Individual differences predict ERP signatures of second language learning of novel grammatical rules*

EMILY S. NICHOLS
The University of Western Ontario
MARC F. JOANISSE
The University of Western Ontario

We investigated the extent to which second-language (L2) learning is influenced by the similarity of grammatical features in one's first language (L1). We used event-related potentials to identify neural signatures of a novel grammatical rule – grammatical gender – in L1 English speakers. Of interest was whether individual differences in L2 proficiency and age of acquisition (AoA) influenced these effects. L2 and native speakers of French read French sentences that were grammatically correct, or contained either a grammatical gender or word order violation. Proficiency and AoA predicted Left Anterior Negativity amplitude, with structure violations driving the proficiency effect and gender violations driving the AoA effect. Proficiency, group, and AoA predicted P600 amplitude for gender violations but not structure violations. Different effects of grammatical gender and structure violations indicate that L2 speakers engage novel grammatical processes differently from L1 speakers and that this varies appreciably based on both AoA and proficiency.

1. Introduction

Learning a second language is considerably easier and more successful when it begins early in life (Johnson & Newport, 1991). Indeed, many studies have shown both behavioral and neural differences in early vs. late L2 learners (Newman, Tremblay, Nichols, Neville & Ullman, 2012; Pakulak & Neville, 2011; Wartenburger, Heekeren, Abutalebi, Cappa, Villringer & Perani, 2003). However, there is ambiguity as to the source of these differences. We might interpret differences in the apparent neural organization of second-language learning as reflecting age-dependent effects of neuroplasticity, or they could simply reflect the general proficiency with which a second language has been learned (Newman et al., 2012; Wartenburger et al., 2003). The difficulty in disentangling the influences of age of acquisition (AoA) and proficiency on L2 learning is the fact that the two tend to correlate to some extent such that earlier L2 learners generally achieve higher proficiency in their second language (Johnson & Newport, 1991; Pakulak & Neville, 2011; Stevens, 1999; Weber-Fox & Neville, 1996).

Bilingualism research has explored many areas of second language acquisition and speaking, in both similar and dissimilar languages; however, one area that is lacking a large body of literature is that of grammatical gender. There is research suggesting that individuals who learn a second language can often use knowledge of their first language to aid them in their second (Foucart & Frenck-Mestre, 2011; Hartsuiker, Beerts, Loncke, Desmet & Bernolet, 2016), but many languages contain aspects of grammar that others do not. Inflectional morphology varies greatly across languages; not only do different morphological systems exist in different languages (e.g., pluralization, gender), but different languages also employ similar morphological systems differently (Aronoff, 1994). The lack of clear mapping from one language to another may be one reason why inflectional morphology tends to be a particularly difficult part of L2 learning (Pakulak & Neville, 2011).

Grammatical gender systems, which classify nouns as masculine, feminine, or sometimes neuter, are present in many of the world's languages. In those languages that contain gender systems, there is sometimes overlap in article-noun gender agreement between languages, which can facilitate learning of noun genders; for instance the word *table* is feminine in both French and Spanish (i.e., *la table/la mesa*; Foucart & Frenck-Mestre, 2011, Paolieri, Cubelli, Macizo, Bajo, Lotto & Job, 2010). However, the situation may be different for a native speaker of a

Address for correspondence:
Emily S. Nichols, The Brain and Mind Institute, Department of Psychology, The University of Western Ontario, London, Ontario, N6A 5B7, Canada
enicho4@uwo.ca

Supplementary material can be found online at https://doi.org/10.1017/S1366728917000566

language that does not have grammatical gender. Current data suggest that L2 speakers with a grammatical gender in L1 show higher accuracy in both gender assignment and pronoun-noun gender agreement in L2 compared to L2 speakers without grammatical gender in L1 (Sabourin, Stowe & de Haan, 2006).

The present study examines the joint contribution of proficiency and AoA to learning grammatical gender in L2. Proficiency is defined as the competence and facility in a second language. It is admittedly correlated with AoA (Stevens, 1999); however, some late learners do achieve high proficiency, and may appear comparable in fluency to early learners and native speakers. For example, highly proficient individuals, regardless of AoA, show an increase in use of discourse markers and conjunctions, and higher fluency when compared to individuals with low proficiency (Neary-Sundquist, 2013). Highly proficient late L2 learners have also shown differences in language-related brain activity from that of low proficiency late learners (Caffarra, Molinaro, Davidson & Carreiras, 2015; Gillon Dowens, Guo, Guo, Barber & Carreiras, 2011; Kotz, 2009; Perani, Paulesu, Sebastian Galles, Dupoux, Dehaene, Bettinardi, Cappa, Fazio & Mehler, 1998; Stowe & Sabourin, 2005; Wartenburger et al., 2003). In addition, when controlling for proficiency, late learners still differ from early learners both in measures of timing (Meulman, Wieling, Sprenger, Stowe & Schmid, 2015; Pakulak & Neville, 2011; Rossi, Kroll & Dussias, 2014) and level of brain activity (Wartenburger et al., 2003). Thus, differences in L2 processing could be due to either behavioral proficiency or true differences in neuroplasticity.

In addition to the effects of proficiency and AoA, some prior work also suggests that L2 learners process inflectional agreement – such as gender – differently from native speakers. Lemhöfer, Spalek, and Schriefers (2008) investigated whether German–Dutch bilinguals performed differently on tasks where the gender of a noun was the same in both languages, compared to when the gender differed. In both a lexical decision task and a picture-naming task, reaction times were faster for gender-congruent trials than for gender-incongruent trials. The authors attributed this to an interaction between grammatical gender systems in the two languages, with facilitation occurring when the genders are congruent. These results have been supported by numerous studies in several different languages (Foucart & Frenck-Mestre, 2011; Paolieri et al., 2010; Salamoura & Williams, 2007), suggesting that the effect is quite robust among languages containing grammatical gender systems, although conflicting results have also been found (Costa, Kovacic, Franck & Caramazza, 2003).

The majority of the behavioral research surrounding grammatical gender in L2 speakers has focused on adult learning of a gender system. In an experiment by Alarcón

(2011), behavioral measures of written comprehension and oral production were used to investigate whether English adult L2 learners of Spanish can acquire gender in their grammar. Results of these measures indicated that, at high proficiencies, late (post-puberty) L2 learners showed no difficulty with grammatical gender, similar to native speakers. Similarly, Keating (2009) found that adult learners of Spanish produce higher rates of gender agreement errors with increasing distance between the adjective and noun. However, other studies have shown conflicting evidence, with adult learners experiencing difficulty in acquiring grammatical gender (Arnon & Ramscar, 2012; Montrul, Foote & Perpiñán, 2008).

The many observed interactions between first and second languages in L2 speakers raise the question of how grammatical gender is learned in individuals whose L1 does not contain a grammatical gender system. Indeed, many of the world's most-spoken languages (e.g., English, Mandarin, Cantonese) do not possess a gender system, and studies examining L2 grammatical gender learners who do not possess a grammatical gender in their L1 have focused on late learners (Foucart & Frenck-Mestre, 2012; Gillon Dowens et al., 2011). The lack of research focusing on early learners leaves open the question of how AoA specifically affects learning of novel syntactic constructions.

ERP provides an ideal mechanism for studying grammatical relationships in first- and second-language processing (Caffarra & Barber, 2015; Foucart & Frenck-Mestre, 2012; Meulman et al., 2015; Morgan-Short, Sanz, & Ullman, 2010; Newman et al., 2012; Osterhout & Holcomb, 1992; Pakulak & Neville, 2011; Rossi et al., 2014; Silva-Pereyra, Gutierrez-Sigut & Carreiras, 2012; Tanner, McLaughlin, Herschensohn & Osterhout, 2013). ERPs represent electroencephalography (EEG) signals that are time-locked to sensory or cognitive events. The high temporal resolution of ERPs allows the researcher to observe neural processing of language as it unfolds over time. This in turn allows us to pinpoint changes in neural processes corresponding to a particular manipulation and isolate the moment at which they occur, typically well before the moment individuals can make an overt judgment of the stimulus or execute a behavioral response. In particular, grammatical violation tasks involve showing subjects sentences which are either grammatically congruent or contain a grammatical violation. For example, the sentence "*He took the whistling teapot off the of stove*" contains a grammatical violation of phrase structure that evokes predictable modulations in ERPs time locked to the onset of the violation. There are several possible grammatical violations, including phrase structure, number, tense, and, most relevant to the present study, grammatical gender. Manipulating the type of violation allows us to isolate processing of specific aspects of grammatical processing.

