SSLA, **27**, 173–204. Printed in the United States of America.
DOI: 10.1017/S0272263105050102

IMPLICIT AND EXPLICIT MEASURES OF SENSITIVITY TO VIOLATIONS IN SECOND LANGUAGE GRAMMAR

An Event-Related Potential Investigation

Natasha Tokowicz
University of Pittsburgh

Brian MacWhinney
Carnegie Mellon University

We used event-related brain potentials (ERPs) to investigate the contributions of explicit and implicit processes during second language (L2) sentence comprehension. We used a L2 grammaticality judgment task (GJT) to test 20 native English speakers enrolled in the first four semesters of Spanish while recording both accuracy and ERP data. Because end-of-sentence grammaticality judgments are open to conscious inspection, we reasoned that they can be influenced by strategic processes that reflect on formal rules and therefore reflect primarily offline explicit processing. On the other hand, because ERPs are a direct reflection of online processing, they reflect

This research was supported by a National Institutes of Health Individual National Research Service Award (NIH HD42948-01) awarded to Natasha Tokowicz and a National Institutes of Health Institutional National Research Service Award (T32 MH19102) awarded to Brian MacWhinney. We thank Beatrice DeAngelis, Dayne Grove, Kwan Hansongkitpong, Katie Keil, Lee Osterhout, Chuck Perfetti, Kelley Sacco, Alex Waid, and Eddie Wlotko for their assistance with this project. We gratefully acknowledge the comments of Rod Ellis, Jan Hulstijn, Albert Valdman, and the two anonymous *SSLA* reviewers on earlier versions of this manuscript. A portion of these results was presented at the 43rd Annual Meeting of the Psychonomic Society (2002, November).

Address correspondence to Natasha Tokowicz, Department of Psychology and Learning Research and Development Center, 3939 O'Hara St., University of Pittsburgh, Pittsburgh, PA 15260; e-mail: tokowicz@pitt.edu; or Brian MacWhinney, Department of Psychology, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213; e-mail: macw@mac.com.

automatic, nonreflective, implicit responses to stimuli (Osterhout, Bersick, & McLaughlin, 1997; Rugg et al., 1998; Tachibana et al., 1999).

We used a version of the GJT adapted for the ERP environment. The experimental sentences varied the form of three different syntactic constructions: (a) tense-marking, which is formed similarly in the first language (L1) and the L2; (b) determiner number agreement, which is formed differently in the L1 and the L2; and (c) determiner gender agreement, which is unique to the L2. We examined ERP responses during a time period between 500 and 900 ms following the onset of the critical (violation or matched control) word in the sentence because extensive past research has shown that grammatical violations elicit a positive-going deflection in the ERP waveform during this period (e.g., the "P600"; Osterhout & Holcomb, 1992).

We found that learners were sensitive (i.e., showed different brain responses to grammatical and ungrammatical sentences) to violations in L2 for constructions that are formed similarly in the L1 and the L2, but were not sensitive to violations for constructions that differ in the L1 and the L2. Critically, a robust grammaticality effect was found in the ERP data for the construction that was unique to the L2, suggesting that the learners were implicitly sensitive to these violations. Judgment accuracy was near chance for all constructions. These findings suggest that learners are able to implicitly process some aspects of L2 syntax even in early stages of learning but that this knowledge depends on the similarity between the L1 and the L2. Furthermore, there is a divergence between explicit and implicit measures of L2 learning, which might be due to the behavioral task demands (e.g., McLaughlin, Osterhout, & Kim, 2004). We conclude that comparing ERP and behavioral data could provide a sensitive method for measuring implicit processing.

Do adult second language (L2) learners process their new language in a native-like way? There is significant debate regarding this issue. Some researchers (e.g., DeKeyser, 2000) believe that adults rely exclusively on explicit knowledge and explicit processing to comprehend sentences in L2. According to this view, the adult L2 learner must use explicit knowledge and processing to speak and comprehend the L2. An alternative view (N. Ellis, 2002; Krashen, 1994) holds that although L2 learners might be exposed to explicit rules in classrooms and textbooks, they rely on implicit knowledge and implicit processing to comprehend sentences in the L2.

