Individual Differences and the Role of the L1 in L2 Processing: An ERP Investigation

Kristi Bond, Alison Gabriele, Robert Fiorentino, and José Alemán Bañón
University of Kansas

1. Introduction

Our study uses event-related potentials (ERP) to investigate two factors in the second language (L2) processing of agreement: the role of number and gender features in the native language (L1), and the impact of individual differences between learners. Previous ERP studies investigating L2 agreement vary as to whether they find native-like processing (e.g., Rossi et al., 2006; Tokowicz and MacWhinney, 2005; Ojima et al., 2005; Sabourin, 2003). Additionally, recent studies have found variability between groups of learners with regard to development of native-like responses to agreement violations (McLaughlin et al., 2010; Tanner et al., 2009). We further examine these issues by testing both learner aptitude and sensitivity to number and gender agreement violations in L2 Spanish by native speakers of English. An analysis of sensitivity to violations in three types of agreement allow the comparison of responses to (a) number features on verbs (*el barco*$_{sg}$*…flota*$_{sg}$/ **flotan*$_{pl}$), which are similar in the L1 English, (b) number features on adjectives (*la isla*$_{sg}$*…rocosa*$_{sg}$ /**rocosas*$_{pl}$), a category where agreement is not instantiated in English, and (c) gender agreement on adjectives (*la isla*$_{fem}$*…rocosa*$_{fem}$/**rocoso*$_{masc}$), which involves a feature that is unique to the L2.

Theories of L2 acquisition differ with regard to claims about the potential for native-like acquisition, a debate that centers on distinct claims regarding the role of a proposed universal set of features and parameters for language, or Universal Grammar (UG) (Chomsky, 1965; 1980; 1981). The *Interpretability Hypothesis* (Tsimpli and Mastropavlou, 2007; Tsimpli and Dimitrakopoulou, 2007) claims that adult learners have access to UG only through their L1 and are unable to incorporate features of the L2 that are not instantiated in the L1 if those features are "uninterpretable". For example, the gender and number features of adjectives in Spanish are argued to be uninterpretable (e.g., Carstens, 2000) and thus should be acquirable by L2 learners only if those features are also present in the L1. For English-speaking learners of Spanish, number agreement should be acquirable to native-like levels, but not gender agreement. The *Full Transfer/Full Access* approach (Schwartz and Sprouse, 1994, 1996), on the other hand, claims that L1 features and parameter settings are influential in the early stages of L2 acquisition, but that L1 properties can be abandoned in favor of other UG-constrained settings as needed in order to accommodate L2 input. Number and gender features in L2 Spanish should both be acquirable by English-speaking learners, but at lower levels of proficiency, native-like processing of gender agreement may not be well-developed.

The extent to which learners exhibit native-like acquisition of agreement features has largely been investigated using behavioral measures (White et al., 2004; McCarthy, 2008; Franceschina, 2002; Montrul et al., 2008), but the availability of brain imaging technology has recently made possible the brain-level comparison of agreement processing in learners and native speakers of a language. Particularly promising is the use of EEG technology to measure event-related potentials (ERPs). ERPs are a measurement of electrical activity in the brain time-locked to a specific stimulus. When averaged

* This study was supported by a National Science Foundation Graduate Research Fellowship to the first author and funding by the University of Kansas General Research Fund to the second author. We would also like to thank the participants, as well as the other members of the Second Language Acquisition Lab and the Neurolinguistics and Language Processing Lab at the University of Kansas.

across multiple participants and multiple tokens of the same type of stimuli, ERPs can reveal very small changes in the brain's electrical signals related specifically to either morphosyntactic violations or semantic anomalies. With regard to the morphosyntactic processing of agreement, a late positivity in the posterior electrodes (the P600) has been observed across a number of L1 studies (e.g., Barber and Carreiras, 2005; Kaan et al., 2000; Osterhout and Mobley, 1995). This response begins around 500ms post-stimulus and lasts for several hundred milliseconds. It is important to contrast the P600 component with an earlier component elicited by other types of stimuli: semantic anomalies and lexical integration tend to elicit a broad, negative-going waveform with onset around 250ms and peaking around 400ms (the N400) (e.g., Kutas and Hillyard, 1980) that is sensitive to factors like semantic congruity in context, lexical frequency and associative probability (e.g., Kutas and Van Petten, 1994; Kutas and Federmeier, 2000; Lau et al., 2008). In summary, the P600 component is a reliable indicator of morphosyntactic processing, while the N400 response indicates lexical/semantic processing.

The results of ERP studies investigating morphosyntactic processes related to features like number and gender agreement are inconsistent with regard to whether the properties of the L1 constrain L2 processing. Some ERP studies find native-like processing in learners in response to agreement violations. With regard to subject-verb agreement, for example, Rossi et al. (2006) found that high (but not low) proficiency bilingual German-Italian participants were equally native-like in both languages. However, it is the case that both languages instantiate subject-verb agreement. In contrast, Ojima et al. (2005) find that even high-proficiency Japanese learners of English do not demonstrate a P600 to subject-verb agreement violations. Ojima et al. conclude that the primary factor in these results is the lack of such agreement in the L1 Japanese. Responses to gender agreement violations are less consistent: Tokowicz and MacWhinney (2005) report a P600 response to gender violations on determiners in L2 Spanish (una_{fem}/*un_{masc} $fiesta_{fem}$) for novice learners whose L1 is English, a language which does not have grammatical gender agreement. However, Sabourin's (2003) L1 English group does not produce a P600 response to Dutch gender violations on determiners (het_{neut}/*de_{com} *kleine*). It would seem, then, that the role of the L1 merits further investigation. Indeed, a review by Kotz (2009) specifically points out the paucity of ERP studies that directly test the effects of L1 transfer.

