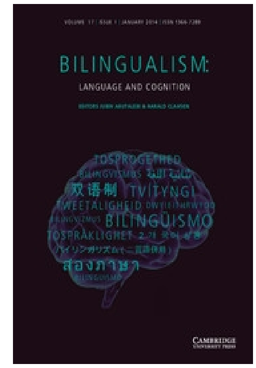

Brain-based individual differences in online L2 grammatical comprehension

DARREN TANNER, KAYO INOUE and LEE OSTERHOUT

Bilingualism: Language and Cognition / Volume 17 / Issue 02 / April 2014, pp 277 - 293
DOI: 10.1017/S1366728913000370, Published online: 13 August 2013

Bilingualism: Language and Cognition 17 (2), 2014, 277–293 © Cambridge University Press 2013 doi:10.1017/S1366728913000370

Brain-based individual differences in online L2 grammatical comprehension*

DARREN TANNER
Department of Linguistics, University of Illinois at Urbana–Champaign
KAYO INOUE
Integrated Brain Imaging Center, Department of Radiology, School of Medicine, University of Washington
LEE OSTERHOUT
Department of Psychology, University of Washington

(Received: January 18, 2013; final revision received: June 18, 2013; accepted: June 19, 2013; first published online 13 August 2013)

Using event-related potentials (ERPs), we investigated the impact of a range of individual difference measures related to L2 learning on proficient L1 Spanish – L2 English bilinguals' brain responses during L2 morphosyntactic processing. Although grand mean ERP analyses revealed a biphasic N400–P600 response to English subject–verb agreement violations, subsequent analyses showed that participants' brain responses varied along a continuum between N400- and P600-dominance. To investigate this pattern, we introduce two novel ERP measures that independently quantify relative brain response type and overall magnitude. Multivariate analyses revealed that larger overall brain responses were associated with higher L2 proficiency, while relative brain response type (N400 or P600) was predicted by a coalition of variables, most notably learners' motivation and age of arrival in an L2 environment. Our findings show that aspects of a learner's background can differentially impact a learner's overall sensitivity to L2 morphosyntax and qualitative use of linguistic cues during processing.

Keywords: ERP, N400, P600, individual differences, morphosyntax, second language acquisition

In an interconnected global society, knowledge of a second language (L2) is an increasingly indispensible skill. Growing globalization in trade, education, and politics requires large numbers of individuals with strong L2 skills, and increasing international immigration is creating a large population of individuals who find themselves needing to master a nonnative language rapidly. However, there is a great deal of variability in the rate, style, trajectory, and ultimate success of L2 learning in adulthood. Understanding what individual-level factors are associated with variability in L2 learning and comprehension is a fundamental question both for cognitive scientists interested in language learning and plasticity in general, as well as for applied researchers interested in identifying cognitive skills or strategies underlying successful learning, identifying gifted language learners for selection in language training programs, or more generally improving learner outcomes. Indeed, some research has sought to characterize predictive cognitive factors or learner strategies associated with rapid and successful learning (e.g., Carroll, 1962; Dörnyei, 2005; Naiman, Fröhlich, Stern & Todesco, 1996; Skehan, 1989). Other studies focusing on experiential factors associated with individuals' long-term learning outcomes have shown that greater success is associated with early immersion, higher motivation to learn, and more frequent L2 use in daily life (e.g., Birdsong & Molis, 2001; Dörnyei, 2005; Flege, Yeni-Komshian & Liu, 1999; Gardner, Tremblay & Masgoret, 1997; Johnson & Newport, 1989).

These studies have largely relied on offline measures such as learners' performance on large test batteries, or subjective measures such as teachers' ratings of learner development over time. However, classroom evaluations and pen-and-paper tests represent a limited way of understanding L2 learning. Proficiency testing provides a continuous, but ultimately unidimensional outcome measure of language knowledge. Moreover, the results of these offline measures may not generalize to learners' real-time language use in context. On the other hand, studying event-related brain potentials (ERPs) allows researchers to investigate the neural mechanisms of real-time language comprehension with millisecond-level temporal resolution. ERPs reflect individuals' brainwave activity that is time- and phase-locked to

* This research was supported by NIDCD grant R01DC01947 to Lee Osterhout and NSF BCS-0951595 to Lee Osterhout and Darren Tanner. Darren Tanner also received support from the William Orr Dingwall Neurolinguistics Dissertation Fellowship and NSF OISE-0968369. We would like to thank Julia Herschensohn, Judith McLaughlin, Janet Nicol, Janet Van Hell and Eleonora Rossi for thoughtful comments and discussion. Our thanks also go to the participants, as well as Geoff Valentine, Kristie Fisher, Shota Moma, and Elliot Collins for their help in acquiring the data. We would also like to thank two anonymous reviewers for helpful comments on a previous version of this manuscript. Any remaining errors are our own.

Address for correspondence:
Darren Tanner, 707 S Matthews Ave., 4080 Foreign Languages Building, MC 168, Urbana, IL 61801, USA
dstanner@gmail.com

the presentation of stimuli, such as words in sentences. Additionally, ERPs are multidimensional, as they provide both quantitative (e.g., effect onset timing and amplitude) and qualitative (e.g., positive or negative effect polarity and scalp distribution) information about processing mechanisms underlying language comprehension (see e.g., Handy, 2005; Luck, 2005, for thorough introductions to ERPs). For example, ERP studies of native language (L1) sentence processing have consistently shown a neurocognitive dissociation between the processing of lexico-semantic and morphosyntactic anomalies. Lexico-semantic manipulations typically trigger an enhanced negativity, peaking around 400 ms poststimulus (the N400 effect: Bentin, 1987; Kutas & Hillyard, 1980, 1984; Osterhout & Nicol, 1999; though see e.g., Kim & Osterhout, 2005; Kuperberg, Kreher, Sitnikova, Caplan & Holcomb, 2007; van de Meerendonk, Kolk, Vissers & Chwilla, 2010), whereas a range of syntactic manipulations, including violations of morphosyntactic rules, typically elicit a large positive-going wave with a maximum around 600 ms poststimulus (the P600 effect: Friederici, Hahne & Mecklinger, 1996; Kaan & Swaab, 2003; Osterhout & Holcomb, 1992; Osterhout & Mobley, 1995).[1]

Based on these L1 findings, N400 and P600 effects can be seen as indices of two independent, but highly interactive "streams" of processing that are differentially sensitive to linguistic cues. Recent conceptions of the N400/P600 dichotomy have posited that dominance of one response or the other reflects a competition between a lexically- or memory-based heuristic processing stream on the one hand, and a combinatorial, algorithmic stream on the other (Kim & Osterhout, 2005; Kuperberg, 2007; Osterhout, Kim & Kuperberg, 2012; see e.g., van de Meerendonk et al., 2010; Van Herten, Chwilla & Kolk, 2006, for a related view). In the context of sentence processing, increased N400 amplitudes are seen when a sentential or discourse context makes a given lexical item more difficult to access from long-term semantic memory or integrate into the still-being-constructed semantic or discourse representation (Hagoort, Hald, Bastiaansen & Petersson, 2004; Kuperberg, 2007; Kutas & Federmeier, 2000; Lau, Phillips & Poeppel, 2008; Van Berkum, Hagoort & Brown, 1999). P600 effects, however, are elicited by engagement of combinatorial processes, which frequently rely on linguistic constraints. These constraints include morphosyntactic rules or predictions (Hagoort & Brown, 2000; Hahne & Friederici, 1999; Osterhout & Mobley, 1995; Osterhout & Nicol, 1999) and verb–argument combinatorial constraints, such as animacy restrictions (Kuperberg, Caplan, Sitnikova, Eddy & Holcomb, 2006; Kuperberg et al., 2007; Paczynski & Kuperberg, 2011). Importantly, this response dichotomy can be used to test whether encountering a particular type of anomaly preferentially engages memory-based lexical or algorithmic combinatorial processing mechanisms.

Given their unique sensitivity to different levels of processing, ERPs can also be useful in characterizing the cognitive mechanisms underlying L2 comprehension. Over the last decade there has been an enormous growth of interest in the neurocognitive substrates of L2 learning and processing (see McLaughlin, Tanner, Pitkänen, Frenck-Mestre, Inoue, Valentine & Osterhout, 2010; Osterhout, McLaughlin, Pitkänen, Frenck-Mestre & Molinaro, 2006; Steinhauer, White & Drury, 2009; Van Hell & Tokowicz, 2010, for reviews of L2 ERP research), and much of this research has focused on identifying whether the neural mechanisms underlying L2 comprehension are fundamentally different from or similar to those observed in L1 populations. As much of the work on L1 processing has assumed that monolinguals always show a P600 effect to syntactic violations (though see below for important caveats to this generalization), one of the driving questions in L2 research on morphosyntactic processing is therefore whether learners can similarly show P600 effects to L2 syntactic violations, and if so, at what point in acquisition they may show them. Although some studies have shown that L2 syntactic violations elicit N400 effects in learners in the earliest stages of acquisition (McLaughlin et al., 2010; Osterhout et al., 2006; Tanner, McLaughlin, Herschensohn & Osterhout, 2013), P600 effects have been observed in relatively low proficiency L2 learners processing violations of syntactic rules common to the L1 and L2 (e.g., McLaughlin et al., 2010; Rossi, Gugler, Friederici & Hahne, 2006; Tokowicz & MacWhinney, 2005), as well as high proficiency learners processing novel L2 features (e.g., Frenck-Mestre, Foucart, Carrasco & Herschensohn, 2009; Gillon Dowens, Guo, Guo, Barber & Carreiras, 2011; Gillon Dowens, Vergara, Barber & Carreiras, 2010; Morgan-Short, Sanz, Steinhauer & Ullman, 2010). Nonetheless, there are some exceptions to this generalization, where P600 effects have failed to be found even in relatively proficient learner groups, usually when L2 features are not realized in the L1 or have different morphological instantiations (e.g., Chen, Shu, Liu, Zhao & Li, 2007; Ojima, Nakata & Kakigi,