Hulstijn (2002) suggested that one way to address this issue would be to measure readers' immediate online neuronal reactions to L2 sentences, using event-related potentials (ERPs). ERPs are electrophysiological brain responses to particular stimulus events (e.g., reading a word) that are derived from the electroencephalographic (EEG) record. The EEG is recorded by electrodes placed on the surface of the scalp. ERPs reflect the synchronous depolarization of populations of neurons in the brain (Fabiani, Gratton, & Coles, 2000) and indicate the brain response that is linked in time either to the presentation of a stimulus (as in the present study) or to a response (e.g., a button press).

As an example, if a literate individual views a word on a computer screen, the viewing of this word will elicit visual processing (which will be indexed by early sensory ERP components for the first 100 ms) and, later, the word's meaning will become activated, which will be indexed by a later, cognitive ERP component. Specific ERP components can be considered indices of specific cognitive events (Fabiani et al., 2000). In particular, an ERP component has been identified that corresponds to syntactic anomalies. This component, termed the P600, has been used with great success to study the degree to which individuals are sensitive to syntactic anomalies (e.g., Osterhout & Holcomb, 1992).

In the current study, we follow Hulstijn's (2002) suggestion by examining ERP data from beginning learners of Spanish as they are engaged in a grammaticality judgment task (GJT). Our findings indicate that beginning L2 learners are implicitly sensitive to grammatical violations in L2, as evidenced by a positive-going brain response that occurs after a grammatical violation. We will refer to a more positive-going brain response to ungrammatical sentences relative to grammatical sentences as indicating *sensitivity* to grammatical acceptability. Moreover, we find that the strength of this implicit processing effect depends on the similarity between the first language (L1) and the L2.

CROSSLANGUAGE DIFFERENCES IN SECOND LANGUAGE LEARNING

When adults attempt to learn a new language, they start with an already-established grammatical system, replete with well-articulated concepts and labels for those concepts. Unlike child language learners, adults are able to transfer large segments of their L1 over to the new L2 (MacWhinney, in press). Not all transfer from the L1 to the L2 is negative. When the two languages are similar, positive transfer will assist learning. However, crosslanguage mismatches might hinder the acquisition of L2 in two ways. First, crosslanguage mismatches can impede the process of learning by leading learners to entertain false hypotheses. For example, learners might erroneously transfer surface cues such as word order or agreement marking as well as deeper structures such as the shape of grammatical classes. Learners eventually revise these L1-like structures to more closely match those appropriate to the L2 (e.g., Zhang, 1995). However, because areas of mismatch coexist with related areas of correct matching, learners often have problems controlling transfer from the L1 to the L2.

Interactive activation models such as the competition model (MacWhinney, in press) typically view transfer as arising from the fact that both the L1 and the L2 remain potentially active during online processing (Frenck-Mestre, in press; Kroll & Tokowicz, in press). When the L1 and L2 provide contrasting interpretations of a given structure, the stronger L1 patterns will typically dominate. In comprehension, this means that learners will attempt to understand L2 information in terms of L1 structures, such as word order patterns or agreement structures (McDonald, 1987). In production, this means that learners will produce sentences in L2 that have a L1 syntactic accent.

Although transfer and competition pose similar challenges to the L2 learner, they have different consequences; transfer from the L1 to the L2 would cause an initial problem that should be resolved as L2 information is learned, whereas online competition between languages is a more pervasive problem that is likely to persist even in later stages of language learning, returning at times when the language system is taxed or processing resources are limited. Eventually, proficient bilinguals must learn to modulate this competition to use the L2 effectively.

THE PRESENT STUDY

The present study addresses two research questions. First, are L2 learners at beginning stages able to process the L2 implicitly online? Second, to what extent do L1 transfer and competition effects modulate implicit processing? Specifically, we hypothesize that learners will show less implicit sensitivity to grammaticality violations for constructions that differ in the two languages than to violations for constructions that are similar in the two languages. We also predict that learners will show more implicit sensitivity to violations of constructions that are unique to the L2 as opposed to constructions that differ in the L1 and the L2. These predictions follow directly from the fundamental principle of online code interaction outlined in the unified version of the competition model (MacWhinney, in press). This updated version of the model extends the notion of competition between cues and devices within a single language to include the more general competition of cues and devices between languages. The mechanism for determining the outcome of these competitions is still based on the Bayesian summation of the Luce Choice Rule (Luce, 1959).