Additionally, some studies have found that native-like processing is not present even for features that are shared between the L1 and L2 in cases in which those features are instantiated differently (Tokowicz and MacWhinney, 2005; Frenck-Mestre et al., 2009; McLaughlin et al., 2010). For example, in a study designed to directly test whether incongruous systems impact processing of similar features in the L2, Frenck-Mestre et al. (2009) targeted gender agreement in German-speaking learners of French, where both the L1 and the L2 instantiate gender agreement but show differences with respect to how the gender feature is realized. Stimuli included gender agreement between determiners and nouns (la_{fem}/*le_{masc} $chaise_{fem}$), a context in which the L1 and L2 are similar, and between nouns and post-nominal plural adjectives ($chaises_{fem}$ $blanches_{fem}$/*$blancs_{masc}$), a context in which agreement is not established in the L1 German. Learners demonstrated native-like processing in the similar condition (determiner-noun), producing a P600 to agreement violations. In contrast, there was no P600 effect in response to noun-adjective agreement violations. Frenck-Mestre et al. interpret this difference in terms of featural instantiations in the L1 and L2, but in another experiment, they achieved the same results with adjectives in a pre-nominal position, a context in which gender agreement is instantiated in the L1 German. Clearly, the extent to which the specific morphological instantiation of a shared feature in the L1 and L2 impacts L2 processing should be examined further.

In addition to testing specific theories regarding the role that differences between languages play in L2 acquisition, a second aim of this project is to investigate differences between learners. Individual differences are believed to contribute to a higher degree of learner variability in adult versus child acquisition, leading Bley-Vroman (1989, 1990) to propose that child L1 and adult L2 acquisition are fundamentally different. The basic claim of his *Fundamental Difference Hypothesis* is that adult L2 acquisition is constrained by general problem-solving capacities helped along by knowledge of the L1. Although these cognitive abilities were originally posited as non-specific to language, subsequent studies have taken this to mean that language aptitude plays a significant role in L2 acquisition (e.g., DeKeyser 2000). In fact, a separate body of literature has shown verbal (language) aptitude to be highly correlated with learner success in L2 acquisition (e.g., Skehan, 1989; Carroll, 1981;

Abrahamsson and Hyltenstam, 2008; Harley and Hart, 1997). Verbal aptitude, however, may not be the best measure of Bley-Vroman's claim regarding *domain-general* cognitive abilities, which remain largely untested due to the high incidence of verbal measures in L2 studies related to individual differences in aptitude, working memory or intelligence (e.g., Robinson, 2002; Sasaki, 1996; McDonald, 2006; Abrahamsson and Hyltenstam, 2008; Sagarra, 2007; Sagarra and Herschensohn, 2010; Juffs, 2004). The question remains, then, as to whether there is independent evidence, *outside the realm of verbal measures,* that general cognitive factors are correlated with L2 performance.

Critically, differences between learners in terms of aptitude - whether domain-specific or domain-general - have not been much investigated in ERP studies to date. However, two recent studies revealed that early learners fall into two groups with regard to whether morphosyntactic violations are treated as syntactic or semantic in nature, resulting in an N400 or P600, respectively. McLaughlin et al. (2010) tested learners of French three times across their first year of study. During the first session, learners demonstrated N400 effects to subject-verb agreement (*Tu*$_{sg}$ *adores*$_{sg}$/**adorez*$_{pl}$), despite the fact that their L1 English also instantiates subject-verb agreement. Based on previous research with pseudowords and beginning learners, these researchers interpret the N400 response here as indexing learners' sensitivity to probabilistic dependencies between morphemes, i.e., "the probability of occurrence of particular pronoun-verb ending combinations" (McLaughlin et al., 2010: 141). By the third session, however, the N400 effect had been replaced by a native-like P600, indicating the grammaticalization of a generalized rule regarding morphosyntactic dependencies. Interestingly, grand averages for the second session showed no grammaticality effects. However, further analysis revealed that the lack of effects was due to averaging a subset of learners who demonstrated N400 effects with another subset that had produced P600 responses. This phenomenon is corroborated by a cross-sectional study that recruited L2 learners in the first and third years of German (Tanner et al., 2009). The first-year learners in this study were further grouped into "quick" and "slow" groups according to scores on a grammaticality judgment task, with the "slow" learners performing at chance. Stimuli included a condition involving grammatical and ungrammatical versions of Subject-Verb agreement (*Ich*$_{1sg}$ *trinke*$_{1sg}$/**trinkt*$_{3sg}$). In this condition, the third-year students demonstrated a native-like P600, although with a broader distribution than that of the native speakers. Critically, the "quick" learners in the first-year group demonstrated a developing P600, while the "slow" learners exhibited an N400 response. The results of these studies demonstrate that group averaging in ERP studies can mask P600 development in a subset of learners. At the same time, these studies indicate an as-yet-uninvestigated role for individual differences between learners in ERP responses. The current study examines factors that may explain individual variability in the P600 response by testing learners using both verbal and nonverbal aptitude measures.

2. Methods

Participants. Our participants included 24 native English speakers enrolled in fourth-semester Spanish classes at the University of Kansas. A questionnaire confirmed that no learners had substantial exposure to Spanish prior to the age of nine, nor had any of them been exposed to other Romance languages prior to their study of Spanish. Learners were tested for Spanish proficiency using a written, 50-item test that included the vocabulary section of the MLA Cooperative Foreign Language test (Educational Testing Service, Princeton, N.J.) and a cloze section from the Diploma de Español como Lengua Extranjera (DELE) test (Spanish Embassy, Washington D.C.). This test was chosen because it has previously been used in several other studies of number and gender agreement in L2 learners and heritage speakers (White et al., 2004; McCarthy, 2008; Montrul et al., 2008). All learners in the current study scored in the low proficiency range (Mean: 19; Range: 13-27). A control group of 8 native Spanish speakers also participated in the study, none of whom had been exposed to other languages before primary school. All participants were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971), had normal or corrected-to-normal vision, and reported no neurological impairment. All participants were financially compensated for their time.

Materials. A grammaticality judgment task was administered to all participants during EEG recording. A total of 240 test sentences were constructed, including 120 in each of two sets of stimuli. In the first set, which was adapted from Alemán Bañón (2010), the grammatical version of the sentence contained an adjectival predicate agreeing in number and gender with the subject of the sentence, which was always singular (see example in (1)). The ungrammatical versions varied with regard to the morphological suffixes on the adjectives in order to provide a mismatch in either gender or number between the subject noun and the adjective in the predicate.

(1) La isla es rocosa/*rocosas/*rocoso y la península también.
The$_{sg.fem}$ island$_{sg.fem}$ is$_{3sg}$ rocky$_{sg.fem/*pl.fem/*sg.masc}$ and the peninsula too.
'The island is rocky and the peninsula too.'