[1] Some studies of morphosyntactic processing have reported an additional negative-going wave prominent over left-anterior portions of the scalp, with an onset around between 100 ms and 400 ms poststimulus (the LAN effect: e.g., Friederici et al., 1996; Osterhout & Holcomb, 1992; Osterhout & Mobley, 1995; Rossi, Gugler, Hahne & Friederici, 2005). However, the LAN has been inconsistent across studies (e.g., Allen, Badecker & Osterhout, 2003; Ditman, Holcomb & Kuperberg, 2007; Kaan, 2002; Kaan, Harris, Gibson & Holcomb, 2000; Kaan & Swaab, 2003; Kuperberg et al., 2003; Molinaro, Kim, Vespignani & Job, 2008; Nevins, Dillon, Malhotra & Phillips, 2007; Osterhout & Mobley, 1995), and more research is needed to precisely identify the experimental conditions under which LANs are reliably elicited. We therefore focus on the P600 as an index of morphosyntactic processing.

2005; Sabourin & Haverkort, 2003; Sabourin & Stowe, 2008).

ERPs' differential sensitivity to levels of linguistic processing makes them an ideal candidate for studying how individual difference measures map on to variation in individuals' ERP responses. However, very little research has taken this approach in either L1 or L2 processing. For example, standard approaches to the quantification of ERPs treat inter-subject variability as a source of noise in statistical analyses (e.g., in the error term in ANOVA statistics), and most published ERP waveforms represent the central tendency after averaging the raw electroencephalogram across both trials and subjects. These waveforms therefore may not accurately depict any individual's brain response to a stimulus on a particular trial. Additionally, ERPs' multidimensional nature may make quantifying the relevant dimension of variation difficult, as individuals' effects may differ in amplitude, polarity, timing, or all three. A limited number of studies have investigated the impact of individual differences using grouped designs, where groups are determined by splits on some relevant background measure (e.g., working memory span, comprehension performance, L2 proficiency, or age of arrival: King & Kutas, 1995; Ojima et al., 2005; Rossi et al., 2006; Vos, Gunter, Kolk & Mulder, 2001; Weber-Fox & Neville, 1996; Weckerly & Kutas, 1999). However, this approach provides little information as to the nature of the relationship between the background and outcome measures (e.g., ERP amplitude or polarity), which may be linear and graded (see Van Hell & Tanner, 2012, for related discussion). On the other hand, multivariate correlation- and regression-based statistics have the power to more accurately capture the continuous nature of individual variation, but have only recently been used to study individual differences in ERPs (e.g., Bond, Gabriele, Fiorentino & Alemán Bañón, 2011; Moreno & Kutas, 2005; Newman, Tremblay, Nichols, Neville & Ullman, 2012; Ojima, Matsuba-Kurita, Nakamura, Hoshino & Hagiwara, 2011; Pakulak & Neville, 2010; Tanner et al., 2013).

Our goal here was to investigate the impact of a range of individual difference measures related to L2 learning on ERP responses to morphosyntactic anomalies in late but highly proficient bilinguals, specifically using a multivariate approach. For example, what factors predict relative reliance on memory-based processes versus combinatorial analyses, and what predicts the magnitude of the relevant effects? One particular individual difference variable of interest in recent language processing studies has been proficiency (e.g., Hopp, 2006, 2010; Jackson & Van Hell, 2011; Newman et al., 2012; Ojima et al., 2005; Pakulak & Neville, 2010; Rossi et al., 2006; Van Hell & Tanner, 2012). In a large-scale review paper, Steinhauer et al. (2009) present a proficiency-based neurocognitive model of L2 development. They propose that low-proficiency learners may show N400 effects to grammatical violations (in line with longitudinal and cross-sectional studies of early-stage L2 learners: McLaughlin et al., 2010; Osterhout et al., 2006; Tanner et al., 2013), but that given high enough proficiency, learners will show large P600 or biphasic LAN-P600 responses, as are assumed to be consistently elicited in native speakers (see also Ullman, 2001, 2005, for a similar proposal). Based on this model, one prediction is that marked inter-subject variability in the type or size of brain response may be present in early-stage learners (e.g., Tanner et al., 2013), but this variability should decrease at high proficiency as learners' brain responses approximate native-like targets. That is, variability in highly proficient bilinguals' brain response type or P600 magnitude should be trivial, and to the extent it exists, should be related to L2 proficiency level.

However, some recent findings from the L1 processing literature suggest that variability exists among even proficient language users, including monolinguals processing their native language – a group that has traditionally been assumed to be relatively homogenous in the ERP literature. In some cases this variability was related to participants' L1 proficiency, supporting Steinhauer et al.'s claims. For example, Newman and colleagues (Newman et al., 2012) showed that both L1 and L2 users' N400 magnitudes to semantic anomalies correlated with proficiency measures. Pakulak and Neville (2010) showed that the laterality of an early negativity and the magnitude of the P600 effect in response to simple English phrase structure violations varied continuously with respect to monolingual English speakers' L1 proficiency. However, proficiency is not always implicated in individuals' processing profiles. Osterhout (1997) showed that, among highly literate university students, certain types of syntactic anomalies elicited P600s in some individuals, but N400s in others. Oines, Miyake and Kim (2012) showed that, after controlling for individuals' language experience, vocabulary, and spatial working memory, verbal working memory measures reliably predicted whether implausible verb-argument relations elicited N400 or P600 effects: those with larger spans showed relatively larger P600s, while those with smaller spans showed relatively larger N400s (see also Nakano, Saron & Swaab, 2010, for similar findings). Overall, this suggests that there are robust individual differences in the size and type of ERP responses among monolinguals processing their L1, such that we might expect to see variability in brain response type and size even among very proficient L2 speakers. It remains open to what extent proficiency is implicated in these differences, however.

In addition to proficiency, a number of other individual difference measures have been associated with L2 learning and processing profiles. For example, Weber-Fox & Neville (1996, 1999) suggest that age of arrival in an

L2 environment may be a crucial determinant in whether syntactic anomalies elicit P600 effects (see Birdsong & Molis, 2001; Johnson & Newport, 1989, for age effects on L2 behavior). Behavioral research has additionally implicated learner motivation, amount of L2 exposure (e.g., length of residence in an L2 environment), and frequency of L2 use in daily in determining learner profiles (e.g., Dörnyei, 2005; Dörnyei & Skehan, 2003; Flege et al., 1999; Gardner et al., 1997; Skehan, 1989), though no ERP research has directly investigated potential impacts of these factors in determining how information is processed online.

To investigate these questions, we recorded ERPs while late, but highly proficient L1 Spanish – L2 English bilinguals processed English sentences involving subject–verb agreement, a morphosyntactic rule shared between Spanish and English. Agreement is well-studied using ERPs in L1 processing (see Molinaro, Barber & Carreiras, 2011, for a recent review), as well as in novice and intermediate L2 learners of typologically similar languages (e.g., Frenck-Mestre, Osterhout, McLaughlin & Foucart, 2008; McLaughlin et al., 2010; Tanner et al., 2013) and proficient L2 learners of typologically distinct languages (Chen et al., 2007; Ojima et al., 2005). Based on previous research investigating the processing of features shared between the L1 and L2 (e.g., Osterhout et al., 2006; Tokowicz & MacWhinney, 2005) we predicted that agreement violations would elicit large P600 effects in our highly experienced learners. However, results showed a great deal of variability in both the type and magnitude of learners' brain responses. Using multiple regression, we investigated this variability by assessing the impact of several factors that are either known from behavioral studies to impact learning outcomes or that have been shown to impact language processing: age of arrival, length of residence, frequency of L2 use, language proficiency, and learner motivation. In doing this, we introduce two new outcome measures for ERP waveforms that independently quantify individuals' relative brain response type (N400 or P600) and overall ERP response magnitude.