To test these predictions, we need to look at three types of construction: first, constructions for which the L1 pattern supports and matches the L2 pattern; second, constructions for which the L2 pattern directly conflicts with or competes against the L1 pattern; and, third, constructions that differ between the two languages without any direct competition or mismatch. Focusing on the comparison between L1 English and L2 Spanish, we can locate examples of these three configurations in the systems for auxiliary marking (matching), number agreement (mismatch), and gender agreement (no match). Therefore, we use these three construction types to evaluate our predictions.

Both English and Spanish form the progressive tenses by placing the auxiliary before the participle. If it is true that L1 online processing strategies can transfer to L2, then L2 learners should be sensitive to violations of this structure in both English and Spanish. In English, we would expect sensitivity in response to auxiliary omission in a sentence such as "*His grandmother cooking very well." Similarly, in Spanish, we would expect sensitivity in response to the parallel translation of that sentence: **Su abuela cocinando muy bien*. We expect that both positive transfer and the absence of online competition between languages for this structure would result in good sensitivity to violations of this type in L2.

The situation is somewhat different for structures that do not match across the languages. English makes no grammatical use of nominal gender. However, in Spanish, determiners and adjectives must always agree with the gender of the noun. Learning to apply this system of gender-marking is a challenge for learners of Spanish. Violations of gender agreement in Spanish are not affected by either negative transfer from English or online competition, because English makes no use of gender in sentence processing. As a result, we would expect brain sensitivity to the violation in (1).

(1) **Ellos fueron a un fiesta*
**They went to a*-MASC *party*-FEM
"They went to a party"

In contrast, there is a mismatch between English and Spanish in the formation of determiner number agreement. In English, the same determiner is used with both singular and plural nouns, yielding both "the boy" and "the boys." This is not true for demonstratives such as "this" and "these," which provide information regarding the number of the following noun. Because the determiner provides no such information, we learn to actively suppress any expectation we might have in the case of demonstratives when we read definite articles. In Spanish, on the other hand, the definite article takes different forms in *el niño* "the boy" and *los niños* "the boys." Because English speakers have learned to not pay attention to the number match between the definite article and the noun, we would expect that they would also tend to ignore this information when processing Spanish. Thus, we expect no sensitivity in response to the violation in (2).

(2) **El niños están jugando*
**The*-SING *boys*-PL *are playing*
"The boys are playing"

In effect, we predict that English comprehenders will transfer their L1 processing strategy to L2. Table 1 shows the sample stimuli.

Table 1. Sample stimuli

Construction	Similarity	Example
Spanish		
Auxiliary omission	Similar	*Su abuela *cocinando/cocina muy bien* "His grandmother *cooking/cooks very well"
Determiner gender	Unique	*Ellos fueron a *un/una fiesta* "They went to *a-MASC/a-FEM party"
Determiner number	Different	**El/Los niños están jugando* "*The-SING/the-PL boys are playing"
English		
Subject-verb	N/A	The boys *makes/make excellent ice cream
Reflexive	N/A	The children enjoyed *himself/themselves
Auxiliary omission	N/A	His grandmother *cooking/cooks very well

Note. Critical words are underlined.

Measuring Explicit and Implicit Processing

To examine explicit processing, we asked participants to produce formal grammaticality judgments after the entire sentence was presented. This type of offline grammaticality judgment allows the learner to use explicit knowledge such as the similarity between the two languages, explicit grammar rules, and novelty of the particular syntactic construction to render a judgment. In the terms of the competition model (MacWhinney, 1987), grammaticality violations arise in comprehension when there is a failure in cue match during the construction of a grammatical attachment between a head and its modifiers. For example, there is an item-based rule in Spanish that is linked to the article *el*. This item-based rule seeks to fill a slot for a head in a following nominal cluster. Additionally, this head must be marked for singular number. If it is marked for plurality, then a grammaticality violation is triggered. It is important to note that the competition model views the initial detection of this violation as arising directly from the normal attachment processes during comprehension and not through some external explicit process. However, when these judgments are rendered offline, the initial comprehension-based processes become intermingled with additional explicit processes derived from both reflection and formal grammaticality training (R. Ellis, this volume). Thus, from the viewpoint of the competition model as well as the proposals of Hulstijn (2002), time delays are likely to increase the relative explicitness of any judgment.