All sentences were presented in the present tense. With regard to gender, approximately half of the sentence subjects were feminine and the rest were masculine. Only nouns displaying canonical gender markings were used in the subject noun phrases, and all were inanimate so as to avoid bias toward natural gender. None of the sentences contained violations of both number and gender agreement.

The second set of stimuli consisted of both grammatical and ungrammatical sentences involving subject-verb (S-V) agreement, along with a filler formed by replacing the lexical verb with a copula:

(2) El barco vacío flota/*flotan/está en la laguna.
The$_{3sg}$ boat$_{3sg}$ empty$_{3sg}$ float$_{3sg/*3pl}$ /be$_{3sg}$ in the lagoon.
'The empty boat floats/is in the lagoon.'

The 240 sentence sets (120 N-Adj, 120 S-V and fillers) were counterbalanced across three lists such that participants read 40 tokens of each condition and never saw more than one version of each target item. Target sentences were mixed with fillers and presented in random order. Critically, all three types of agreement (N-Adj NUM, N-Adj GEN, and S-V) must be calculated across a phrase boundary, thus all three conditions were controlled with respect to structural distance. By including an adjective in the subject DP for the subject-verb stimuli (as in (2) above), the stimuli were also matched across experiments in terms of the linear distance between the subject and target word (one intervening word). Across both experiments, critical words (subjects, adjectives, and verbs) were followed by additional material in order to avoid interference from end-of-sentence ERP effects. In an attempt to eliminate any effects of unfamiliar vocabulary, the nouns, adjectives, and verbs used across all conditions were chosen either because they appear in textbooks or because they are cognates with English. In a further effort to ensure that learners had been familiar with these words, all adjectives and verbs appearing in critical regions of the sentences were included in a computerized vocabulary recognition task completed by learners after the experimental task was done. The EEG recording and grammaticality judgment task were also followed by an offline gender assignment task that included the 60 nouns that appeared as subjects in the stimuli targeting N-Adj Gender agreement.

Importantly, our stimuli allow us to test the role of the L1 in the processing of L2 agreement. With regard to number agreement on verbs versus adjectives, both the *Interpretability Hypothesis* (Tsimpli and Mastropavlou, 2007; Tsimpli and Dimitrakopoulou, 2007) and *Full Transfer/Full Access* (Schwartz and Sprouse, 1994, 1996) predict learners will exhibit P600 effects in response to violations of number agreement in the S-V and N-Adj conditions, regardless of differences in how the number feature is realized. However, the results of previous ERP studies (e.g., Tokowicz and MacWhinney, 2005; Frenck-Mestre et al., 2009) seem to suggest that no P600 will be found in the N-Adj condition, where the L1 does not instantiate number agreement. With regard to gender agreement, the L2 theories under investigation make different predictions. The *Interpretability Hypothesis* predicts that learners whose L1 English does not instantiate gender agreement will not be able to acquire the uninterpretable gender feature on Spanish adjectives. Thus, learners are expected to diverge from native speakers in that learners will not demonstrate native-like P600 responses to gender agreement violations. It should be noted that if learners attempt to establish associations between nouns and adjectives based on orthographic regularities in the input, they may demonstrate sensitivity in behavioral results and in ERP responses resulting in an N400 effect, as predicted in initial stages of learning by McLaughlin et al. (2010) and similar to what was found by Barber and Carreiras (2005) for native speakers in word-pair conditions where associations were tested without reference to a syntactic context. On the other hand, *Full Transfer/Full Access* allows that learners could demonstrate sensitivity to gender agreement. Under this hypothesis, gender should ultimately be acquirable, and sensitivity to gender agreement should be largely a function of proficiency. In this case, given the low proficiency of the learners being tested, it is still possible to find no grammaticality effects or even N400 effects, but a P600 is possible.

Aptitude Tests. In addition to investigating the role of the L1 in the L2 processing of number and gender agreement, we examine individual differences between learners, using both verbal and nonverbal aptitude tests. Verbal aptitude was measured by Short Version of the Modern Language Aptitude Test (MLAT) (Carroll and Sapon, 1959). The first subtest (*MLAT3: Spelling Clues*) is a multiple-choice test that asks participants to select appropriate synonyms of words that are spelled as they are pronounced rather than by conventional orthography (e.g., an appropriate synonym for *ritn* might be 'printed', since *ritn* can represent the pronunciation of *written*). This test is designed to assess phonetic coding ability and memory for vocabulary. There are 50 items in this section, but only 5 minutes are allowed. The *MLAT4: Words in Sentences* is a test of grammatical sensitivity or a broader language analytic ability (Skehan, 1998), including sensitivity to grammatical roles and the ability to make analogies at a grammatical level. Participants must select from among several underlined words in each test sentence the one choice that functions in the same way as the underlined word in an example sentence. For example, if the subject is underlined in the example sentence, the participant's correct response would also indicate the subject of the test sentence. Participants are allowed 20 minutes to complete 45 items. The *MLAT5: Paired Associates* requires the rapid learning of a list of 24 vocabulary words in an adapted language, along with their associated English meanings. Following a 2-minute period of vocabulary memorization and a subsequent 2-minute practice period, participants have 4 minutes to choose the correct meaning from multiple choices given for each of the 24 items. The MLAT5 indexes associative memory and requires the storage and retrieval of a large amount of material in a short period of time.

In order to contrast the contributions of skills that are domain-specific to language with those that are more domain-general, a test of nonverbal aptitude was also conducted. The Raven's Advanced Progressive Matrices (RAVEN) (Raven, 1965) is a multiple-choice test of nonverbal intelligence and reasoning skills, including the ability to decompose complex problems, search for rules and manage those rules in working memory (Carpenter, Just, and Shell, 1990). In a multidimensional scaling analysis by Snow, Kyllonen, and Marshalek (1984), where a wide variety of domain-general and domain-specific tests were placed in concentric circles representing the closeness of correlations among the tests, the RAVEN occupied a central position, demonstrating that it is an optimal test for measuring domain-general reasoning. Each of the 12 practice items in Set I of the RAVEN and the 36 test items in Set II consists of a visual array of patterns with one item missing, followed by eight choices of patterns to fill the missing item. The test set can be administered as timed (usually for 40 minutes) or untimed; here it was administered in a fairly short amount of time due to time constraints and the possibility of participant fatigue during the testing session. The time allowed for the practice set was five minutes, and twenty minutes were allowed for the test set[1].