Table 1. *Background information for participants.*

Measure	Mean	St. Dev.	Range
Age at testing (years)	35.2	6.9	24–49
AoE (years)	11.6	7.1	5–30
AoA (years)	23.9	6.4	15–40
LoR (years)	10.6	6.4	5–27
Self-rated Proficiency (Spanish)			
Speaking	7	0	7–7
Listening	7	0	7–7
Reading	7	0	7–7
Writing	6.95	0.2	6–7
Self-rated Proficiency (English)			
Speaking	5.8	0.7	4–7
Listening	6.2	0.7	4–7
Reading	6.3	0.6	5–7
Writing	5.8	0.9	4–7
Michigan ECPE Proficiency Scores			
Vocabulary (30 possible)	27.7	2.7	20–30
Grammar (20 possible)	17.4	2.3	11–20
Total (50 possible)	45.1	4.2	36–50
L2 Motivation	6.4	0.8	5–7
Frequency use English (overall)	4.6	0.9	3–7

Method

Participants

Our participants included 24 native Spanish speakers who had acquired English as an L2. Data from four participants were excluded from final analysis due to excessive eye movement or other artifact in the raw EEG. Thus, data from 20 participants (7 male) are reported here. All participants were strongly right-handed as assessed by an abridged version of the Edinburgh Handedness Inventory, and had normal or corrected-to-normal vision. In order to ensure that participants had acquired the L2 after childhood, but had sufficient exposure to achieve high proficiency, participants were screened such that they had not been exposed to English in the home, had first moved to an English-speaking country at age 15 or later, and had lived immersed in an English-speaking environment for a minimum of five years. Participants completed a language background questionnaire, which included self-reports on age of initial exposure to English (AoE), age of arrival in an English-speaking environment (AoA), and total length of residence in an English-speaking environment (LoR), as well as proficiency self-ratings for their L1 Spanish and L2 English on a Likert-scale between 1 (no proficiency) and 7 (perfect proficiency). Other questions asked participants about their frequency of use of English in various contexts in daily life, an overall estimate of English use between 1 (never use English) and 7 (always use English), as well as their motivation to speak English like a native speaker between 1 (not important to sound like a native speaker at all) and 7 (extremely important to sound like a native speaker). Participants also completed a pen-and-paper proficiency test consisting of 50 questions selected from the Michigan Examination for the Certificate of Proficiency in English (ECPE). Participants' responses are reported in Table 1. Additionally, eight of the participants reported no significant competency in languages other than Spanish and English. The remaining participants reported varying non-native competence in other European languages, including French, Italian, German, Portuguese, and Catalan. Participants provided informed consent and received a small amount of cash for taking part.

Materials

Stimuli were sentences that were either grammatically correct or contained a violation of subject–verb agreement. All sentences contained a singular subject noun, followed by a prepositional phrase modifier containing another singular noun, followed by a verb that either agreed or disagreed with the subject noun in number (*is*/*are*, *was*/*were*, or *has*/*have*), followed by a short predicate (e.g., *The winner of the big trophy has*/**have proud parents*). Two hundred and forty sentence frames were constructed with eight versions of each sentence. Two versions corresponded to the grammatical–ungrammatical sentence pairs reported here; the other six versions investigated agreement in more complex syntactic configurations (see Tanner, 2011; Tanner, Nicol, Herschensohn & Osterhout, 2012). Sentences were distributed across eight experimental lists in a Latin square design, such that each list contained only one version of each sentence frame. Each list thus contained 30 grammatical and 30 ungrammatical sentences in the experiment reported here. Experimental sentences were pseudo-randomized among other sentences belonging to other experiments not reported here. Half of the filler sentences were anomalous, either with an anomaly of verb tense (e.g., **The man will cooked the food in the refrigerator*) or lexical semantics (e.g., **Jane will bake a book in her spare time*). Each list contained 540 sentences, half of which contained either a syntactic or semantic anomaly.

Procedure

Participants took part in three sessions. The first session involved completion of the background questionnaires and proficiency test; ERPs were recorded during the two subsequent sessions. Because of the large number of experimental sentences, participants saw half of the sentences at each session; experimental sentences were balanced across sessions, such that each participant saw an equal number of sentences from each condition at both session. ERPs were first averaged within sessions. No between-session differences were found, so ERPs were subsequently averaged across sessions. All reported results reflect these cross-session averages. During ERP recording participants were seated in a comfortable recliner in front of a CRT monitor. Participants were instructed to relax and minimize movements and eye blinks while reading and to read each sentence as normally as possible. Each trial consisted of the following events: each sentence was preceded by a blank screen for 1000 ms, followed by a fixation cross, followed by a stimulus sentence, presented one word at a time. The fixation cross and each word appeared on the screen for 400 ms followed by a 200 ms interstimulus interval. Sentence-ending words appeared with a full stop. This screen was followed by a "yes/no" prompt asking for a sentence acceptability judgment. Participants were instructed to respond "yes" to sentences that were well-formed and semantically coherent and "no" to sentences that were ungrammatical or semantically incoherent. Participants were randomly assigned to use either their left or right hand for the "yes" response.

Data acquisition and analysis

Continuous EEG was recorded from 19 tin electrodes attached to an elastic cap (Electro-cap International) in accordance with the 10–20 system (Jasper, 1958). Eye movements and blinks were monitored by two electrodes, one placed beneath the left eye and one placed to the right of the right eye. Electrodes were referenced to an electrode placed over the left mastoid. EEG was also recorded from an electrode placed on the right mastoid to determine of there were experimental effects detectable on the mastoids. No such effects were found. EEG signals were amplified with a bandpass of 0.01–100 Hz (3db cutoff) by an SAI bioamplifier system. ERP waveforms were filtered offline below 15 Hz. Impedances at scalp and mastoid electrodes were held below 5 kΩ and below 15 kΩ at eye electrodes.

Continuous analog-to-digital conversion of the EEG and stimulus trigger codes was performed at a sampling frequency of 250 Hz. ERPs, time-locked to the onset of the critical verb (underlined in the examples above), were averaged offline for each participant at each electrode site in each condition. Trials characterized by eye blinks, excessive muscle artifact, or amplifier blocking were not included in the averages; 5.8% and 5.5% of grammatical and ungrammatical trials were rejected, respectively. ERPs were quantified as mean amplitude within a given time window. All artifact-free trials were included in the ERP analyses. Because of a small amount of noise in the prestimulus interval, a 50 ms prestimulus to 50 ms poststimulus baseline was used to mitigate early differences in the waveforms. In accordance with previous literature and visual inspection of the grand mean waveforms, the time windows 400–500 ms and 500–1000 ms were chosen, as they correspond roughly to the N400 and P600 effects, respectively.[2] For grand mean analyses, ANOVAs were calculated within

[2] The 400–500 ms time window begins somewhat later than the 300–500 ms time window frequently used to quantify the N400 in studies of monolingual language processing. However, N400 effects have reported to be somewhat delayed in some studies of L2 processing, relative to L1 processing (Hahne, 2001).

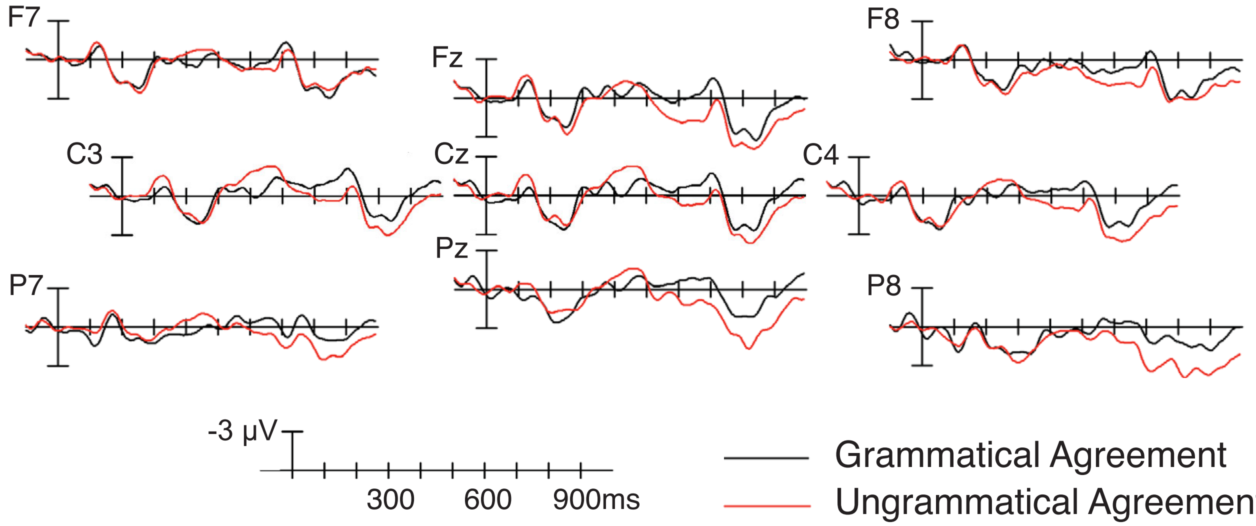

Figure 1. (Colour online) Grand mean waveforms for grammatical (black line) and ungrammatical (red line) verbs from nine representative electrodes. Onset of the critical verb is indicated by the vertical bar. Calibration bar shows 3 μV of activity; each tick mark represents 100 ms of time. Positive voltage is plotted down.

each time window with grammaticality (grammatical, ungrammatical) as a within-subjects factor. Data from midline (Fz, Cz, Pz), medial–lateral (right hemisphere: Fp2, F4, C4, P4, O2; left hemisphere: Fp1, F3, C3, P3, O1), and lateral–lateral (right hemisphere: F8, T8, P8; left hemisphere: F7, T7, P7) electrode sites were treated separately in order to identify topographic and hemispheric differences. ANOVAs on midline electrodes included electrode as an additional within-subjects factor (three levels), ANOVAs on medial–lateral electrodes included hemisphere (two levels) and electrode pair (five levels) as additional within-subjects factors, and ANOVAs over lateral–lateral electrodes included hemisphere (two levels) and electrode pair (three levels) as additional within-subjects factors. The Greenhouse-Geisser correction for inhomogeneity of variance was applied to all repeated measures on ERP data with greater than one degree of freedom in the numerator. In such cases, the corrected p-value is reported. Regression models will be described with the results.