One ERP component that is sensitive to grammatical violations is the Left Anterior Negativity (LAN), a negative-going component with a left-anterior distribution (Molinaro, Vespignani & Job, 2008; Neville, Nicol, Barss, Forster & Garrett, 1991; Pakulak & Neville, 2011). The LAN is thought to reflect early syntactic integration or first-pass grammatical processing (Friederici, Pfeifer & Hahne, 1993; Rösler, Pütz, Friederici & Hahne, 1993). Although the time-course is similar to the N400, the LAN possesses a different topography and is evoked in response to syntactic rather than semantic errors (however see Tanner, 2014 for a discussion). The LAN is often followed by a P600, a positive going component with a centro-parietal distribution that occurs approximately 600 ms post-stimulus onset. It is thought to reflect second-pass grammatical processing (Hahn & Friederici, 1999) or syntactic reanalysis (Kaan, Harris, Gibson & Holcomb, 2000). By manipulating the grammaticality of a sentence, the P600 has been shown to vary in its amplitude as well as its scalp distribution (Kotz & Friederici, 2003; Molinaro et al., 2008; Pakulak & Neville, 2011).

ERP markers of grammatical gender processing have been widely explored in monolingual speakers of languages that incorporate grammatical gender. Gender agreement violations have been found to elicit both a LAN and a P600 in native speakers of numerous languages including German, Italian, Dutch, and Spanish (Barber & Carreiras, 2005; Gunter, Friederici & Schriefers, 1996; Molinaro et al., 2008; Sabourin & Stowe, 2008). As the LAN and the P600 are markers of syntactic violation processing, it can be concluded that the brain processes grammatical gender agreement violations much like other forms of syntactic violations, though the timing and scalp distribution of these effects has been found to vary (Barber & Carreiras, 2005; Gillon Dowens et al., 2011; Molinaro et al., 2008).

Syntactic ERPs described above have also been used to evaluate the time-course and native-like characteristics of L2 syntactic processing. Of note to the present study, some researchers have used these effects to argue for differences in how L2 learners detect grammatical violations. For instance, L2 learners might tend to show reduced or absent LAN and/or P600 effects in response to violations in grammatical structures known to be difficult for these individuals. Although results previously attributed to AoA may in fact be due to proficiency, several L2 ERP studies attribute these results solely to AoA. Pakulak and Neville (2011) investigated whether AoA affects syntactic processing, holding proficiency constant. A native English group and a high proficiency, late acquisition L2 English group performed a sentence comprehension task with phrase structure violations while their EEG was being recorded. The researchers found both a LAN and a P600 in response to syntactic violations in the native group, but found only a P600 in the late-learners,

suggesting that late learners are not integrating incoming syntactic information in the same way as native speakers, perhaps relying on different neural mechanisms due to maturational constraints.

Similarly, some studies have specifically used ERPs to study grammatical gender in L2 speakers. A study by Morgan-Short and colleagues (2010) examined second language learning of gender using an artificial grammar, in both implicit-learning (immersion-like) and explicit-learning (classroom-like) settings. The researchers tested subjects first at low proficiency and again at high proficiency, and found that when subjects viewed article-noun gender agreement violations at low proficiency, an N400 component, a negative going ERP component typically thought to reflect lexical-semantic violations, was elicited in only the implicit-learning group. At high proficiency, however, noun-article gender agreement violations elicited P600 in both groups. The authors suggest that from these results, it can be inferred that both proficiency and training affect inflectional morphological processing in L2 learners. Evidence from this study suggests that level of proficiency in late learners affects how the brain processes grammatical gender, implying that it may be possible to attain native-like processing of grammatical gender regardless of AoA, depending on the level of proficiency attained.

These results are supported by findings from Gillon Dowens et al. (2011), in which gender processing was studied using a group of late acquisition Spanish learners who spoke Mandarin as a first language. The authors sought to characterize gender processing in proficient L2 speakers who did not have a gender system in their L1. Subjects viewed sentences containing gender agreement violations while their EEG was recorded. As in the Morgan-Short et al. (2010) study, results indicated that a P600 component was elicited for gender agreement violations in this group. However, neither experiment had an L1 group to which they could compare their L2 results. This leaves undetermined how L2 speakers' ERPs response to gender agreement violations compares to those of native speakers. More recently, Meulman and colleagues (2015) found that AoA influences the ERP response to grammatical gender violations but not to verb agreement violations, suggesting that similarities between grammatical constructs in L1 and L2 may drive differences in the effect of AoA on grammatical processing.

That said, there are few studies directly comparing grammatical gender in L2 and L1 speakers of the same language. Foucart and Frenck-Mestre (2011) compared German–French bilinguals and native French speakers on a grammatical gender task. The authors manipulated gender agreement in French sentences, and found that violations elicited similar P600 effects in both groups, and also found that the P600 was larger for words whose

gender was the same across languages. The authors suggest that syntactic processing in a second language is affected by similarities between L1 and L2. While these findings describe language transfer effects between two languages that possess grammatical gender systems, these results cannot be generalized to second language speakers who do not have a gender system in their native language. However, a follow-up study in 2012 by the same authors found that both native French speakers and high proficiency, late acquisition English–French learners showed P600s in response to grammatical gender violations in spite of the fact that English does not have grammatical gender. The authors concluded that late L2 learners are able to acquire grammatical features not present in L1 (Foucart & Frenck-Mestre, 2012).

1.1 Rationale for the Current Study

While previous research suggests that high proficiency L2 or early AoA speakers process gender agreement violations differently from low proficiency or late AoA speakers respectively, there has been very little research comparing gender processing in L2 speakers to that in L1 speakers, especially across languages that do not both have a gender system. Further, research to date has tended to examine AoA and proficiency in isolation, leaving open the question of which of the two factors can best explain apparent differences, or whether maturational constraints and proficiency interact (Nichols & Joanisse, 2016). We addressed this issue by examining ERP indices of grammatical gender agreement violations in L2 learners of differing proficiencies and AoAs, compared to those of native speakers. Additionally, a grammatical word order (i.e., structural) violation condition was used in order to determine whether the similarity of a grammatical feature in L1 and L2 affects acquisition of L2 grammar. Because the structural violations we employed here can exist in both English and French, it was possible to compare the effect of AoA and proficiency on grammatical gender to their effect on a rule system that is similar across both L1 and L2, allowing us to determine whether there is a difference between learning a novel rule vs. simply learning a new language.

In line with previous studies showing independent effects of AoA and proficiency but also of group (e.g., bilingual vs. monolingual; Newman et al., 2012; Nichols & Joanisse, 2016), we predicted that in native and high proficiency L2 speakers, gender agreement violations would elicit both a LAN and a P600, and that the amplitudes of these effects would decrease with decreasing proficiency. We also predicted that at earlier AoA, L2 speakers would have large LANs and P600s, again similar to native speakers. But as AoA increases, amplitude would decrease (Meulman et al., 2015; Hahne & Friederici, 2001; Chen, Shu, Liu, Zhao & Li, 2007;

Ojima, Nakata & Kakigi, 2005; Weber-Fox & Neville, 1996). Such findings of separable contributions of AoA and proficiency would lend support to the theory that both AoA and proficiency play independent roles in the processing of grammatical gender in L2.

Structure violations were predicted to evoke both a LAN and P600, but AoA should not modulate the magnitude of these effects (Hahne, 2001; Hahne & Friederici, 2001; Neville et al., 1991, Newman, Ullman, Pancheva, Waligura & Neville, 2007; Rossi, Gugler, Friederici & Hahne, 2006; Weber-Fox & Neville, 1996). The reason for this is that this type of syntactic error is possible in both English and French, thus AoA of L2 should not influence processing (MacWhinney, 1987, 2005). We also predicted an increase in LAN and P600 amplitude with proficiency regardless of L1/L2 status, as the error would be more egregious to higher proficiency French speakers. If there are indeed different effects of AoA and proficiency between grammatical gender and structure violations, it would suggest that while it is possible for L2 speakers to acquire novel grammatical rules, this process is different from learning grammatical rules that are present in L1.

2. Method

2.1 Subjects

Forty right-handed neurologically healthy adults were recruited from the University of Western Ontario community. Twenty L1 speakers (16 female) were individuals who reported learning French as their first language, ranging in age from 18 to 38 ($M = 23$, $SD = 5.3$). An additional 20 (14 female) L2 speakers were individuals who reported learning English as their first language and French at any point after English, ranging in age from 18 to 33 ($M = 21$, $SD = 3.8$). A summary of group descriptives is provided in Table 1, and an extended description of the L2 speakers is available in Appendix A (Supplementary Materials).