Procedures. Participants in the learner group attended two experimental sessions for a total of approximately five hours. During the first session, they provided informed consent for the entire study and completed a language background survey. The MLAT, RAVEN, and proficiency tests were then administered in pseudorandomized order, with breaks between each test. The second session included completing the Edinburgh Handedness Inventory (Oldfield, 1971) in order to verify right-handedness. Participants were then fitted with an electrode cap and additional electrodes above, below, and to the outside of each eye and behind each ear. Participants were seated comfortably in front of a CRT monitor in a dimly-lit, sound-attenuated experiment room. The Paradigm experimental control system designed by Perception Research Systems, Inc. (Tagliaferri, 2005) was employed for randomized stimulus presentation. In order to familiarize the participant with the task, each recording session began with a practice session of nine trials, which included items targeting areas of grammar not

[1] The decision to shorten the amount of time allowed for the RAVEN was made after piloting the test on 13 native speakers of English recruited for extra credit in a linguistics class. Recorded data included the numbers of test items in Set II completed at 20 minutes (Mean=23.82, SD=4.21) and at 30 minutes (Mean=26.18, SD=5.00). While a paired-samples t-test did show that the scores after 30 minutes were significantly higher than the scores after 20 minutes, $t(12)=.580$, $p<.001$, there was a high level of correlation between the two sets of scores ($r=.94$).

investigated in the experiment. No words appeared in both the practice and experimental items. Participants received feedback after each of the first three practice items. For all practice and experimental items, each trial was preceded by a blank screen for 500ms, allowing the participant time to blink, followed by a fixation cross for 500ms and then a 300-ms pause, after which the stimulus sentence was presented word-by-word in an RSVP (Rapid Serial Visual Presentation) paradigm. Each word appeared in black text on a dark gray screen for 450ms, with a pause between each word lasting 300ms. A pause of 1000ms followed the final word of the sentence before a response prompt appeared on the screen, asking participants to indicate whether the sentence was "Bien" (good) or "Mal" (bad). Participants received instructions prior to the test to respond by button push for "Bien" if they felt the Spanish sentence was grammatical and "Mal" if they felt it was ungrammatical. After every 40 trials, participants were prompted to take a break. The task took participants about an hour to complete, including breaks.

Following the experimental task, participants were given a break, followed by a short computerized test that included both the gender assignment and the vocabulary recognition tasks. Instructions and six practice items immediately preceded each task. No feedback was provided for the practice items. For the gender assignment task, each trial was preceded by a 300-ms pause and 500-ms fixation cross. In order to mimic the conditions of the grammaticality judgment task, each Spanish noun was presented for 450ms in black text on a dark gray screen. Presentation of the noun was followed by a prompt that presented the masculine determiner "El" on the left of the screen and the feminine determiner "La" on the right. Participants responded by mouse click in the area of the determiner of choice. For each item in the vocabulary recognition task, a Spanish word was presented in lower-case letters, with two possible English translations for that word in capital letters below it. Participants were asked to use the mouse to select the appropriate translation for each item. A 1000-ms pause occurred between each trial. Participants finished these computerized tasks in approximately 10 minutes. The total time for this session was around 2½ hours.

The native speakers in this study were not required to take the aptitude and proficiency tests, so they attended only one session. At the beginning of this session, they provided informed consent and completed a background questionnaire and the Edinburgh Handedness Inventory (Oldfield, 1971), followed by the experimental task and the gender assignment task, which were administered in the same way as described for learners above.

Data Acquisition. Continuous EEG was recorded using an electrode cap (Electro-Cap International, Inc.) containing 32 Ag/AgCl scalp electrodes arrayed in a modified 10-20 layout (midline: FPZ, FZ, FCZ, CZ, CPZ, PZ, OZ; lateral: F7/8, F3/4, FT7/8, FC3/4, T3/4, C3/4, TP7/8, CP3/4, T5/6, P3/4, O1/2). Additional electrodes were placed on the left and right outer canthus to monitor eye-movements and above and below each eye to monitor eye-blinks. Electrode AFZ serves as the ground, and an electrode at the left mastoid provides an online reference. Impedances for each scalp electrode are kept below 5 kOhms. The recordings are amplified by a Neuroscan Synamps2 amplifier (Compumedics Neuroscan, Inc.) with an online bandpass filter of 0.1 to 200 Hz and digitized at a sampling rate of 1000 Hz. Data were re-referenced offline to the linked mastoid electrodes.

Pre-processing of the EEG data included the rejection of trials characterized by excessive muscle artifacts or eye movements and blinks, resulting in 81% data retention in the learner group and 84% for native speakers. Event-related potentials, time-locked to the onset of the target word in each sentence, were averaged off-line for each subject at each electrode site. A 200ms pre-stimulus interval was included in each epoch, which extended to 1000ms post-stimulus, as determined by visual inspection of the grand average waveform (averaged across participants). Following artifact rejection, baseline correction, and averaging, data was filtered using a 30 Hz low-pass filter.

3. Results

The mean acceptance rates for native speakers and learners are presented in Figure 1 for each condition [grammatical Subject-Verb (S-V), ungrammatical Subject-Verb (*S-V), grammatical N-Adj (N-Adj), ungrammatical N-Adj with respect to number (*N-Adj NUM), ungrammatical N-Adj with respect to gender (*N-Adj GEN)].

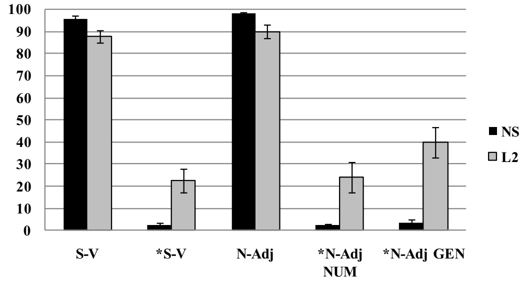

Figure 1. **Mean acceptance rates for native speakers (NS) and learners (L2) across conditions**

A repeated-measures analysis of variance (ANOVA) was conducted separately for each ungrammatical condition against its grammatical counterpart. Grammaticality effects were found for all three types of agreement (S-V: $F(1,30)=192.484$, $p<.001$; N-Adj Number: $F(1,30)=149.916$, $p<.001$; N-Adj Gender: $F(1,30)=130.436$, $p<.001$). A significant Group effect was present only for N-Adj Gender agreement ($F(1,30)=7.146$, $p=.012$). For all three types of agreement, a significant Group x Grammaticality interaction was present (S-V: $F(1,30)=6.148$, $p=.019$; N-Adj Number: $F(1,30)=5.194$, $p=.030$; N-Adj Gender: $F(1,30)=12.487$, $p=.001$), but paired-samples *t*-tests confirmed grammaticality effects (*p* values <.001) for both groups for all three types of agreement. In summary, both groups generally demonstrated the same pattern of high acceptance rates for grammatical sentences and significantly lower acceptance rates for ungrammatical sentences across all three types of agreement. Although learner acceptance rates for ungrammatical sentences were higher than those of the native speakers, this difference was only significant between groups for gender agreement.

