Robert V. Reichle*, Annie Tremblay and Caitlin Coughlin

# Working memory capacity in L2 processing

DOI 10.1515/probus-2016-0003

**Abstract:** In this paper, we review the current state of the second language (L2) processing literature and report data suggesting that this subfield should now turn its attention to working memory capacity as an important factor modulating the possibility of (near)-native-like L2 processing. We first review three major overarching accounts of L2 processing (Clahsen et al. 2006a, Grammatical processing in language learners. *Applied Psycholinguistics* 27. 3–42; Ullman 2001, The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research* 30. 37–69; McDonald 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; Hopp 2006, Syntactic features and reanalysis in near-native processing. *Second Language Research* 22. 369–397, and Hopp 2010, Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua* 120. 901–931) and frame their predictions in terms of the qualitative and quantitative differences in processing expected between native speakers and L2 learners. We next review event-related potential (ERP) research on L2 processing and argue that the field's current understanding of qualitative and quantitative differences in ERPs warrants an additional focus on variables other than L2 proficiency that can also predict individual differences in L2 processing. Recent L2 research (relying on ERPs, self-paced reading, and other online measures) suggests that the most promising such variable is working memory (WM) capacity. We summarize results from our recent L2 WM studies – and report new ERP findings – that point to the possibility of a modulatory effect of WM capacity on the nativelikeness of L2 processing. We conclude that the study of WM capacity is the logical next step for this L2 processing subfield.

**Keywords:** ERPs, second language processing, working memory, grammatical processing, focus structure

***Corresponding author: Robert V. Reichle,** Northern Illinois University, DeKalb, IL, USA; University of Texas at Austin, Austin, TX, USA, E-mail: robert.reichle@austin.utexas.edu
**Annie Tremblay:** E-mail: atrembla@ku.edu, **Caitlin Coughlin:** E-mail: c.coughlin@ku.edu, University of Kansas, Lawrence, KS, USA

# 1 Introduction

Over the past decade, different theoretical concerns have driven second/foreign language (L2) processing research. One primary question is whether "late" L2 learners (i. e., individuals who have begun learning the L2 after the ages of 10–12) are capable of processing the target language in a way that resembles native speakers' processing. Evidence for or against the possibility of native-like L2 processing is of inherent interest to the greater second language acquisition field and must be integrated with the predictions of dominant L2 processing accounts, thereby tying experimental findings to theory. While current L2 processing accounts differ in the details of their predictions, these predictions can be distilled to claims of qualitative or quantitative processing differences between native and L2 processing. Time-sensitive measures of L2 processing, such as reading times and event-related potentials (ERPs), provide critical evidence for or against such differences.

The influence of age of acquisition on facets of second language acquisition was once the primary motivators for L2 processing studies. However, in more recent years, the modulating effect of L2 proficiency in online sentence processing has become an increasingly important focus for investigation. As this effect has become better established, the field is now turning to previously unexplored variables that may predict additional individual variability in L2 sentence processing. One variable that has received considerable attention in recent L2 processing research, and on which we focus herein, is verbal working-memory (WM) capacity, as measured by reading-span tasks. As we will show, WM capacity can shed new light on whether or not L2 learners can reach native-like processing in the target language.

In this paper, we review the current state of the L2 processing literature – with special attention to ERP findings – and report data suggesting that this subfield of research should now turn its attention to WM capacity as an important factor modulating the possibility of near-native L2 processing. In Section 2, we review three major overarching accounts of L2 processing (Clahsen et al. 2006a; Ullman 2001; McDonald 2006; Hopp 2006; Hopp 2010) and frame their predictions in terms of the qualitative and quantitative differences in processing expected between native speakers and L2 learners. Section 3 reviews the ERP research on L2 processing and argues that the field's current understanding of qualitative and quantitative differences in ERPs warrants an additional focus on variables other than L2 proficiency that might also predict individual differences in L2 processing. In Section 4, we discuss recent L2 ERP and self-paced reading research suggesting that one promising such variable is WM capacity. In Section 5, we summarize results from our recent L2 WM studies, and in Section 6, we report new ERP findings that illustrate the modulatory effect of WM capacity on morphosyntactic processing.

We conclude by advancing the argument that the study of WM capacity is the logical next step for research on L2 processing.

## 2 Theoretical accounts of L2 processing

Three theoretical accounts deserve attention for motivating and expanding the current directions of L2 processing research. Clahsen and Felser's (2006a) Shallow Structure Hypothesis entails that grammatical processing in late L2 learners is qualitatively different from grammatical processing in native speakers, irrespective of learners' L1 or their L2 proficiency. Native speakers are said to be capable of constructing detailed, hierarchical syntactic parses, whereas L2 learners are said to rely on lexical and semantic information, building "local" (i. e., short-distance, linear) morphosyntactic dependencies but not complex (i. e., long-distance, hierarchical) ones. In other words, L2 parses are flat and "shallow" compared to the more complex parses of native speakers. For morphosyntactic agreement, although L2 learners are predicted to compute it only between adjacent or nearby constituents (Clahsen et al. 2006b), they are not predicted to decompose morphologically complex words (e. g., Clahsen et al. 2010; Neubauer and Clahsen 2009; Silva and Clahsen 2008). These predictions are all qualitative inasmuch as they predict a difference between native speakers and L2 learners in the type of processes underlying sentence comprehension rather than a difference in one of its quantifiable dimensions (e. g., latency and/or size of processing effects). This account can thus be refuted by data showing that the same types of processes underlie native and L2 sentence comprehension (i. e., behavioral or brain-response data that pattern similarly for native speakers and L2 learners).

Second, Ullman's (2001) Declarative/Procedural model of lexicon and grammar proposes that native speakers draw upon two separate domain-general memory systems in language processing: Declarative memory is relied on for the storage and retrieval of monomorphemic words, multi-word chunks (such as idioms), and words with irregular inflection, whereas procedural memory is relied on for the decomposition of morphologically complex words and the application of regular inflection. Ullman proposes that late L2 learners, on the other hand, have recourse to procedural memory only once they reach sufficiently high levels of L2 proficiency. This account then predicts a qualitative change in processing: As L2 proficiency increases, L2 learners' parsing of regularly inflected words relies less on declarative memory and more on procedural memory. This account also predicts a quantitative component: With increasing L2 proficiency and an increasing role of procedural memory, language processing effects gradually become more native-like. This account can be refuted by data showing that the same types of processes underlie native speakers' and low-level L2 learners’ sentence comprehension, as well as by data showing that different types of processes underlie native speakers’ and high-level L2 learners’ sentence comprehension (i. e., behavioral or brain-response data that pattern differently for native speakers and L2 learners).

The third relevant account is oriented toward the role of the cognitive processes called upon in language processing. McDonald (2006) and Hopp (2006, 2010) propose that the cognitive demands of L2 processing hinder access to WM and other attentional and decoding abilities, therefore leading to quantitatively non-native-like processing in the L2. These cognitive-resource-centric accounts do not propose that different processes or systems are engaged in L2 processing; instead, they suggest that the same systems used to process the native language (L1) are reduced in their effectiveness when processing the target language, because processes in the L2 are less automatized than native processes and L2 learners’ knowledge of another linguistic system (that of the L1) interferes with their use of these processes. In other words, this account predicts quantitative differences between native and L2 processing, with factors such as L2 proficiency, WM capacity, and other attentional and decoding processes modulating the size of these differences. This account can be refuted by data showing that different types of processes underlie native and L2 sentence comprehension.