2.2 Materials and Procedures

In order to assess AoA, all subjects completed a detailed language history questionnaire in French, which inquired about past and present exposure in both their first and any second languages. To assess proficiency, all subjects completed an intensive proficiency test which assessed both grammar and vocabulary proficiency. The French proficiency test was issued by pen-and-paper, and consisted of 100 questions. The test was designed by the French department at the University of Western Ontario to place non-native French speakers in the appropriate class. Scores correspond to the Common European Framework of Reference for Languages, levels A-C, with a score of

Table 1. *Group demographic and behavioral measures*

Measure		Group Mean (SD)		*t*(df)	*p*
		L1 French	L2 French		
AoA in years		0 (0)	7.20 (3.85)	−8.37(19)	< .001
Proficiency (% correct)		87.65 (9.76)	64.05(15.69)	5.71(31)	< .001
Sentence Comprehension Accuracy	Control	86.00 (5.22)	79.88 (8.98)	2.64(30)	.013
(% correct)	Gender	83.88 (21.62)	26.00 (20.12)	8.76(37)	< .001
	Structure	95.33 (6.14)	83.72 (16.25)	2.99(24)	.006
	Filler	87.36 (6.94)	77.60 (11.89)	3.17(30)	.003
Number of trials included in	Control	32.2 (3.97)	29.55 (4.03)	2.09(38)	.043
final analysis (out of 40)	Gender	31.5 (8.62)	9.80 (6.77)	8.85(36)	< .001
	Structure	36.15 (2.80)	32.10 (6.46)	2.57(26)	.016

Note. One-sample t-test used to test L2 AoA against 0. Welch's t-test used to test L1 vs. L2 for all other measures due to unequal variances between groups.

78% or greater corresponding to native-like proficiency, and 88% or greater corresponding to high native-like proficiency. Sixty-one questions were on grammar; this section had participants complete sentences by choosing the correct grammatical form, covering the eight parts of speech (e.g., noun, verb, adjectives) as well as three grammatical tenses; the *passé composé*, *participle passé*, and *présent de l'indicatif*. The grammar section also covered the negative form, requiring answers to questions in the negative. Thirty-nine questions were on vocabulary. This section had several subsections in which participants completed sentences by choosing the correct, noun or verb to fit the context, perform verb-to-noun and noun-to-verb conversion, complete the opposite logical expression of a given statement, choose the correct name to describe inhabitants of a certain city or country, and finally to complete common proverbs. Completion of the test took approximately 50 minutes. An abridged version of the Edinburgh Handedness Inventory (Oldfield, 1971) was used to verify handedness.

Stimuli in the experimental task consisted of 160 sentences, with 40 containing article-noun gender agreement violations (*J'ai nagé dans le^m piscine^f tous les jours /I swam in the pool every day*), 40 well-formed sentences containing no violations (*J'ai nagé dans la^f piscine^f tous les jours /I swam in the pool every day*), and 40 sentences containing structural violations in which two words were switched such that the grammatical structure was incorrect but the gender agreement was intact (*J'ai nagé dans piscine^f la^f tous les jours /I swam in pool the every day*). Experimental sentences were counterbalanced across subjects, with the sentences that contained gender violations for a third of the subjects being the sentences that contained either no violations or structure violations for the other two thirds. An additional 40 well-formed filler sentences were used which remained the same between counter-balanced lists and were used to ensure equal numbers of violation and well-formed

sentences. Experimental items are available in Appendix B (Supplementary Materials), with seven of the sentences taken from Baudiffier, Caplan, Gaonac'h, and Chesnet (2011).

Sentences were presented visually in the center of a CRT screen, word-by-word, using rapid serial visual presentation (RSVP). Words were on-screen for 300 ms with a 200 ms gap, and following each sentence subjects were asked whether the sentence was well-formed via a visual cue *"Est-ce une bonne phrase Française?"* (*"Is this a good French sentence?"*). Yes/no responses were made via button-press. Sentences were presented over four blocks of 40 sentences each, with half containing violations. Prior to the experimental trials, subjects completed a practice block of 5 sentences, which they were allowed to complete as many times as they wished.

2.3 EEG Recording and Preprocessing

Stimuli were presented using the E-Prime 2.0 software package (Schneider, Eschman & Zuccolotto, 2002). Continuous EEG data was collected using BioSemi software from 32 scalp electrodes (Fp1/2, AF3/4, F7/8, F3/4, T7/8, C3/4, CP5/6, CP1/2, P7/8, P3/4, PO3/4, O1/2, Fz, Cz, Pz, Oz) and two mastoid electrodes, and electrooculogram (EOG) was recorded from four face electrodes placed above and below the left eye and on the outer canthus of each eye using the BioSemi ActiveTwo EEG system consisting of amplifier-embedded Ag/AgCl electrodes arranged according to the International 20–30 system. A Common Mode Sense active electrode and a Driven Right Leg passive electrode were used as the ground. Data was recorded in the frequency range of 0.1-100 Hz at a 512 Hz sampling rate, with impedances below 20 kΩ.

ERP data was processed using EEGLAB software (Delorme & Makeig, 2004) and the ERPLAB add-on

software (Lopez-Calderon & Luck, 2014). After importing the data, EEG data underwent a .1 – 30 Hz bandpass filter with a 60 Hz notch filter to remove line and muscle noise. EEG data was segmented into -200 – 1000 ms single-trial epochs of each critical word in each condition of interest (gender violation, structure violation, control) and baseline corrected to a pre-stimulus baseline (-200 – 0 ms). Critical words consisted of the noun immediately following the gender cue (correct vs. incorrect), and the first word in a syntactically reversed grammatical violation. Artifacts were removed by excluding epochs from analysis in which voltage exceeded -100 – 100 μV at any scalp electrode. In order to ensure that we were analyzing sentences in which the violation was detected, only sentences that were responded to correctly were included in analyses. Filler sentences were used in order to equate the number of correct sentences with the number of violation sentences and were thus excluded from analysis. Total number of trials included in the final analysis after rejecting artifacts and incorrect trials are described in Table 1.

To examine the LAN, mean amplitude between 300 – 500 ms was computed for each electrode. Electrodes were grouped into regions of interest arranged in a 3 x 3 grid over the scalp (left/midline/right and anterior/center/posterior), and data from each electrode within an ROI were treated as repeated measures of that ROI. To ensure that the violation conditions were eliciting the LAN, difference waves were computed from each type of violation minus the control condition and amplitudes were submitted to linear mixed effects (LME) analysis with condition (gender violation minus control/structure violation minus control), group (L1 speaker/L2 speaker), and ROI as fixed effects and subjects as a random effect. We then assessed AoA and proficiency on the amplitude of the LAN. A forward stepwise procedure was then performed on mean amplitude of the difference waves, examining the independent contributions to a LME model with ROI and group as fixed effects, AoA and proficiency as continuous effects, and subjects as a random effect. The predictor explaining the most variability was assessed using AIC values, and a drop-one procedure was used to compute whether a single term could be removed from the model at each step without significantly reducing the model's explanatory value. The final model contained (stepwise) the variables that explained significant variability in the data, excluding variables that could be removed without influencing the model.

To examine the P600, mean amplitude between 500 – 800 ms was computed for each condition over the same nine ROIs. Similar to the LAN, we submitted difference wave amplitudes to LME analysis with condition (gender violation minus control/structure violation minus control),

group (L1 speaker/L2 speaker), and ROI as fixed effects and subjects as a random effect in order to ensure that the violation conditions were eliciting a P600. Again, to assess the effect of AoA and proficiency on the amplitude of the LAN, a forward stepwise procedure was performed on difference wave amplitudes, examining the independent contributions of electrode, group, AoA, and proficiency to an LME model.

Because our participants ranged in proficiency, we expected a large range of accuracy in performance on the violation detection task, leading to some participants having more trials than others included in the analysis. The LME modeling approach used here helped address potential issues this might raise with some types of statistical analyses; LME models include both fixed effects and random effects and can account for unbalanced data and nonsphericity (Baayen, Davidson & Bates, 2008; Bagiella, Sloan & Heitjan, 2000). For these reasons they are ideal for ERP data, especially in designs that lead to necessarily unbalanced data (Tibon & Levy, 2015). The present study used the *lme4* (Bates et al., 2015, version 1.1-12) and *LMERConvenienceFunctions* (version 2.10) packages for R (R Core Team, 2015, version 3.2.2).