ERP Results. Mean averages of amplitude were calculated for each participant for each condition across four regions of electrodes (Left Anterior, Right Anterior, Left Posterior, Right Posterior). Each ungrammatical condition was compared to its grammatical counterpart using repeated-measures Analysis of Variance (ANOVA), including the between-subjects factor Group and within-subjects factors of Grammaticality (Grammatical, Ungrammatical), Hemisphere (Left, Right), and Anteriority (Anterior, Posterior). Greenhouse-Geisser corrected *p*-values will be reported in all cases due to lack of homogeneity of variance. Significant effects were followed up by adjusted paired-samples *t*-tests as needed. Here, results will be reported only for the overall P600 time window of 450-950ms, which was chosen based on previous literature and visual inspection of the waveforms for native speakers.

ERP responses to Subject-Verb agreement are plotted for representative electrodes in Figure 2 for both native speakers and learners. A Grammaticality effect was present in the 450-950ms time window, $F(1,30)=6.202$, $p=.019$, along with a Grammaticality x Hemisphere interaction, $F(1,30)=8.212$, $p=.008$, and a marginal Grammaticality x Anteriority interaction, $F(1,30)=3.803$, $p=.061$, reflecting greater positivities in response to ungrammatical versus grammatical stimuli in the

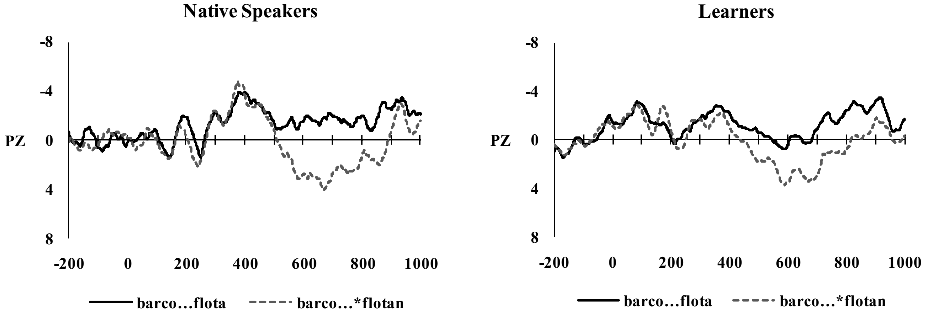

Figure 2. **Grand average ERPs for native speakers (left) and learners (right) in response to grammatical (solid line) and ungrammatical (dashed line) Subject-Verb agreement at representative electrodes.**

posterior regions, with a larger difference in the right versus the left hemisphere. There was a marginal effect for Group, $F(1,30)=3.854$, $p=.059$, but no interactions involving Group (p values >.05).

A Grammaticality effect was also present for N-Adj Number agreement, $F(1,30)=9.127$, $p=.005$, as well as a Grammaticality x Anteriority interaction, $F(1,30)=6.411$, $p=.017$, and a marginal Grammaticality x Hemisphere interaction, $F(1,30)=3.451$, $p=.073$, both consistent with greater positivities in response to ungrammatical versus grammatical stimuli in the posterior regions, with a larger difference in the right versus the left hemisphere. There was no effect for Group, $F(1,30)=.892$, $p=.353$, nor were there any Group interactions (p values >.05). Figure 3 shows ERPs in response to N-Adj Number agreement at representative electrodes for both native speakers and learners.

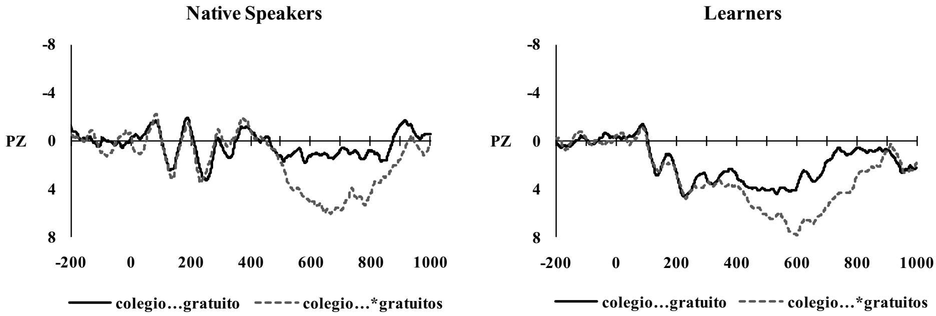

***Figure 3.* Grand average ERPs for native speakers (left) and learners (right) in response to grammatical (solid line) and ungrammatical (dashed line) N-Adj Number agreement at representative electrodes.**

Finally, for N-Adj Gender agreement, there was also a Grammaticality effect, $F(1,30)=4.417$, $p=.044$, and a Grammaticality x Hemisphere interaction, $F(1,30)=4.223$, $p=.049$, along with a marginal Grammaticality x Anteriority interaction, $F(1,30)=3.510$, $p=.071$. There was no effect for Group, $F(1,30)=.209$, $p=.651$, nor any interaction of Group and Grammaticality, $F(1,30)=1.642$, $p=.210$, indicating that both groups were also sensitive to violations of N-Adj Gender agreement. However, there was a Group x Grammaticality x Hemisphere interaction, $F(1,30)=5.518$, $p=.026$. Native speakers demonstrated a greater positivity to ungrammatical stimuli only in the right posterior region, $t(7)=2.431$, $p=.045$, while the greater positivity to ungrammatical stimuli was evident in the left posterior region for the learner group, $t(23)=2.437$, $p=.023$. Representative electrodes for each group with regard to N-Adj Gender agreement are plotted in Figure 4.

