investigated the sensitivity of mid- and high-level English L2 learners of French and native French speakers to number agreement violations between object clitics that were close to or distant from their left-dislocated antecedents. The results of an acceptability judgment task and a self-paced reading task showed that whereas L2 learners and native speakers were sensitive to agreement violations in the offline task, only the high-level L2 learners showed sensitivity to agreement violations in the online task, and these L2 learners did so in both the short- and long-distance agreement dependencies. The results also showed a weak relationship between L2 learners' sensitivity to agreement violations and their WM capacity. These findings suggest that proficiency and WM capacity can modulate whether L2 learners are sensitive to agreement morphology in sentence processing. These factors should thus play a central role in theories of L2 morphological processing.

APPENDIX A

Experimental sentences

1	2	3	4
Short Distance			
Ce bonbon	Annie le choisira	à la grande confiserie	pour les étudiants.
Ce cadeau	Cécile le donnera	à sa petite fille	pour sa graduation.
Ce chèque	Laura le déposera	dans son compte épargne	tôt cette semaine.
Ce citron	Sarah le pressera	au nouveau restaurant	tôt cet après-midi.
Ce crayon	Carla le taillera	avant ses examens	tôt demain matin.
Ce devoir	Claire le finira	pour bien comprendre	avant l'examen.
Ce fruit	Marie le mangera	pour sa collation	avant l'entretien.
Ce gâteau	Lola le décorera	pour la grande fête	dans trois jours.
Ce jouet	Annie le donnera	aux enfants pauvres	dans le village.
Ce livre	Martine le lira	à la bibliothèque	pendant tout l'été.
Ce masque	Marie le portera	pendant toute la pièce	pour son rôle.
Ce poème	Anna le récitera	pendant ses vacances	dans les Caraïbes.
Cette banane	Henri la mangera	pendant la discussion	après la conférence.
Cette banque	Alain la gèrera	avec son partenaire	à la réouverture.
Cette boîte	Jean la laissera	contre le bâtiment	la semaine suivante.
Cette carotte	Adrien la mettra	dans la salade de chou	avant la réception.
Cette chaise	Louis la jettera	après l'année scolaire	sans hésitation.
Cette chanson	Alain la chantera	au festival de musique	le weekend prochain.
Cette chemise	Rémi la trouvera	au centre d'achats	un peu plus tard.
Cette fleur	Paul la cueillera	pour sa belle copine	mardi après-midi.
Cette maison	Jules la vendra	avant sa retraite	l'année prochaine.
Cette photo	Henri la montrera	aux jeunes étudiants	pendant le cours.
Cette pomme	Carole la coupera	pour les étudiants	avant le pique-nique.
Cette voiture	Chris la réparera	dans la cours arrière	pendant ses loisirs.
Long Distance			
Ce bateau	pendant l'été	Adèle le pilotera	sur la longue rivière.
Ce billet	après-demain	Simone le vendra	à l'entrée du stade.
Ce cadre	en après-midi	Sandra le vendra	au marché aux puces.
Ce cahier	tôt en matinée	Rose le révisera	avant de faire l'examen.
Ce cigare	vendredi soir	Céleste le fumera	après la cérémonie.
Ce disque	tôt en soirée	Karine le jouera	en lisant les paroles.
Ce légume	dimanche soir	Laura le mangera	dans la grande cuisine.
Ce projet	en septembre	Léa le présentera	devant le conseil.
Ce stylo	en avant-midi	Lise le laissera	dans la grande salle.
Ce texte	très bientôt	Eve le révisera	pour la revue sport.
Ce verre	demain matin	Lise le mettra	dans l'autre armoire.
Cette boisson	en après-midi	Matthieu la boira	après le match de foot.
Cette caméra	très bientôt	Henri la voudra	pour sa collection.
Cette canne	pendant l'été	Michel la donnera	à la dame très âgée.
Cette carte	au mois de mai	David la postera	pour la fête des mères.

APPENDIX A (cont.)