3. Results

3.1 Behavioral

Group measures of AoA, proficiency, and ERP task accuracy are reported in Table 1. Results from the language background questionnaire confirmed that all L1 speakers reported learning French from birth, while L2 speakers learned French from a range of 0 – 16 years of age. Although one L2 speaker reported learning French from birth, they reported living in an English-speaking household in Montreal, and considered themselves an L2 speaker of French. L1 speakers' proficiency scores ranged from 63 – 100%, and L2 speakers' proficiency scores ranged from 32 – 91%. L1 speakers were significantly more accurate on all sentence types in the ERP task than were L2 speakers. There was a significant correlation between AoA and proficiency when both groups were combined ($r = -.62$, $p < .001$). However, this effect was not evident for the L2 speaker group alone ($r = -.21$, $p = .380$). L2 performance on the gender violation sentences (i.e., detecting the error in the gender violation sentence) ranged from 2.5 – 92.5% correct ($M = 26\%$, $SD = 20.12$), indicating that some L2 speakers had difficulty detecting grammatical gender violations while performing well above chance on the rest of the task. Because some participants performed especially poorly on the gender violation detection task, additional analyses were run excluding those scoring below 25% accuracy on all violation conditions, as discussed further below.
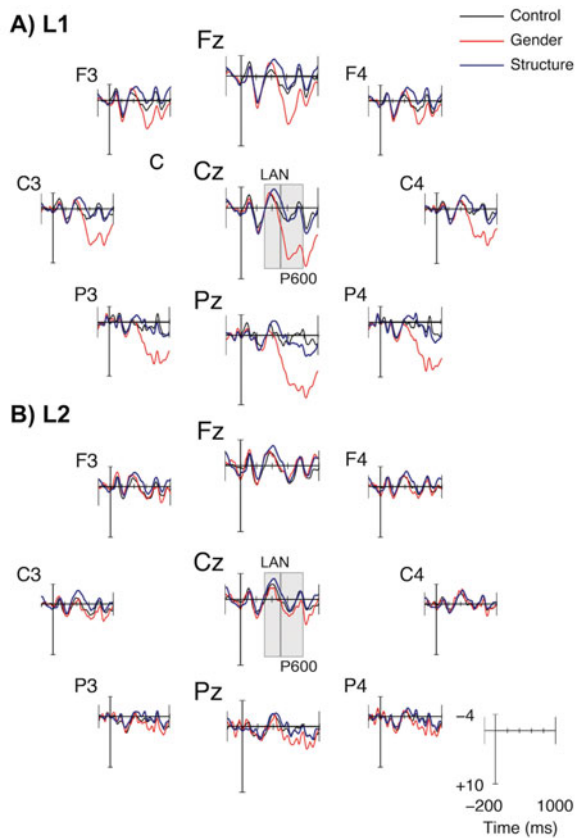
Figure 1. (Colour online) Group average ERPs for A) L1 French speakers and B) L2 French speakers for the control, gender violation, and structure violation sentences in the -200–1000 ms time window, between -4–10 µV. Negative is plotted upward.

3.2 Left Anterior Negativity

ERPs for control, gender and structure violations are shown in Figure 1; difference waveforms and topographic maps for the L1 and L2 groups are shown in Figure 2. To confirm that the grammatical gender and structure violations produced a LAN, we first examined the violation condition subtraction waves (i.e., gender – control and structure – control) within the 300 – 500 ms time window, across groups. A mixed ANOVA with violation, group, and ROI as fixed factors revealed a main effect of ROI type ($F(8, 304) = 4.08$, $p < .001$), no main effect of group ($F(1, 38) = .01$, $p = .907$, *ns*), no main effect of violation ($F(1, 38) = 1.19$, $p = .283$, *ns*), and no significant interactions. Post-hoc Bonferroni-corrected paired t-tests between violation types revealed that left frontal, right frontal, and left center ROI amplitudes differed significantly from posterior electrodes, with the most negative amplitudes in the left center ($M = -.91$ µV, $SD = 3.10$) and left frontal ($M = -.78$ µV, $SD = 3.23$) ROIs. These results suggest that, as a whole, when not accounting for the variability in AoA

or proficiency, our combined L1 and L2 sample showed a left anterior component in the LAN time window, and that this effect did not differ based on language status.

Next we examined whether AoA or proficiency might modulate the amplitude of the LAN in response to gender or structure violations versus control sentences. A forward stepwise procedure to determine the best-fit LME model with group, proficiency, AoA, and ROI revealed that three factors, proficiency, AoA, and group predicted LAN amplitude, and the interactions are shown in Figure 3A. A significant violation type × proficiency interaction ($F(1,2536) = 38.13$ $p < .001$) was found and appeared to be driven by structure violations, in that the amplitude of the LAN became more negative with increasing proficiency, while LAN amplitude in response to gender violations increased by .20 µV. A significant violation type × AoA interaction was also found ($F(1,2536) = 76.35$, $p < .001$). Unlike proficiency, the effect of AoA appeared to be driven by gender violations rather than structure violations, with more negative LAN amplitudes for earlier AoAs. Finally, a significant violation type × group interaction was found ($F(1,2536) = 14.41$, $p < .001$), with L2s showing greater disparity between violation conditions.

In order to ascertain that the effect of AoA is not being driven by the L1 group whose AoA was uniformly zero, we repeated the same forward stepwise procedure with only individuals in the L2 group. Results were similar to the prior analysis: AoA was found to be the largest predictor as demonstrated by a significant violation type × AoA interaction ($F(1,1257) = 63.38$, $p < .001$; Figure 3B), followed by a violation type × proficiency interaction ($F(1,1257) = 31.25$, $p < .001$; Figure 3B). Again, the effect of AoA appears to be driven by gender violations, with more negative LAN amplitudes for earlier AoAs.

Because some participants performed especially poorly on the gender violation detection task, there was the concern that the signal-to-noise ratio for those individuals may have been extremely low due to the inclusion of very few accurate trials in their mean ERPs. This could in turn have artificially deflated the effect of grammaticality on observed ERP waveforms, which then could explain the individual differences effects observed above. To address this, data were reanalyzed including only individuals who performed with 25% or greater accuracy on all conditions, with a total of 12 participants being removed, all from the L2 group. Results of the best-fit LME model with group, proficiency, AoA, and ROI did not differ from the initial LME LAN analysis. Thus, excluding participants with fewer correct trials yielded the same pattern of significance as with the entire L2 sample.

These results indicate that both proficiency and AoA affect early syntactic integration, as indexed by LAN, however the type of syntax matters. AoA modulated LAN
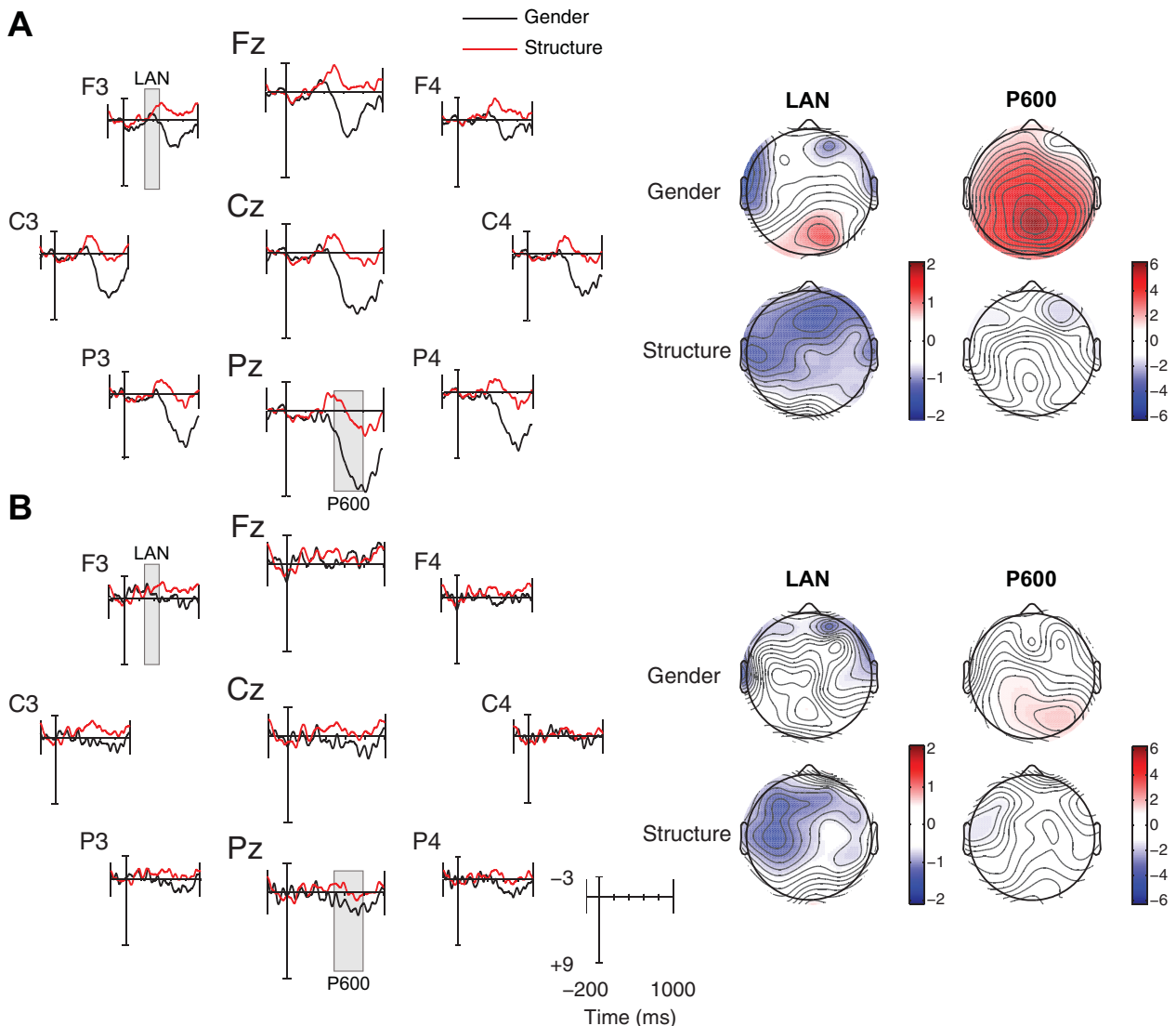
Figure 2. (Colour online) ERP differences waves and topographic plots for native and L2 French speakers. (A) ERP difference waves for L1 French speakers, computed as gender violation minus control in black and structure violation minus control in red. Negative is plotted upward. Topographic maps show the difference between gender violation minus control and structure violation minus control over LAN and P600 time windows for L1 French speakers. (B) ERP difference waves for L2 French speakers.