## 3 Event-related potentials and L2 proficiency

If the above predictions can be falsified by evidence that signatures of L2 processing are similar to or differ in kind from those of native processing, then it is crucial to employ online measures that can specify the nature of the processes underlying L2 learners’ sentence comprehension. One such technique is the event-related potential (ERP) technique. This technique has been increasingly used in recent years to study L2 processing.

The ERP technique links electroencephalographic data (EEG) to the presentation of a stimulus. EEG data are recorded at the scalp using an array of electrodes. These data reflect the summation of post-synaptic potentials in the cortex. Data from this technique are measurements of voltage over time and are often represented as figures plotting voltage over time as a waveform. Patterns in EEG data that are reliably associated with specific language-related cognitive processes are referred to as ERP components. When aggregated EEG data display a significant difference between conditions, it is said to show an ERP effect. Comparisons of ERP effects can reveal whether differences are qualitative (e. g., ERP effects that differ in polarity) or quantitative (e. g., the amplitude or latency of effects differ, but their overall polarity is the same). Note that this technique provides data with an extremely high temporal resolution (on the order of milliseconds), but low spatial resolution.

Three ERP effects are particularly relevant to linguistic processing. The first is the N400, which is a negative increase in voltage with a right posterior or bilateral distribution on the scalp approximately 400 ms after the presentation of a lexical or semantic violation (e. g., *I took a bite of the *windmill*; Kutas and Hillyard 1980). In sentence processing, the amplitude of this effect is modulated by the cloze probability of the anomalous word, with more improbable words yielding larger N400 effects and with semantic priming reducing the size of these effects (Kutas and Hillyard 1984). Other sources of contextual information (e. g., discourse context, world knowledge, attitudes) also modulate N400 size (e. g., Hald et al. 2007; Nieuwland and Van Berkum 2006). The N400 has also been portrayed as a reflection of the effortful retrieval of conceptual information from long-term semantic memory (e. g., Kutas and Federmeier 2000).

The P600, on the other hand, is generally believed to index syntactic, rather than semantic, processes. The P600 (also sometimes referred to as the syntactic positive shift) is a late increase in positivity in the centro-parietal region of the scalp observed approximately 600 ms after the presentation of morphosyntactic violations (e. g., *The cats won't *eating the food*; Osterhout and Hagoort 1999; Osterhout and Holcomb 1992). A wide array of other morphosyntactic violations have been observed to elicit this effect, including phrase structure violations (e. g., *The scientist criticized Max's *of proof the theorem*), subject-verb agreement violations (e. g., *The elected officials *hopes to succeed*), and others (e. g., Friederici et al. 1996; Hagoort et al. 1993; Neville et al. 1991; Osterhout and Holcomb 1993; Osterhout et al. 1994, 1997; Osterhout and Mobley 1995; Osterhout and Nicol 1999).

Additionally, phrase structure and morphosyntactic violations are often associated with a left anterior negativity effect (LAN) between 150 and 500 ms (e. g., Friederici et al. 1993; Kluender and Kutas 1993; Münte et al. 1998a; Neville et al. 1991; Osterhout and Holcomb 1992). The LAN is thought to be an automatic initial response to morphosyntactic violations, and the later P600 is believed to be a function of controlled processing reflecting conscious reanalysis of those same violations (e. g. Ullman 2012). Notably, the LAN is present with less reliability and uniformity than the P600 (e. g., Friederici 2004; Neville et al. 1991; Steinhauer and Connolly 2008). It has also been suggested that the LAN effect reflects increased WM load stemming from the storage of fillers or other syntactic dependencies in WM (e. g., Cowles 2003; Kluender and Kutas 1993; Münte et al. 1998b).

The usefulness of the ERP technique lies not only in the insight it yields into language-related cognitive processes and the relative temporal order of language processing stages but also in the comparisons that can be drawn between native speakers and L2 learners. Recent work suggests that adult late L2 learners are indeed able to display native-like ERP effects at sufficiently high levels of proficiency (e. g., Gillon Dowens et al. 2010; McLaughlin et al. 2004, 2010; Ojima et al. 2005; Osterhout et al. 2006, 2008; Steinhauer et al. 2006; Tanner et al. 2009).

In a cross-sectional study, Steinhauer et al. (2006) investigated the processing of syntactic information by English monolinguals, French L2 learners of English, and Chinese L2 learners of English at different proficiencies. The L2 groups were split into low- and high-proficiency subgroups, and participants read English sentences containing phrase structure violations (e. g., *The man hoped to *<u>meal</u> the *enjoy with friends*). These phrase structure violations, specifically the first syntactically illicit word in the sentence, elicited a LAN effect followed by a P600 in the native speaker group. A LAN effect was also present for the high-proficiency French L2 learners of English, but not for the low-proficiency French-speaking subgroup. Both of these subgroups exhibited a P600 effect. For the Chinese L2 learners of English, however, the low-proficiency subgroup showed no signs of differential ERPs for phrase structure violations, and the high-proficiency subgroup showed a P600 but no LAN. Steinhauer et al. interpreted these results as supporting the possibility of native-like L2 processing among highly proficient adult learners. At the same time, given the greater typological similarity between English and French than between English and Chinese, the lack of a LAN for the highly proficient Chinese speakers raised the possibility that L1 transfer modulates nativelikeness of processing (for similar findings, see Foucart and Frenck-Mestre 2011; Rossi et al. 2006).

Recent work has also linked different L2 ERP effects to structure-specific L2 proficiency. Using a cross-sectional design, Tanner at al. (2009) found that English L2 learners of German who made more accurate grammaticality judgments on sentences containing verb agreement violations (e. g., *Ich *<u>trinkt</u> nie Alkohol* 'I never *drinks alcohol') exhibited P600 effects for these violations, whereas learners who were less accurate in their judgments exhibited N400 effects. These results suggest that the two subgroups of L2 learners may have obtained different ERP signatures as a result of attaining different structure-specific proficiencies. McLaughlin et al. (2010), Osterhout et al. (2006), and Osterhout et al. (2008) found similar ERP differences when dividing participants into subgroups of more accurate and less accurate L2 learners based on grammaticality judgment tasks. Similarly, Reichle (2010a) found increased positivities for correct responder subgroups versus incorrect responder subgroups on an acceptability task for anomalies in the syntactic constructions used to mark focus structure in native and L2 French. Importantly, Tanner et al. (2013) report results from a cross-sectional study of English L2 learners of German wherein an examination of individual responses to subject-verb agreement violations on a grammaticality judgment task revealed that most subjects showed either an N400 or P600 effect (but not both), whereas the same data in a group-level analysis showed a biphasic N400–P600 pattern. These results not only suggest that L2 proficiency on specific structures may correspond to L2 processing differences but also highlight the importance of closely examining individual variability in L2 ERP data.

Further evidence for structure-specific L2 proficiency effects comes from White et al. (2012). This study tracked L2 learners' ERP signatures in response to past tense violations (e. g., *The teacher did not *started the lesson*) as participants took part in an intensive English course. As expected, learners went from showing no ERP effect at the beginning of the course to showing a P600 effect at the end of the course 9 weeks later. Notably, inter-learner variability was predicted by structure-specific proficiency: Compared to low scorers on a behavioral task, high scorers displayed a P600 effect that more closely resembled that of native speakers. These and the above findings suggest that L2 learners' ERP signatures are closely linked to their knowledge of particular grammatical structures, as evidenced in their performance on behavioral tasks.

