CrossMark

RESEARCH ARTICLE

Variability in native and nonnative language: An ERP study of semantic and grammar processing

Sarah Grey* (ORCID)

Fordham University, Bronx, NY, USA
*Corresponding author. E-mail: sgrey4@fordham.edu.

Abstract
This study examined individual-level variability in N400 and P600 ERP correlates of native and nonnative language sentence processing of semantic and grammar information. Twenty-six native English-speaking learners of Spanish as a second language were tested. Participants completed sentence reading tasks in English and Spanish during EEG recording. The group-level results for grammar showed P600s in the native and nonnative language. For semantics, there was an N400 only in the native language. Individual-level ERP patterns revealed that, for native language semantics, about two thirds of participants showed N400s, but approximately one-third showed P600s. For native language grammar, approximately one third of participants exhibited N400s instead of P600s. Individual-level ERPs showed similar variability in the nonnative language for semantics and grammar, and N400/P600 variability in Spanish related to Spanish verbal fluency and grammar knowledge. This contributes knowledge of how nonnative and native language neurocognition compare regarding the processing routes that individuals use during comprehension.

Introduction

Is nonnative language processing similar to native language processing? This question has been pursued from both psycholinguistic and neurocognitive perspectives, and insights into the question have implications for theory and research across a number of disciplines including cognitive psychology, cognitive neuroscience, bilingualism, and second language acquisition (e.g., Abutalebi, 2008; Caffarra et al., 2015; Clahsen & Felser, 2006b; Kotz, 2009; Morgan-Short, 2014). Within the last 15 years, there has been increasing interest in using the Event-Related Potential (ERP) technique to study nonnative language processing. ERPs reflect real-time electrophysiological brain activity and their excellent temporal precision enables researchers to examine aspects of language in close detail. As such, they are a useful tool for studying the neural correlates of nonnative language (L2) processing, and its comparability to native language (L1) processing. With the expansion of this research area, new questions are emerging and revealing limits in our knowledge of L2 neurocognitive processing. This study addressed gaps in L2 ERP research by examining individual differences in the N400 and P600 ERP correlates of L1 and L2 sentence processing of semantic and grammar information, within-subjects.

ERPs and language

Neurocognitive measures of language processing, and in particular ERPs, are able to provide novel, disambiguating insight into group-level and individual-level differences, and they are also useful for uncovering qualitative differences in how language information is processed (e.g., Grey et al., 2018; Morgan-Short, Finger et al., 2012; Morgan-Short, Steinhauer et al., 2012; Tanner et al., 2014; Tanner & Van Hell, 2014).

ERPs are derived through recording naturally occurring electroencephalogram (EEG) data and consist of changes in the brain's electrical activity in response to a time-locked external event, such as a word in a sentence. ERP language studies often use "violation" paradigms in which the neural activations of correct (or standard) stimuli are compared to matched "violation" stimuli, for example a grammar error or semantic error,[1] often called a semantic anomaly, as in the examples in Table 3 (Grey & Tagarelli, 2018; Kaan, 2007). ERPs have been used to study language processing for more than 40 years and the present study focused on two well-researched and reasonably well-understood ERP effects: N400 and P600.

The N400 is a negative-going waveform with a broad central-posterior scalp distribution; it is typically understood to reflect lexical/semantic processing in the brain (e.g., Kaan, 2007; Kutas & Federmeier, 2011; Lau et al., 2008). The amplitude of the N400 covaries with a number of lexical and semantic factors, including word frequency, the predictability and contextual felicity of a word in a sentence or discourse, and interpretative relevance of the target (Choudhary et al., 2009; Kutas & Federmeier, 2011; Nieuwland & Van Berkum, 2006b, 2006a; Van Berkum et al., 1999). Since its discovery, the N400 has been widely accepted as indexing difficulty in accessing semantic information (Kutas & Hillyard, 1980; Kutas & Federmeier, 2011; for an alternative account suggesting that semantic inhibition underlies the N400, see Debruille, 2007; Debruille et al., 2008).

The second ERP effect of interest, the P600, is a positive-going waveform with a posterior scalp distribution. P600s are typically elicited in response to violations of morphosyntax, for example in verb tense (e.g., Grey et al., 2017; Osterhout & Nicol, 1999) and subject-verb agreement (Coulson et al., 1998; Kaan, 2002; Osterhout & Mobley, 1995; Silva-Pereyra & Carreiras, 2007; Tanner et al., 2017), and they have also been elicited with semantic manipulations, such as thematic role or animacy errors and semantic anomalies (e.g., Chow & Phillips, 2013; Kim & Osterhout, 2005; Kuperberg et al., 2007). The precise nature of P600 effects is a topic of discussion in the literature (e.g., Bornkessel-Schlesewsky & Schlesewsky, 2008; Chow & Phillips, 2013; Coulson et al., 1998; Friederici et al., 2002; Kolk & Chwilla, 2007; Osterhout et al., 2012; van de Meerendonk et al., 2010), but several decades of research demonstrate that the P600 is sensitive to and reliably elicited by morphosyntactic (i.e., grammar) violations. And despite differences in current theoretical descriptions of P600 effects, there is general agreement that P600s reflect processing of a stimulus in conflict with an expected linguistic representation and a late attempt at repair or reanalysis, and that the P600 effect reflects a set of processes that are neurocognitively distinct from the processes reflected in the N400 (Allen et al., 2003; Bornkessel-Schlesewsky & Schlesewsky, 2008;

---

[1]Violations of expected semantic or grammar patterns are referred to as "errors" to differentiate them from sentences where the expected patterns are found (termed "correct"). This follows from the violation paradigm used in psycholinguistics and ERP language research. However, this is not intended as a judgment of language production that might be attested in language learners or bilinguals.

Chow & Phillips, 2013; DeLong et al., 2014; Kolk & Chwilla, 2007; Kuperberg, 2007; Osterhout & Nicol, 1999; van de Meerendonk et al., 2010).

In L2, N400s have been observed for semantic processing and both N400s and P600s have been reported for grammar processing. One trend in the L2 ERP grammar processing literature is that N400s, instead of P600s, are occasionally reported for grammar at lower L2 proficiency (but see Gabriele et al., 2021 for P600s in novice learners). With increasing L2 proficiency, P600s are more likely to be found. For reviews, see Caffarra et al. (2015), Morgan-Short (2014), Steinhauer (2014), Steinhauer et al. (2009), and Van Hell and Tokowicz (2010).

Individual differences in the ERP correlates of L1 and L2 sentence processing

Most ERP research on L2 sentence processing has either examined L2 compared to L1 literature, that is, compared L2 results to previous findings reported in the ERP literature on L1 processing (e.g., Batterink & Neville, 2013), or compared the L2 group to a separate L1 group (e.g., Bowden et al., 2013; Fromont et al., 2020). Using these between-subjects designs, deviations in L2 from L1 patterns have led to conclusions about qualitative and quantitative differences in L2 (for reviews, see e.g., Morgan-Short, 2014; Steinhauer, 2014; Van Hell & Tokowicz, 2010), which have informed theoretical models of L2 neurocognition (e.g., Clahsen & Felser, 2006b, 2006a; Paradis, 2009; Ullman, 2001, 2020). However, ERP language research increasingly demonstrates that even L1 processing deviates from "the norm." This work indicates that L1 processing shows individual-level variability and highlights that L1 groups are not, as is often assumed, homogenous in their L1 processing (e.g., Grey et al., 2017; Kim et al., 2018; Kos et al., 2012; Nakano et al., 2010; Pakulak & Neville, 2010; Tanner & Van Hell, 2014).

For example, building off of Osterhout (1997), Tanner and Van Hell (2014) demonstrated that L1 English speakers exhibited individual variability in ERP responses to subject-verb agreement and verb tense errors during sentence processing. In the study, the grand mean analyses suggested a biphasic left anterior negativity (LAN)-P600 effect, a pattern that has been implicated in theoretical accounts of native and nonnative language morphosyntactic processing (e.g., Caffarra et al., 2019; Clahsen & Felser, 2006b, 2006a; Friederici, 2002; Molinaro et al., 2015; Molinaro et al., 2011). However, examination of individual-level ERP responses with a Response Dominance Index (RDI) indicated that the apparent LAN-P600 biphasic pattern was a product of some individuals showing N400-dominant effects and others showing P600-dominant effects. N400 and P600 effects are both characterized by central-posterior scalp distributions and Tanner and Van Hell (2014) argued that this overlap masked part of the N400 in the grand mean analyses, leaving the nonoverlapping distribution that appeared to be a left anterior negativity (for more detailed discussion on this perspective, see Caffarra et al., 2019; Molinaro et al., 2015; Tanner, 2015, 2019). In sum, the group-level grand mean information obscured individual differences in N400- and P600-related mechanisms employed to process grammar information during L1 sentence comprehension (see also Grey et al., 2017). For summary information on this study and the related research reviewed in this section, see Table 1.

Tanner (2019) further examined N400/P600 variability during grammar processing in a large cohort of L1 English speakers and obtained similar results. Grand mean analyses suggested a biphasic LAN-P600 pattern but when examining individual-level processing ERP responses to subject-verb agreement errors systematically varied, with some individuals showing N400-dominant effects and others showing P600-dominant

Table 1. Summary of ERP studies on N400/P600 individual differences during sentence processing

Study	N	L1	L2	L2 prof. or experience†	Design	Grammar target	Semantic target	Task	Group ERPs	Indiv.-level ERPs	Additional details
Studies on L1											
Kim et al. (2018)	60	English	–	–	–	–	Semantic attraction vs. no attraction anomalies	AGJT	N400 to no attraction, P600 to attraction	Variation in N400/P600 responses	Higher verbal working memory related to larger P600, smaller N400
Tanner (2019)	137	English	–	–	–	Subject-verb agreement	–	AGJT	Biphasic LAN-like negativity-P600	Variation in N400/P600 responses	No relation between cognitive variables and ERP variation
Tanner & Van Hell (2014)*	40	English	–	–	–	Subject-verb agreement & verb tense	–	AGJT	Biphasic LAN-like negativity-P600	Variation in N400/P600 responses	Familial left-handedness related to greater N400 dominance
Studies on L2											
Bice & Kroll (2021)	25 early bilinguals 12 late bilinguals	Spanish (nondominant language) English (dominant language)	English (dominant language) Spanish (nondominant language)	Proficient in both languages	Within-subjects	Subject-verb agreement	–	AGJT	–	Variation in N400/P600 responses in L1 and L2	L1 and L2 ERP dominance was related; Higher composite proficiency performance predicted P600-dominance across languages
Finestrat Martinez et al. (2018)	8	English	Spanish	Low/intermediate	Within-subjects	Subject-verb agreement, phrase structure, & number agreement	–	AGJT	L1: N400 to phrase structure; P600 to Subject-verb agreement L2: Anterior positivity to subject-verb agreement	Variation in N400/P600 responses in L1 and L2 across all structures	L1 and L2 P600 dominance was related only for number agreement