Results

Grand mean results

Results from the end-of-sentence judgment task showed that participants were highly accurate in judging the acceptability of the sentences (grammatical mean proportion correct = .90, SE = .02; ungrammatical mean proportion correct = .88, SE = .04). Grand mean ERP results for grammatical and ungrammatical verbs are depicted in Figure 1. Visual inspection of the waveforms showed that, relative to grammatical verbs, ungrammatical verbs elicited a small biphasic waveform characterized by a centrally-distributed negativity between approximately 400 ms and 500 ms poststimulus (N400), followed by a broadly-distributed positivity (P600). Statistical analysis confirmed these observations. In the 400–500 ms window there was an effect of grammaticality that was strongest over central electrodes (grammaticality $\times$ electrode interaction: midline, $F(2,38) = 4.967, p < .03$; medial–lateral, $F(4,76) = 4.721, p < .02$), and which showed a stronger left-hemisphere distribution over lateral–lateral sites (grammaticality $\times$ hemisphere interaction, $F(1,19) = 9.190, p < .01$). Between 500 ms and 1000 ms the positivity reached significance over a broad portion of the scalp (main effect of grammaticality: midline, $F(1,19) = 8.719, p < .01$; medial–lateral, $F(1,19) = 10.292, p < .01$; lateral–lateral. $F(1,19) = 7.519, p < .02$). Over lateral–lateral sites the positivity showed a stronger posterior (grammaticality $\times$ electrode interaction: $F(2,38) = 4.299, p < .03$) and right-hemisphere distribution (grammaticality $\times$ hemisphere interaction: $F(1,19) = 10.776, p < .01$).

Individual differences analyses

Despite the significant biphasic N400–P600 results in the omnibus ANOVA, inspection of individuals' ERP waveforms showed that some individuals showed primarily either an N400 or P600 to ungrammatical verbs, but not both, while others showed a biphasic response. Following Inoue & Osterhout (2013) and Tanner et al. (2013), we regressed individuals' N400 effect magnitude onto their P600 effect magnitude. To compute these measures, we calculated participants'

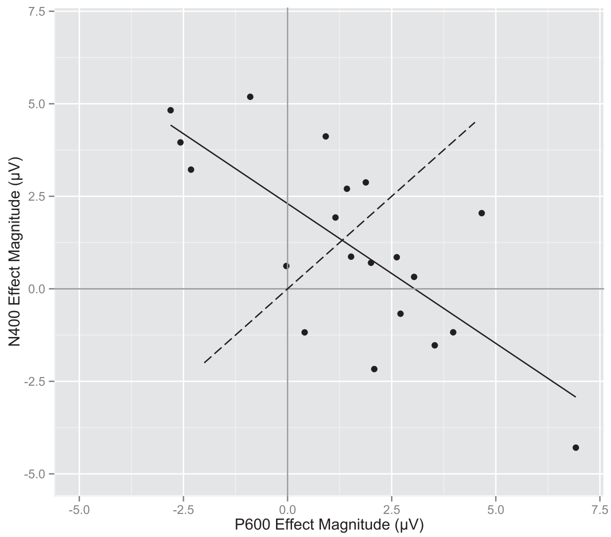

Figure 2. Scatterplot showing the distribution of N400 and P600 effect magnitudes across participants, averaged within a centro-parietal ROI (C3, Cz, C4, P3, Pz, P4). Each dot represents a data point from a single participant. The solid line shows the best-fit line for the data from the regression analysis. The dashed line represents equal N400 and P600 effect magnitudes: individuals above/to the left of the dashed line showed primarily an N400 effect to agreement violations, while individuals below/to the right of the dashed line showed primarily a P600 effect.

mean amplitudes in a centro-parietal region of interest (electrodes C3, Cz, C4, P3, Pz, P4), where N400 and P600 effects are typically largest. N400 effect magnitude was calculated as mean activity in the grammatical minus ungrammatical condition between 400 ms and 500 ms; P600 effect magnitude was calculated as mean activity in the ungrammatical minus grammatical conditions between 500 ms and 1000 ms. Results showed the two effects to be highly negatively correlated, $r = -.737$, $p = .0002$, Figure 2. Brain responses showed a continuous distribution between negativity-dominance, biphasic, and positivity-dominance across individuals; as one effect increased, the other decreased to a similar degree. To illustrate these differences, we averaged ERPs separately for those who showed a negativity-dominance (e.g., those above/to the left of the dashed line in Figure 2) and those who showed a positivity-dominance (e.g., those below/to the right of the dashed line in Figure 2). Figure 3 shows mean waveforms for these two groups. Those in the negativity-dominant group showed a large, centrally-distributed negativity to ungrammatical verbs between approximately 400 ms and 600 ms (an N400), but no later positivity; those in the positivity-dominant group showed no earlier negativity, but instead a large, posteriorly distributed positivity between 500 ms and 1000 ms (a P600). However, as can be seen in Figure 2, the true distribution of brain response dominance across individuals is continuous. Thus, grand mean waveforms depicted an effect that was not representative of most individuals' ERP profiles.

To investigate what factors predict the type and magnitude of ERP response to agreement violations in this group of learners, we computed two new measures to serve as the dependent variables (DVs) in multiple regression models. For the first measure, the Response Dominance Index (RDI), we measured each individual's relative response dominance (N400 or P600) by fitting the individual's least squares distance from the equal effect sizes line (the dashed line in Figure 2) using perpendicular offsets. RDI values near zero reflect relatively equal-sized N400 and P600 effects, while more negative or positive values reflect relatively larger negativities or positivities across both time windows, respectively. While the RDI gives an indication about response dominance, it does not provide a strong measure of overall response size. For example, it would not differentiate between two individuals who each showed equal-sized N400 and P600 effects, but where one individual showed relatively small effects and the other showed a large biphasic response. Both individuals would have an RDI value near zero, though the second individual is clearly more sensitive to the agreement violations than the first. We therefore computed a second measure, the Response Magnitude Index (RMI), by calculating each individual's Euclidian distance from zero in Figure 2. The RMI gives a measure of the overall level of sensitivity an individual shows to the agreement violations within the N400 and P600 time windows. Larger values of the RMI indicate relatively greater neural responses to the agreement violations across both time windows, regardless of the type of the response. Equations (1) and (2) show how the RMI and RDI were computed, where N400 and P600 refer to mean amplitude between 400 ms and 500 ms, and 500 ms and 1000 ms, respectively, averaged within the centro-parietal ROI (C3, Cz, C4, P3, Pz, P4).

$$(1)\quad RMI = \sqrt{N400_{Gram} - N400_{Ungram})^2 + (P600_{Ungram} - P600_{Gram})^2}$$

$$(2)\quad RDI = \frac{P600_{Ungram} - P600_{Gram}) - (N400_{Gram} - N400_{Ungram})}{\sqrt{2}}$$

We fit two multiple regression models with the RMI and RDI serving as the respective DVs. Five background measures known to affect learning outcomes and processing profiles were selected as independent variables for the models: AoA, LoR, frequency of L2 use, L2 proficiency (ECPE scores), and motivation to speak like a native speaker (see Table 1). Because of right-skewed distributions, AoA and LoR were log-transformed prior to entry into the regression models. Additionally,

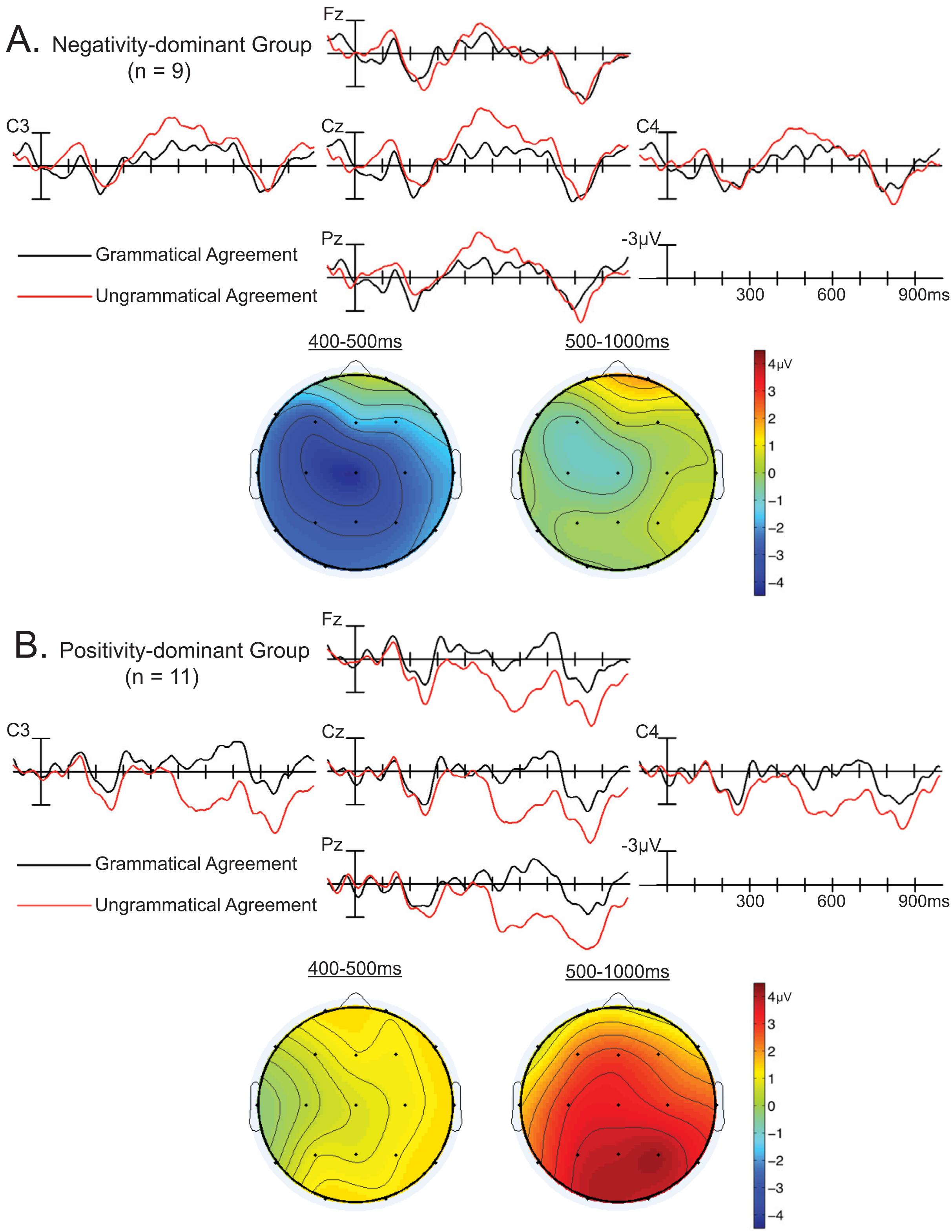

Figure 3. (Colour online) Grand mean waveforms and scalp topographies of grammaticality effects (i.e., mean amplitude in ungrammatical minus grammatical condition) presented separately for those who showed primarily N400 (Panel A) and P600 (Panel B) effects to subject–verb agreement violations. Onset of the critical verb is indicated by the vertical bar. Calibration bar shows 3μV of activity; each tick mark represents 100 ms of time. Positive voltage is plotted down. Topographic map scale is in microvolts.