In light of the developmental sequences observed in L2 ERP studies, two overlapping accounts of L2 processing have been proposed. Steinhauer et al. (2009) characterize L2 processing in terms of ERPs that change dynamically as a function of L2 proficiency. This account highlights the role of L2 proficiency in modulating brain responses, in contrast with age of acquisition, which does not necessarily guarantee nativelikeness in processing. First, novice learners with no prior exposure to the target language do not display differential ERP effects when presented with morphosyntactic errors, because they do not perceive differences between grammatical and ungrammatical stimuli – a natural consequence of their lack of experience with the L2. As learners increase to a low level of proficiency, they begin to exhibit an N400 for morphosyntactic errors (though an N400 stage was not seen in White et al. 2012). At that stage, they rely on mechanisms of semantic or pragmatic plausibility or lexical retrieval to process morphosyntactic violations, which results in ERP effects associated with lexical-semantic processes. As proficiency continues to increase, L2 learners begin to show a small P600 and then a larger P600, suggesting that they rely increasingly on proceduralization and rule-governed processes and use these processes for late repair of morphosyntactic violations. Finally, at the highest levels of proficiency, L2 learners show both a LAN and a P600, a pattern that is identical to that typically obtained for native speakers. At that stage, L2 learners are fully engaging the same early automatic and late controlled processes of repair as native speakers. Note that this account posits the presence of LAN effects at fully native-like levels of processing. Since the laterality, and even the presence, of the LAN is not completely uniform among native speakers, this account predicts the presence of an LAN in L2 learners only when one is found for native speakers.

Osterhout et al. (2006) and McLaughlin et al. (2010) have interpreted the existing L2 ERP literature in a similar way and have posited a similar sequence of changing ERPs, wherein morphosyntactic violations elicit N400s prior to eliciting P600s throughout development. Crucially, in light of the highly variable findings relating to the presence of LAN effects in the L1, this account does not take the presence of a LAN in the L2 as a requisite condition for nativelikeness. McLaughlin et al. also highlight low-proficiency learners' greater reliance on chunking and declarative memory, and contrast this with the increasing role of proceduralization among learners at higher levels of proficiency. They present results from a longitudinal study in which L2 learners of French read real words and pseudowords that were incorrectly inflected using French's regular verb inflection paradigm (e. g., *ils *parnons*). Their results indicate that some learners displayed a P600 effect for incorrectly inflected pseudowords after only 9 months of instruction. McLaughlin et al. interpret this as evidence that learners may apply inflectional morphology rules at earlier stages of learning than previously thought.

These accounts contrast with Clahsen and Felser's (2006a) Shallow Structure Hypothesis, which predicts that L2 learners may not be able to attain native-like processing for areas of syntax that rely on complex hierarchical structures or long-distance dependencies. Importantly, they make this prediction irrespective of learners' L1 or their L2 proficiency. However, several ERP studies of long-distance dependency violations have found P600s in L2 learners (e. g., Gillon Dowens et al. 2010; Tanner et al. 2014), and ERP evidence on long-distance wh-dependencies in the L1 suggests that processing is not qualitatively different for dependencies of different lengths (Phillips et al. 2005).

## 4 The role of working memory in L2 processing

As illustrated above, it is now well established that important L2 processing changes take place with increasing L2 proficiency. Research on this topic is thus turning away from sole reliance on L2 proficiency as predictor of nativelikeness and is expanding its scope of investigation to other, finer-grained factors. One such factor is WM capacity. WM is defined as a collection of systems responsible for holding information in short-term memory while performing other tasks.

The first WM model, proposed by Baddeley and Hitch (1974), was comprised of two temporary store systems (i. e., a phonological loop for speech-based information and a visuospatial sketchpad for visuospatial information) and a central executive system in charge of allocating attention and controlling information in each store. In its most recent version, this WM model also includes an episodic buffer, which serves as interface between each system and long-term memory (Baddeley 2000). Research investigating WM effects on language-related cognitive processes distinguishes between two types of WM: non-verbal WM, which typically involves holding digits, letters, or words in short-term memory while performing an unrelated, non-verbal task (e. g., arithmetic), and verbal WM, which involves holding digits, letters, or words in short-term memory while processing language (also called a reading- or listening-span task).

In this paper, we focus on verbal WM. Furthermore, we assume that WM capacity is language independent – that is, it is a cognitive capacity that varies from individual to individual but that remains constant across languages (e. g., Osaka and Osaka 1992). Reading- and listening-span scores are lower in the L2 than in the L1, because processing the target language is less automatic and thus more effortful and taxing than processing the L1; however, everything else being equal, L2 reading- and listening-span scores co-vary with L1 reading- and listening-span scores, as we should expect if WM capacity is language independent (Alptekin and Erçetin 2010; Coughlin and Tremblay 2013; Harrington and Sawyer 1992; Service et al. 2002; Van den Noort et al. 2006).

Caplan and Waters (1999) propose that only post-interpretive processes (i. e., processes that use the meaning of a sentence to perform other operations) degrade as WM capacity decreases and as WM load increases; according to their separate-sentence-interpretation-resource model, interpretive processes (i. e., processes that assign a syntactic structure to sentences) are not differentially affected by WM capacity or WM load (see also Caplan and Waters 1995, 1996; Waters et al. 1996a, 1996b, 1996c). Caplan and Waters (1999) review a series of studies showing that readers with lower WM capacity or under a higher WM load do not show greater effects of syntactic complexity than readers with higher WM capacity or under no WM load. This view contrasts with that of the single-resource model, according to which all verbal tasks are performed by the same set of verbal processing resources (e. g., Just and Carpenter 1992; King and Just 1991; MacDonald et al. 1992; Miyake et al. 1994). Hence, the single-resource model entails that both interpretive and post-interpretive processes should be differentially affected by WM capacity or WM load.

The relevance of WM capacity for understanding whether or not L2 learners can show sensitivity to morphosyntactic violations such as agreement errors is clear: L2 learners should have to hold the first word of an agreement dependency (and its corresponding features) in WM and keep it activated until they encounter the second word in the agreement dependency. Notice that this should be true whether or not L2 learners decompose regularly inflected words into stem and affix or retrieve them as a whole word from their lexicon.

Studies on the processing of agreement morphology indicate that WM capacity can indeed predict native speakers' sensitivity to agreement violations. An ERP study in Dutch by Vos et al. (2001) found that manipulations of WM load and between-subject differences in verbal WM span (as assessed by a listening-span task) modulated the amplitude of the LAN in response to subject-verb agreement violations, with higher-span participants showing a larger amplitude LAN than lower-span participants; the same WM factors did not modulate the amplitude of the P600 in response to the same violations, but they did modulate its latency. Behavioral measures in the same study were not sensitive to WM manipulations, a discrepancy that highlights the utility of ERPs in this type of investigation.

WM capacity can also predict sensitivity to agreement violations among L2 learners. McDonald (2006) found that high-WM L2 learners of English from various language backgrounds performed better on an aural grammaticality judgment task containing several types of morphosyntactic violations, including agreement violations between adjacent words, than low-WM L2 learners; in her study, WM was assessed with a non-verbal WM task. Similarly, Sagarra (2007) found that English L2 learners of Spanish with higher WM scores, as measured by a reading-span task in English, understood Spanish sentences with within-phrase gender agreement violations (e. g., *La mujer lava la blusa *<u>blanco</u> en la cocina* 'The woman washes the *white-masc blouse-fem in the kitchen') more accurately than L2 learners with lower WM scores. However, WM was not found to be a reliable predictor of L2 learners' sentence comprehension for sentences that contained gender agreement violations across clauses (e. g., *La mujer lava la blusa que es *<u>blanco</u> en la cocina* 'The woman washes the blouse-fem that is *white-masc in the kitchen'). Note, however, that on these sentences, L2 learners also did not show a significant effect of grammaticality. Likewise, Sagarra and Herschensohn (2010) report that the WM scores of intermediate English L2 learners of Spanish (also assessed by a reading-span task in English) correlated positively with these L2 learners' reading times at the critical region of Spanish sentences with within-phrase gender agreement violations (e. g., *El ingeniero presenta el prototipo *<u>famosa</u> en la conferencia* 'The engineer presents the prototype-masc famous-fem at the conference') and with their grammaticality judgments on the same sentences. This indicates that these L2 learners were more sensitive to gender agreement violations with increasing WM capacity, suggesting that WM capacity can be a reliable predictor of L2 learners' sensitivity to agreement violations between adjacent words, at least for within-phrase agreement dependencies.