(Continued)

Table 1. (*Continued*)

Study	N	L1	L2	L2 prof. or experience†	Design	Grammar target	Semantic target	Task	Group ERPs	Indiv.-level ERPs	Additional details
Pelissier 2020	32	French	English	Intermediate	Between-subjects	Past tense with auxiliaries	–	AJT	No sig. L2 ERP effects	Variation in N400/P600 responses in L2	Better post-EEG GJT performance related to greater N400-dominance
Tanner et al. (2013)	20 beginning learners 13 advanced learners	English	German	Beginning and advanced	Between-subjects	Subject-verb agreement	–	GJT	Beginning learners: biphasic N400-P600 Advanced learners: P600	Variation in N400/P600 responses in L2	Better performance on GJT related to larger P600
Tanner et al. (2014)	20	Spanish	English	Unlabeled	Between-subjects	Subject-verb agreement	–	AGJT	Biphasic N400-P600	Variation in N400/P600 responses in L2	Earlier age of arrival and motivation to master L2 were related to greater P600 dominance
Wampler et al. (2014)	20	English	French	2nd year of instruction	Within-subjects	Subject-verb agreement	–	No report	No report	Variation in N400/P600 responses in L1 and L2	Females showed greater P600 dominance; L1 and L2 response dominance was not related
Present study	*26*	*English*	*Spanish*	*Intermediate*	*Within-subjects*	*Verb tense*	*Semantic anomaly*	*Word probe*	*L1: N400 to semantic anomalies; P600 to verb tense errors L2: P600 to verb tense errors*	*Variation in N400/P600 responses in L1 and L2 for semantics and verb tense*	*Greater N400 dominance related to higher L2 verbal fluency and higher L2 grammar knowledge; L1 and L2 response dominance was not related*

†The information in this column is based on the descriptions and/or labels provided in the original studies. Prof. = proficiency; Indiv. = individual. *See also Grey et al. (2017); AJT = acceptability judgment task, decision is meaning-focused; AGJT = acceptability/grammaticality judgment task, decision includes both meaning and form; GJT = grammaticality judgment task, decision is form-focused.

effects. This pattern of N400/P600 individual differences was sustained regardless of whether agreement was based on inflectional morphology or lexical information and when using different sentence reading paradigms (Rapid Serial Visual Presentation or Self-Paced Reading). Tanner (2019) also gathered measures of verbal working memory and language experience to examine whether these factors could help explain N400/P600 variability in grammar processing, but no relationships were found.

This individual-level N400/P600 variation in L1 is not limited to grammar processing. Kim et al. (2018) investigated individual differences in L1 English speakers' processing of verb-based semantic anomalies with or without semantic attraction (semantic attraction example: *The hearty meal was *devouring with gusto*; no semantic attraction example: *The dusty tabletops were *devouring with gusto*; *marks the anomaly; examples are from Kim et al., 2018). The group-level results indicated N400s for no-attraction sentences and P600s for attraction sentences, but at the individual-level the authors observed similar patterns of variability for semantic processing as observed for grammar by Tanner (2019) and Tanner and Van Hell (2014). Some individuals exhibited N400 effects, as would typically be expected for semantic anomalies, and others showed P600 effects. Kim et al. (2018) assessed verbal and nonverbal working memory as well as language experience in relation to individuals' N400/P600 effects and found that higher verbal working memory capacity was related to larger P600s and smaller N400s during semantic processing.

Notably, these individual differences in N400/P600 responses have also been observed for L2 learners. Tanner, McLaughlin, Herschensohn, and Osterhout (2013) investigated the ERP correlates of subject-verb agreement processing in L2 German learners of varying proficiency compared to L1 German speakers and probed individual differences in processing. Grand mean results suggested that the L1 group and more advanced L2 learner group exhibited P600s during sentence processing whereas the less advanced L2 group showed a biphasic N400-P600 response. Individual-level analyses demonstrated that these learners varied along a similar N400/P600 dominance continuum as observed in the L1 studies discussed previously: some learners were N400-dominant and others were P600-dominant. Further, the magnitude of less and more advanced learners' P600 effect was positively related to their behavioral ability to detect the L2 grammar errors on a grammaticality judgment test (GJT); this relationship was not observed for the L1 group.

In a subsequent study, Tanner, Inoue, and Osterhout (2014) examined N400/P600 individual differences in L1 Spanish–L2 English late bilinguals' processing of L2 English grammatical agreement. The grand mean group-level ERP results indicated a biphasic N400-P600 response, but examination of individual-level ERP responses showed that participants varied in their L2 processing along an N400/P600 dominance continuum, as in Tanner et al. (2013). The authors additionally assessed whether a host of language experience variables could explain this variability and found that earlier age of arrival to the L2 environment and higher motivation to master the L2 were associated with greater P600 dominance.

Finally, Pélissier (2020) examined individual differences in intermediate L1 French–L2 English learners' and a group of L1 English speakers' processing of past tense verb morphology with auxiliaries. The group-level results indicated that only the L1 English speakers showed a P600. At the individual level, the results showed that both the L1 group and L2 learner group varied in their processing along an N400/P600 continuum, showing either an N400 in response to the grammar errors or a P600, but not both. The author also examined whether performance on a GJT administered after the EEG/ERP task related to ERP responses and found that performance on the GJT was positively

correlated with N400 amplitude in the L2 learners, with no relationships observed for the L1 group.

Overall, these L2 ERP studies parallel the L1 work and, when considered together, this research indicates that native as well as nonnative language users exhibit systematic individual variation in whether they rely more on lexical/semantic access (N400) or repair/reanalysis (P600) mechanisms while processing semantic and grammar information during sentence comprehension. With respect to L2 ERP research, this complicates the general tendency of interpreting between-subjects L1 versus L2 differences as being due to a "deviant," "nonnativelike" L2 system.

To develop a more sophisticated understanding of how L2 processing compares to L1, there is a need for work that examines L2 and L1 processing *within*-subjects during sentence processing. Very little research has done this, let alone investigate individual-level variability. In fact, of the 41 L2 grammar ERP articles published between 2002–2016, identified by and discussed in Bowden et al. (in preparation; see also Morgan-Short, 2014), only one employed a within-subjects design (Tokowicz & MacWhinney, 2005).

There are two unpublished reports and one published study that have tested L1 and L2 together using a within-subjects approach (Bice & Kroll, 2021; Finestrat Martinez et al., 2018; Wampler et al., 2014); each of the studies tested grammar processing but did not test semantic processing. Because these studies measured individual-level variability in N400/P600 ERP responses across L1 and L2 they are informative points of reference for the present work.

Wampler et al. (2014) measured L1 and L2 processing of subject-verb agreement in L1 English–L2 French learners in their second year of L2 instruction and found that participants demonstrated variability in their N400/P600 dominance for both their L1 and L2. This variability was not related between L1 and L2; that is, N400 (or P600)-dominant processing for L1 grammar was not related to N400 (or P600)-dominant processing for L2 grammar. Finestrat Martinez et al. (2018) tested a small group of L1 English-low/intermediate L2 Spanish learners and, like Wampler et al. (2014), observed variability in N400/P600 dominance for both L1 and L2. Of the three grammar structures they examined (word order, subject-verb agreement, noun phrase number agreement), L1 and L2 variability was related only for P600 magnitudes for noun-phrase number agreement. Finally, Bice and Kroll (2021) investigated ERP individual differences in L1 Spanish–L2 English early heritage bilinguals' and late L1 English–L2 Spanish bilinguals' subject-verb agreement processing; they also included an L1 English monolingual control group. The results showed variability in N400/P600 responses in the bilinguals' L1 and L2, in line with Wampler et al. (2014) and Finestrat Martinez et al. (2018). Further, L1/L2 proficiency was related to N400/P600 variation, with higher proficiency relating to more P600-dominant responses and smaller N400-dominant responses. The authors also note that bilinguals for whom L1 and L2 proficiency was roughly equivalent tended to show similar brain responses. To summarize, this small body of work indicates that N400/P600 processing of grammatical information varies systematically within individuals' native and nonnative language, though the research is inconclusive regarding whether such variability is related across individuals' two languages because the studies report different findings.

As mentioned in the preceding text, these L2 studies tested grammar, but not semantic, processing. In the L2 ERP literature as a whole, there are remarkably few reports of L2 sentence-level semantic processing (for review, see Bowden et al., 2013) and conclusions from these studies are difficult to reconcile. This is partly because some of the work tested fluent bilinguals rather than L2 learners (Ardal et al., 1990; Moreno & Kutas, 2005) and partly because the studies that tested L2 learners compared L1 and L2

patterns between subjects (Bowden et al., 2013; Fromont et al., 2020; Hahne, 2001; Hahne & Friederici, 2001; Newman et al., 2012; Ojima et al., 2005). Thus, although most of the small amount of work reports differences for L2 semantic processing, for example in N400 latency or distribution, explanations for these differences are poorly understood.

The present study

This study addressed two main issues in current L2 ERP sentence processing research. First, the research has centered on between-subjects L2 comparisons to L1 control groups or L1 ERP literature, largely with the assumption that the L1 is a homogenous point of comparison and group-level grand mean information is representative of the individuals within the group. Given that group-level grand mean-based L1 ERP work seems to not always accurately capture L1 processing, it seems likely that previous L2 comparisons to the L1 ERP literature or monolingual L1 control groups have not always accurately characterized L2 processing. Further, comparing L2 speakers to monolingual L1 groups (the "monolingual bias") may not be an appropriate analytic approach more generally because bilinguals/L2 learners and monolinguals are fundamentally different in their language use and linguistic experiences (Ortega, 2013b, 2013a).

A second issue is that the majority of L2 ERP sentence processing research has tested grammar only, with remarkably few reports of sentence-level semantic processing. This is surprising given that prominent theoretical discussions of L2 processing posit differential reliance on lexical/semantic versus structural, morphosyntactic processes during L2 learning and development (Clahsen & Felser, 2006b; Paradis, 2009; Ullman, 2001, 2020). Furthermore, as the N400/P600 individual differences research continues to mount, including for L2, it is increasingly important to examine semantic processing patterns together with grammar processing patterns because the two ERP effects that are showing this systematic variation are qualitatively different, functionally dissociable brain responses generally linked with lexical/semantic access (N400) or structural repair/reanalysis (P600) processes. To address these issues, the study investigated the ERP correlates of L1 English and L2 Spanish semantic and grammar processing, within-subjects and from both group-level and individual-level perspectives.