Table 2. *Mean, standard deviation, and range for DVs and background measures used in regression models.*

Measure	Mean	St. Dev.	Range
RDI	0.25	3.28	−5.40–7.93
RMI	3.55	1.71	0.62–8.14
LogAoA	1.36	0.11	1.18–1.60
LogLoR	1.00	0.22	0.70–1.43
Freq.	4.60	0.88	3–7
Proficiency	45.05	4.17	36–50
Motivation	0.55	0.51	0–1

RDI = Response Dominance Index; RMI = Response Magnitude Index; LogAoA = Log-Age of Arrival; LogLoR = Log-Length of Residence; Freq. = Frequency of L2 use; Proficiency = Total ECPE score; Motivation = Dichotomized Motivation self-report

since all participants reported having relatively high motivation to speak like a native English speaker, this measure was dichotomized into those reporting very high motivation (i.e., those reporting 7 on the Likert-scale, n = 11) and those reporting high motivation (i.e., those reporting 5 or 6 on the Likert-scale, n = 9). "High" motivation to speak like a native speaker set as the reference level (0) and "very high" motivation set to 1 in the models. We also fit regression models with the motivation variable undichotomized. Results were qualitatively similar. However, we chose to include the dichotomized variable in the final model because the distribution of the variable was highly skewed and there were a large number of observations at one end of the distribution. Dichotomization is an appropriate transformation in such circumstances (MacCallum, Zhang, Preacher & Rucker, 2002). Distributional statistics for the DVs and background measures entered into the models are reported in Table 2; the correlation matrix for the measures is reported in Table 3. Collinearity diagnostics showed that multicollinearity among predictor variables was not a problem (tolerances > .7, VIFs < 1.5). Residuals of both models were approximately normally distributed.

The total model predicting the RMI was not significant (R^2 = .371, Adjusted R^2 = .146, $F(5,14) = 1.650$, $p = .211$), though one individual predictor reached significance. After controlling for the effects of AoA, LoR, frequency of use, and motivation to speak like a native speaker, English proficiency showed a positive association with the RMI ($B = .209$, $\beta = .509$, partial-$r = .517$, $p = .04$). This shows that those who were more proficient in an offline pen-and-paper test also showed larger overall neural sensitivity to agreement violations. In order to specifically test the prediction made by Steinhauer et al. (2009; see also Rossi et al., 2006) that N400 magnitude would decrease and P600 magnitude would increase with increasing proficiency, we correlated individuals' respective effect magnitudes with ECPE proficiency measures. Neither N400 magnitudes ($r = .070$) nor P600 magnitudes ($r = .084$) correlated with individuals' L2 proficiency, suggesting that proficiency is most associated with overall response magnitude, as measured by the RMI, rather than the magnitude of specifically the N400 or P600 response in particular. The model predicting brain response dominance (the RDI) was highly significant (Table 4), such that the linear combination of AoA, LoR, frequency of use, L2 proficiency and motivation to speak like a native speaker accounted for approximately 54% of the variance in response dominance. Two individual predictors also reached significance. AoA and motivation to speak like a native speaker each uniquely predicted response dominance: those who experienced earlier immersion in an English-speaking environment and who reported higher motivation to speak in a native-like way showed stronger P600-dominant brain responses, while those who arrived later and reported lower motivation to speak like a

Table 3. *Correlation matrix for DVs and background measures used in regression models.*

	RDI	RMI	LogAoA	LogLoR	Freq.	Proficiency	Motivation
RDI	–						
RMI	.076	–					
LogAoA	−.277	−.327	–				
LogLoR	.445*	.097	−.367	–			
Freq.	.286	.104	−.164	.235	–		
Proficiency	.007	.478*	−.003	.264	.077	–	
Motivation	.517*	.031	.385	.016	.047	.209	–

$p < .05$
RDI = Response Dominance Index; RMI = Response Magnitude Index; LogAoA = Log-Age of Arrival; LogLoR = Log-Length of Residence; Freq. = Frequency of L2 use; Proficiency = Total ECPE score; Motivation = Dichotomized Motivation self-report

Table 4. *Multiple regression coefficients for model with RDI as DV.*

Predictor	B	SE	β	t	Partial-r	*p*
LogAoA	−13.20	5.43	−0.42	−2.25	−.515	.041
LogLoR	4.608	2.64	0.31	1.75	.423	.103
Freq.	0.478	0.60	0.13	0.80	.208	.439
Proficiency	−0.186	0.13	−0.24	−1.43	−.357	.175
Motivation	4.60	1.28	0.72	4.08	.737	.001

Note: $R^2 = .660$, Adjusted $R^2 = .539$, $F(5,14) = 5.434$, $p = .006$. B and β represent the unstandardized and standardized regression coefficients, respectively.

native speaker were more likely to show N400-dominant responses. To test whether the linear combination of the non-significant predictor variables (LoR, proficiency, and frequency of use) together provided predictive power over and above motivation and AoA alone, we subdivided the RDI model into two hierarchically related models: the first (nested) model included only the two significant predictors (AoA and motivation), while the second (full) model added the remaining three predictors. In the first model, AoA and motivation to speak like a native speaker were individually significant (partial-*r*s = −.603 and .703, respectively, *p*s < .007), and together accounted for a significant amount of variance in the RDI, $R^2 = .533$, Adjusted $R^2 = .479$, $F(2,17) = 9.718$, $p = .002$. Adding LoR, proficiency, and frequency of use did not add a significant amount of predictive power to the model, however, R^2-change = .127, $F(3,14) = 1.737$, $p = .205$.

Discussion

We investigated the impact of a number of variables related to L2 learning on proficient L1 Spanish – L2 English bilinguals' brain responses to violations of English subject–verb agreement. Although grand mean analyses showed a reliable biphasic N400–P600 pattern, individual analyses showed that this pattern was not representative of most learners' brain responses. Individuals' brain responses varied along a continuum between N400- and P600-dominant effects, with most participants showing dominance in one response or the other and some showing relatively equal dominance. Thus, contrary to some models of L2 processing, a large amount of variability in brain responses existed among the bilinguals we tested, despite relatively high L2 proficiency and long-term L2 exposure. Importantly, multivariate analyses showed that a large portion of this between-subject variability was systematic. By differentially quantifying the type and magnitude of brain responses, we showed that different aspects of the ERP signal (e.g., amplitude and polarity) were associated with different aspects of a learner's background. Earlier age of arrival and higher motivation to speak like a native English speaker were associated with greater dominance of the P600 effect, relative to the N400 effect. Additionally, the linear combination of age of arrival, length of residence, frequency of use, and motivation to speak in a native-like way accounted for a majority of the overall variance in the type of brain response individuals showed to agreement violations. However, contrary to predictions made by the model proposed by Steinhauer and colleagues (Steinhauer et al., 2009), proficiency had little impact on relative response dominance, as measured by the RDI, nor on individuals' N400 or P600 magnitudes specifically. Instead, higher performance on the pen-and-paper proficiency test was instead associated with a larger overall brain response across both time windows, regardless of the type of response. It is important to note that we found these partial correlations and multivariate effects despite a relatively small sample size. This indicates exceptionally strong and systematic relationships, but also introduces a cautionary note in interpreting the null effects seen in the regression analyses.

Our findings add to a growing number of L2 studies reporting N400 responses to syntactic violations and extend these findings in important ways. Several longitudinal studies have reported that syntactic anomalies elicit N400 effects in early-stage L2 learners, but elicit P600 effects after increased L2 instruction (McLaughlin et al., 2010; Morgan-Short et al., 2010; Morgan-Short, Steinhauer, Sanz & Ullman, 2012; Osterhout et al., 2006). Furthermore, Tanner and colleagues (Tanner et al., 2013) report cross-sectional results from novice L2 learners who showed a nearly identical distribution of brain responses to that reported here. In most of these studies, N400 effects are associated with the earliest stages of learning and are replaced by P600 effects within a single year of classroom L2 instruction. However, in the present study, N400 effects to L2 agreement violations persisted in some of the bilinguals, despite long-term L2 immersion and high L2 proficiency. There is therefore a seeming inconsistency between the current results from immersed, proficient bilinguals and reports of low- and intermediate-proficiency learners showing robust P600 (and no N400) effects to inflectional anomalies (Bond et al., 2011; Frenck-Mestre et al., 2008; McLaughlin et al., 2010; Rossi et al., 2006; Tokowicz & MacWhinney, 2005). Since N400 effects to syntactic violations have been associated with the earliest stage of grammatical learning, it is conceivable that processing systems in the N400-dominant learners in the current study may have somehow fossilized at an earlier stage of acquisition. One argument might therefore be that N400 effects in late-stage bilinguals represent instances of poor L2 learning, at least at a neurocognitive level.