amplitude in response to gender violations, which are novel to L2 speakers, suggesting that learning the rule earlier leads to more native-like syntactic processing. Alternatively, proficiency modulated the LAN in response to structure violations, which are not unique to French, supporting the hypothesis that the structure errors are more egregious to higher proficiency speakers, while remaining unaffected by AoA.

3.3 P600

To confirm that violations were producing a P600, we examined the violation subtraction waves within the 500

– 800 ms time window across groups. A mixed ANOVA with violation type, group, and ROI revealed a main effect of violation type ($F(1, 34) = 22.66$, $p < .001$) and a main effect of ROI ($F(8, 272) = 8.33$, $p < .001$). There was a significant violation type × group interaction ($F(1, 34) = 6.77$, $p = .014$), as well as a significant violation type × ROI interaction ($F(8, 272) = 3.97$, $p < .001$). Post-hoc Bonferroni-corrected Welch's t-tests revealed that groups differed in their response to gender violations, with L1s producing larger amplitudes to gender violations than L2s ($t(1114) = 11.53$, $p < .001$). Only the gender violations showed an effect of ROI, with the posterior left, center, and right, and the mid center ROIs differing from the frontal
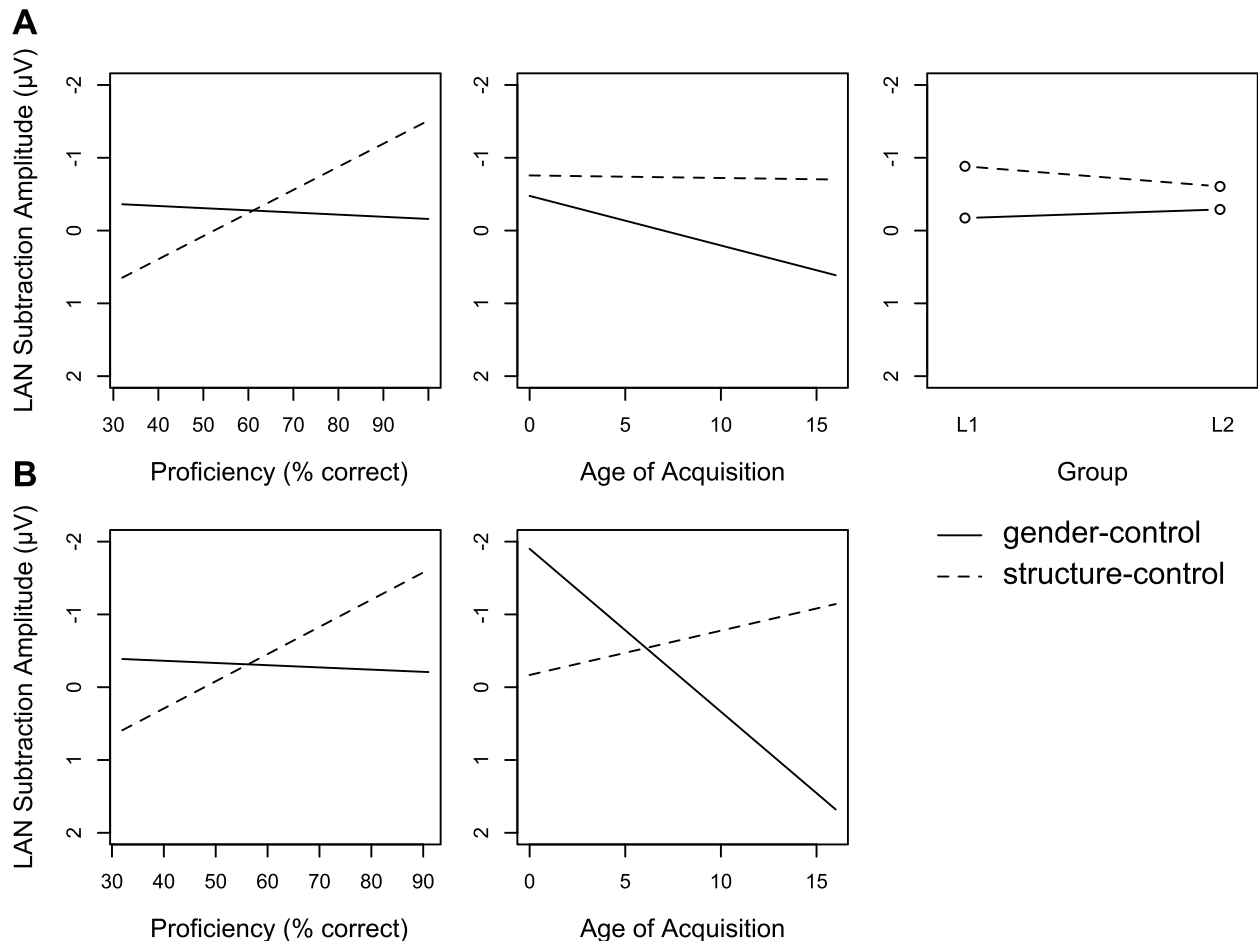
A

B

Figure 3. Proficiency, age of acquisition, and group interactions with sentence type in the LAN time window. (A) Violation type × Proficiency and violation type × AoA interactions across L1 and L2 French speakers. The x-axis is proficiency and AoA respectively, and the y-axis is mean LAN amplitude of the difference wave between 300–500 ms, with negative plotted up. The solid line is the gender violation minus control condition, and the dotted line is the structure violation minus control condition. B) Violation type × Proficiency and Violation type × AoA interactions across L2 French speakers only.

left and right ROIs ($p < .001$ in all comparisons). These results indicate that only L1 speakers produced a P600 to the gender violations, while no P600 was produced in either group to the structure violations.

We next examined whether AoA and/or proficiency modulated the P600 in response to gender and structure violations, beyond the group effect observed above. Forward stepwise LME analysis revealed that four factors predicted P600 subtraction amplitude: proficiency, ROI, group, and AoA. Results revealed significant violation type × proficiency ($F(1, 2280) = 142.92, p < .001$), violation type × ROI ($F(1, 2280) = 4.22, p < .001$), violation type × group ($F(1, 2280) = 14.93, p < .001$), and violation type × AoA ($F(1, 2280) = 8.37, p < .001$) interactions. As can be seen in Figure 4 (see Figure 2 for topographic differences between violation types), the proficiency, group, and AoA interactions appear to be driven by gender violations. This suggests that proficiency,

group, and AoA contributed independently to the P600 response. In contrast, amplitudes to structure violations were not modulated by these variables, suggesting that there is a difference between how proficiency, group, and AoA modulate grammatical gender violations and structure violations. As with the LAN, data were re-analyzed removing individuals scoring below 25% accuracy. Results showed the same pattern as the previous P600 analysis.