Methods

Participants

Participants were recruited at a private university in a large city in the Northeast region of the United States of America. Thirty-eight participants were tested after providing written informed consent. Of these, 12 were excluded from data analysis: 3 were pilot participants, 8 had excessive EEG artifact (i.e., recorded activity that is not generated by the brain, such as eye blinks), and 1 set of data was lost due a testing session error. Data from 26 participants (4 males, 22 females; M age = 19.3 years) were therefore included in the analyses. This sample size is comparable to or exceeds samples sizes in related L2 ERP research, including the studies that inform the present work (see the Studies on L2 section in Table 1). According to calculations based on medium-sized effects in $G*Power$, 27 participants would be needed to satisfy 80% statistical power and the final sample size of 26 participants satisfies 78.8% statistical power.

All participants were right-handed (Oldfield, 1971) native speakers of English who had learned Spanish as a nonnative second language beginning after the age of 6 years

old in a foreign language classroom setting (*M* age of Spanish language learning onset = 11.3 years; *SD* = 2.8). On a language background questionnaire, 14 participants reported only having studied Spanish and 12 reported having studied other languages in addition to Spanish, including French, German, Hebrew, Italian, Latin, Mandarin, and Russian, with self-described proficiency in these languages ranging from beginner to advanced.

Participants were recruited mainly from the final course in the university's required foreign language sequence, equivalent to an upper-intermediate Spanish language course, which utilized a content-based instructional approach that included grammar and vocabulary practice as well short-form literary reading comprehension and analysis. All participants had studied Spanish in high school and they had taken an average of 2.7 university-level Spanish language courses at the time of testing. Four participants were pursuing Spanish majors at the time of testing and nine were pursing Spanish minors. Although proficiency can be a challenging construct, a series of additional measures were gathered to help ascertain aspects of participants' L1 and L2 abilities. Participants self-rated their Spanish and English proficiency and additionally completed independent tests of language ability consisting of Spanish and English semantic verbal fluency tasks and Spanish and English grammar knowledge tests, see Table 2 (for task administration, description, and coding information, see Online Supplementary Materials). Based on this information, participants are considered to be intermediate-level Spanish L2 learners.

Materials

Stimuli and sentence reading task

The study employed a sentence reading task with word-probe verification while EEG/ERP data were recorded. Stimuli for the sentence reading task were designed following the well-established violation paradigm (Grey & Tagarelli, 2018; Kaan, 2007), wherein the electrophysiological brain response to language errors, such as a semantic error or grammar error, is compared to correct items. Sentences were extended from and based on the stimuli used in the ERP study by Osterhout and Nicol (1999).

Table 2. Descriptive information on language ability

	L1 English	L2 Spanish
	M (SD)	*M (SD)*
N = 26	*95% CI*	*95% CI*
Verbal fluency	49.6 (7.5)	24.9 (4.7)
	46.6–52.6	23.0–26.8
Grammar knowledge	45.2 (4.9)	23.2 (5.1)
	43.2–47.2	21. 25.2
Speaking	7.0 (0)	4.3 (1.0)
	7.0–7.0	3.8–4.7
Listening	7.0 (0)	4.6 (1.1)
	7.0–7.0	4.2–5.1
Reading	7.0 (0)	4.7 (0.9)
	7.0–7.0	4.4–5.2
Writing	7.0 (0)	4.8 (1.0)
	7.0–7.0	4.4–5.1

Notes. Grammar knowledge assessed using the Michigan English Language Institute College Entrance Test (MELICET) for English and the Diploma de Español como Lengua Extranjera (DELE) for Spanish, max. score = 50 each. Speaking, listening, reading, and writing were self-rated on a scale of 1–7 where 7 = "nativelike" and 1 = "beginning learner."

Two sets of stimuli were used, one in English and one in Spanish. Stimuli across the two languages were equivalent in sentence length (6–15 words; English $M = 10.3$, Spanish $M = 9.4$) and they were all novel sentences, that is, they were not translations of each other. In both English and Spanish, the critical word (bold underlined words in Table 3) was in sentence-medial position to avoid potential end-of-sentence wrap-up effects (for a recent reflection on sentence wrap-up effects in ERP language research, see Stowe et al., 2018).

The target items in the study were correct declarative sentences and declarative sentences with a semantic error or grammar error in verb tense (see Table 3 for examples; for a complete list of sentence stimuli see Online Supplementary Materials). In each language, correct/semantic error/verb error sentences were distributed across four experimental lists in a Latin-square design such that no list contained two versions of the same experimental sentence. Each experimental list contained 240 sentences with 40 correct items, 40 semantic error items, 40 verb tense error items, and an additional 120 sentences (80 filler items and 40 combined semantic + verb tense error items that are not discussed here).

For each sentence that participants read, they performed a word-probe verification by indicating using a button press whether a visually presented word had appeared in the preceding sentence or not. For similar procedures, see Schacht et al. (2014). In each list, 50% of the sentences were followed by a word (noun, verb, or adjective) that had appeared in the sentence (yes-probe), and 50% were followed by a word (noun, verb, or adjective) that had not appeared in the sentence (no-probe). Yes/no probes were evenly distributed across experimental conditions. For yes-probes, half the words had appeared before the critical word and half had appeared after; the word probe was never the critical word. Participants were instructed to read each sentence carefully and answer the word probe as quickly and accurately as possible. Sentences were presented in randomized order across six blocks of 40 sentences each.

Each trial began with a fixation cross for 350 ms followed by a sentence presented one word at a time in Rapid Serial Visual Presentation (RSVP) format. Words were presented for 350 ms each with a 200 ms ISI (interstimulus interval). After each sentence, the word probe appeared with a question mark; this remained on screen for 5,000 ms or until participants made their response, whichever event occurred first. The response hand (left/right) for "yes" was counterbalanced across participants. Between sentence trials, a screen with the word "Ready?" (or "¿Listo/a?" in Spanish) appeared. Participants were instructed to use this screen to blink or rest their eyes and were asked not to blink during the presentation of the sentences. Participants self-initiated the next trial with a button press.

Note that word-probe verification diverges from the predominance of acceptability/grammaticality judgment tasks in L2 ERP sentence processing research. Although

Table 3. Example English and Spanish sentences used in the study

	English	Spanish
Correct	The sea lions can **bask** on the beach all day.	Los peces van a **nadar** en el mar. *The fish are going to swim in the sea.*
Grammar error	The sea lions can **basking** on the beach all day.	Los peces van a **nadando** en el mar. *The fish are going to swimming in the sea.*
Semantic error	The sea lions can **edit** on the beach all day.	Los peces van a **caminar** en el mar. *The fish are going to walk in the sea.*

acceptability/grammaticality judgment tasks are solidly grounded in the historical context of Second Language Acquisition research and appropriate for ERPs, asking participants to make such explicit evaluations of the language stimuli may affect ERP outcomes in nonnegligible ways (Martin-Loeches et al., 2006; Schacht et al., 2014). Furthermore, all the previous work on individual differences in N400/P600 effects has asked participants to provide explicit judgments about the sentences using acceptability/grammaticality judgment tasks (see Table 1). It is worthwhile and important to determine whether the systematic variation observed across those studies is related to processing the sentences in the context of an explicit judgment task and implementing different task-solving strategies, as opposed to reflecting a natural tendency to employ different processing mechanisms during sentence comprehension (Pélissier, 2020). By using a task that does not require explicit metalinguistic evaluation, the present study can begin to address this point.

Procedure

Participants were tested in a single session lasting approximately 3 hours. Following the completion of a language background survey, participants were seated in a comfortable chair in a sound-attenuated room to complete the L1 English and L2 Spanish sentence reading tasks while EEG/ERP data were recorded. Practice was completed prior to the experimental lists and the order in which the L1 and L2 tasks were administered was counterbalanced across participants. After the L1 and L2 sentence reading/EEG tasks, participants completed a suite of behavioral tasks that assessed L1 and L2 abilities (see Table 2). Participants received $30 in Amazon gift cards for participating in the study.

EEG acquisition and analysis

Scalp EEG was recorded at a sampling rate of 1,000 Hz from 32 Ag/AgCl active electrodes (extended 10–20 system; Jasper, 1958) mounted in an elastic cap (Brain Products ActiCap, Germany). EEG was amplified with a BrainVision BrainAmp DC system (Brain Products, Germany) and filtered online with a .016–250 Hz bandpass filter. Scalp electrodes were referenced online to a vertex reference. Electrodes above and below the left eye and at the outer canthus of each eye, referenced in bipolar montages, monitored eye movements. Impedances were kept below 10kΩ.

EEG/ERP processing was carried out using EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014) plugins in MATLAB (Version R2016_b). At the first stage of processing, a 30 Hz half-amplitude low-pass filter (24dB/octave roll-off) was applied to the data and scalp electrodes were re-referenced to the average activity from the right and left mastoids. Trials that were characterized by eye or muscle artifacts were excluded from analyses. ERPs were time-locked to the onset of the critical word for each sentence (the bold underlined words in Table 3) and averaged for the correct, semantic, and grammar (i.e., verb tense) conditions in each participant (200 ms prestimulus baseline) for L1 English and L2 Spanish.

To be included in the analysis, datasets had to retain at least 60% of trials (24/40 trials) in each condition after artifact rejection. Included trials represented both correct behavioral responses to the yes/no probes on the word-probe verification task and incorrect responses (note that performance was very high for this task in both languages, see "Results"). Trials included in the analysis were similar across English and Spanish: On average in L1 English, each participant's final dataset contained

32 correct trials ($SD = 5.06$), 31 semantic error trials ($SD = 4.82$), and 31 grammar error trials ($SD = 5.99$). In L2 Spanish, each participant's final dataset contained 32 correct trials ($SD = 5.30$), 32 semantic error trials ($SD = 6.01$), and 31 grammar error trials ($SD = 5.76$). Informed by previous research and visual evidence from the ERP waveforms, 300–500 ms and 500–900 ms time-windows were selected to capture N400 and P600 effects, respectively. These time-windows are representative of the ERP effects of interest and attested in the ERP language processing literature for L1 and L2 (e.g., Batterink & Neville, 2013; Gillon Dowens et al., 2010; Grey et al., 2019; Osterhout & Nicol, 1999; Tanner et al., 2017).