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 Dutch, where verbal agreement disambiguated between the two types of clauses. The researchers manipulated the length of agreement dependencies, such that an adverbial phrase intervened or did not intervene between the second noun in the relative clause and the embedded 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 showed that only the L2 learners with a high span (as measured by Dutch and German versions of a reading-span task) had more difficulty processing object relative clauses as compared to 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 when the two agreeing words were relatively close, but not when they were distant. One possible explanation for the absence of WM effects in long-agreement dependencies is that they may have been too taxing for WM to be a reliable predictor of processing difficulty.

More recently, Coughlin and Tremblay (2013) examined the role of proficiency (as assessed by a cloze test; for details, see Tremblay 2011) and WM capacity in the processing of number violations between object clitics and their left-dislocated antecedents by L1 English L2 learners of French. They manipulated the length of the agreement dependencies such that the clitic was close to or distant from its antecedent (e. g., *Ces fruits (avant l'entretien) Marie *le mangera (avant l'entretien) pour sa collation* 'These fruits [before the interview] Marie will eat *it [before the interview] for her snack'). WM was assessed in both English and French, and word-recall scores in French – which correlated with the word-recall scores in English but not with proficiency in French – were used as predictor of the grammaticality effect. Coughlin and Tremblay found that high-proficiency L2 learners showed more sensitivity to agreement violations than mid-proficiency L2 learners for both short- and long-agreement dependencies. They also reported that WM capacity was a modest predictor of the effect of grammaticality for short dependencies with plural clitics and long-agreement dependencies with singular clitics. They attributed the lack of significant effects for the remaining two conditions (singular clitics in short dependencies and plural clitics in long dependencies) to the fact that these two conditions may have been, respectively, too easy and too taxing for the L2 learners. In other words, simple structures may not generate sufficient variability to be predicted by WM capacity, whereas complex structures may be calling upon additional cognitive resources that are not captured by WM capacity.

The previous studies indicate that WM capacity can be a reliable predictor of L2 learners' sensitivity to agreement violations at least in short dependencies, suggesting that processing agreement dependencies places a small burden on WM even when the two agreeing words are within the same syntactic phrase. Note, however, that these results cannot be taken as evidence for separate-sentence-interpretation-resource WM models or single-resource WM models, in part because most significant relationships were found in short-agreement dependencies, for which neither WM theory predicts any difficulty. While not all agreement studies have shown a significant relationship between L2 processing and WM capacity (e. g., Foote 2011), the fact that a number of studies have indeed found such a relationship suggests that the role of WM capacity in L2 processing warrants further investigation.

# 5 The effect of WM capacity on L2 ERP signatures of subject-verb agreement

In our work, we further investigated whether L2 processing is quantitatively or qualitatively different from native processing (as a function of L2 proficiency) and whether WM capacity is a reliable predictor of L2 learners' sensitivity to agreement. We did so by examining L2 learners' ERP signatures. Here, we report new analyses of data previously reported by Reichle et al. (2013), in which we recorded ERPs for subject-verb agreement violations in French.

## 5.1 Methodology

### 5.1.1 Participants

Participants ($n$ = 10; 7 females) were mid-proficiency English-speaking L2 learners of French. We assessed participants' L2 proficiency with a self-report language background questionnaire and a cloze test (Tremblay 2011), and we assessed WM capacity in both French and English with a reading-span task (Coughlin and Tremblay 2013). We used the participants' word-recall scores (shown in Table 1) as index of WM capacity, as these scores did not correlate with proficiency in French. Language background information, proficiency

**Table 1:** L2 Learners' language background information (adapted from Reichle et al. 2013).

| | First exposure to French (age) | Instruction on French (years) | Residence in French-Speaking Environment (months) | Weekly use of French (%) | Cloze test scores (/45) | Word span English | Word span French |
|---|---|---|---|---|---|---|---|
| Mean | 13.5 | 6.8 | 0.75 | 15.8 | 20.8 | 3.25 | 2.65 |
| SD | 2.6 | 3.1 | 1.3 | 8.7 | 5.7 | 1.44 | 0.97 |
| Range | 10–18 | 3–12 | 0–4 | 3–30 | 12–31 | 1–5 | 1–4.5 |

scores, and reading-span scores are provided in Table 1. The participants' average proficiency score was that typically obtained by third-year university students in French. Hence, our participants were roughly at intermediate levels of proficiency.

### 5.1.2 Materials

In the ERP experiment, participants read sentences that contained subject-verb agreement errors either between adjacent constituents in short-agreement dependencies (e. g., *Pour ce grand poète, les filles *rôtit un gros jambon* 'For this great poet, the daughters *roast a big ham') or between non-adjacent constituents in long-agreement dependencies (e. g., *Les filles de ce grand poète *rôtit un gros jambon* 'The daughters of this great poet *roast a big ham'). A true-or-false comprehension question (rather than an acceptability judgment) appeared after 25 % of the test items. The experiment included 120 experimental items and 120 filler sentences.

### 5.1.3 Procedures

We recorded participants' ERPs over two sessions – the first one in L2 French and the second one in L1 English. In the first session, participants saw all ERP test items in L2 French; they also completed a language background questionnaire, a cloze test (Tremblay 2011), and a verbal WM task in French (Coughlin and Tremblay 2013). The cloze test was a fill-in-the-blank task where participants completed a paragraph requiring a range of French lexical and grammatical knowledge. In the WM task (adapted from Waters et al. 1996a), participants read 56 sentences and made semantic acceptability judgments about them while simultaneously keeping the sentence-final words in WM for variable spans. Approximately two weeks later, participants returned to the lab for a second session in English, during which ERPs were recorded for equivalent items in English and their WM capacity in English was assessed. The L2 version of the task was administered first so that participants would not obtain any metalinguistic information about the task by doing it first in their L1 and then be influenced by it in the L2.

Scalp EEG data were recorded from nine electrode channels (Fz, F3, F4, Cz, C3, C4, Pz, P3, P4) referenced to the linked average of the right and left mastoid. The data were acquired using amplifiers and software from Biopac Systems, Inc.; additional analyses were conducted using EEGLAB (Delorme and Makeig 2004). Recordings from electrodes placed above and below the right eye (VEOG channel) were used to detect trials with eye blinks, which were rejected from analysis. A right earlobe electrode served as ground, and all data were subjected to an analog bandpass filter from 0.1 to 35 Hz. All impedances were kept below 5 kΩ, and the data were sampled at a rate of 250 Hz. Mean amplitude EEG data were analyzed in the 300–500 ms and 500–800 ms time windows to look for the presence of N400 and P600 responses, respectively.