4. Discussion

The present study used event-related potentials to examine the effects of AoA and proficiency on grammatical gender processing in second language speakers whose first language does not possess a gender system. We measured brain responses in native French speakers and L2 French speakers as they read control sentences and sentences
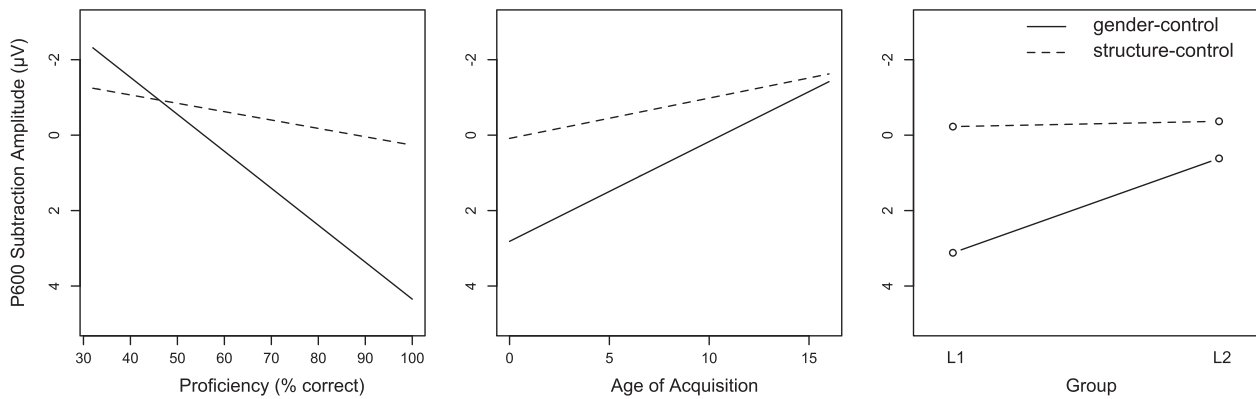
Figure 4. P600 interactions with proficiency, age of acquisition, and group. (A) Violation type × Proficiency interaction across L1 and L2 French speakers. The *x*-axis is proficiency, and the *y*-axis is mean P600 amplitude of the difference wave between 500–800 ms, with negative plotted up. The solid line is the gender violation minus control condition, and the dotted line is the structure violation – control condition. (B) Violation type × AoA interaction across L1 and L2 French speakers. (C) Violation type × Group interaction.

containing syntactic violations that evoked the LAN and P600 components. Our unique sample of participants allowed us to treat both proficiency and AoA as continuous variables, providing a more complete description of how both variables predict grammatical gender processing. Results indicated that, at what we would argue to be first-pass stages of grammatical processing, AoA predicted LAN amplitude to gender but not structure violations, while proficiency predicted LAN amplitude to structure but not gender violations. L2 speakers also showed a greater disparity between LAN responses to gender and structure violations. However, at later stages of grammatical processing, proficiency, group, and AoA each independently predict P600 amplitude to gender violations, while there was no P600 elicited to structure violations.

4.1 Left Anterior Negativity

When only considering group membership, a LAN was elicited to both structure violations and gender violations in both groups. However, when including group membership, proficiency, and AoA in the model, proficiency predicted an increase in amplitude to structure violations, while AoA predicted a decrease in amplitude to gender violations. The LAN is thought to represent early syntactic integration or first-pass grammatical processing reflecting detection of a syntactic violation (Bornkessel & Schlesewsky, 2006; Friederici, 2002; Friederici et al., 1993; Rösler et al., 1993). These results thus suggest that, as AoA increases, individuals increasingly fail to exhibit this early-stage marker of grammatical gender processing. This effect holds despite participants' overt detection of errors – as marked by affirmative behavioral response. Moreover, this effect also holds when controlling for

second-language proficiency level as measured offline by a standardized measure. This finding is supported by the existing literature, in which late AoA learners were found to have reduced or absent neural markers of early syntactic processing when compared to native speakers or early L2 learners (Pakulak & Neville, 2011; Wartenburger et al., 2003; Yan, Zhang, Xu, Chen & Wang, 2016). The present results confirm that this effect does owe to age-dependent effects and is not strictly due to these individuals' overall proficiency in their second language.

In the present study proficiency did not predict LAN amplitude in response to gender agreement violations, with only AoA explaining significant variance in amplitude. Thus, it is possible that changes in LAN amplitude previously attributed to proficiency may in fact be due to AoA, which is often highly correlated with proficiency. Proficiency explained significant variance in LAN amplitude to structure violations, with larger LANs as proficiency increased in the combined L1 and L2 sample. This positive relationship suggests that with regards to structure violations, increased proficiency predicts stronger early syntactic processing in both L1 and L2 speakers. These results replicate research showing that higher proficiency monolinguals show greater LAN amplitude to syntactic errors than low proficiency monolinguals (Pakulak & Neville, 2010), suggesting that proficiency is a major contributor to syntactic processing in both L1 and L2.

The difference in contributing factors to structure violation and gender violation processing indicates that L2 processing of these two forms of syntactic processing relies on dissociable neurocognitive mechanisms. The present results indicate that early syntactic integration of structure errors to the rest of the sentence depends on proficiency; the more proficient, the more difficult it is

to integrate the error into the sentence, indexed by larger LAN responses. This is in contrast to gender violations, which are more sensitive to AoA than to proficiency, with early integration processes becoming less involved as AoA increases, indexed by decreasing LAN amplitude as AoA increases.

AoA is thought to affect syntactic processing more than proficiency (Pakulak & Neville, 2011; Weber-Fox & Neville, 1996; Wartenburger et al., 2003); proficiency has also been argued to affect semantic processing more than AoA (Weber-Fox, Davis & Cuadrado, 2003; Wartenburger et al., 2003). Given the lack of P600 to structure violations and the negativity between 300–500 ms, it could be proposed that the evoked effect is in fact an N400 rather than a LAN. This would imply that participants were treating structure violations as semantic violations rather than syntactic violations, and producing an N400, and would support the hypothesis that proficiency predicts semantic processing. Indeed, recently some have suggested that the LAN may in fact be an N400, which has a skewed topography due to the following P600 (Tanner, 2014; Tanner & Van Hell, 2014). However, in the present study the LAN is not followed by a P600, and the topography of the evoked response to structure violations reflects that of the LAN, with the signal appearing greatest over left anterior electrodes. This is in contrast to the N400, which has a signal appearing greatest over midline centro-parietal electrodes. Additionally, Newman and colleagues (2012) found no relationship between proficiency in L2 and N400 amplitude. Thus, the evoked response appears characteristic of the LAN and it is not likely that the effect is in fact an N400. Although the LAN and P600 often occur together to form a LAN/P600 biphasic response (Gunter et al., 1996; Hahne & Friederici, 1999; Kim & Sikos, 2011; Molinaro et al., 2008), many studies have produced one effect without the other (Foucart & Frenck-Mestre, 2012; Friederici et al., 1993; Gillon-Dowens et al., 2011; Meulman et al., 2015; Schacht, Sommer, Shmuilovich, Martíenz & Martín-Loeches, 2014; Silva-Pereyra et al., 2012).

Despite the lack of a biphasic response in the structure violation condition, several conclusions can be drawn from the comparison to gender violations. We hypothesized that because syntactic structure is relevant to both English and French, knowledge of those syntactic rules in a specific language will be modulated by proficiency in that language. This is in contrast to AoA, which should not influence processing of structure violations because learning that rule (which is not novel) is not subject to neuroplastic effects (MacWhinney, 1987, 2005). Instead, AoA influences gender because the age at which L2 is learned determines the extent of neuroplastic effects, as the speaker has no foundation from L1 on which to build. At this early stage of syntactic processing, gender

processing did not appear to be sensitive to proficiency. Thus, while it is possible for L2 speakers to acquire novel grammatical rules, this process is different to learning grammatical rules that are present in L1.

4.2 P600

When compared to the L2 learner group as a whole, only the L1 group yielded a significant P600 to gender violations, and neither group produced a P600 to structure violations. However, delving deeper into the L2 group data revealed a more nuanced set of results. When we included proficiency and AoA in the statistical model, proficiency, group, and AoA each contributed to P600 amplitude in response to gender violations, but not to structure violations. As predicted, P600 amplitude increased with increasing proficiency and decreased with increasing AoA. These results thus suggest that there are multiple contributing factors that influence late-stage syntactic processing. The P600 is thought to represent second-pass grammatical processing (Hahn & Friederici, 1999) or syntactic reanalysis (Kaan et al., 2000); suggesting that, as a group, L1 speakers reanalyzed the gender violations more reliably than L2 speakers as a group. Interestingly, the results suggest that the proposed reanalysis stage indexed by the P600 is sensitive to the type of syntactic violation being induced. As a result, violations of word order yielded only a LAN and not a P600. Given this, it seems too simplistic to assume that any violation in syntactic structure invokes the same syntactic error detection and/or reanalysis mechanisms, and that this process may in fact be multifactorial. Indeed, the present findings lend further support to dissociable syntactic processes characterized by the LAN and P600 (Molinaro, Barber, Caffarra & Carreiras, 2014).

L2 speakers as a group did not produce a significant P600 in response to gender violations; however, further inspection revealed that this reflected the large variability in proficiency and AoA in our sample. Closer inspection revealed that both these factors predicted significant variance in P600 amplitude such that higher proficiency and earlier AoA both yielded larger P600 violation effects. That said, the group still contributed significant variance in our analysis of the combined L1 and L2 samples. This supports the view that regardless of other factors, this aspect of L2 language processing is still qualitatively different to L1 language processing. This is concordant with the view that even early L2 learners show differences in neural markers of syntactic processing (Hernandez & Li, 2007; Kotz, 2009; Weber-Fox & Neville, 1996).