Group-level analysis

The main interest of the present work is individual-level ERP patterns. Nonetheless, to parallel the analytical approaches conducted in the related research in this area (e.g., Tanner & Van Hell, 2014; Kim et al., 2018), group-level grand mean analyses were performed. Group-level analyses were conducted on data from lateral electrodes that were grouped into four regions of interest (ROIs): left frontal (F7, F3, FC1), right frontal (F8, F4, FC2), left posterior (CP5, CP1, P7, P3), and right posterior (CP6, CP2, P8, P4). This ROI approach aligns with analytical approaches in L1 and L2 ERP research, see, for example, Chow and Phillips (2013), Grey et al. (2017), Tanner and Van Hell (2014) and the specific ROIs selected cover a similar distributional area as Tanner and Van Hell (2014) and Grey et al. (2017), who also examined N400 and P600 individual differences in sentence processing.

To minimize Type 1 error rate and following the suggestions of Luck and Gaspelin (2017), the global ANOVAs contained a limited number of factors and step-down ANOVAs were only conducted for the main effects and interactions of theoretical interest in the study. Global ANOVAs included the factors Language (L1 English, L2 Spanish), Condition (correct, semantic error, verb tense error), Anterior/posterior (anterior, posterior), and Hemisphere (left, right) as within-subjects factors. Greenhouse–Geisser corrected p-values are reported for data with more than one degree of freedom in the numerator. Partial-eta squared effect sizes are reported with the ANOVA outcomes. Significant main effects of or interactions with the factors of Language or Condition were followed-up with step-down ANOVAs to clarify the effect.

Individual-level analysis

To examine patterns of individual-level variability in L1 and L2 semantic and grammar processing, each participant's N400 and P600 effect magnitudes were calculated for both the semantic and grammar conditions. These effect magnitudes were calculated over a central-posterior ROI consisting of electrodes CP1, CP2, Cz, C3, C4, Pz, P3, and P4, where N400 effects and P600 effects are typically largest (for similar approaches, see e.g., Grey et al., 2017; Kim et al. 2018; Tanner & Van Hell, 2014). The effect magnitudes were then used to calculate each individual's relative response dominance for N400 or P600 effects during L1 and L2 semantic and grammar processing. This metric, termed a Response Dominance Index (RDI), is made by fitting the individual's least squares distance from the equal effect sizes line (the dashed line in Figure 4) with perpendicular offsets. An RDI value near zero indicates that the individual showed N400 and P600 effects of relatively equal size. More negative RDIs (above and to the left of the dashed line) indicate that an individual is N400-dominant in their processing of the linguistic

target whereas more positive RDIs (below and to the right of the dashed line) indicate that an individual is P600-dominant. The equation for calculating the RDI follows (see also Grey et al., 2017; Tanner & Van Hell, 2014).

$$\text{RDI} = \frac{(\text{P600}_{\text{Error}} - \text{P600}_{\text{Correct}}) - (\text{N400}_{\text{Correct}} - \text{N400}_{\text{Error}})}{\sqrt{2}}$$

Results

Behavioral results for word-probe verification

For both L1 English and L2 Spanish, participants were highly accurate in deciding whether or not the word probes had appeared in the sentences. For English, mean word probe accuracy was 96.80 (SD = 1.81; *95% CI* 96.1 – 97.7) and for Spanish, mean accuracy was 96.65 (SD = 2.28; *95% CI* 95.7 – 97.6) with no significant difference between the two, $t(25)$ = .530, p = .601.

Group-level ERP results

Grand mean ERP waveforms for L1 and L2 semantic and grammar conditions are presented in Figure 1 and Figure 2; Figure 3 presents topographic maps for these conditions. Visual examination of the ERP waveforms for L1 English suggested an N400 effect for semantics and P600 effect for grammar. For L2 Spanish, visual examination suggested no significant ERP effects for semantics and a P600 effect for grammar.

Global ANOVA 300–500 ms time-window

The statistical results from the global ANOVA in the 300–500 ms time-window are summarized in Table 4. As can be seen, there were significant main effects for Language and Condition as well as significant interactions for Condition × Anterior/posterior, Language × Hemisphere, Language × Condition × Hemisphere, and Condition × Anterior/posterior × Hemisphere.

Examining the significant main effect of language

Separate analyses were conducted for L1 English and L2 Spanish to examine the main effect of Language. Follow-up analysis for L1 English sentences showed a significant Condition × Anterior/posterior interaction, $F(2,50)$ = 4.269, p = .033, ηp^2 = .15 with no other significant effects. This interaction was due to ERPs elicited by semantic errors being more negative than ERPs elicited by correct items and verb tense errors, and ERPs elicited by verb tense errors being more positive than ERPs elicited by correct items and semantic errors, specifically in the posterior region: posterior main effect of Condition, $F(2,50)$ = 11.320, p < .001, ηp^2 = .31 (semantic error M μV = –1.311, SE = .295; verb tense error M μV = .236, SE = .200; correct M μV = –.455, SE = .354; anterior main effect of Condition, $F(2,50)$ = 3.258, p = .052, ηp^2 = .12). This outcome indicated an N400 effect for L1 English semantic errors and the onset of a P600 effect for L1 English verb tense errors.
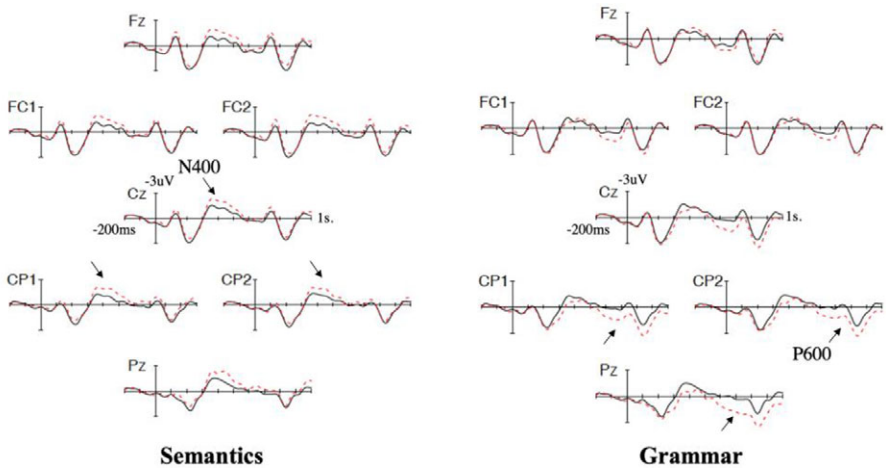
Figure 1. Grand mean ERP waveforms for L1 English semantic and grammar processing for correct (solid black line) and error conditions (semantic anomaly or verb tense; red dotted line). Each tick mark represents 100 ms; negative voltage is plotted up. These and all subsequent waveforms were filtered with a 15 Hz low-pass filter for presentation purposes only. The black arrows indicate the N400 (Semantics panel) and P600 (Grammar panel) effects.
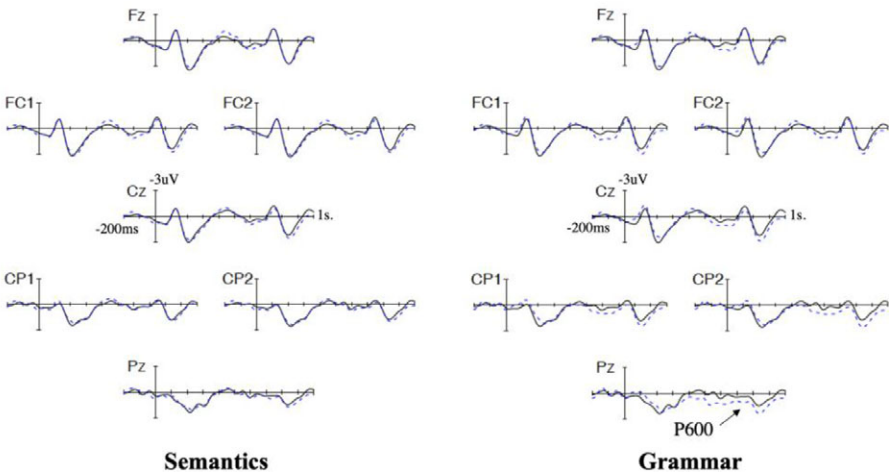

Figure 2. Grand mean ERP waveforms for L2 Spanish semantic and grammar processing for correct (solid black line) and error conditions (semantic anomaly or verb tense; blue dotted line). Each tick mark represents 100 ms; negative voltage is plotted up. The black arrow indicates the P600 effect.

For L2 Spanish sentences, the follow-up ERP analysis showed a significant Condition × Hemisphere interaction, $F(2,50) = 4.757, p = .014, \eta p^2 = .16$, which was qualified by a significant three-way Condition × Anterior/posterior × Hemisphere interaction, $F(2,50) = 4.84, p = .015, \eta p^2 = .16$. Further analysis on the three-way interaction indicated a significant Condition × Hemisphere interaction in the anterior region, $F(2,50) = 11.53, p < .001, \eta p^2 = .32$ that did not lead to any further significant effects in
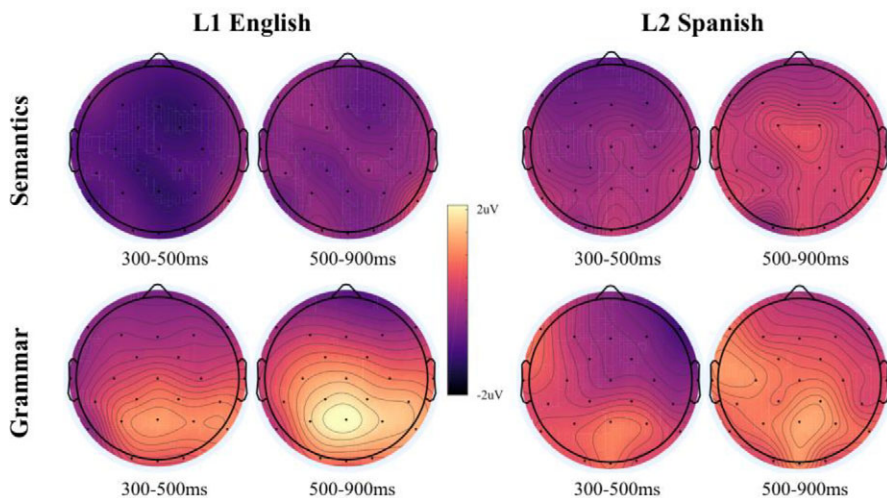
Figure 3. Topographical maps showing the scalp distribution of activity in the semantic anomaly minus matched correct condition and grammar error (verb tense) minus matched correct condition for L1 English and L2 Spanish, averaged for the 300–500 and 500–900ms time windows. Calibration scale is $\pm 2\mu V$.