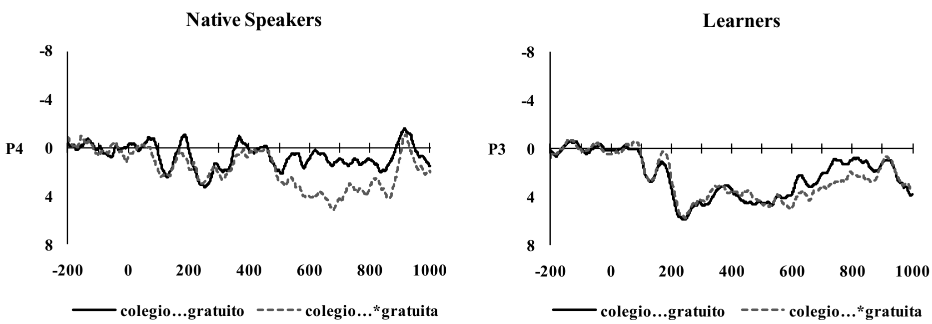

***Figure 4.* Grand average ERPs for native speakers (left) and learners (right) in response to grammatical (solid line) and ungrammatical (dashed line) N-Adj Gender agreement at representative electrodes.**

It was not statistically expedient to include a within-subjects factor of Agreement Type in the previous analysis, since two types of agreement (N-Adj NUM, N-Adj GEN) rely on comparison with the same grammatical stimuli. Accordingly, a separate statistical test was performed for Agreement Type using the effect sizes (mean amplitudes for ungrammatical minus grammatical conditions) of the components in each time window and a between-subjects factor of Group. This analysis was performed separately for two comparisons of interest. First, the two conditions involving number agreement (Subject-Verb, N-Adj Number) were compared. No effects of Agreement Type or Group were present, nor any Group x Agreement Type interaction (p values >.05). Next, N-Adj Number and N-Adj Gender agreement were compared. No Group effect or interaction was present (p values >.05), but there was an effect of Agreement Type, $F(1,30)=4.262$, $p=.048$, indicating that responses to ungrammatical versus grammatical stimuli for N-Adj Number agreement were significantly more positive than for N-Adj Gender agreement. Follow-up paired-samples t-tests indicate a significant difference between N-Adj Number and N-Adj Gender agreement for learners, $t(23)=3.249$, $p=.004$, but not native speakers, $t(23)=.560$, $p=.593$.

Overall, learners demonstrated similar responses to native speakers with regard to all three types of agreement. Number agreement violations on adjectives produced greater P600 responses than violations of gender agreement, but learners demonstrated sensitivity to the gender feature even though it is unique to the L2. No differences were found between the two groups for either of the conditions involving number agreement.

Correlations with aptitude and proficiency measures. Aptitude scores were recorded for the Raven Advanced Progressive Matrices (RAVEN), the total score on the Modern Languages Aptitude Test (MLAT Total), as well as each of its subtests (MLAT 3: Spelling Clues; MLAT4: Words in Sentences; MLAT 5: Paired Associates). The mean scores, standard deviations, and range for each measure are presented in Table 1.

Table 1. Observed measures of aptitude and proficiency, along with the maximum possible scores for each measure

Test	Mean	SD	Range	Max
MLAT3	19.25	5.97	8-30	out of 50
MLAT4	21.96	5.64	14-35	out of 45
MLAT5	20.21	4.48	9-24	out of 24
MLAT Total	61.42	9.67	40-83	out of 119
RAVEN	21.92	2.98	16-28	out of 36

Pearson correlation coefficients and their p-values were then calculated in order to assess the relationships between these independent variables and sensitivity to violations of each type of agreement (S-V, N-Adj NUM, N-Adj GEN), both in terms of grammaticality judgments and ERP responses in the P600 time window. For grammaticality judgments, sensitivity to violations was measured by d' scores, which take into account acceptance rates for both the grammatical and ungrammatical conditions. For this analysis[2], a d' score near zero represents performance at chance, while perfect performance in this analysis results in a d' score of approximately 4.0.

Pearson correlation coefficients for the relationships between d' scores and aptitude tests are summarized here and presented in Table 2 below. Verbal aptitude as measured by the MLAT Total score was found to be correlated only with sensitivity to violations of Subject-Verb agreement, $r=.446$, $p=.029$. This correlation was driven by a significant correlation with the MLAT4 Words in Sentences, $r=.466$, $p=.022$, and a marginal correlation with the MLAT5 Paired Associates, $r=.400$, $p=.053$. These moderate correlations were present to the same degree as has been found in other measures of L2 outcomes, as reported by Carroll (1981). A marginal correlation between the MLAT4 and d' scores for

[2] The formula used here is specific to forced-choice experimental paradigms, where H = the Hit rate for choosing Grammatical when the stimulus is grammatical, and FA = the False Alarm rate for choosing Grammatical when the stimulus is ungrammatical: NORMSINV(H)-NORMSINV(FA))/SQRT(2). Following standard procedures, Hit rates of 1 (corresponding to 100% acceptance rates) are corrected to 1 – 1/480, where 480 represents twice the number of items in the test. False Alarm rates of zero are corrected to 1/480.

N-Adj Number agreement was also present, *r*=.384, *p*=.064. Nonverbal aptitude, as measured by the Raven Advanced Progressive Matrices, was not found to be significantly correlated with any behavioral measure of L2 sensitivity to agreement violations (*p* values >.575).

Table 2. Correlations between d' scores and measures of verbal (MLAT) and nonverbal (Raven) aptitude

		S-V	N-Adj NUM	N-Adj GEN
MLAT3	*r*	-.018	-.132	-.055
	(*p*)	(.933)	(.539)	(.799)
MLAT4	*r*	**.466***	**.384†**	.298
	(*p*)	**(.022)**	**(.064)**	(.157)
MLAT5	*r*	**.400†**	.177	.115
	(*p*)	**(.053)**	(.407)	(.593)
MLAT Total	*r*	**.446***	.225	.193
	(*p*)	**(.029)**	(.291)	(.366)
Raven	*r*	.121	.042	.021
	(*p*)	(.575)	(.844)	(.922)

* Correlation is significant at the *p*<.05 level.
† Correlation is marginally significant (.10>*p*>.05).