4.3 Theoretical considerations

Different theories have been put forward to account for the differences observed in the AoA and proficiency

literature. The declarative/procedural model (Ullman, 2001a; Ullman, 2001b) suggests that the processing of semantics in both L1 and L2 relies on declarative memory, and has shared neural bases. In contrast, syntactic processing in L1 and initial L2 learning are proposed to have different neural bases. In L1, grammar is subserved by procedural memory, which allows rules or sequences to be applied to semantic content. In L2 however, the procedural system is not initially available to the learner, who must instead rely on declarative memory processes for grammar processing. This reliance on declarative memory is proposed to be dependent on both L2 proficiency and AoA. At earlier AoAs, speakers are less dependent on declarative memory than at later AoAs, and as a speaker becomes more proficient in L2, the underlying neural processes regulating grammar shift to a more native-like state, relying more on procedural functions. This difference is proposed to account for why L2 learning is appreciably more difficult than L1 learning, even though it is still possible for some individuals to achieve high proficiency in their L2. Additionally, the declarative/procedural model highlights the interaction of AoA and proficiency.

In contrast, connectionist-based models of second language processing assume that L1 and L2 are processed by the same brain structures in similar fashions, albeit with L2 requiring greater processing resources within these regions (Abutalebi, 2008; Indefrey, 2006). For instance, Indefrey (2006) has suggested that L1 and L2 rely on similar neurocognitive mechanisms, but lower processing efficiency in late-learning or low-proficiency L2 speakers leads to different patterns of activity. As L2 speakers become more proficient in their L2, their neural language function becomes more efficient, leading to more native-like processing. Similarly, Abutalebi (2008) has proposed that L2 grammar and vocabulary are acquired through structures similar to those in L1. The author suggested that the neural representation of language processing is more extended in L2 speakers, in part due to competition between L1 and L2, but also that, as they become more proficient, processing becomes more automatic and native-like.

Similarly, MacWhinney's Unified Competition Model (2005) posits that, although weakened in L2 acquisition, L1 and L2 acquisition share core learning mechanisms. Linguistic similarity between L1 and L2 is known to affect L2 processing (Jeong, Sugiura, Sassa, Haji, Usui, Taira, Horie, Sato & Kawashima, 2007; Sabourin et al., 2006; Sabourin & Stowe, 2008), and the Competition Model states that both in semantics and syntax, any item that can transfer from L1 and L2, will. However, transfer is most effective earlier in life, when the brain is more plastic.

The effect of AoA, proficiency, and group (i.e., native or L2 speaker) on syntactic processing markers is consistent with the declarative/procedural model of

L1 and L2 syntactic processing. Although both the declarative/procedural model and connectionist models suggest an effect of proficiency, we observed this effect a) on structure violations – which are similar in L1 and L2, thus would not be affected by AoA or group – and b) on later-stage processing of gender violations, suggesting that earlier, automatic syntactic processing depends on different neural mechanisms between L1 and L2 speakers. Thus, while rules that are similar between L1 and L2 may share neural bases, AoA largely predicts how novel syntactic rules are processed. The Competition model predicts an effect of AoA on syntactic processing; however, we also observed an effect of proficiency and group independently of AoA. Additionally, it has been argued that the early age of typical L1 acquisition can itself explain L2 learning outcomes, regardless of L2 AoA (Mayberry & Lock, 2003); yet, while the present study did not examine L1 AoA, we have demonstrated clear influences of both L2 AoA and proficiency on grammatical processing in L2.

Finally, there remain potential confounds with respect to the differences in response to the phrase structure and gender violations. First, phrase structure and gender violations are two different forms of grammatical violation, with phrase structure being purely syntactic and gender violations being morphological. Phrase structure violations were used due to their similarity across English and French. Second, in addition to their cross-language similarity, phrase structure violations may simply be more disruptive and easy to learn than gender violations, resulting in differences in processing. Finally, gender violations were always indexed to nouns; in contrast, structure violations, while usually indexed by a noun, sometimes occurred relative to other types of words instead. This raises the concern that ERP differences between the two violation types might reflect the type of word they occurred in rather than a morphosyntactic process. That said, this explanation seems unlikely given that previous research has identified the LAN/P600 complex in response to both phrase structure violations and gender violations (Barber & Carreiras, 2005; Newman et al., 2007; but see Steinhauer & Drury, 2012). Additionally, number and gender violations have been compared previously in Chinese–Spanish learners, with no difference being found between the two forms of violation (Gillon Dowens et al., 2011). However, the different results between the two violation conditions cannot be solely assigned to cross-language similarities/dissimilarities, and future research should seek to disentangle these potential confounds.

In conclusion, the present study investigated how individual differences in L2 proficiency and AoA (AoA) influenced ERP markers of both familiar and novel grammatical processing. We found that while AoA predicted LAN amplitude in response to

novel grammatical rules, AoA, proficiency, and group membership (L1 vs. L2) predicted P600 amplitude. In contrast, proficiency predicted LAN amplitude to familiar grammatical rules, with no P600 effect. The results of this study highlight the importance of examining individual differences in understanding neural markers of L2 language processing. It similarly highlights the utility of considering similarities and differences between L1 and L2 in this respect. Different effects of AoA and proficiency between gender and structure violations indicate that while it is possible for L2 speakers to acquire novel grammatical rules, this process is different to learning grammatical rules that are present in L1. Additionally, while second language speakers can approach what looks like native-like processing, the fact that they are L2 speakers still affects syntactic resolution independently of both proficiency and AoA, suggesting differing neural mechanisms for syntactic processing of L1 and L2.

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1017/S1366728917000566

References

Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychologica, 128*(*3*)*,* 466–478.

Alarcón, I. V. (2011). Spanish gender agreement under complete and incomplete acquisition: Early and late bilinguals' linguistic behavior within the noun phrase. *Bilingualism: Language and Cognition, 14*(*3*)*,* 332–350.

Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: how order-of-acquisition affects what gets learned. *Cognition, 122*(*3*)*,* 292–305.

Aronoff, M. (1994). *Morphology by Itself: Stems and Inflectional Classes*. The MIT Press, Cambridge, MA.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(*4*)*,* 390–412.

Bagiella, E., Sloan, R. P., & Heitjan, D. F. (2000). Mixed-effects models in psychophysiology. *Psychophysiology, 37*(*1*)*,* 13–20.

Barber, H. A., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: an ERP comparison. *Journal of Cognitive Neuroscience, 17*(*1*)*,* 137–53.

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

Baudiffier, V., Caplan, D., Gaonac'h, D., & Chesnet, D. (2011). The effect of noun animacy on the processing of unambiguous sentences: Evidence from French relative clauses. *The Quarterly Journal of Experimental Psychology, 64*(*10*)*,* 1896–1905.

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychological Review, 113*(*4*)*,* 787–821.

Caffarra, S., & Barber, H. A. (2015). Does the ending matter? The role of gender-to-ending consistency in sentence reading. *Brain Research, 1605*(*1*)*,* 83–92.

Caffarra, S., Molinaro, N., Davidson, D., & Carreiras, M. (2015). Second language syntactic processing revealed through event-related potentials: An empirical review. *Neuroscience and Biobehavioral Reviews, 51*, 31–47.

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

Costa, A., Kovacic, D., Franck, J., & Caramazza, A. (2003). On the autonomy of the grammatical gender systems of the two languages of a bilingual. *Bilingualism: Language and Cognition, 6*(*3*)*,* 181–200.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(*1*)*,* 9–21.

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*(*3*)*,* 379–399.

Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language, 66*(*1*)*,* 226–248.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(*2*)*,* 78–84.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cognitive Brain Research, 1*, 183–192.

Gillon Dowens, M., Guo, T., Guo, J., Barber, H. a., & Carreiras, M. (2011). Gender and number processing in Chinese learners of Spanish - evidence from Event Related Potentials. *Neuropsychologia, 49*(*7*)*,* 1651–9.

Gunter, T. C., Friederici, A. D., & Schriefers, H. (1996). Syntactic Gender and Semantic Expectancy: ERPs Reveal Early Autonomy and Late Interaction. *Journal of Cognitive Neuroscience, 12*(*4*)*,* 556–568.

Hahne, A. (2001). What's Different in Second-Language Processing? Evidence from Event-Related Brain Potentials. *Journal of Psycholinguistic Research, 30*(*3*)*,* 251–266.

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis. Early automatic and late controlled processes. *Journal of Cognitive Neuroscience, 11*(*2*)*,* 194–205.