Table 4. F-statistics from the grand mean ANOVA on mean ERP amplitudes in the 300–500 ms time window

$N = 26$	df	F	η^2_p	p
Language	1, 25	**30.651**	**.551**	**<.001**
Condition	2, 50	**5.767**	**.187**	**.007**
Language × Condition	2, 50	2.165	<.10	.126
Language × AP	1, 25	1.993	<.10	.170
Condition × AP	2, 50	**4.632**	**.156**	**.015**
Language × Cond. × AP	2, 50	0.676	<.10	.497
Language × Hemi.	1, 25	**5.371**	**.177**	**.029**
Condition × Hemi.	2, 50	1.049	<.10	.356
Language × Cond. × Hemi.	2, 50	**5.185**	**.172**	**.009**
Language × AP × Hemi.	1, 25	0.584	<.10	.452
Cond. × AP × Hemi.	2, 50	**6.796**	**.214**	**.003**
Language × Cond. × AP × Hemi.	2, 50	2.251	<.10	.116

Notes. Significant effects are in bold text. Cond. = Condition (correct, verb tense error, semantic error); AP = Anterior/posterior; Hemi. = Hemisphere.

follow-up analyses (anterior left, main effect of Condition, $F(2,50) = 1.637$, $p = .205$, $\eta p^2 < .10$; anterior right, main effect of Condition, $F(2,50) = 1.661$, $p = .200$, $\eta p^2 < .10$; posterior region, main effect of Condition, *ns*; Condition × Hemisphere, *ns*). In sum, there were no discernible significant ERP effects for L2 Spanish sentences in this analysis.

Examining the significant Language × Condition × Hemisphere interaction
Analysis of the Language × Condition × Hemisphere interaction in the global ANOVA produced no significant effects, all $ps > .10$.

Examining the significant Condition × Anterior/posterior × Hemisphere interaction
Analysis of the three-way Condition × Anterior/posterior × Hemisphere interaction revealed in the global ANOVA showed a marginally significant main effect of Condition, $F(2,50) = 3.135$, $p = .055$, $\eta p^2 = .11$ and a significant Condition × Hemisphere interaction at the anterior region, $F(2,50) = 3.897$, $p = .031$, $\eta p^2 = .14$, which appeared to be due to ERPs elicited by semantic errors being more negative than ERPs elicited by correct items and verb tense errors, as indicated by a marginally significant main effect of Condition, $F(2,50) = 3.230$, $p = .051$, $\eta p^2 = .11$ (semantic error M μV $= -0.822$, $SE = .328$; verb tense error M μV $= -.094$, $SE = .385$; correct M μV $= -.095$, $SE = .295$). At the posterior region, there was a significant main effect of Condition, $F(2,50) = 16.708$, $p < .001$, $\eta p^2 = .40$, due to ERPs elicited by semantic errors being more negative than ERPs elicited by correct items and verb tense errors, and ERPs elicited by verb tense errors being more positive than correct items and semantic errors (semantic error M μV $= -0.478$, $SE = .199$; verb tense error M μV $= .769$, $SE = .252$; correct M μV $= .247$, $SE = .246$). This outcome indicated an N400 effect for semantic errors and the onset of a P600 effect for verb tense errors, effects that were driven by responses to L1 English sentences, as shown in the analysis on each language separately, reported in the preceding text.

Global ANOVA 500–900 ms time-window

The statistical results from the global ANOVA in the 500–900 ms time-window are summarized in Table 5. There was a significant main effect of Condition, a significant interaction for Condition × Anterior/posterior and a marginally significant Condition × Anterior/posterior × Hemisphere interaction.

Examining the significant Condition × Anterior/posterior interaction
Follow-up analysis of the Condition × Anterior/posterior interaction demonstrated a significant main effect of Condition at the posterior region, $F(2,50) = 11.101$, $p < .001$, $\eta p^2 = .31$. This was due to ERP effects elicited by verb tense errors being more positive than ERPs elicited by correct sentences and semantic errors (semantic error M μV $= .347$, $SE = .230$; verb tense error M μV $= 1.553$, $SE = .292$; correct M μV $= .478$, $SE = .231$). This result confirms a P600 effect in response to verb tense errors, across both L1 English and L2 Spanish sentences.

To summarize, the group-level ERP results indicated an N400 effect for semantic errors in L1 English, with no significant ERP effects for L2 Spanish in this condition. For verb tense errors, results showed a P600 effect for L1 English and L2 Spanish.

Individual-level ERP results

As discussed, group-level information does not always reflect individual ERP responses within the group, and individuals can systematically vary in the processing approaches they employ during sentence comprehension (Grey et al., 2017; Kim et al., 2018; Tanner, 2019; Tanner & Van Hell, 2014). N400 and P600 effect magnitudes were

Table 5. F-statistics from the grand mean ANOVA on mean ERP amplitudes in the 500–900 ms time window

N = 26	df	F	η_p^2	p
Language	1, 25	1.134	<.10	.297
Condition	2, 50	**5.177**	**.172**	**.009**
Language × Condition	2, 50	1.728	<.10	.192
Language × AP	1, 25	0.034	<.10	.855
Condition × AP	2, 50	**6.303**	**.201**	**.004**
Language × Cond. × AP	2, 50	1.655	<.10	.204
Language × Hemi.	1, 25	0.505	<.10	.484
Condition × Hemi.	2, 50	0.451	<.10	.632
Language × Cond. × Hemi.	2, 50	0.615	<.10	.542
Language × AP × Hemi.	1, 25	0.490	<.10	.490
Cond. × AP × Hemi.	2, 50	**3.221**	**.114**	**.050**
Language × Cond. × AP × Hemi.	2, 50	1.290	<.10	.284

Notes. Significant effects are in bold text. Cond. = Condition (correct, verb tense error, semantic error); AP = Anterior/posterior; Hemi. = Hemisphere.

significantly negatively correlated with each other for L1 semantics ($r = -.897, p < .001$), L1 grammar ($r = -.815, p < .001$), L2 semantics ($r = -.722, p < .001$), and L2 grammar ($r = -.778, p < .001$). This indicates that, within each language and linguistic domain, individuals tended to exhibit either a P600 or N400 response, but not both. Thus, they exhibited a trade-off between either showing P600-dominant responses or N400-dominant responses. This trade-off in N400/P600 effects parallels previous reports (e.g., Bice & Kroll, 2021; Kim et al., 2018) and the RDI information calculated from these effect magnitudes can shed further light on individual-level processing approaches that may be hidden in group-level analyses.

Figure 4 presents RDI scatterplots for the L1 and L2 semantic and grammar conditions. As can be seen, variability in whether participants exhibited N400 or P600 responses was present for both languages and linguistic domains. For L1 grammar, many individuals were P600-dominant in their ERP responses to verb tense errors, which indicates that they were relying on repair/reanalysis mechanisms while processing grammar information, but some individuals were N400-dominant, suggesting they were relying lexical/semantic mechanisms. Variability for L2 grammar was slightly higher: about half of individuals were P600-dominant and about half were N400-dominant.

For semantic processing in L1, most individuals were N400-dominant, indicating they were relying on lexical/semantic mechanisms to process semantic information. However, approximately one third of individuals were P600-dominant, which implies these individuals relied more on repair/reanalysis mechanisms to process the same semantic information. Variability for L2 semantics was slightly higher than for L1, with about half of individuals being N400-dominant and about half P600-dominant.

To examine whether individuals' N400/P600 dominance was related between linguistic domains or L1 and L2, RDI values for L1 and L2 semantics and grammar were entered into a correlation analysis (two-tailed). A positive correlation between RDI values would suggest that individuals tended to employ similar processing approaches (N400 or P600) between the linguistic domains, and/or between L1 and L2. There was a positive correlation between semantic and grammar RDIs for L1, $r = .608, p = .001$. No other significant relationships were observed (L2 semantics and grammar, $r = .332, p = .097$; L1 and L2 semantics, $r = .015, p = .942$; L1 and L2 grammar, $r = .180, p = .378$).

Exploring relationships between individual-level ERPs and language ability measures
Some of the previous work on N400/P600 individual differences has shown that language experience/ability factors can (Bice & Kroll, 2021; Tanner et al., 2014) but do not always (Kim et al., 2018; Tanner, 2019) relate to the observed N400/P600 variability. To examine whether N400/P600-dominant L1 or L2 semantic or grammar processing was related to participants' L1 or L2 abilities, participants' RDI values were entered into a correlation analysis with L1 and L2 verbal fluency and L1 and L2 grammar knowledge; for descriptive information on these tasks, see Table 2.

The results showed that L2 semantic RDIs were significantly negatively correlated with L2 verbal fluency, $r = -.543$, $p = .004$. Thus, higher L2 verbal fluency was associated with greater N400 dominance (i.e., more negative RDIs) during L2 semantic processing. There were also significant negative correlations between L2 semantic RDIs and L2 grammar knowledge, $r = -.464$, $p = .017$, as well as between L2 grammar RDIs and L2 grammar knowledge, $r = -.585$, $p = .002$. This indicates that individuals with greater L2 grammar knowledge tended to show greater N400 dominance during both L2 semantic and L2 grammar processing. There were no significant relationships between L1 processing and language ability measures (L1 semantic RDIs and L1 verbal fluency, $r = -.092$, $p = .654$; L2 verbal fluency, $r = .316$, $p = .115$; L1 grammar knowledge, $r = -.011$, $p = .565$; L2 grammar knowledge, $r = .331$, $p = .099$; L1 grammar RDISs and L1 verbal fluency, $r = .293$, $p = .147$; L2 verbal fluency, $r = .195$, $p = .339$; L1 grammar knowledge, $r = -.019$, $p = .352$; L2 grammar knowledge, $r = -.199$, $p = .330$). For a report on correlations between the behavioral language ability measures of verbal fluency and grammar knowledge see Online Supplementary Materials.
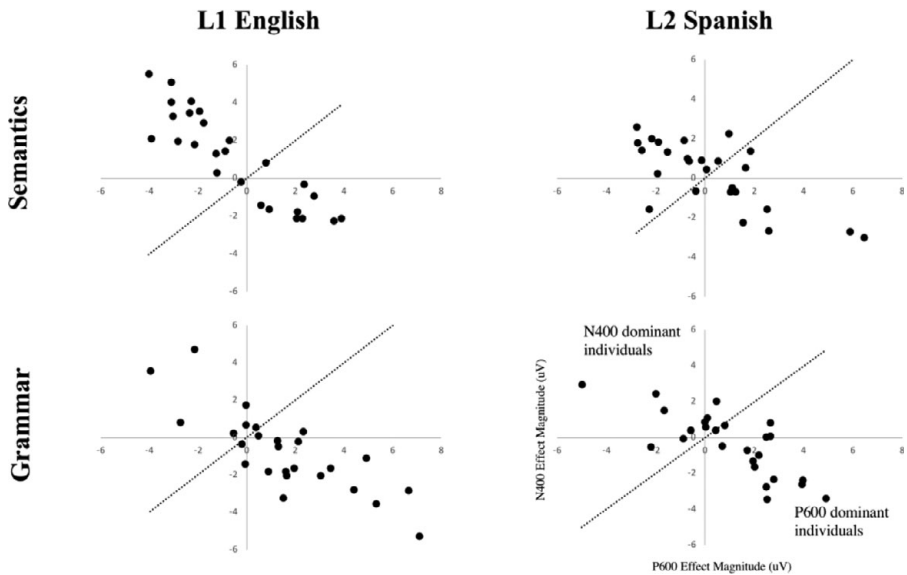

Figure 4. Plot of individuals' N400/P600 response dominance for L1 English semantic and grammar processing and L2 Spanish semantic and grammar processing. More negative RDIs (above and to the left of the dashed line) indicate that an individual is N400-dominant in their processing of the linguistic target whereas more positive RDIs (below and to the right of the dashed line) indicate that an individual is P600-dominant.