The significant positive partial correlations between the response dominance index and learners' age of arrival and motivation speak like a native speaker are consistent with this hypothesis, as behavioral research has long noted relationships between these variables and long-term learning outcomes. For example, Birdsong and Molis (2001) showed that among L1 Spanish speakers of L2 English – a learner population similar to that tested here – increasing age of arrival in an English-speaking environment was negatively correlated with performance on a grammaticality judgment task, even among late learners who arrived after a putative critical period for language learning (see also Flege et al., 1999; compare Johnson & Newport, 1989). A similar effect was found among our late learners; however, instead of showing a relationship with behavioral measures of L2 knowledge, age of arrival correlated with the quality of brain responses to syntactic anomalies in our ERP data (i.e., N400- or P600-dominance). Additionally, our results provide evidence that motivation can impact learners' online brain responses during language comprehension in a similar way, where P600-dominance correlated with higher motivation to speak like a native speaker. Behavioral research has shown that learner motivation correlates with increased L2 achievement (Gardner et al., 1997), even after controlling for confounding factors such as age of acquisition and length of residence (Flege et al., 1999). However, our measure did not differentiate between sub-types of learner motivation or provide a detailed quantification of learner identity (see e.g., Dörnyei, 2005, for a thorough overview of these issues), and future research may wish to more systematically investigate the role of motivation on the neurocognition of an L2 (see Pulvermüller & Schumann, 1994, for one such account). Nonetheless, our data show that higher motivation to speak like a native speaker, along with an earlier age of arrival in an L2 environment, are associated with ERP responses that more closely approximate those typically seen in grand mean waveforms in L1 experiments investigating similar phenomena (e.g., Hagoort & Brown, 2000; Hagoort, Brown & Groothusen, 1993; Osterhout & Mobley, 1995; Rossi, Gugler, Hahne & Friederici, 2005).

However, a broader view of the relevant literature raises doubts about the "fossilization" interpretation of the N400 effect in our bilinguals. Syntactic violations have sometimes elicited N400 or biphasic N400–P600 responses similar to those reported here in L1 sentence processing studies (e.g., Deutsch & Bentin, 2001; Mancini, Molinaro, Rizzi & Carreiras, 2011a, b; Osterhout, 1997; Severens, Jansma & Hartsuiker, 2008). As was the case in the current study, some L1 studies have reported biphasic responses that were shown to be artifacts resulting from averaging over individuals with different response profiles (Inoue & Osterhout, 2013; Nieuwland & Van Berkum, 2008; Osterhout, 1997; Osterhout, McLaughlin, Kim, Greewald & Inoue, 2004; see also Kos, van den Brink & Hagoort, 2012; Nakano, et al., 2010). Moreover, the "response dominance continuum" reported here in bilinguals has also been observed in studies of L1 syntactic processing involving Japanese-speaking adults (Inoue & Osterhout, 2013; see also Osterhout et al., 2004), as well as English-speaking adults processing subject–verb agreement violations similar to those studied here. For example, Tanner and Van Hell (2012) present findings from monolingual native English speakers processing morphosyntactic violations similar to those used here. Although grand mean waveforms in that study showed only a P600 effect to agreement anomalies, subsequent analysis showed a similar continuum of brain responses to that reported here: although most individuals showed a P600 effect to English morphosyntactic violations, nearly one-third of participants showed an N400-dominant response. However, the early onset of the P600 effect in some individuals cancelled out the N400 seen in others, resulting in little negativity in the grand mean. Collectively, these results indicate that the individual differences reported here exist in early-stage L2 learners, later-stage bilinguals, as well native speakers. That is, individual differences in the N400–P600 continuum are not restricted to L2 learners, but instead seem to be a property of language processing in general. Although age of arrival and motivation were implicated in predicting brain response type in the L2 learners studied here, these must be two of many interacting variables associated with ERP response profiles across the spectrum of language users.

Individual differences in adult language processing have previously been linked to variation in working memory capacity, inhibitory control, neural efficiency, and cortical connectivity, among other variables (e.g., Bates, Devescovi & Wulfeck, 2001; Gernsbacher, 1993; Just & Carpenter, 1992; Prat, 2011; Prat, Keller & Just, 2007; Van Petten, Weckerly, McIsaac & Kutas, 1997; Vos et al., 2001). Although these and other possibilities should be explored further, the sentence processing model proposed by Ferreira, Christianson and colleagues (Christianson, Hollingworth, Halliwell & Ferreira, 2001; Ferreira, 2003) might prove to be particularly efficacious for explaining the results reported here. Ferreira, Christianson and colleagues argue that, in certain cases, comprehenders fail to compute detailed syntactic representations of incoming linguistic information. In these circumstances individuals may instead rely on shallower, lexically- or plausibility-based "good enough" processing heuristics. For example, in complex syntactic configurations such as passives or garden path sentences, individuals may place greater weight on lexical or semantic than syntactic cues, especially when the two sets of cues conflict (Ferreira, 2003; see also Kim & Osterhout, 2005). Seen in this context, the processing of linguistic anomalies could proceed along two possible paths, one based on lexical associations stored in memory (i.e., a shallower, "good enough" analysis) or one based on combinatorial, syntactic analyses. Relative reliance on one stream or the other might then depend on the relative strength of lexical and syntactic cues in the sentence. The N400 and P600 effects may therefore be seen as indexing the preferential engagement of one stream or the other (Kuperberg, 2007; Osterhout et al., 2012; see Jackendoff, 2007; MacWhinney, Bates & Kliegl, 1984, for processing models that similarly propose a competitive dynamic between lexical/conceptual and syntactic cues; see also Morgan-Short, Faretta-Stutenberg, Brill, Carpenter & Wong (published online March 1, 2013); Ullman, 2004, 2005, for memory-based models that propose that separable declarative and procedural memory systems may subserve these two processing streams and that have implications for L2 learning).

Our results suggest further that cue strength and reliance on "good enough" processing heuristics may differ across individuals. Some individuals seem to rely more heavily on memory-based heuristics (i.e., they are N400-dominant) and others on combinatorial information during sentence processing (i.e., they are P600-dominant). Under this interpretation, the RDI represents a metric of cue strength, a value that can vary across individuals at all levels of language proficiency. For morphosyntactic dependencies like agreement, the relative weighting of the cues (i.e., response dominance) seems to be especially malleable in many early-stage L2 learners, such that additional instruction can produce a within-learner shift from an N400 response to a P600 response to the same syntactic anomalies (McLaughlin et al., 2010). Presumably this reflects the gradual acquisition of a grammatical rule and subsequent reduced reliance on lexical information for dealing with that particular aspect of the L2. In L1 adults and later-stage L2 learners, response dominance may reflect more stable processing biases, though this interpretation has yet to be verified empirically.

From this perspective, between-subject variability in the response dominance continuum for both L1 and L2 processing would be minimized by narrowing the set of relevant cues and maximized by broadening the set of cues. For example, the syntactic violations in the current study were realized with suppletive lexical alternations between short, high-frequency verbs (e.g., *is*/*are*, *has*/*have*). Those showing N400-dominant effects may have relied upon lexical form-based expectations of verbal agreement features, while those showing P600-dominant responses may have been more sensitive to the syntactic features carried on the verbs. The availability of multiple cues to subject–verb agreement violations (wordform-based versus syntactic-feature–based) may allow for variability in the exact cue that individuals attend to. Conversely, in contexts where syntactic relationships are marked morphologically, the morphosyntactic cue to the agreement dependency would be strengthened and violations should be more likely to elicit P600 effects across all individuals. One potential test of this hypothesis is provided by Foucart and Frenck-Mestre (2012), who investigated grammatical gender processing in L1 English learners of L2 French. In L2 learners, gender agreement violations elicited P600 effects when the anomaly was detectable on adjectives, which are morphologically marked for gender, but an N400 effect when the anomaly was detectable on nouns, where gender is encoded lexically and not morphologically. The relative strength of phonological cues to morphosyntactic well-formedness may also play a role in determining the neural signatures of processing agreement morphology. Frenck-Mestre and colleagues (Frenck-Mestre et al., 2008) showed in L1 German learners of L2 French that visually-presented subject–verb agreement violations elicited a robust P600 effect when there was a strong phonological cue to the violation, but a trend toward an N400 effect when there was no phonological cue. Thus, a coalition of morphosyntactic and morphophonological cues provided by target stimulus may play a role in preferentially engaging one stream or the other.

Indeed, this may explain the discrepancy between the current findings and other L2 studies that have found P600 (but no N400) effects to morphosyntactic anomalies in learners across the proficiency spectrum. P600 effects have been found most strongly in studies that used agreement relationships marked by phonologically-realized, decomposable inflectional morphemes (e.g., Frenck-Mestre et al., 2008; McLaughlin et al., 2010; Rossi et al., 2006; Sabourin & Stowe, 2008; Tokowicz & MacWhinney, 2005; see also Hahne, Mueller & Clahsen, 2006). Taken in this context, our findings suggest that later and less motivated learners may prioritize lexical cues to syntactic relationships when they are available. However, given that P600 effects are readily found in L2 learners across the proficiency spectrum, the broader L2 processing literature suggests that late learners are in no way restricted to these shallow processes.