## 5.2 Results and discussion

Our results showed different ERP effects for the L1 (English) and the L2 (French): In English, mixed-effects models revealed a P600 effect for agreement violations in short dependencies (grammaticality × dependency length: $t[16,192] = -28.73$, $p < 0.001$) and a N400–P600 effect for agreement violations in long dependencies (grammaticality × dependency length: $t[16,192] = -20.19$, $p < 0.001$). In French, group-level analyses showed N400 effects for both types of violations, and for non-adjacent violations, the size of these effects increased marginally as participants' WM capacity in English increased ($t[4,495] = -1.69$, $p < 0.091$) (for a visual presentation of the results, see Reichle et al. 2013).

We now report additional analyses of the Reichle et al. (2013) data, examining individual differences in ERPs. Mean amplitude data were calculated for the 300–500 ms and 500–800 ms time windows for all four conditions (short- vs. long-agreement dependency, violation vs. no violation). The difference in mean amplitude between violation and no-violation conditions was calculated for each scalp electrode site, and a grand average of these values across all electrode sites was calculated for each participant, yielding a measure of the overall positive or negative shift for each individual participant, one that reflects qualitative individual differences regardless of scalp distribution (see Table 2; note that English data are

**Table 2:** Individual differences in ERPs to subject-verb agreement violations (new analyses of data from Reichle et al. 2013).

| **Part. #** | **English** | | | | | **French** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Word span English** | **Short dependency** | | **Long dependency** | | **Word span French** | **Short dependency** | | **Long dependency** | |
| | | **300–500 ms** | **500–800 ms** | **300–500 ms** | **500–800 ms** | | **300–500 ms** | **500–800 ms** | **300–500 ms** | **500–800 ms** |
| 1 | 3.5 | +2.79 | +4.21 | −2.28 | < \|2\| | 2.5 | −2.39 | < \|2\| | < \|2\| | −3.57 |
| 2 | 5.0 | < \|2\| | +2.21 | < \|2\| | < \|2\| | 4.5 | < \|2\| | −2.54 | −2.86 | −4.07 |
| 3 | 1.0 | < \|2\| | +6.52 | < \|2\| | +5.78 | 1.0 | < \|2\| | +3.02 | −2.31 | −3.82 |
| 4 | 1.5 | −4.47 | −3.83 | < \|2\| | +2.25 | 1.5 | −5.54 | −7.66 | < \|2\| | < \|2\| |
| 5 | 5.0 | +3.17 | +5.36 | < \|2\| | < \|2\| | 2.5 | < \|2\| | < \|2\| | −4.62 | −3.55 |
| 6 | 2.0 | < \|2\| | +4.08 | < \|2\| | < \|2\| | 3.0 | < \|2\| | −2.55 | < \|2\| | < \|2\| |
| 7 | 2.5 | < \|2\| | +3.09 | −2.25 | −2.67 | 2.5 | < \|2\| | < \|2\| | −2.45 | < \|2\| |
| 8 | | | | | | 3.5 | −8.23 | −7.59 | −5.23 | −7.59 |
| 9 | 3.5 | < \|2\| | +2.87 | < \|2\| | < \|2\| | 3.0 | −4.96 | −4.84 | < \|2\| | < \|2\| |
| 10 | 4.0 | < \|2\| | +4.42 | −2.93 | < \|2\| | 2.5 | −4.92 | −3.32 | < \|2\| | < \|2\| |

not reported for Participant 8 due to an EEG recording error). For the sake of categorizing the shifts as positive or negative, we were conservative and considered any difference smaller than the absolute value of 2 μv not meaningful.

The English data were fairly homogenous for agreement violations in short dependencies, with 8 participants displaying a positivity in the 500–800 ms time window and 2 of them displaying a positivity also in the 300–500 ms time window. For agreement violations in long dependencies, 2 participants showed a negativity in the 300–500 ms time window, 2 participants showed a positivity in the 500–800 ms time window, and 1 participant showed a negativity in both time windows. This suggests that the N400–P600 effects observed in the group results came from different participants rather than being biphasic in nature. Furthermore, across agreement conditions, not all participants showed the same ERP effect – that is, a late positivity for short-distance agreement violations often did not correspond to a late positivity for long-distance agreement violations. While these between-subject differences in L1 ERPs differ from the dependency-driven latency modulations of Phillips et al. (2005), they are consistent with the finding that ERPs for grammatical errors often have individual differences that are masked by grand averages (e. g. Kutas and Hillyard 1983; King and Kutas 1995; Vos et al. 2001).

By contrast, the L2 French data for agreement violations in short dependencies show that 5 participants displayed a negativity in the 300–500 ms time window, 6 participants displayed a negativity in the 500–800-ms time window, and only 1 participant displayed a positivity in the 500–800 ms time window; for long-distance agreement violations, 5 participants showed a negativity in the 300–500 ms time window and 5 participants showed a negativity in the 500–800 ms time window. These results suggest that only one of the L2 learners (Participant 3, in bold) may be in the process of developing target-like sensitivity to agreement violations in the short-distance condition (though not in the long-distance condition). The individual variability observed herein is consistent with Tanner et al.'s (2013) and Reichle's (2010a) work on individual differences in L2 ERPs and reinforces the notion that finer-grained measures and analyses of individual variables (such as structure-specific proficiency and WM capacity) may be necessary to more fully account for variability in L2 processing.

L2 learners' N400 effects in response to morphosyntactic violations in L2 French but their P600 in response to morphosyntactic violations in L1 English (at least in short-agreement dependencies) represents a qualitative difference between L1 and L2 processing. Assuming that most of our L2 learners were not sufficiently proficient to show a P600 effect, these results can be argued to be compatible with Ullman's (2001) Declarative/Procedural model for lower proficiency L2 learners. These results would also not falsify Clahsen and Felser's (2006a) Shallow Structure Hypothesis. Another contributing factor for the absence of P600s may be the lack of acceptability judgment accompanying the EEG recording. Such tasks have been seen to modulate P600 effect sizes compared to a no-task condition (e. g., Gascon et al. 2011). It is thus possible that the true/false task presented to participants after 25 % of questions yielded weaker effects than an acceptability task appearing after 100 % of items would have.

Although the N400 effect we found is not compatible with the predictions of McDonald (2006), we interpreted the relationship between WM capacity and the size of the N400 effect in L2 French as evidence that access to cognitive resources modulates L2 processing. Here, WM tested in the L1 was a better predictor of ERPs than WM tested in the L2. This is perhaps due to the taxing nature of verbal WM tasks, which further increase processing demands when L2 learners already have difficulty parsing sentences in the L2. Our participants indeed showed lower and less variable word-recall scores in French ($M$ = 2.55/5; SD = 0.96) than in English ($M$ = 3.37/5, SD = 1.42). Such a problem could be avoided by examining higher-proficiency learners. Since all of our participants lived in the United States at the time of testing, and in light of previous research suggesting that L2 learners can eventually display native-like L2 ERP signatures, we concluded that our data were most in line with an account that posits both a qualitative change in L2 processing across proficiency levels and quantitative differences between native and L2 processing as a function of cognitive-resource variables such as WM capacity.