Discussion

This study extended research on the ERP correlates of sentence processing in native and nonnative language by (a) examining L1 and L2 processing within-subjects, (b) probing individual-level variability in reliance on lexical/semantic or repair/reanalysis mechanisms, and (c) testing semantic in addition to grammar processing. The group-level results suggested that participants were engaging the commonly expected processing mechanisms for L1 grammar (P600) and L1 semantics (N400). For L2, the group-level results indicated that participants employed P600-related mechanisms for grammar, which has been observed more frequently for L2 at higher than lower proficiency levels. For L2 semantics, which is rarely examined in L2 ERP sentence processing research, no significant ERP effects were observed at the group level.

Analysis of individual-level ERP patterns provided more nuanced details on participants' language processing. For both languages and linguistic domains, individuals varied in whether they relied on lexical/semantic (N400) or repair/reanalysis (P600) mechanisms. This variability was not found to be related between L1 and L2 and was slightly higher for L2 than L1, with some of the L2 variability relating to L2 verbal fluency and grammar knowledge abilities.

The finding that the group-level grand mean ERP analyses did not fully represent the L1 or L2 processing profiles of the individuals within the group adds to the growing number of studies that have obtained similar results for L1 and L2 grammar processing (e.g., Bice & Kroll, 2021; Finestrat Martinez et al., 2018; Grey et al., 2017; Osterhout, 1997; Tanner, 2019; Tanner et al., 2014; Tanner et al., 2013; Tanner & Van Hell, 2014; Wampler et al., 2014) and the small amount of work that has been done on semantic processing, in L1 only (Kim et al., 2018). This body of research contributes importantly to our understanding of L2 processing and how it compares to L1.

As discussed in the "Introduction," nearly all existing L2 ERP research has compared L2 processing to L1 ERP literature or separate L1 control groups. With this approach, deviations from benchmark L1 patterns have been interpreted as reflecting a non-optimal, "nonnativelike," "immature," or "deficient" L2 system. For example, one predominant trend that has emerged in L2 ERP literature over the last 15 years is that at lower L2 proficiency, learners exhibit N400s to L2 grammar (e.g., McLaughlin et al., 2010)—potentially because they are processing grammar features using lexical associations or pattern matching—whereas at higher proficiency, they are considered more likely to show P600s during L2 grammar processing (e.g., Morgan-Short, 2014; Van Hell & Tokowicz, 2010). This has implied that the process of adult nonnative language learning involves a neurocognitive progression from employing "nonnativelike" lexical/semantic N400 mechanisms during L2 grammar processing to using "nativelike" P600 repair/reanalysis processes. However, as has been demonstrated here and evidenced in the related reviewed research, highly proficient L1 speakers show N400s for their processing of grammar information. Thus, the perspective that N400s for L2 grammar processing underlie a "deviant" or nonoptimal language system is difficult to clearly support.

Variability during grammar processing

The finding that individuals demonstrated systematic variability in whether they were N400-dominant or P600-dominant during grammar processing parallels the previous research for L1 (Tanner, 2019; Tanner & Van Hell, 2014) and for L2 (Tanner et al., 2014; Tanner et al., 2013), including the few L1/L2 within-subjects sentence processing

studies (Bice & Kroll, 2021; Finestrat Martinez et al., 2018; Wampler et al., 2014). The patterns observed here reinforce that individuals do not uniformly engage the oft-expected structural repair/reanalysis P600 processes when they encounter language grammar information, in L2 or L1. Rather, some individuals, and in this study approximately one third of individuals for the L1 and 1/2 for their L2, employed lexical/semantic processes to handle the same grammar information. Although this study has framed the observed N400/P600 variation in terms of lexical/semantic access versus repair/reanalysis mechanisms, the results can be interpreted more broadly within other theoretical frameworks.

Fromont, Steinhauer, and Royle (2020), for example, used a similar stimuli design as the present study and observed biphasic N400-P600 effects across all of their experimental conditions (syntactic violation, semantic anomaly, combined syntactic + semantic anomaly) in L1 French speakers. They suggest that N400s in response to syntactic violations reflect a mismatch between a predicted word stem and the presented stimulus, and that this mechanism takes place simultaneously with lexical/semantic processing. Ultimately, they argue that results from balanced ERP stimuli designs (as was used in the present study) provide evidence against "syntax-first" approaches to sentence processing. This has implications for our understanding of what to expect from L1 speakers' sentence processing and, by extension, what to expect from L2 learners' sentence processing.

The finding that some individuals exhibited N400s in response to L1 and L2 grammar information may alternatively be a reflection of these individuals using a "good enough" processing approach wherein they rely on shallow lexical or heuristic processing more than on detailed, rule-based grammatical processing (for a review on "good enough" language processing, see Karimi & Ferreira, 2016). Aligning with this, the N400 has been theoretically linked with shallower lexical and heuristic processing and the P600 with more detailed, rule-based processing (e.g., Kuperberg, 2007; Tanner, 2011). This may thus be one theoretical dimension along which the participants in the present study varied in their grammar processing.

The fact that RDIs in L1 and L2 processing of grammar information were not related to each other indicates that N400-dominant participants did not have an individual tendency to default to "good enough" processing of grammar across their two languages. In fact, shallow lexical and heuristic processing seemed to be more likely for L2 than L1 grammar as indicated by the greater number of individuals who showed N400-dominant responses to L2 grammar than L1 grammar (see Figure 4). This suggests that L2 learners are more likely use "good enough" processing for grammar and interpreting the results from a "good enough" approach paves the way for research on what factors differentiate whether and when individuals use "good enough" processing in their L1 and L2.

Another theory that the present results relate to is the Declarative/Procedural (DP) Model of Language (e.g., Ullman, 2001, 2004, 2016, 2020). In the DP model, L2 learners at lower levels of proficiency are predicted to rely on the declarative memory system to process both lexical/semantic and grammar information, and the model links N400s to declarative memory processing. With increased proficiency, L2 learners are predicted to gradually shift reliance to procedural memory for grammar processing, with procedural memory processing linked to the elicitation of LANs (with P600s) in response to grammar violations. Within the model, then, individuals who were N400-dominant in response to L1 or L2 verb tense errors may have been relying on declarative memory. However, aside from linking N400s to declarative memory, the model does not seem to account for the observation in this study and related research (see Table 1)

of individual variability and N400s in response to L1 grammar. In particular, it will be interesting to consider what the findings for individual-level variability in N400/P600 dominance may entail for the model's L1 and L2 ERP predictions, which have been informed mainly by between-subjects, group-level L1 and L2 ERP language research.

This study's findings can also be interpreted with respect to the Shallow Structures Hypothesis (SSH; Clahsen & Felser, 2006b, 2006a, 2018). The basic premise of the SSH is that L2 grammar may not be specified enough to provide the kinds of detailed structural information that are needed to process L2 grammar in an L1-like way. Thus, L2 learners parse L2 sentences shallowly—relying more on lexical, pragmatic, semantic, and other information, but less so or not at all on the more detailed, complex grammatical information that L1 speakers would use. Notably, shallow processing is not unique to the L2. L1 speakers can also use shallow processing, as noted by Clahsen and colleagues and shown in the "good enough" processing research (e.g., Karimi & Ferreira, 2016). Overall, the SSH claims that grammatical processing in L2 learners may be less robust than that of L1 speakers and that "L2 processing tends to rely more on nongrammatical information than on the grammatical route to interpretation" (Clahsen & Felser, 2018, p. 701).

From an SSH perspective, the study's findings indicate that, for L2 grammar, some participants exhibited shallow processing as evidenced by their N400-dominance when processing verb tense errors. However, L2 grammar processing was not restricted to shallow processing and was as robust as what might be expected from L1 speakers, as shown in the group-level P600 effect for L2 grammar and the P600-dominant effects observed at the individual level. The SSH, like the DP model, has been formulated based on L1/L2 between-subjects comparisons (Clahsen & Felser, 2018) and it therefore does not seem to be quite articulated enough to capture the present study's within-subjects, individual-level findings for L1 and L2 processing. With increased N400/P600 individual differences research in the second language acquisition and bilingualism fields, theoretical models of adult L2 learning and processing may need to incorporate more nuanced predictions that can more directly account for the patterns of systematic interindividual variability being observed within and across L1 and L2 (e.g., Bice & Kroll, 2021; Finestrat Martinez et al., 2018; Wampler et al., 2014).

Variability during semantic processing

In addition to grammar, the study also tested semantic processing. The results of variability in whether individuals relied on N400 or P600 mechanisms to process L1 semantic information are in line with Kim et al. (2018). These results show that individuals varied in whether they relied on lexical/semantic access mechanisms or structural reanalysis mechanisms to process the semantic information. Observing P600s to the semantic anomalies may indicate that the same underlying neurocognitive resources are recruited for syntactic and semantic integration (Fromont et al., 2020). The finding that individuals also varied in their processing of L2 semantic information, which had not yet been examined from an L1/L2 within-subjects, individual-level perspective, highlights the general systematicity of individual differences in the ERP correlates of sentence processing. Overall, the study's findings underscore that language users follow multiple routes to comprehension, in the L1 as well as L2, and for both semantic and grammar domains of language.

Of note also is that the present study used a word-probe verification task, which diverges from the acceptability/grammaticality judgment tasks used in all of the

previous reports on this systematic variability. This suggests, similar to the findings of Tanner (2019) for stimuli presentation type, that the variability is not tied to the type of task participants are asked to carry out but rather seems to reflect a natural tendency for individuals to vary in the mechanisms they rely on during sentence processing.