1	2	3	4
Long Distance			
Cette corde	en après-midi	Bernard la nouera	à l'arrière du garage.
Cette cravate	demain matin	Fabien la portera	au mariage de sa sœur.
Cette facture	d'ici janvier	Benoît la paiera	avec sa carte de crédit.
Cette guitare	tôt en soirée	Andy la laissera	dans la salle de concert.
Cette lettre	d'ici octobre	Roger la signera	devant le président.
Cette poupée	le lendemain	Marc la laissera	à sa petite fille.
Cette recette	avant la fête	Jean la copiera	pour son meilleur ami.

Note: The items are presented in their version with singular grammatical clitics.

ACKNOWLEDGMENTS

This research was supported in part by the Department of French and the College of Liberal Arts and Science at the University of Illinois. We thank Sarah Brown-Schmidt, Zsuzsanna Fagyal, Peter Golato, Duane Watson, the members of the CaLL and Conversation Labs, and two anonymous reviewers for their valuable comments, as well as all of the participants who have taken part in this research.

NOTES

1. Paradis (2004, 2009) proposed a model that is similar to but also different from Ullman's (2001, 2004) model. Like Ullman, Paradis assumes a distinction between declarative and procedural knowledge, with L2 learners relying more on the former in early stages of development and more on the latter as their proficiency in the target language increases. Unlike Ullman, however, Paradis proposed that declarative and procedural knowledge are, respectively, explicit and implicit in nature, and he characterizes higher level L2 learners' increasing reliance on procedural knowledge in terms of automaticity and acquisition of implicit knowledge. In this paper, we focus on Ullman's model rather than on Paradis's, because only the former explicitly discusses the processing of agreement morphology.
2. Foote's (2011) sentences in the nonadjacent and adjacent conditions could not be compared directly, because they contained different lexical items. The effect of grammaticality in the nonadjacent conditions should therefore not have been subtracted from that in the adjacent conditions.
3. The example here and elsewhere uses the same lexical words for illustration purposes; the test items used different lexical words for the short- and long-distance conditions.
4. The L2 learners were not required to take a standardized test of French proficiency when entering the French program, and administering such a test for the purpose of the present study would not have been feasible due to the amount of time it would have taken. Hence, standardized proficiency scores were not available for the present L2 learners.

5. The response times were put on a negative scale before they were converted into Z scores, because they are inversely proportional to processing efficiency.
6. One L2 learner was not included in this analysis, as her data were lost due to a computer problem.

REFERENCES

- Alptetin, C., & Erçetin, G. (2010). The role of L1 and L2 working memory in literal and inferential comprehension in L2 reading. *Journal of Research in Reading, 33*, 206–219.
- Ardila, A. (2003). Language representation and working memory with bilinguals. *Journal of Communication Disorders, 36*, 233–240.
- Bachman, L. F. (1985). Performance on cloze tests with fixed-ratio and rational deletions. *TESOL Quarterly, 16*, 61–70.
- Baddeley, A. D. (2000). The Phonological loop and the irrelevant speech effect: Some comments on neath. *Psychonomic Bulletin and Review, 7*, 544–549.
- Baddeley, A. D. (2007). *Working memory, thought and action*. Oxford: Oxford University Press.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press.
- Baddeley, A. D., & Wilson, B. A. (1985). Phonological coding and short-term memory in patients without speech. *Journal of Memory and Language, 24*, 490–502.
- Birdsong, D., & Flege, J. E. (2001). Regular-irregular dissociations in L2 acquisition of English morphology. In A. H.-J. Do, L. Domínguez, & A. Johansen (Eds.), *Proceedings of the 25th Annual Boston University Conference on Language Development* (pp. 123–132). Somerville, MA: Cascadilla Press.
- Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and language, 44*, 235–249.
- Bowden, H., Gelfand, M., Sanz, C., & Ullman, M. (2010). Verbal inflectional morphology in L1 and L2 Spanish: A frequency effects study examining storage versus composition. *Language Learning, 60*, 44–87.
- Caplan, D., & Waters, G. S. (1995). Aphasic disturbances of syntactic comprehension and working memory capacity. *Cognitive Neuropsychology, 12*, 637–649.
- Caplan, D., & Waters, G. S. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 49A*, 51–79.
- Clahsen, H., & Felser, C. (2006a). Grammatical processing in language learners. *Applied Psycholinguistics, 27*, 3–42.
- Clahsen, H., & Felser, C. (2006b). Continuity and shallow structures in language processing. *Applied Psycholinguistics, 27*, 107–126.
- Clahsen, H., Felser, C., Neubauer, K., Sato, M., & Silva, R. (2010). Morphological structure in native and nonnative language processing. *Language Learning, 60*, 21–43.
- Coughlin, C. E., & Tremblay, A. (2011). Native and non-native processing of short and long agreement dependencies in French. In L. Armstrong (Ed.), *Proceedings of the 2011 Annual Conference of the Canadian Linguistic Association*. Retrieved from http://homes.chass.utoronto.ca/~clacl/actes2011/Coughlin_and_Tremblay_2011.pdf
- Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450–466.
- Felser, C., & Roberts, L. (2007). Processing wh-dependencies in a second language: A cross modal priming study. *Second Language Research, 23*, 9–36.
- Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics, 24*, 453–489.
- Foote, R. (2011). Integrated knowledge of agreement in early and late English–Spanish bilinguals. *Applied Psycholinguistics, 32*, 187–220.