# 6 The effect of WM capacity on L2 ERP signatures of focus processing

We now present results from a recent, related study that also used the ERP technique to investigate the relationship between WM capacity and the ERP signatures of L2 processing. This time, we examined L2 learners' processing of focus marking in L2 French. Focus marking indicates new or relevant information about a discourse topic, and in French it is very frequently indicated by the use of syntactic structures such as the c'est 'it is' cleft, in which a focal argument is coindexed by a copula clause and a relative clause (e. g., *C'est le cuisinier qui fait la soupe* 'it is the chef who makes the soup'). Focus marking is a useful structure for investigation in L2 processing for multiple reasons. As a syntax-pragmatics interface phenomenon, it is likely to cause processing difficulties for L2 learners (following Sorace and Serratrice 2009, *inter alia*). It also places a load on working memory as focal status is tracked across discourse. Cowles (2003) found that native English speakers showed a sustained LAN effect corresponding to an increased WM load for contrastive focus, which is more WM-intensive than informational focus. Reichle (2010b) found an early negativity under similar conditions in L2 French, and this negativity was present for L2 learners of high proficiency (Reichle and Birdsong 2014). For the present study, we predicted that in discourse contexts where focus is licensed, a similar LAN would index the increased WM load required to track the focal referent across discourse, as compared to sentences with extraneous focus marking. We also predicted that these LANs would be modulated by WM capacity and L2 proficiency, such that learners with higher WM capacity and higher L2 proficiency would show larger LANs than learners with lower WM capacity and lower L2 proficiency.

## 6.1 Methodology

### 6.1.1 Participants

The same English-speaking L2 learners of French from Reichle et al. (2013) participated in this study (see Section 5.1). All participants completed the same proficiency and WM tasks (see Table 1).

### 6.1.2 Materials

The filler items for Reichle et al.'s agreement study served as experimental items for this focus study. In the ERP task, participants saw sentences instantiating licensed and extraneous focus. Test items, based on items used by Cowles (2003) and exemplified in Table 3, consisted of a context paragraph followed by a question that either overtly licensed focus in the response or did not; a target

**Table 3:** Experimental items.

| Context | Focus contexts | | Response (Target sentence) |
|---|---|---|---|
| | Licensed | Extraneous | |
| A baker, a chef, and a waiter are employed in a big restaurant. At 11 pm, they all want to go home, but someone needs to make a special soup for the next day. | Who is it exactly who makes the soup? | Who makes the soup? | It is the chef who makes the soup. |
| Un boulanger, un cuisinier et un serveur sont employés dans un grand restaurant. A 23h, ils veulent rentrer chez eux, mais quelqu'un doit faire une soupe spéciale pour le lendemain. | Qui c'est qui fait la soupe? | Qui fait la soupe? | C'est le cuisinier qui fait la soupe. |

sentence answering the question was then presented visually one word at a time. There were 30 items in the licensed focus condition and 30 items in the extraneous focus condition, as well as 180 fillers (which included the agreement sentences from Reichle et al. 2013). A true-or-false comprehension question appeared after 25 % of items.

### 6.1.3 Procedures

The EEG recording procedures were identical to those for the experiment reported in Section 5.1.3. Mean amplitude EEG data were analyzed in the 300–500 ms time window following the visual presentation of the clefted noun in the target sentence (e. g., *C'est le* <u>*cuisinier*</u> *qui fait la soupe*). The data were subjected to repeated-measures 2 (condition) × 3 (laterality) × 3 (anteriority) ANOVAs, with Greenhouse–Geisser correction applied in cases of violated sphericity. We also used L2 proficiency and word span in English and French as covariates in separate analyses, in which we looked for linear relationships between these individual variables and the effect of condition.

## 6.2 Results and discussion

Our results for the English data showed that the processing of licensed focus compared to extraneous focus elicited a lateralized and marginally anterior effect of condition (condition × hemisphere: $F[1.532,12.253] = 8.233$, $p < 0.008$, partial $\eta^2 = 0.507$; condition × anteriority: $F[1.201,9.610] = 3.980$, $p < 0.07$, partial $\eta^2 = 0.332$). The effect was largest at sites F3 and Fz, as shown in Figure 1.

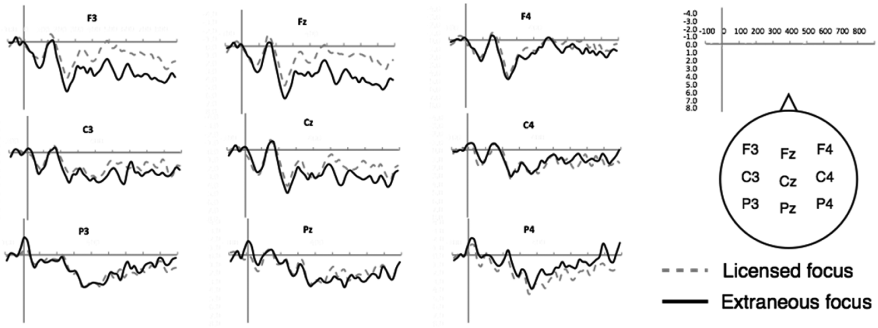


**Figure 1:** Average ERPs at clefted English nouns in licensed versus extraneous focus conditions.

We interpret this lateral and anterior negative shift as a LAN effect, which supports our hypothesis that a LAN would index the increased WM load required to track focal status across discourse. While the sustained nature of the negativity differs from the usual latency of morphosyntax-related LANs, its classification as a LAN is consistent with Cowles' (2003) finding of an anterior negative shift that was sustained for the duration of the first three words in instantiations of contrastive versus informational focus, and which she also interpreted as being a LAN reflecting increased WM load. Notably, a covariate analysis revealed a marginal modulatory effect of English WM capacity (condition × hemisphere × English word span: $F[1.786,12.505] = 3.664$, $p < 0.06$, partial $\eta^2 = 0.344$), a finding that is consistent with the notion that this individual variable has a linear relationship with indices of language processing.

The L2 French data for the equivalent focus items revealed the presence of a widespread early negative shift in voltage for licensed focus. Average waveforms are presented in Figure 2 (we do not report individual results due to the fact that there was only one effect of interest, as opposed to the competing N400 and P600s of the previous study). This effect trended toward significance (condition: $F[1,9] = 3.362$, $p < 0.1$, partial $\eta^2 = 0.272$) and was neither lateralized nor anterior (for condition × hemisphere and for condition × anteriority, $F < 1$). There was no relationship between L2 WM capacity and effect size ($F < 1$).

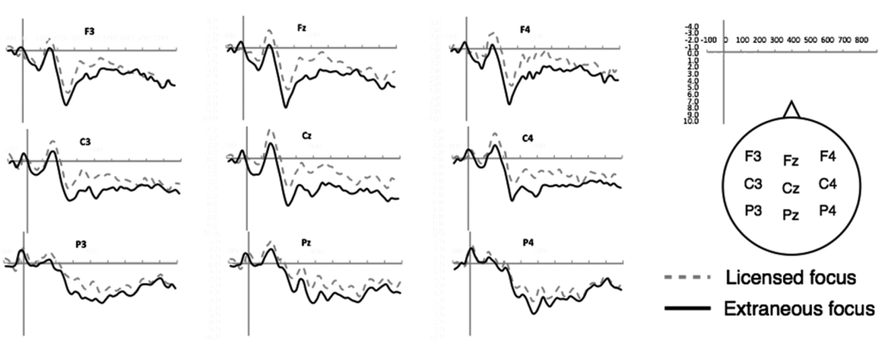


**Figure 2:** Average ERPs at clefted French nouns in licensed versus extraneous focus conditions.

The finding that participants displayed a LAN in their L1 but a more widespread, bilateral negativity trending toward significance in the target language is consistent with findings from the L2 morphosyntactic processing ERP literature. Steinhauer et al. (2009) and Pakulak and Neville (2010), among others, point out that there is a high degree of variability in the lateralization of LANs as a function of both native and L2 proficiency. The widespread nature of this early L2 negativity may therefore reflect the participants' intermediate level of proficiency. The negativity's bilateral distribution also reinforces the notion that phenomena that are more difficult to process, such as syntax-pragmatics interface phenomena, may elicit more variable (and possibly less-native-like) ERPs in the target language.