A similar picture emerges when effect sizes of P600 responses are considered. The MLAT Total score was found to be correlated only with mean amplitude differences in P600 responses to N-Adj Number agreement, *r*=.538, *p*=.007. This correlation was driven by a significant correlation with the MLAT4 Words in Sentences, *r*=.418, *p*=.042, and a marginal correlation with the MLAT5 Paired Associates, *r*=.352, *p*=.092. Raven scores were not significantly correlated with mean amplitude differences in the P600 response to any type of agreement violations (*p* values >.237). Table 3 presents Pearson correlation coefficients representing the relationships between mean amplitude differences in the P600 time window and the MLAT and its subtests, as well as the Raven.

Table 3. Correlations between aptitude measures and mean amplitude differences from 450-950ms in the posterior regions for each violation type

		S-V	N-Adj NUM	N-Adj GEN
MLAT3	*r*	-.115	.213	.251
	(*p*)	(.594)	(.317)	(.238)
MLAT4	*r*	.101	**.418***	.001
	(*p*)	(.640)	**(.042)**	(.996)
MLAT5	*r*	.043	**.352†**	.244
	(*p*)	(.841)	**(.092)**	(.251)
MLAT Total	*r*	.008	**.538***	.268
	(*p*)	(.971)	**(.007)**	(.205)
Raven	*r*	-.117	-.016	-.251
	(*p*)	(.586)	(.943)	(.237)

* Correlation is significant at the *p*<.05 level.
† Correlation is marginally significant (.10>*p*>.05).

Overall, the above analysis uncovered significant positive correlations between verbal aptitude and measures of sensitivity to violations of number agreement, which is also present in the L1 of the participants. The subtest of the MLAT that seemed most related to d' scores was the MLAT4 Words in Sentences, which tests language analytic ability. No correlations were found between verbal aptitude and sensitivity to gender agreement violations, nor was any correlation present for nonverbal aptitude.

4. Conclusion

This study contributes to the growing body of literature investigating the role of the L1 in L2 morphosyntactic processing, providing evidence that adult learners even at low proficiency can exhibit development of native-like processing of (a) uninterpretable features that are not present in their L1, as well as (b) novel instantiations of features that are shared between the L1 and L2, supporting claims of full-access theories (e.g., Schwartz and Sprouse, 1994, 1996). In particular, the study tests number and gender agreement across phrases in well-controlled contexts, allowing the comparison of featural instantiations that vary parametrically in their similarity to the L1. The development of native-like processing was also evident for gender agreement on adjectives, even though grammatical gender agreement is not present in the L1. In contrast to earlier ERP studies (e.g., Tokowicz and MacWhinney, 2005; Frenck-Mestre et al., 2009). Our results showed that learners were sensitive to number agreement violations on both verbs and adjectives, despite differences between these two categories in the realization of number in the L1.

With regard to individual differences between learners, the study does not find any evidence thus far that domain-general aptitude is predictive of success in adult L2 acquisition (cf. Bley-Vroman, 1990). However, correlations were present between aspects of verbal aptitude and both online and offline measures of sensitivity to agreement violations. The finding that verbal aptitude scores can be correlated to specific properties of brain responses to language demonstrates that ERP is a promising methodology for investigating the role of individual differences in L2 processing. It is hoped that the results of this study will prompt further research into the effects of individual differences on brain responses in second language learners, and, ultimately, provide brain-based research that can inform pedagogical approaches to second language acquisition.

References

Abrahamsson, Niclas, & Hyltenstam, Kenneth. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition, 30,* 481-509.

Alemán Bañón, José. (2010). The processing of number and gender agreement in Spanish: An ERP investigation. MA Thesis, University of Kansas, 2010.

Barber, Horacio, & Carreiras, Manuel. (2005). Grammatical gender and number agreement in Spanish: an ERP comparison. *Journal of Cognitive Neuroscience, 17*, 137-153.

Bley-Vroman, Robert. (1989). What is the logical problem of foreign language learning? In S. Gass & J. Schacter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 41-68). Cambridge: Cambridge University Press.

Bley-Vroman, Robert. (1990). The logical problem of foreign language learning. *Linguistic Analysis, 20,* 3-49.

Carroll, John B. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in foreign language aptitude* (pp. 83-118). Rowley, MA: Newbury House.

Carroll, John B., & Sapon, Stanley. (1959). *Modern Language Aptitude Test: Form A.* New York: Psychological Corporation.

Carpenter, Patricia A., Just, Marcel Adam, & Shell, Peter. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97,* 404-431.

Chomsky, Noam. (1965). *Aspects of the Theory of Syntax.* Cambridge, MA: MIT Press.

Chomsky, Noam. (1980). *Rules and Representations*. Oxford: Blackwell.

Chomsky, Noam. (1981). Principles and parameters in syntactic theory. In N. Hornstein & D. Lightfoot (Eds.), *Explanation in linguistics: the logical problem of language acquisition* (pp. 32-75). London: Longman.

DeKeyser, Robert M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22,* 499-533.

Franceschina, Florencia. (2002). Case and (phi)-feature agreement in advanced L2 Spanish grammars. *EUROSLA Yearbook, 2,* 71-86.

Frenck-Mestre, Cheryl, Foucart, Alice, Carrasco, Haydee, & Herschensohn, Julia. (2009). Processing of grammatical gender in French as a first and second language. *EUROSLA Yearbook, 9,* 76-106.

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

Hahne, Anja. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistics Research, 30*, 251–66.

Hahne, Anja, & Friederici, Angela D. (1999). Electrophysiological evidence for two steps in syntactic analysis: early automatic and late controlled processes. *Journal of Cognitive Neuroscience, 11*, 194+.

Harley, Birgit, & Hart, Doug. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition, 19,* 379-400.

Juffs, Alan. 2004. Representation, processing, and working memory in a second language. *Transactions of the Philological Society, 102,* 199–225.

Kaan, Edith, Harris, Anthony, Gibson, Edward, & Holcomb, Phillip. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15,* 159-201.

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

Kutas, Marta, & Hillyard, Steven A. (1980). Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science* 207: 203-205.

Kutas, Marta, & Federmeier, Kara D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science, 4*, 463-470.