- Fotos, S. S. (1991). The cloze test as an integrative measure of EFL proficiency: A substitute for essays on college entrance examinations? *Language Learning, 41*, 313–336.
- Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Bilingualism: Language and Cognition, 14*, 379–399.
- Frenck-Mestre, C. (2002). An on-line look at sentence processing in the second language. In R. Heredia & J. Altarriba (Eds.), *Bilingual sentence processing* (pp. 217–236). Amsterdam: Elsevier.
- Gathercole, S., & Baddeley, A. D. (1993). Phonological working memory: A critical building block for reading development and vocabulary acquisition? *European Journal of Psychology of Education, 8*, 259–272.
- Hahne, A., Mueller, J., & Clahsen, H. (2006). Morphological processing in a second language: Behavioral and event-related brain potential evidence for storage and decomposition. *Journal of Cognitive Neuroscience, 18*, 121–134.
- Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition, 14*, 25–38.
- Havik, E., Roberts, L., Van Hout, R., Schreuder, R., & Haverkort, M. (2009). Processing subject-object ambiguities in the L2: A self-paced reading study with German L2 learners of Dutch. *Language Learning, 59*, 73–112.
- Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research, 22*, 369–397.
- Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua, 120*, 901–931.
- Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics, 25*, 603–634.
- Jiang, N. (2007). Selective integration of linguistics knowledge in adult second language learning. *Language Learning, 57*, 1–33.
- Johnson, J., & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*, 60–99.
- Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society, 102*, 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*, 121–151.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122–149.
- Keating, G. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning, 59*, 503–535.
- King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The rate of working memory. *Journal of Memory and Language, 30*, 580–602.
- Lardiere, D. (2006). *Ultimate attainment in second language acquisition: A case study*. Mahwah, NJ: Erlbaum.
- MacDonald, J. L. (2000). Grammaticality judgments in a second language: Influences of age of acquisition and native language. *Applied Psycholinguistics, 21*, 395–423.
- MacDonald, M., Just, M., & Carpenter, P. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology, 24*, 56–98.
- Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition, 27*, 53–78.
- Mayo, L., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research, 40*, 686–693.
- 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.
- Meador, D., Flege, J., & MacKay, I. (2000). Factors affecting the recognition of words in a second language. *Bilingualism: Language and Cognition, 3*, 55–67.