WM capacity positively correlated with the size of the LAN effect in English. This modulatory effect of WM has also been seen for verb agreement violations in L1 Dutch by Vos et al. (2001). The results from Reichle et al. (2013) similarly showed a positive relationship between an ERP effect size for L2 morphosyntactic violations (in this case, an N400) and WM in English. Taken together, these results suggest that higher WM participants may be more successful in keeping the status of a referent in WM. Such an interpretation is consistent with the cognitive-resource-centric accounts of McDonald (2006) and Hopp (2006, 2010). Notice that the relationship between WM and ERP amplitude may have been predicted to pattern in the opposite direction: The size of ERP effects may have been expected to decrease as WM increases, with higher-span individuals being better at processing the ungrammatical or more resource-intensive structures than lower-span individuals. The positive relationship we observed is instead in line with the notion that higher-span individuals may be more sensitive to linguistic information and make more efficient use of it when it is congruent with expectations; as a result of this sensitivity, they may be more easily disrupted by linguistic information that is incongruent with expectations (i. e., agreement violations, extraneous focus). This of course raises the possibility that verbal WM tasks such as that used in this and other studies are a proxy for experienced-based measure of language-processing skills (MacDonald and Christiansen 2002), something that should be investigated in further research.

# 7 Conclusion

The accounts put forward by Clahsen and Felser (2006a), Ullman (2001), and McDonald (2006) and Hopp (2006, 2010) have made valuable contributions to our understanding of L2 processing and, more generally, to the field of second language acquisition. The ensemble of evidence coming from ERP and other online measures paints a picture of L2 processing wherein L2 learners at sufficiently high proficiencies and native speakers may show qualitatively similar processing signatures for morphosyntax, with the degree of nativelikeness being contingent on global L2 proficiency and on language-related cognitive resources such as WM capacity. We have summarized the results of two of our studies examining whether WM capacity modulates the ERP responses that L2 learners show in sentence comprehension, specifically in their computation of agreement dependencies and their sensitivity to agreement violations, as well as in their computation of more complex syntactic structures such as cleft sentences. These findings suggest that WM capacity, which is not language-specific, may affect L2 learners' processing of agreement dependencies. These findings highlight the importance of investigating WM capacity in L2 sentence comprehension. They are also most parsimonious with L2 processing accounts that posit quantitative modulations in the efficiency of processing mechanisms. With further research, our understanding of WM effects on sentence processing will continue to improve, as will our greater knowledge of the role of individual variables in L2 processing.

**Acknowledgments:** This work was supported by the Northern Illinois University Center for the Interdisciplinary Study of Language and Literacy, the Northern Illinois University Center for the Study of Family Violence and Sexual Assault, and the Department of French at the University of Illinois. We would like to thank two anonymous reviewers for their valuable comments.

## References

Alptekin, Cem & Gülcan Erçetin. 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.

Baddeley, Alan D. & Graham Hitch. 1974. Working memory. In G. Bower (ed.), *The psychology of learning and motivation* 8. 47–90. New York: Academic Press.

Baddeley, Alan D. 2000. The Phonological loop and the irrelevant speech effect: Some comments on Neath (2000). *Psychonomic Bulletin and Review* 7. 544–549.

Caplan, David & Gloria S. Waters. 1999. Verbal working memory and sentence comprehension. *Behavioral Brain Science* 22. 77–94.

Caplan, David & Gloria S. Waters. 1995. Aphasic disturbances of syntactic comprehension and working memory capacity. *Cognitive Neuropsychology* 12. 637–649.

Caplan, David & Gloria S. Waters. 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, Harald & Claudia Felser. 2006a. Grammatical processing in language learners. *Applied Psycholinguistics* 27. 3–42.

Clahsen, Harald & Claudia Felser. 2006b. Continuity and shallow structures in language processing. *Applied Psycholinguistics* 27. 107–126.

Clahsen, Harald, Claudia Felser, Kathleen Neubauer, Mikako Sato & Renita Silva. 2010. Morphological structure in native and nonnative language processing. *Language Learning* 60. 21–43.

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

Cowles, H. Wind. 2003. *Processing information structure: Evidence from comprehension and production*. La Jolla, CA: University of California, San Diego dissertation.
Delorme, Arnaud & Scott Makeig. 2004. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods* 134. 9–21.
Foucart, Alice & Cheryl Frenck-Mestre. 2011. Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1-L2 syntactic similarity. *Bilingualism: Language and Cognition* 14. 379–399.
Foote, Rebecca. 2011. Integrated knowledge of agreement in early and late English–Spanish bilinguals. *Applied Psycholinguistics* 32. 187–220.
Friederici, Angela D., Erdmut Pfeifer & Anja Hahne. 1993. Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research* 1. 183–192.
Friederici, Angela D. 2004. The neural basis of syntactic processes. In M. S. Gazzaniga (ed.), *The cognitive neurosciences III*, 789–801. Cambridge, MA: MIT Press.
Friederici, Angela D., Anja Hahne & Axel Mecklinger. 1996. Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology. Learning, Memory, and Cognition* 22. 1219–1248.
Gascon, Alex, Véronique Lebel, Phaedra Royle, John E. Drury & Karsten Steinhauer. 2011. Task effects on ERPs for semantics and gender in French. Presented at NTPLX, Donostia-San-Sebastian.
Gillon Dowens, Margaret, Marta Vergara, Horacio Barber & Manuel Carreiras. 2010. Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience* 22. 1870–1887.
Hagoort, Peter, Colin Brown & Jolanda Groothusen. 1993. The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes* 8. 439–483.
Hald, Lea A., Esther G. Steenbeek-Planting & Peter Hagoort. 2007. The interaction of discourse context and world knowledge in online sentence comprehension: Evidence from the N400. *Brain Research* 1146. 210–218.
Harrington, Michael & Mark Sawyer. 1992. L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition* 14. 25–38.
Havik, Else, Leah Roberts, Roeland Van Hout, Robert Schreuder, & Marco Haverkort. 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, Holger. 2006. Syntactic features and reanalysis in near-native processing. *Second Language Research* 22. 369–397.
Hopp, Holger. 2010. Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua* 120. 901–931.
Just, Marcel Adam & Patricia A. Carpenter. 1992. A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* 99. 122–149.
King, Jonathan & Marcel Adam Just. 1991. Individual differences in syntactic processing: The rate of working memory. *Journal of Memory and Language* 30. 580–602.
King, Jonathan W. & Marta Kutas. 1995. Who did what and when? Using word-and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience* 7(3). 376–395.
Kluender, Robert & Marta Kutas. 1993. Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience* 5. 196–214.