This is not to say, however, that task effects can be excluded as such effects on ERP correlates of language are well-documented (e.g., Hahne & Friederici, 2002; Osterhout & Mobley, 1999; Nieuwland, 2014; Schacht et al., 2014). It could be the case, for example, that the lack of group-level N400 effects for L2 semantics was related to not drawing participants' attention to well-formedness of the sentences within an acceptability/grammaticality judgment paradigm. For example, Schacht et al. (2014) found that the N400 response to semantic anomalies in Spanish L1 speakers was smaller in amplitude when using a word-probe verification task than a grammaticality judgment task. However, as has been highlighted throughout this article, caution seems warranted in interpreting ERP effects—or a lack of ERP effects—at the group level because group-level results may not accurately reflect the ERP responses of all individuals in the group. Indeed, the individual-level results in this study show that participants were sensitive to L2 Spanish semantics in the word-probe task, even in the absence of being oriented to well-formedness—they were simply variable in whether they showed N400-dominant or P600-dominant brain responses, and this variability was comparable to that found in related research that has used acceptability/grammaticality judgment paradigms (see Table 1). Nonetheless, further examination of potential task effects on the patterns of N400/P600 variability in L1/L2 sentence processing will be warranted in future research.

Relationships between variability in brain responses and other factors

The negative correlation observed between N400 and P600 effect magnitudes indicates a trade-off relationship between the two ERP effects (Kim et al., 2018; Tanner & Van Hell, 2014; Tanner, 2019). Kim et al. (2018) hypothesize that this trade-off arises due to brain responses competing with each other for expression, such that one occurs when the other does not, or stronger expression of one (i.e., larger effect magnitudes) corresponds to weaker expression of the other (i.e., smaller effect magnitudes). Similarly, Tanner et al. (2013) suggest that this trade-off may be explained by processing models that posit a competitive dynamic between a shallower lexical processing stream and a more abstract, rule-based processing stream. The current study's results support these perspectives of a competitive trade-off and provide evidence that the trade-off manifests within individuals' L1 and L2 and for both semantic and grammar processing.

Although systematic interindividual variability in ERP responses was observed for both languages and linguistic domains, this variability was not related between individuals' native and nonnative language. That is, individuals who demonstrated N400-dominance for L1 grammar processing did not also tend to demonstrate N400-dominance for L2 grammar processing (and vice versa for P600 responses). This aligns with Wampler et al. (2014) and Finestrat Martinez et al. (2018; interpret with caution because they report only eight participants; see Table 1). However, the lack of a relationship contrasts partially with Bice and Kroll (2021) who found that individual-level ERP variation showed some overlap between L1 and L2 grammar processing.

Learners' L2 proficiency level may explain these outcomes. Bice and Kroll tested fluent heritage bilinguals and proficient late bilinguals and suggested that language

processing proceeds similarly in an individual's two languages once proficiency in each language approaches similar levels. In contrast, the learners in Wampler et al. were only in their second year of L2 instruction, those in Finestrat Martinez et al. were classified as low/intermediate proficiency, and those tested in the present study were classified as intermediate proficiency. These lower proficiency learners may have not yet reached an L2 proficiency level at which their L2 and L1 processing preferences, be they N400-based or P600-based, converge. This lack of an L1–L2 relationship, for grammar or semantics, suggests that individual language processing preferences for L1 are not automatically transferred to the L2 and that L2 processing exhibits its own pattern of differential reliance on lexical/semantic or repair/reanalysis mechanisms during sentence processing, at least when L1/L2 proficiency is unbalanced and/or not at similarly high levels.

Recall that the group-based L2 ERP research has indicated a potential neurocognitive progression from N400 to P600-based processing (of grammar). The findings from the present study and future within-subjects, individual-level research may lead to a modification of that perspective. Specifically, rather than a progression from N400s to P600s as indicating that L2 grammar processing becomes more "nativelike," it may be that a fruitful framework for elucidating similarities or differences between L1 and L2 is probing whether, and under what conditions, individuals' language processing preferences align between their two languages, that is, whether and under what conditions individuals employ similar mechanisms for L2 comprehension as they do for L1, be they N400-based mechanisms or P600-based mechanisms. This individualized language processing perspective could apply to semantic processing as well and thus offers a broad yet comprehensive electrophysiological framework for understanding L1 and L2 sentence processing, for which L2 research and theory has focused mainly on explaining grammar. Continued research in this area will shed further light on the relationship between L1 and L2 at the level of individual differences in brain processes related to real-time language processing.

From the perspective of individual variation, it is of interest to not only illuminate the systematic individual differences in neural correlates of language but also attempt to identify the sources of this variation. Indeed, several of the N400/P600 individual differences studies discussed in this article attempted to uncover potential sources. For L1 semantic processing, Kim et al. (2018) provided evidence that verbal working memory may be one source underlying N400/P600 variability. In their study, higher verbal working was linked with larger P600s and smaller N400s during semantic processing. For grammar processing, Bice and Kroll (2021) observed that higher working memory was related to larger N400s and smaller P600s in the monolingual L1 control group whereas Tanner (2019) did not find any relationships between N400/P600 responses and language experience/knowledge or cognitive factors in L1 speakers, including for working memory. The present study, though it did not focus on the potential sources of individual variation, similarly did not observe relationships between N400/P600 responses and L1 experience/knowledge.

In some studies, higher variability and an increased likelihood of N400 effects during L1 grammar processing has been related to the biological factor of self or familial left-handedness (Grey et al., 2017; Kos et al., 2012; Tanner & Van Hell, 2014). In this study, all participants were right-handed; seven reported having a left-handed blood relative (i.e., familial left-handedness). Of the seven, three were N400-dominant during L1 grammar processing. Only one of these three was also N400-dominant during L2 grammar processing. This information suggests that familial left-handedness was

unlikely to be a contributor to the variability observed here for L1 or L2 grammar processing.

Although L1 processing was not linked with any of the assessed language experience/knowledge factors in the present study, L2 processing did show relationships with these factors. Related individual differences work (which studied L2 grammar only) has observed relationships between cognitive measures and N400/P600 individual differences. For example, Faretta-Stutenberg and Morgan-Short (2018a) found that working memory performance and procedural memory ability related to increased ERP magnitudes following a study abroad experience. These relationships were observed regardless of whether individual learners showed P600s or N400s to L2 syntactic violations. However, in an examination of the same learners' processing of morphosyntactic gender agreement, Faretta-Stutenberg and Morgan-Short (2018b) observed no significant relationships between L2 processing and the language experience factors of initial L2 proficiency measured before beginning study abroad or reports of L2 contact during study abroad.

The current study revealed relationships between L2 processing and L2 verbal fluency and L2 grammar knowledge. The interpretations offered here are tentative and work with larger sample sizes will help to confirm or complicate them. The relationship between L2 verbal fluency and N400-dominance for L2 semantics indicates that individuals with higher L2 verbal fluency tended to be more N400-dominant when processing semantic information. The verbal fluency task used in the study was a semantic fluency task, which represents lexical knowledge and indexes word retrieval efficiency under time pressure (Portocarrero et al., 2007; Rosselli et al., 2000; Santilli et al., 2019). From this perspective, the observed relationship suggests that individuals with greater L2 lexical knowledge and more efficient lexical access are more likely to exploit these skills and employ lexical/semantic mechanisms to process semantic information during L2 sentence comprehension.

Regarding the factor of L2 grammar knowledge, higher L2 grammar knowledge was related to greater N400-dominance for both L2 semantic and L2 grammar processing. This implies that with increasing grammar knowledge/ability in the L2, individuals are more likely to rely on lexical/semantic mechanisms to process L2 semantic as well as L2 grammar information. This finding may seem counterintuitive because one might expect that higher grammatical abilities correspond with more reliance on structural reanalysis P600 processes, at least for grammar. This is not a straightforward expectation for L2, however. Tanner et al. (2013), for example, observed that higher behavioral ability to detect L2 grammar errors during a grammaticality judgment test was related to higher P600 magnitudes, but Pelíssier (2020) found that post-EEG GJT performance was related to greater N400-dominance.

The relationship found here for L2 semantics and grammar may be related to the L2 grammar knowledge test used in the study. In the test, participants selected responses to each item from a list of options in multiple-choice format (see Online Supplementary Materials for details). This means they only had to recognize the correct response, rather than produce/compute it on their own. This format arguably lends itself to using lexically based recognition processes over grammar-based, structural processes. Therefore, participants may have been using lexical knowledge of grammar-related information to complete the test. Taken together, the results for L2 verbal fluency and L2 grammar knowledge (which were positively correlated, $R = .440$, $p = .024$; see Online Supplementary Materials) suggest that individuals with greater L2 lexical abilities are more likely to rely on lexical/semantic processes during

real-time L2 comprehension. Linking this back to the theoretical perspectives discussed earlier, this points to individuals with greater L2 lexical abilities being more likely to employ "good enough" processing during sentence comprehension or, within the DP model, being more likely to rely on declarative memory.

To summarize, this study contributes to the small amount of existing within-subjects ERP work on L1/L2 sentence processing, and it provides intriguing information on potential links between behavioral L2 abilities and individual differences in the neural correlates of L2 sentence processing. Nonetheless, as noted by Tanner (2019), there is not yet "a clear picture of what cognitive, experiential, or neuroanatomical differences might underlie the individual differences" (p. 223). Future studies in this area will help to clarify the various sources that contribute to individual variability in ERP responses during language processing, for L2 as well as L1.

Finally, it is of note that these relationships were observed for individuals' L2 processing but not their L1. This may be explained by higher variability in L2 making relationships between behavioral individual differences and neural individual differences more pronounced. For L2 semantics and grammar, the RDI results showed that about half of individuals were N400-dominant and about half were P600-dominant (see Figure 4); this variation was slightly higher than that observed for L1. L2 processing in general is considered to be more variable than L1 processing and that held true in this study. Interestingly, L2 grammar and L2 semantic processing were comparable in their variation; that is, N400/P600 response dominance variation was not notably higher in one domain or the other. L2 grammar processing is generally viewed as being more variable during sentence comprehension than L2 semantic processing (e.g., Bowden et al., 2013; Morgan-Short, 2014). However, the current results demonstrate that this is not necessarily the case and point to a need for more ERP research on L2 semantic processing in sentence-level contexts to better understand the neurocognition of this language domain in L2 users.

Conclusion

This is the first ERP study, to the author's knowledge, to examine L1 and L2 semantic and grammar processing, together, using a within-subjects, individual-level approach for sentence comprehension. This study and the work that inspired it offer a compelling foundation upon which to continue to examine individual differences in the electrophysiological correlates of native and nonnative sentence processing. Future research should aim to illuminate whether and under what conditions the neural processes underlying individual processing preferences between languages align. Overall, the findings of systematic individual differences in the neural correlates of sentence processing, across both languages and linguistic domains, offer new insight into native and nonnative language processing and how they compare, particularly with respect to the different processing routes that individuals take during language comprehension.

experimental design and to the members of the EEG lab for Language and Multilingualism Research at Fordham University for assistance with data collection and coding. Preparation of the manuscript was supported by a Fordham University Faculty Fellowship.