Hahne, A., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition, 4*(*2*)*,* 123–141.

Hartsuiker, R. J., Beerts, S., Loncke, M., Desmet, T., & Bernolet, S. (2016). Cross-linguistic structural priming in multilinguals: Further evidence for shared syntax. *Journal of Memory and Language, 90*, 14–30.

Hernandez, A. E., & Li, P. (2007). Age of acquisition: Its neural and computational mechanisms. *Psychological Bulletin, 133*(*4*), 638–650.

Indefrey, P. (2006). A Meta-analysis of Hemodynamic Studies on First and Second Language Processing: Which Suggested Differences Can We Trust and What Do They Mean? *Language Learning, 56*, 279–304.

Jeong, H., Sugiura, M., Sassa, Y., Haji, T., Usui, N., Taira, M, Horie, K., Sato, S., & Kawashima, R. (2007). Effect of syntactic similarity on cortical activation during second language processing: A comparison of English and Japanese among native Korean trilinguals. *Human Brain Mapping, 28*(*3*), 194–204.

Johnson, J. S., & Newport, E. L. (1991). Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language. *Cognition, 39*(*3*), 215–258.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. J. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15*(*2*), 159–201.

Keating, G. (2009). Sensitivity to Violation of Gender Agreement in Native and Nonnative Spanish. *Language Learning, 59*(*September*), 503–535.

Kim, A., & Sikos, L. (2011). Conflict and surrender during sentence processing: An ERP study of syntax-semantics interaction. *Brain and Language, 118*(*1–2*), 15–22.

Kotz, S. A. (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain and Language, 109*(*2–3*), 68–74.

Kotz, S. A., & Friederici, A. D. (2003). Electrophysiology of normal and pathological language processing. *Journal of Neurolinguistics*.

Lemhöfer, K., Spalek, K., & Schriefers, H. (2008). Cross-language effects of grammatical gender in bilingual word recognition and production. *Journal of Memory and Language, 59*(*3*), 312–330.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8*(*April*), 1–14.

MacWhinney, B. (1987). Applying the Competition Model to bilingualism. *Applied Psycholinguistics, 8*(*4*), 315–327.

MacWhinney, B. (2005). Extending the Competition Model. *International Journal of Bilingualism, 9*(*1*), 69–84.

Mayberry, R. I., & Lock, E. (2003). Age constraints on first versus second language acquisition: Evidence for linguistic plasticity and epigenesis. *Brain and Language, 87*(*3*), 369–384.

Meulman, N., Wieling, M., Sprenger, S. A., Stowe, L. A., & Schmid, M. S. (2015). Age effects in L2 grammar processing as revealed by ERPs and how (not) to study them. *PLoS One,* 1–31.

Molinaro, N., Barber, H. a., Caffarra, S., & Carreiras, M. (2014). On the left anterior negativity (LAN): The case of morphosyntactic agreement. *Cortex,* 4–7.

Molinaro, N., Vespignani, F., & Job, R. (2008). A deeper reanalysis of a superficial feature: an ERP study on agreement violations. *Brain Research, 1228*, 161–76.

Montrul, S., Foote, R., & Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning, 58*(*3*), 503–553.

Morgan-Short, K., Sanz, C., & Ullman, M. T. (2010). Second Language Acquisition of Gender Agreement in Explicit and Implicit Training Conditions: An Event-Related Potential Study. *Language Learning, 60*(*1*), 154–193.

Neary-Sundquist, C. a. (2013). The development of cohesion in a learner corpus. *Studies in Second Language Learning & Teaching, 3*(*1*), 109–130.

Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically Based Sentence Processing Classes: Evidence fiom Event-Related Brain Potentials. *Journal of Cognitive Neuroscience, 3*(*2*), 151–165.

Newman, A. J., Tremblay, A., Nichols, E. S., Neville, H. J., & Ullman, M. T. (2012). The Influence of Language Proficiency on Lexical Semantic Processing in Native and Late Learners of English. *Journal of Cognitive Neuroscience, 24*(*5*), 1205–1223.

Newman, A. J., Ullman, M. T., Pancheva, R., Waligura, D. L., & Neville, H. J. (2007). An ERP study of regular and irregular English past tense inflection. *NeuroImage, 34*(*1*), 435–45.

Nichols, E. S., & Joanisse, M. F. (2016). Functional activity and white matter microstructure reveal the independent effects of age of acquisition and proficiency on second-language learning. *NeuroImage, 143*, 15–25.

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

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(*1*), 97–113.

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

Osterhout, L., & Mobley, L. a. (1995). Event-Related Brain Potentials Elicited by Failure to Agree. *Journal of Memory and Language*.

Pakulak, E., & Neville, H. J. (2011). Maturational constraints on the recruitment of early processes for syntactic processing. *Journal of Cognitive Neuroscience, 23*(*10*), 2752–65.

Pakulak, E., & Neville, H. J. (2010). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience, 22*(*12*), 2728–44.

Paolieri, D., Cubelli, R., Macizo, P., Bajo, T., Lotto, L., & Job, R. (2010). Grammatical gender processing in Italian and Spanish bilinguals. *Quarterly Journal of Experimental Psychology (2006), 63*(*8*), 1631–45.

Perani, D., Paulesu, E., Galles, N. S., Dupoux, E., Dehaene, S., Bettinardi, V., Cappa, S. F., Fazio, F., & Mehler, J. (1998). The bilingual brain proficiency and age of acquisition of the second language. *Brain, 121*(*10*), 1841–1852.

R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rösler, F., Pütz, P., Friederici, A., & Hahne, A. (1993). Event-Related Brain Potentials while encountering semantic and syntactic contraint violations. *Journal of Cognitive Neuroscience, 5*(*3*), 345–362.

Rossi, E., Kroll, J. F., & Dussias, P. E. (2014). Clitic pronouns reveal the time course of processing gender and number in a second language. *Neuropsychologia, 62*(*1*)*,* 11–25.

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.

Sabourin, L., & Stowe, L. A. (2008). Second language processing: when are first and second languages processed similarly? *Second Language Research, 24*(*3*)*,* 397–430.

Sabourin, L., Stowe, L. A., & de Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research, 22*(*1*)*,* 1–29.

Salamoura, A., & Williams, J. N. (2007). The representation of grammatical gender in the bilingual lexicon: Evidence from Greek and German. *Bilingualism: Language and Cognition, 10*(*3*)*,* 257–275.

Schacht, A., Sommer, W., Shmuilovich, O., Martienz, P. C., & Martin-Loeches, M. (2014). Differential task effects on N400 and P600 elicited by semantic and syntactic violations. *PLoS ONE, 9*(*3*)*,* 1–8.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime reference guide. *Psychology Software Tools, 3*(*1*)*,* 1.

Silva-Pereyra, J., Gutierrez-Sigut, E., & Carreiras, M. (2012). An ERP study of coreference in Spanish: Semantic and grammatical gender cues. *Psychophysiology, 49*(*10*)*,* 1401–1411.

Stevens, G. (1999). Age at immigration and second language proficiency among foreign-born adults. *Language in Society, 28*(*4*)*,* 555–578.

Stowe, L. A., & Sabourin, L. (2005). Imaging the processing of a second language: Effects of maturation and proficiency on the neural processes involved. *International Review of Applied Linguistics in Language Teaching, 43*(*4*)*,* 329–353.

Tanner, D. (2014). On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement. *Cortex,* 1–7.

Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition, 16*(*2*)*,* 367–382.

Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia, 56*(*1*)*,* 289–301.

Tibon, R., & Levy, D. A. (2015). Striking a balance: analyzing unbalanced event-related potential data. *Frontiers in Psychology, 6*(*JAN*)*,* 1–4.

Ullman, M. T. (2001a). The neural basis of lexicon and grammar in first and second language: the declarative/procedural model. *Bilingualism: Language and Cognition, 4*(*1*)*,* 105–122.

Ullman, M. T. (2001b). A neurocognitive perspective on language: the declarative/procedural model. *Nature Reviews. Neuroscience, 2*(*10*)*,* 717–26.

Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., Perani, D., & Olgettina, V. (2003). Early Setting of Grammatical Processing in the Bilingual Brain. *Neuron, 37*(*1*)*,* 159–170.

Weber-Fox, C. M., Davis, L. J., & Cuadrado, E. (2003). Event-related brain potential markers of high-language proficiency in adults. *Brain and Language, 85*(*2*)*,* 231–244.

Weber-Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience, 8*(*3*)*,* 231–256.

Yan, H., Zhang, Y. M., Xu, M., Chen, H. Y., & Wang, Y. H. (2016). What to do if we have nothing to rely on: Late bilinguals process L2 grammatical features like L1 natives. *Journal of Neurolinguistics, 40*, 1–14.