Kutas, Marta & Kara D. Federmeier. 2000. Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science* 4. 463–470.
Kutas, Marta & Steven A. Hillyard. 1980. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* 207. 203–205.
Kutas, Marta & Steven A. Hillyard. 1983. Event-related brain potentials to grammatical errors and semantic anomalies. *Memory and Cognition* 11(5). 539–550.
Kutas, Marta & Steven A. Hillyard. 1984. Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307. 161–163.
MacDonald, Maryellen C., Marcel Adam. Just & Patricia A. Carpenter. 1992. Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology* 24. 56–98.
MacDonald, Maryellen C. & Morten H. Christiansen. 2002. Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review* 109. 35–54.
McDonald, Janet 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.
McLaughlin, Judith, Lee Osterhout & Albert Kim. 2004. Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nature Neuroscience* 7. 703–704.
McLaughlin, Judith, Darren Tanner, Ilona Pitkänen, Cheryl Frenck-Mestre, Kayo Inoue, Geoffrey Valentine, Lee Osterhout. 2010. Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning* 60. 123–150.
Miyake, Akira, Patricia A. Carpenter & Marcel Adam Just. 1994. A capacity approach of syntactic comprehension disorders: Making normal adults perform like aphasic patients. *Cognitive Neuropsychology* 11. 671–717.
Münte, Thomas F., Hans-Jochen Heinze, Mike Matzke, Bernadina M. Wieringa & Sönke Johannes. 1998a. Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia* 36. 217–226.
Münte, Thomas F., Kolja Schiltz & Marta Kutas. 1998b. When temporal terms belie conceptual order. *Nature* 395. 71–73.
Neubauer, Kathleen & Harald Clahsen. 2009. Decomposition of inflected words in a second language: An experimental study of German participles. *Studies in Second Language Acquisition* 31. 403–435.
Neville, Helen, Janet L. Nicol, Andrew Barss, Kenneth I. Forster & Merrill F. Garrett. 1991. Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience* 3. 151–165.
Nieuwland, Mante S. & Jos J. A. Van Berkum. 2006. When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience* 18. 1098–1111.
Ojima, Shiro, Hiroki Nakata & Ryusuke Kakigi. 2005. An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience* 17. 1212–1228.
Osaka, Mariko & Naoyuki Osaka. 1992. Language-independent working memory as measured by Japanese and English reading span tests. *Bulletin of the Psychonomic Society* 30. 287–289.
Osterhout, Lee, Phillip J. Holcomb & David A. Swinney. 1994. Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. *Journal of Experimental Psychology. Learning, Memory, and Cognition* 20. 786–803.

Osterhout, Lee, Judith McLaughlin, Ilona Pitkänen, Cheryl Frenck-Mestre & Nicola Molinaro. 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, Lee, Andrew Poliakov, Kayo Inoue, Judith McLaughlin, Geoffrey Valentine, Ilona Pitkanen, Cheryl Frenck-Mestre & Julia Hirschensohn. 2008. Second-language learning and changes in the brain. *Journal of Neurolinguistics* 21. 509–521.

Osterhout, Lee, Judith McLaughlin & Michael Bersick. 1997. Event-related brain potentials and human language. *Trends in Cognitive Sciences* 1. 203–209.

Osterhout, Lee & Peter Hagoort. 1999. A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King and Kutas (1998) in the P600/SPS-P300 debate. *Language and Cognitive Processes* 14. 1–14.

Osterhout, Lee & Phillip J. Holcomb. 1992. Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language* 31. 785–806.

Osterhout, Lee & Phillip J. Holcomb. 1993. Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes* 8. 413–438.

Osterhout, Lee & Linda A. Mobley. 1995. Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language* 34. 739–773.

Osterhout, Lee & Janet Nicol. 1999. On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes* 14. 283–317.

Pakulak, Eric & Helen J. Neville. 2010. Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience* 22. 2728–2744.

Phillips, Colin, Nina Kazanina & Shani H. Abada. 2005. ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research* 22(3). 407–428.

Reichle, Robert V. 2010a. Judgments of information structure in L2 French: Nativelike performance and the critical period hypothesis. *International Review of Applied Linguistics in Language Teaching* 48. 53–85.

Reichle, Robert V. 2010b. Near-nativelike processing of contrastive focus in L2 French. In B. VanPatten & J. Jegerski (eds.), *Research in second language processing and parsing*. 321–344. Amsterdam: John Benjamins.

Reichle, Robert V. & David Birdsong. 2014. Processing focus structure in L1 and L2 French: L2 proficiency effects on ERPs. *Studies in Second Language Acquisition* 36(3). 535–564.

Reichle, Robert V., Annie Tremblay & Caitlin E. Coughlin. 2013. Working-memory capacity effects in the processing of non-adjacent subject-verb agreement: An event-related brain potentials study. In E. Voss, S. J. D. Tai & Z. Li (eds.), *Selected proceedings of the 2011 second language research forum: Converging theory and practice*. Somerville, MA: Cascadilla Proceedings Project.

Rossi, Sonja, Manfred F. Gugler, Angela D. Friederici & Anja Hahne. 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, Nuria. 2007. Online processing of gender agreement in low proficient English–Spanish late bilinguals. In M. J. Cabrera, J. Camacho, V. Deprez, N. Flores & L. Sanchez (eds.), *Romance linguistics 2006: Selected papers from the 36th Linguistic Symposium on Romance Languages*. 240–253. Amsterdam: John Benjamins.

Sagarra, Nuria & Julia Herschensohn. 2010. The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua* 120. 2022–2039.
Service, Elisabet, Marjut Simola, Oili Mesänheimo & Sini Maury. 2002. Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology* 14. 383–408.
Silva, Renita, & Harald Clahsen. 2008. Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition* 11. 245–260.
Sorace, Antonella & Ludovica Serratrice. 2009. Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism* 13. 195–210.
Steinhauer, Karsten, Erin White, Erin King, Sonia Cornell, Fred Genesee & Lydia White. 2006. The neural dynamics of second language acquisition: Evidence from Event-Related Potentials. *Journal of Cognitive Neuroscience* supplement. 99.
Steinhauer, Karsten & John F. Connolly. 2008. Event-related potentials in the study of language. In B. Stemmer & H. A. Whitaker (eds.), *Handbook of the neuroscience of language*. 91–104. New York: Elsevier.
Steinhauer, Karsten, Erin J. White & John E. Drury. 2009. Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research* 25. 13–41.
Tanner, Darren, Lee Osterhout & Julia Herschensohn. 2009. Snapshots of grammaticalization: Differential electrophysiological responses to grammatical anomalies with increasing L2 exposure. In J. Chandlee, M. Franchini, S. Lord & G.-M. Rheiner (eds.), *Proceedings of the 33rd Boston university conference on language development*. 528–539. Somerville, MA: Cascadilla Press.
Tanner, Darren, Judith McLaughlin, Julia Herschensohn & Lee Osterhout. 2013. Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition* 16. 367–382.
Tanner, Darren, Kayo Inoue & Lee Osterhout. 2014. Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition* 17. 277–293.
Tremblay, Annie 2011. Proficiency assessment standards in second language acquisition research: “Clozing” the gap. *Studies in Second Language Acquisition* 33. 339–372.
Ullman, Michael T. 2001. The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research* 30. 37–69.
Ullman, Michael T. 2012. The declarative/procedural model. In P. Robinson (ed.), *Routledge encyclopedia of second language acquisition*, 160–164. New York: Routledge.
Van den Noort, Maurits, Peggy Bosch & Kenneth Hugdahl. 2006. Foreign language proficiency and working memory capacity. *European Psychologist* 11. 289–296.
Vos, Sandra H., Thomas C. Gunter, Herman H. J. Kolk & Gijsbertus Mulder. 2001. Working memory constraints on syntactic processing: An electrophysiological investigation. *Psychophysiology* 38. 41–63.
Waters, Gloria S. & David Caplan. 1996a. The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology* 49A. 51–79.
Waters, Gloria S. & David Caplan. 1996b. Processing resource capacity and the comprehension of garden path sentences. *Memory and Cognition* 24. 342–355.

Waters, Gloria S. & David Caplan. 1996c. The capacity theory of sentence comprehension: critique of Just & Carpenter 1992. *Psychological Review* 103. 761–772.
White, Erin Jacquelyn, Fred Genesee & Karsten Steinhauer. 2012. Brain responses before and after intensive second language learning: Proficiency based changes and first language background effects in adult learners. *PloS One* 7. e52318.