Instead, our findings are consistent with the notion that individuals may modulate their relative reliance on the lexical/semantic or morphosyntactic processing streams based on the cues provided by the target stimulus.

Studies of L1 processing further corroborate this proposal. For example, Osterhout (1997) showed that syntactically anomalous function words (which play a grammatical role but have little referential meaning) elicited P600 effects in most participants. By contrast, syntactically anomalous content words (which have rich referential meanings as well as grammatical roles) elicited robust P600 effects in some subjects and N400 effects in others (Osterhout, 1997). Similarly, anomalous Japanese case markers, which simultaneously indicate the thematic and grammatical roles of each noun, elicit robust P600 effects in some Japanese-speaking adults and N400 effects in others (Inoue & Osterhout, 2013; Osterhout et al., 2004). An important implication of all of these results is that the grand mean brain response observed in any given experiment will be a function of not only the manipulated properties of the stimulus (e.g., its linguistic category, morphological complexity, well-formedness) but also systematic properties of the participants (e.g., each subject's weighting of the relevant linguistic cues), and an interaction between the two. Repeated sampling from a population would by chance produce different outcomes. Collectively, these results illustrate the dangers inherent in the exclusive use of grand average ERPs to characterize sentence processing. In some cases (as in the present study), a thorough investigation of between-subject variability can provide an important complement to traditional grand mean analyses.

In all of these experiments (including the present study), we have observed a robust negative correlation between N400 and P600 magnitudes across individuals: as the magnitude of one effect increased in size, the magnitude of the other effect decreased to a similar extent (Fig. 2). We have observed the negative correlation in L2 learners and in native speakers (Inoue & Osterhout, 2013; Tanner, 2011; Tanner et al., 2013; Tanner & Van Hell, 2012; see also Nakano et al, 2010; Nieuwland & Van Berkum, 2008; Osterhout 1997; Zhang, Li, Piao, Liu, Huang & Shu, 2013). Conceivably, the negative correlation manifests a dynamic interaction between negatively correlated cortical networks involved in different aspects of sentence comprehension (for example, lexico-semantic and grammatical aspects). Evidence to support this notion has been acquired recently in an experiment that combined an ERP linguistic anomaly paradigm with distributed source analysis (Inoue & Osterhout, 2013). This experiment demonstrated a negative correlation in activity within anterior and posterior regions of the left-hemisphere perisylvian cortex when participants read well-formed sentences. Individual differences in the "balance of activation" within these areas when reading well-formed sentences predicted individual differences in the ERP response to anomalous versions of those sentences, providing a neurobiological explanation for the individual differences.

A further implication of the present findings follows from the observation that our two ERP indices, the Response Magnitude Index and the Response Dominance Index, were uncorrelated ($r = .076$). That is, variation in the quantity of the ERP response to linguistic anomalies was uncorrelated with the quality of the response. Moreover, variability in each of these measures independently was systematically associated with different subject-level covariates, suggesting that the RDI and RMI together provide a more complete account of how individual difference measures impact processing than traditional ERP measures alone, such as mean amplitude. However, it remains to be seen how some of the individual variables mentioned above (e.g., working memory, neural efficiency, cortical connectivity) map onto the neurocognitive correlates of language processing that we report here (see Bond et al., 2011, for a first attempt to link individuals' specific language aptitude and non-verbal reasoning ability with L2 ERP effects). A further matter is the relationship between L1 and L2 processing mechanisms within individuals. Given the evidence reviewed above that individuals vary along the N400–P600 continuum even in their L1, it may be the case that an individual's location along the continuum when processing his or her L1 may account for additional variance in L2 brain responses. That is, do bilinguals show similar ERP response biases in their L1 and L2, or are the mechanisms separable by language? While our current data cannot speak to this issue, it remains a potentially fruitful area for future research.

Finally, our findings show that ERP studies of L2 learning can inform theory beyond simply identifying whether L2 syntactic processing can become "native-like", where native-like is defined as the presence of some specific brain response to a linguistic anomaly. While traditional grand mean analyses are useful and informative about the central tendency and normative brain response seen in a given population, a broad view of the L1 ERP and neuroimaging literature indicates that native-likeness can encompass a continuum of brain activation and ERP response profiles (Mason & Just, 2007; Nakano et al., 2010; Nieuwland & Van Berkum, 2008; Osterhout, 1997; Pakulak & Neville, 2010; Prat, 2011; Prat, Keller & Just, 2007; Prat, Long & Baynes, 2007; Tanner, 2011; Tanner & Van Hell, 2012). Taken in this context, our finding that variation exists in proficient L2 learners is unsurprising. The measures we introduced here (the RDI and RMI) provide a new and useful way of understanding this variation, particularly with regard to the multiple dimensions of processing that ERP recordings provide. Although the pen-and-paper proficiency measures correlated with participants' brain response magnitudes, the combination of response magnitude with response dominance measures provided a much richer characterization of language comprehension profiles than either offline test batteries or traditional ERP quantification metrics would have provided alone. Although further research is needed to better understand these relationships, as well as the extent to which these findings will generalize to other learner populations or linguistic manipulations, the present study provides yet another indication of the unique sensitivities of event-related potentials as tools for studying language learning.

References

Allen, M., Badecker, W., & Osterhout, L. (2003). Morphological analysis in sentence processing: An ERP study. *Language and Cognitive Processes, 18,* 405–430.

Bates, E., Devescovi, A., & Wulfeck, B. (2001). Psycholinguistics: A cross-language perspective. *Annual Review of Psychology, 52,* 369–396.

Bentin, S. (1987). Event-related potentials, semantic processes, and expectancy factors in word recognition. *Brain and Language, 31,* 308–327.

Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language, 44,* 235–249.

Bond, K., Gabriele, A., Fiorentino, R., & Alemán Bañón, J. (2011). Individual differences and the role of the L1 in L2 processing: An ERP investigation. In J. Herschensohn & D. Tanner (eds.), *Proceedings of the 11th Generative Approaches to Second Language Acquisition Conference (GASLA 2011)*, pp. 17–29. Somerville, MA: Cascadilla Proceedings Project.

Carroll, J. B. (1962). The prediction of success in intensive foreign language training. In R. Glaser (ed.), *Training research and education*, pp. 87–36. Pittsburgh, PA: University of Pittsburgh Press.

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,* 161–174.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology, 42,* 368–407.

Deutsch, A., & Bentin, S. (2001). Syntactic and semantic factors in processing gender agreement in Hebrew: Evidence from ERPs and eye movements. *Journal of Memory and Language, 45,* 200–224.

Ditman, T., Holcomb, P. J., & Kuperberg, G. R. (2007). An investigation of concurrent ERP and self-paced reading methodologies. *Psychophysiology, 44,* 927–935.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty & M. H. Long (eds.), *The handbook of second language acquisition*. Malden, MA: Blackwell.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47,* 164–203.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language, 41,* 78–104.

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,* 226–248.

Frenck-Mestre, C., Foucart, A., Carrasco, H., & Herschensohn, J. (2009). Processing of grammatical gender in French as a first and second language. In L. Roberts, D. Véronique, A. Nilsson & M. Tellier (eds.), *Eurosla Yearbook 9*, pp. 76–106. Amsterdam: John Benjamins.

Frenck-Mestre, C., Osterhout, L., McLaughlin, J., & Foucart, A. (2008). The effect of phonological realization of inflectional morphology on verbal agreement in French: Evidence from ERPs. *Acta Psychologica, 128,* 528–536.

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic processing: Early and late event-related potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1219–1248.

Gardner, R. C., Tremblay, P. F., & Masgoret, A.-M. (1997). Towards a full model of second language learning: An empirical investigation. *The Modern Language Journal, 81,* 344–362.

Gernsbacher, M. A. (1993). Less skilled readers have less efficient suppression mechanisms. *Psychological Science, 4,* 294–298.

Gillon Dowens, M., Guo, T., Guo, J., Barber, H., & Carreiras, M. (2011). Gender and number processing in Chinese learners of Spanish – evidence from Event Related Potentials. *Neuropsychologia, 49,* 1651–1659.

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

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia, 38,* 1531–1549.

Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8,* 439–484.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304,* 438–441.

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research, 30,* 251–266.

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

Hahne, A., Mueller, J. L., & Clahsen, H. (2006). Morphological processing in a second language: Behavioral and event-related brain potential evidence for storage and decomposition. *Journal of Cognitive Neuroscience, 18,* 121–134.

Handy, T. C. (2005). *Event-related potentials: A methods handbook*. Cambridge, MA: MIT Press.

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research, 22,* 369–397.

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

Inoue, K., & Osterhout, L. (2013). Sentence comprehension as a neural seesaw. Ms., University of Washington.

Jackendoff, R. (2007). A Parallel Architecture perspective on language processing. *Brain Research, 1146,* 2–22.

Jackson, C. N., & Van Hell, J. G. (2011). The effects of L2 proficiency level on the processing of *wh*-questions among Dutch second language speakers of English. *International Review of Applied Linguistics in Language Teaching, 49,* 195–219.

Jasper, H. H. (1958). The ten–twenty system of the International Federation. *Electroencephalography and Clinical Neurophysiology, 10,* 371–375.

Johnson, J., & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21,* 60–99.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122–149.