References

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

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

Ardal, S., Donald, M. W., Meuter, R., Muldrew, S., & Luce, M. (1990). Brain responses to semantic incongruity in bilinguals. *Brain and Language*, *39*, 187–205.

Batterink, L., & Neville, H. (2013). Implicit and explicit second language training recruit common neural mechanisms for syntactic processing. *Journal of Cognitive Neuroscience*, *25*, 936–951.

Bice, K., & Kroll, J. F. (2021). Grammatical processing in two languages: How individual differences in language experience and cognitive abilities shape comprehension in heritage bilinguals. *Journal of Neurolinguistics*, *58*, 100963.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain Research Review*, *59*, 55–73.

Bowden, H. W., Grey, S., Reichle, R., & Ullman, M.T. (In preparation). *In-depth quantitative review of event-related potential research on L2 grammar processing*. Presented in part at the 2018 meeting of the American Association for Applied Linguistics.

Bowden, H. W., Steinhauer, K., Sanz, C., & Ullman, M. T. (2013). Native-like brain processing of syntax can be attained by university foreign language learners. *Neuropsychologia*, *51*, 2492–2511.

Caffarra, S., Mendoza, M., & Davidson, D. (2019). Is the LAN effect in morphosyntactic processing an ERP artifact? *Brain and Language*, *191*, 9–16.

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

Choudhary, K. K., Schlesewsky, M., Roehm, D., & Bornkessel-Schlesewsky, I. (2009). The N400 as a correlate of interpretatively relevant linguistic rules: Evidence from Hindi. *Neuropsychologia*, *47*, 3012–3022.

Chow, W.-Y., & Phillips, C. (2013). No semantic illusions in the "Semantic P600" phenomenon: ERP evidence from Mandarin Chinese. *Brain Research*, *1506*, 76–93.

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

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

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

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, *13*, 21–58.

Debruille, J. B. (2007). The N400 potential could index a semantic inhibition. *Brain Research Review*, *56*, 472–477.

Debruille, J. B., Ramirez, D., Wolf, Y., Schaefer, A., Nguyen, T.-V., Bacon, B., & Brodeur, M. (2008). Knowledge inhibition and N400: A within- and between-subjects study with distractor words. *Brain Research*, *1187*, 167–183.

DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, *61*, 150–162.

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

Faretta-Stutenberg, M., & Morgan-Short, K. (2018a). The interplay of individual differences and context of learning in behavioral and neurocognitive second language development. *Second Language Research*, *34*, 67–101.

Faretta-Stutenberg, M., & Morgan-Short, K. (2018b). Contributions of initial proficiency and language use to second-language development during study abroad: Behavioral and event-related potential evidence. In *The Routledge Handbook of Study Abroad Research and Practice* (pp. 421–435). Routledge.

Finestrat Martinez, I., Luque, A., Abugaber, D., & Morgan-Short, K. (2018). *Native-language processing as an individual difference explaining variability in L2 processing: An event-related potential study*. American Association for Applied Linguistics, Chicago, IL.

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

Friederici, A., Hahne, A., & Saddy, D. (2002). Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. *Journal of Psycholinguistic Research*, *31*, 45–63.

Fromont, L. A., Royle, P., & Steinhauer, K. (2020). Growing Random Forests reveals that exposure and proficiency best account for individual variability in L2 (and L1) brain potentials for syntax and semantics. *Brain and Language*, *204*, 104770.

Fromont, L. A., Steinhauer, K., & Royle, P. (2020). Verbing nouns and nouning verbs: Using a balanced design provides ERP evidence against "syntax-first" approaches to sentence processing. *PloS One*, *15*, e0229169.

Gabriele, A., Alemán Bañón, J., Hoffman, L., Covey, L., Rossomondo, A., & Fiorentino, R. (2021). Examining variability in the processing of agreement in novice learners: Evidence from event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication.

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

Grey, S., Sanz, C., Morgan-Short, K., & Ullman, M. T. (2018). Bilingual and monolingual adults learning an additional language: ERPs reveal differences in syntactic processing. *Bilingualism: Language and Cognition*, *21*, 970–994.

Grey, S., Schubel, L. C., McQueen, J. M., & Van Hell, J. G. (2019). Processing foreign-accented speech in a second language: Evidence from ERPs during sentence comprehension in bilinguals. *Bilingualism: Language and Cognition*, *22*, 912–929.

Grey, S., & Tagarelli, K. M. (2018). Psycholinguistic methods. In A. Phakiti, P. de Costa, L. Plonsky, & S. Starfield (Eds.), *Palgrave Handbook of Applied Linguistics Research Methodology* (pp. 287–312). Palgrave MacMillan.

Grey, S., Tanner, D., & Van Hell, J. G. (2017). How right is left? Handedness modulates neural responses during morphosyntactic processing. *Brain Research*, *1669*, 27–43.

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

Hahne, A., & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research*, *13*, 339–356.

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

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, Edith. (2007). Event-related potentials and language processing: A brief overview. *Language and Linguistics Compass*, 571–591.

Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly Journal of Experimental Psychology*, *69*, 1013–1040.

Kim, A. E., Oines, L., & Miyake, A. (2018). Individual differences in verbal working memory underlie a tradeoff between semantic and structural processing difficulty during language comprehension: An ERP investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 406.

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

Kolk, H., & Chwilla, D. (2007). Late positivities in unusual situations. *Brain and Language*, *100*, 257–261.

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

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

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

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., & 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. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential. *Annual Review of Psychology*, *62*, 621–647.

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

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

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 213.

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, *54*, 146–157.

Martin-Loeches, M., Nigbur, R., Casado, P., Hohlfeld, A., & Sommer, W. (2006). Semantics prevalence over syntax during sentence processing: A brain potential study of noun–adjective agreement in Spanish. *Brain Research*, *1093*, 178–189.

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

Molinaro, N., Barber, H. A., Caffarra, S., & Carreiras, M. (2015). On the left anterior negativity (LAN): The case of morphosyntactic agreement. *Cortex*, *66*, 156–159.

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

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. (2014). Electrophysiological approaches to understanding second language acquisition: A field reaching its potential. *Annual Review of Applied Linguistics*, *34*, 15–36.

Morgan-Short, K., Finger, I., Grey, S., & Ullman, M. T. (2012). Second language processing shows increased native-like neural responses after months of no Exposure. *PLoS One*, *7*(3), e32974.

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.

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.

Newman, Aaron. 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. (2014). "Who's he?" Event-related brain potentials and unbound pronouns. *Journal of Memory and Language*, *76*, 1–28.

Nieuwland, M. S., & Van Berkum, J. J. A. (2006a). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. *Brain Research*, *1118*, 155–167.

Nieuwland, M. S., & Van Berkum, J. J. A. (2006b). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, *18*, 1098–1111.

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

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

Ortega, L. (2013a). SLA for the 21st century: Disciplinary progress, transdisciplinary relevance, and the bi/multilingual turn. *Language Learning*, *63*, 1–24.

Ortega, L. (2013b). Ways forward for a bi/multilingual turn in SLA. In *The multilingual turn* (pp. 42–63). Routledge.

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., Kim, A., & Kuperberg, G. R. (2012). The neurobiology of sentence comprehension. In M. Spivey, M. Joannissee, & K. McCrae (Eds.), *The Cambridge Handbook of Psycholinguistics* (pp. 365–389). Cambridge University Press.

Osterhout, L., & Mobley, L. A. (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.

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.

Paradis, M. (2009). *Declarative and procedural determinants of second languages* (Vol. *40*). John Benjamins Publishing Company.

Pélissier, M. (2020). Comparing ERPs between native speakers and second language learners: Dealing with individual variability. In A. Edmonds, P. Leclerq, & A. Gudmestead (Eds.), *Interpreting language-learning data* (pp. 39–69). Language Science Press.

Portocarrero, J. S., Burright, R. G., & Donovick, P. J. (2007). Vocabulary and verbal fluency of bilingual and monolingual college students. *Archives of Clinical Neuropsychology*, *22*, 415–422.

Rosselli, M., Ardila, A., Araujo, K., Weekes, V. A., Caracciolo, V., Padilla, M., & Ostrosky-Solí, F. (2000). Verbal fluency and repetition skills in healthy older Spanish-English bilinguals. *Applied Neuropsychology*, *7*, 17–24.

Santilli, M., Vilas, M. G., Mikulan, E., Caro, M. M., Muñoz, E., Sedeño, L., … & García, A. M. (2019). Bilingual memory, to the extreme: Lexical processing in simultaneous interpreters. *Bilingualism: Language and Cognition*, *22*, 331–348.

Schacht, A., Sommer, W., Shmuilovich, O., Martíenz, P. C., & Martín-Loeches, M. (2014). Differential task effects on N400 and P600 elicited by semantic and syntactic violations. *PloS One*, *9*, e91226.

Silva-Pereyra, J. F., & Carreiras, M. (2007). An ERP study of agreement features in Spanish. *Brain Research*, *1185*, 201–211.

Steinhauer, K. (2014). Event-related potentials (ERPs) in second language research: A brief introduction to the technique, a selected review, and an invitation to reconsider critical periods in L2. *Applied Linguistics*, *35*, 393–417.

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.

Stowe, L. A., Kaan, E., Sabourin, L., & Taylor, R. C. (2018). The sentence wrap-up dogma. *Cognition*, *176*, 232–247.

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

Tanner, D. (2015). On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement. *Cortex*, *66*, 149–155.

Tanner, D. (2019). Robust individual differences in grammatical agreement processing: A latent variable approach. *Cortex*, *111*, 210–237.

Tanner, D., Grey, S., & van Hell, J. G. (2017). Dissociating retrieval interference and reanalysis in the P600 during sentence comprehension. *Psychophysiology*, *54*, 248–259.

Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in on-line L2 grammatical comprehension. *Bilingualism: Language and Cognition*, *17*, 277–293.

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

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. (2020). The declarative/procedural model: A Neurobiologically-motivated theory of first and second language. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in Second Language Acquisition* (3rd ed., pp. 128–161). Routledge.

Ullman, M. T. (2016). The declarative/procedural model: A neurobiological model of language learning, knowledge, and use. In G. Hickok & S. A. Small (Eds.), *The Neurobiology of Language* (pp. 953–968). Elsevier.

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

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.

Van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, *41*, 147–182.

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., & 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.

Wampler, E. K., McLaughlin, J., & Osterhout, L. (2014). *How Gender, Handedness, and L1 Processing Strategy Influence L2 Grammatical Processing.* Annual Meeting of the Society for the Neurobiology of Language.