- Miyake, A., Carpenter, P. A., & Just, M. A. (1994). A capacity approach of syntactic comprehension disorders: Making normal adults perform like aphasic patients. *Cognitive Neuropsychology*, *11*, 671–717.
- Miyake, A., & Friedman, N. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy & L. E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339–364). Mahwah, NJ: Erlbaum.
- Montgomery, J. (2000). Relation of working memory to offline and real time sentence processing in children with specific language impairment. *Applied Psycholinguistics*, *21*, 117–148.
- Neubauer, K., & Clahsen, H. (2009). Decomposition of inflected words in a second language. *Studies in Second Language Acquisition*, *31*, 403–435.
- Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience*, *17*, 1212–1228.
- Osaka, M., & Osaka, N. (1992). Language-independent working memory as measured by Japanese and English reading span tests. *Bulletin of the Psychonomic Society*, *30*, 287–289.
- Osterhout, L., McLaughlin, J., Pitkanen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A means for exploring the neurocognition of second language processing. *Language Learning*, *56*, 199–230.
- Osterhout, L., Poliakov, A., Inoue, K., McLaughlin, J., Valentine, G., Pitkanen, I., et al. (2008). Second-language learning and changes in the brain. *Journal of Neurolinguistics*, *21*, 509–521.
- 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.
- Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. Amsterdam: John Benjamins.
- Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Amsterdam: John Benjamins.
- Pinker, S. (1999). *Words and rules: The ingredients of language*. New York: Basic.
- Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Science*, *6*, 456–463.
- Roberts, R., & Gibson, E. (2002). Individual differences in sentence memory. *Journal of Psycholinguistics Research*, *31*, 573–598.
- Rossi, S., Gugler, M. F., Friederici, A. D., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, *18*, 2030–2048.
- Sagarra, N. (2007). Online processing of gender agreement in low proficient English–Spanish late bilinguals. In M. J. Cabrera, J. Camacho, V. Déprez, N. Flores, & L. Sánchez (Eds.), *Romance linguistics 2006: Selected papers from the 36th Linguistic Symposium on Romance Languages* (pp. 240–253). Amsterdam: 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*, 2022–2039.
- Salthouse, T. A. (1990). Working memory as a processing resource in cognitive aging. *Developmental Review*, *10*, 101–124.
- Sato, M., & Felser, C. (2010). Sensitivity to morphosyntactic violations in English as a second language. *Second Language*, *9*, 101–118.
- Segalowitz, N., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371–388). Oxford: Oxford University Press.
- Service, E., Simola, M., Mesanheimo, O., & Maury, S. (2002). Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology*, *14*, 383–408.
- Silva, R., & Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition*, *11*, 245–260.
- Steinhauer, K., White, E., & Drury, J. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research*, *25*, 13–41.
- Tirre, W. C., & Peña, C. M. (1992). Investigation of functional working memory in the reading span test. *Journal of Educational Psychology*, *84*, 462–472.
- Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar. *Studies in Second Language Acquisition*, *27*, 173–204.

- Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research: "Clozing" the gap. *Studies in Second Language Acquisition*, 33, 339–372.
- Tremblay, A., & Garrison, M. D. (2010). Cloze tests: A tool for proficiency assessment in research on L2 French. In M. T. Prior, Y. Watanabe, & S.-K. Lee (Eds.), *Selected proceedings of the 2nd Language Research Forum 2008* (pp. 73–88). Somerville, MA: Cascadia Press.
- Turner, M., & Engle, R. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127–154.
- Ullman, M. T. (2001). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research*, 30, 37–69.
- Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/ procedural model. *Cognition*, 92, 231–270.
- Van den Noort, M., Bosch, P., & Hugdahl, K. (2006). Foreign language proficiency and working memory capacity. *European Psychologist*, 11, 289–296.
- Van Wijngaarden, S. J., Steeneken, H. J. M., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *Journal of the Acoustical Society of America*, 111, 1906–1916.
- 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, 315–339.
- Waters, G., & Caplan, D. (1996a). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology*, 49A, 51–79.
- Waters, G., & Caplan, D. (1996b). Processing resource capacity and the comprehension of garden path sentences. *Memory and Cognition*, 24, 342–355.
- White, L., Valenzuela, E., Kozłowska-MacGregor, M., & Leung, I. Y.-K. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics*, 25, 105–133.
- Williams, J. N. (2011). Working memory. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 427–441). London: Routledge.