Kaan, E. (2002). Investigating the effects of distance and number interference in processing subject–verb dependencies: An ERP study. *Journal of Psycholinguistic Research, 31,* 165–193.

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

Kaan, E., & Swaab, T. Y. (2003). Repair, revision and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience, 15,* 98–110.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52,* 205–225.

King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7,* 376–395.

Kos, M., van den Brink, D., & Hagoort, P. (2012). Individual variation in the late positive complex to semantic anomalies. *Frontiers in Psychology*, *3*:318. doi:10.3389/fpsyg.2012.00318.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research, 1146,* 23–49.

Kuperberg, G. R., Caplan, D., Sitnikova, T., Eddy, M., & Holcomb, P. J. (2006). Neural correlates of processing syntactic, semantic, and thematic relationships in sentences. *Language and Cognitive Processes, 21,* 489–530.

Kuperberg, G. R., Holcomb, P. J., Sitnikova, T, Greve, D., Dale, A. M., & Caplan, D. (2003). Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *Journal of Cognitive Neuroscience, 15,* 272–293.

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain and Language, 100,* 223–237.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4,* 463–470.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic anomaly. *Science, 207,* 203–205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307,* 161–163.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience, 9,* 920–933.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7,* 19–40.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior, 23,* 127–150.

Mancini, S., Molinaro, N., Rizzi, L., & Carreiras, M. (2011a). A person is not a number: Discourse involvement in subject–verb agreement computation. *Brain Research, 1410,* 64–76.

Mancini, S., Molinaro, N., Rizzi, L., & Carreiras, M. (2011b). When persons disagree: An ERP study of Unagreement in Spanish. *Psychophysiology, 48,* 1361–1371.

Mason, R. A., & Just, M. A. (2007). Lexical ambiguity in sentence comprehension. *Brain Research, 1146,* 115–27.

McLaughlin, J., Tanner, D., Pitkänen, I., Frenck-Mestre, C., Inoue, K., Valentine, G., & Osterhout, L. (2010). Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning, 60 (Suppl. 2),* 123–150.

Molinaro, N., Barber, H., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex, 47,* 908–930.

Molinaro, N., Kim, A., Vespignani, F., & Job, R. (2008). Anaphoric agreement violation: An ERP analysis of its interpretation. *Cognition, 106,* 963–974.

Moreno, E. M., & Kutas, M. (2005). Processing semantic anomalies in two languages: An electrophysiological exploration in both languages of Spanish–English bilinguals. *Cognitive Brain Research, 22,* 205–220.

Morgan-Short, K., Faretta-Stutenberg, M., Brill, K., Carpenter, H., & Wong, P. C. M. Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism: Language and Cognition*, doi:10.1017/S1366728912000715. Published online by Cambridge University Press, March 1, 2013.

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potentials study. *Language Learning, 60,* 154–193.

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience, 24,* 933–947.

Naiman, N., Fröhlich, M., Stern, H. H., & Todesco, A. (1996). *The good language learner*. Philadelphia, PA: Multilingual Matters.

Nakano, H., Saron, C., & Swaab, T. Y. (2010). Speech and span: Working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *Journal of Cognitive Neuroscience, 22,* 2886–2898.

Nevins, A., Dillon, B., Malhotra, S., & Phillips, C. (2007). The role of feature-number and feature-type in processing Hindi verb agreement violations. *Brain Research, 1164,* 81–94.

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,* 1205–1223.

Nieuwland, M. S., & Van Berkum, J. J. A. (2008). The interplay between semantic and referential aspects of anaphoric noun phrase resolution: Evidence from ERPs. *Brain and Language, 106,* 119–1131.

Oines, L., Miyake, A., & Kim, A. (2012). Individual differences in verbal working memory predict reanalysis vs. integration in syntax–semantics conflict scenarios. Poster presented at the Architectures and Mechanisms for Language Processing Conference, Riva del Garda, Italy.

Ojima, S., Matsuba-Kurita, H., Nakamura, N., Hoshino, T., & Hagiwara, H. (2011). Age and amount of exposure to a foreign language during childhood: Behavioral and ERP data on the semantic comprehension of spoken English by Japanese children. *Neuroscience Research, 70,* 197–205.

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of Proficiency. *Journal of Cognitive Neuroscience, 17,* 1212–1228.

Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language, 59,* 494–522.

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

Osterhout, L., Kim, A., & Kuperberg, G. R. (2012). The neurobiology of sentence comprehension. In M. Spivey, M. Joannisse & K. McCrae (eds.), *The Cambridge handbook of psycholinguistics*, pp. 365–389. Cambridge: Cambridge University Press.

Osterhout, L., McLaughlin, J., Kim, A., Greewald, R., & Inoue, K. (2004). Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. In M. Carreiras & C. Clifton (eds.), *The on-line study of sentence comprehension: Eyetracking-ERPs, and beyond*, pp. 271–308. New York: Psychology Press.

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A means for exploring the neurocognition of second language processing. *Language Learning, 56 (Suppl. 1),* 199–230.

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

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

Paczynski, M., & Kuperberg, G. R. (2011). Electrophysiological evidence for use of the animacy hierarchy, but not thematic role assignment, during verb–argument processing. *Language and Cognitive Processes, 26,* 1402–1456.

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,* 2728–2744.

Prat, C. S. (2011). The brain basis of individual differences in language comprehension abilities. *Language & Linguistics Compass, 5,* 635–649.

Prat, C. S., Keller, T. A., & Just, M. A. (2007). Individual differences in sentence comprehension: A functional Magnetic Resonance Imaging investigation of syntactic and lexical processing demands. *Journal of Cognitive Neuroscience, 19,* 1950–1963.

Prat, C. S., Long, D. L., & Baynes, K. (2007). The representation of discourse in the two hemispheres: An individual differences investigation. *Brain and Language, 100,* 283–294.

Pulvermüller, F., & Schumann, J. H. (1994). Neurobiological mechanisms of language acquisition. *Language Learning, 44,* 681–734.

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.

Rossi, S., Gugler, M. F., Hahne, A., & Friederici, A. D. (2005). When word category information encounters morphosyntax: An ERP study. *Neuroscience Letters, 384,* 228–233.

Sabourin, L., & Haverkort, M. (2003). Neural substrates of representation and processing of a second language. In R. van Hout, A. Hulk, F. Kuiken & R. Towell (eds.), *The lexicon–syntax interface in second language acquisition*, pp. 175–195. Amsterdam & Philadelphia, PA: John Benjamins.

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

Severens, E., Jansma, B. M., & Hartsuiker, R. J. (2008). Morphophonological influences on the comprehension of subject–verb agreement: An ERP study. *Brain Research, 1228,* 135–144.

Skehan, P. (1989). *Individual differences in second-language learning*. New York: Arnold.

Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research, 25,* 13–41.

Tanner, D. (2011). Agreement mechanisms in native and nonnative language processing: Electrophysiological correlates of complexity and interference. Ph.D. dissertation, University of Washington.

Tanner, D., & Van Hell, J. G. (2012). ERPs reveal individual differences in syntactic processing strategies. Poster presented at the 2012 Psychonomics Society Meeting, Minneapolis, MN.

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

Tanner, D., Nicol, J., Herschensohn, J., & Osterhout, L. (2012). Electrophysiological markers of interference and structural facilitation in native and nonnative agreement processing. In A. K. Biller, E. Y. Chung & A. E. Kimball (eds.), *Proceedings of the 36th Annual Boston University Conference on Language Development*, pp. 594–606. Somerville, MA: Cascadilla Press.

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar – an event-related potential investigation. *Studies in Second Language Acquisition, 27,* 173–204.

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

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92,* 231–270.

Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In C. Sanz (ed.), *Mind and context in adult second language acquisition*, pp. 141–178. Washington, DC: Georgetown University Press.

Van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience, 11,* 657–671.

van de Meerendonk, N., Kolk, H. H. J., Vissers, C. T. W. M., & Chwilla, D. J. (2010). Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience, 22,* 67–82.

Van Hell, J. G., & Tanner, D. (2012). Second language proficiency and cross-language lexical activation. *Language Learning, 62 (Suppl. 2),* 148–171.

Van Hell, J. G., & Tokowicz, N. (2010). Event-related brain potentials and second language learning: Syntactic processing in late L2 learners at different L2 proficiency levels. *Second Language Research, 26,* 43–74.

Van Herten, M., Chwilla, D. J., & Kolk, H. H. J. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience, 18,* 1181–1197.

Van Petten, C. K., Weckerly, J., McIsaac, H. K., & Kutas, M. (1997). Working memory capacity dissociates lexical and sentential context effects. *Psychological Science, 8,* 238–242.

Vos, S. H., Gunter, T. C., Kolk, H. H. J., & Mulder, G. (2001). Working memory constraints on syntactic processing: An electrophysiological investigation. *Psychophysiology, 38,* 41–63.

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,* 231–256.

Weber-Fox, C. M., & Neville, H. J. (1999). Functional neural subsystems are differentially affected by delays in second language immersion: ERP and behavioral evidence in bilinguals. In D. Birdsong (ed.), *Second language acquisition and the critical period hypothesis*. Mahwah, NJ: Lawrence Erlbaum.

Weckerly, J., & Kutas, M. (1999). An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology, 36,* 559–570.

Zhang, Y., Li, P., Piao, Q., Liu, Y., Huang, Y., & Shu, H. (2013). Syntax does not necessarily precede semantics in sentence processing: ERP evidence from Chinese. *Brain and Language, 126,* 8–19.