Lau, Ellen, Phillips, Colin, & Poeppel, David. (2008). A cortical network for semantics: (de)constructing the N400. *Nature, 9,* 920-933.

McCarthy, Corrine. (2008). Morphological variability in the comprehension of agreement: An argument for representation over computation. *Second Language Research, 24,* 459-486.

McDonald, Janet L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language, 55,* 381-401.

McLaughlin, Judith, Tanner, Darren, Pitkänen, Ilona, Frenck-Mestre, Cheryl, Inoue, Kayo, Valentine, Geoffrey, & Osterhout, Lee. (2010). Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning, 60 (Suppl. 2),* 123-150.

Montrul, Silvina, Foote, Rebecca, & Perpiñán, Silvia. (2008). Gender Agreement in Adult Second Language Learners and Spanish Heritage Speakers: The Effects of Age and Context of Acquisition. *Language Learning, 58*, 503–553.

Ojima, Shiro, Nakata, Hiroki, & Kakigi, Ryusuke. (2005). An ERP study of second language learning after childhood: effects of proficiency. *Journal of Cognitive Neuroscience, 17*, 1212-28.

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

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

Osterhout, Lee, McLaughlin, Judith, Kim, Albert, Greenwald, Ralf, & Inoue, Kayo. (2004). Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. In M. Carreiras & C. Clifton, Jr. (Eds.), *The on-line study of sentence comprehension* (pp. 271–308). New York: Psychology Press.

Raven, John C. (1965). *Advanced Progressive Matrices, Sets I and II.* London: H. K. Lewis. (Distributed in the United States by The Psychological Corporation, San Antonio, TX).

Robinson, Peter. (2002). Effects of individual differences in intelligence, aptitude and working memory on adult incidental SLA. In P. Robinson (Ed.), *Individual Differences and Instructed Language Learning*. (pp. 211-266). Amsterdam: John Benjamins.

Rossi, Sonja, Gugler, Manfred F., Friederici, Angela D., & Hahne, Anja. (2006). The impact of proficiency on syntactic second language processing of German and Italian: evidence from event-related potentials. *Journal of Cognitive Neuroscience, 18*, 2030-2048.

Sabourin, Laura. (2003). Grammatical Gender and Second Language Processing: An ERP study. Ms., *GRODIL: Groningen Dissertations in Linguistics, 42.*

Sagarra, Nuria. (2007). Online processing of gender agreement in low proficient English–Spanish late bilinguals. In M. J. Cabrera, J. Camacho, V. Déprez, N. Flores, L. Sánchez (Eds.), *Current Issues in Linguistic Theory Series* (pp. 240–253). John Benjamins, Amsterdam.

Sagarra, Nuria, & Herschensohn, Julia. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua, 120,* 2022-2039.

Sasaki, Miyuki. (1996). *Second language proficiency, foreign language aptitude, and intelligence: Quantitative and qualitative analyses.* New York: Peter Lang.

Schwartz, Bonnie D., & Sprouse, Rex A. (1994). Word order and nominative case in nonnative language acquisition. In T. Hoekstra & B. Schwartz (Eds.), *Language acquisition studies in generative grammar* (pp. 317–368). Amsterdam: John Benjamins.

Schwartz, Bonnie D., & Sprouse, Rex A. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research, 12,* 40–72.

Skehan, Peter. (1989). *Individual differences in second language learning.* London: Arnold.

Skehan, Peter. (1998). *A Cognitive Approach to Language Learning.* New York: Oxford University Press.

Snow, Richard. E., Kyllonen, Patrick. C., & Marshalek, Brachia. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the Psychology of Human Intelligence, Vol. 2* (pp. 47-103). Hillsdale, NJ: Erlbaum.

Tagliaferri, Bruno. (2005). *Paradigm*. Perception Research Systems, Inc., www.perceptionresearchsystems.com.

Tanner, Darren, Osterhout, Lee, & Herschensohn, Julia. (2009). Snapshots of grammaticalization: Differential electrophysiological responses to grammatical anomalies with increasing L2 exposure. In J. Chandlee, M. Franchini, S. Lord, & G-M. Rheiner (Eds.), *Proceedings of the 33rd Boston University conference on language development (pp. 528–539). Somerville, MA: Cascadilla Press.*

Thoma, Robert J., Lewine, Jeffrey D., Davis, John T., & Paulson, Kim. (2000). Duration of parietal MEG source dipole correlated with Ravens advanced progressive matrices score. *Journal of Cognitive Neuroscience, (Suppl. S),* 86–87.

Thompson, Amy S. (2009). The multilingual/bilingual dichotomy: An exploration of individual differences. PhD Dissertation, Michigan State University.

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

Tsimpli, Ianthi Maria, & Dimitrakopoulou, Maria. (2007). The interpretability hypothesis: Evidence from *wh*-interrogatives in second language acquisition. *Second Language Research, 23,* 215-242.

Tsimpli, Ianthi Maria, & Mastropavlou, Maria. (2007). Feature interpretability in L2 acquisition and SLI: Greek clitics and determiners. In J. Liceras, H. Zobl, & H. Goodluck, (Eds.), *The Role of Formal Features in Second Language Acquisition*. Mahwah, NJ: Lawrence Erlbaum.

Weber-Fox, Christine M., & Neville, Helen J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience, 8*, 231-56.

White, Lydia, Valenzuela, Elena, Kozlowska-Macgregor, Martyna, & Leung, Yan-Kit Ingrid. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics, 25,* 105–33.

Ye, Zheng., Luo, Yue-jia, Friederici, Angela D., & Zhou, Xiaolin. (2006). Semantic and syntactic processing in Chinese sentence comprehension: evidence from event-related potentials. *Brain Research, 1071*, 186-196.

Proceedings of the 11th Generative Approaches to Second Language Acquisition Conference (GASLA 2011)

edited by Julia Herschensohn and Darren Tanner

Cascadilla Proceedings Project Somerville, MA 2011

Copyright information

ISBN 978-1-57473-445-4 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Bond, Kristi, Alison Gabriele, Robert Fiorentino, and José Alemán Bañón. 2011. Individual Differences and the Role of the L1 in L2 Processing: An ERP Investigation. In *Proceedings of the 11th Generative Approaches to Second Language Acquisition Conference (GASLA 2011)*, ed. Julia Herschensohn and Darren Tanner, 17-29. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2542.