To examine implicit processing of L2 syntax, we used ERPs to measure comprehension and attachment processes as they unfold over a very short period of time (less than 800 ms). In particular, we focused our attention on the late positivity in the ERP waveform—the P600—that peaks at approximately 600 ms

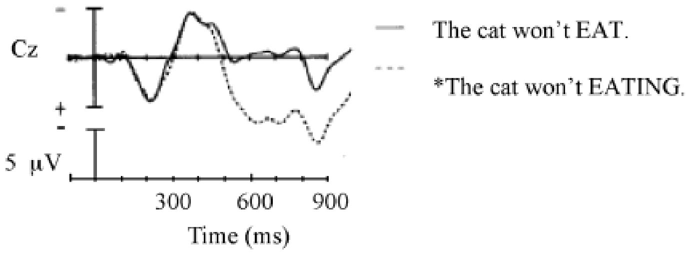

Figure 1. Event-related brain potentials elicited by syntactic anomalies (adapted from Osterhout & Nicol, 1999) recorded from an electrode at the surface of the scalp over the vertex (Cz). Onset of the critical words is indicated by the vertical bar.

poststimulus and is centroparietally distributed (see Figure 1), which is an index of syntactic anomaly (Osterhout & Holcomb, 1992). For example, a P600 can be observed in response to the sentence "*The cat won't eating" (e.g., Osterhout & Nicol, 1999). This ERP reflects online processing of a stimulus. Although there are both early (sensory) and late (cognitive) components of ERPs, all of these components relate to various properties of the stimulus, and none of them involve metacognition, which would take considerably more time. Because ERPs measure implicit processing, researchers who believe that L2 learners use only explicit processing should predict that ERPs from L2 learners would show no sensitivity to grammatical violations. Moreover, they should also predict that learners would show better sensitivity to syntactic violations in offline grammaticality judgments that allow for explicit knowledge to be used than in the online ERP measure.

There is now evidence that directly contradicts these predictions. Osterhout, McLaughlin, Inoue, and Loveless (2000, Experiment 2) have shown that brain responses indicate better comprehension in L2 learners than would be suggested by overt responses obtained from accuracy to offline grammaticality judgments. Overt grammaticality judgments obtained at the ends of sentences showed that, as early as the fourth week of study, L2 learners could not determine the grammaticality of a sentence with better than chance accuracy. However, their covert ERP responses to such syntactic violations suggested that they were sensitive to the violations as comprehension occurred. This sensitivity to grammatical violations was initially observed as an N400 effect (which is typically associated with semantic rather than syntactic anomalies), but later shifted to a P600 effect (after 16 weeks of instruction).

Furthermore, McLaughlin, Osterhout, and Kim (2004) used ERPs to examine L2 word learning in individuals with various amounts of experience with French as the L2 (14, 63, or 138 hours of exposure). While ERPs were recorded, participants viewed prime-target pairs and indicated whether the target item of each pair was a real French word. The stimuli included related word pairs, unrelated word pairs, and word-pseudoword pairs. McLaughlin et al. found that even individuals with only 14 hours of French instruction were sensitive to word versus pseudoword differences (as indicated by ERPs) and that individuals with 63 or 138 hours of exposure were also sensitive to related versus unrelated word differences. Most relevant to the present study is that these brain responses were found in the absence of accurate word versus nonword judgments. Thus, there was a divergence between the implicit measure of ERPs and the explicit measure of overt lexicality judgments: This divergence underscores the use of ERPs as an index of implicit knowledge. McLaughlin et al. stated that "ERPs might more accurately reflect implicit learning and continuous change in knowledge than do explicit categorical judgments. The method used here could be extended to examine the effects of L1-L2 similarity, instructional methods, and learners' age on L2 acquisition" (p. 704).

Using ERPs to Measure Implicit Processing

Event-related potentials have been used extensively to study implicit processing in other domains. For example, Tachibana et al. (1999) considered the N400 repetition effects they observed to be a measure of implicit memory processing. Furthermore, Rugg et al. (1998) demonstrated that ERPs vary with other measures of implicit memory, suggesting that ERPs are a valid measure of implicit processing. Koelsch, Gunter, Schröger, and Friederici (2003) used ERPs as a measure of implicit knowledge of musical regularities in nonmusicians. Morris, Squires, Taber, and Lodge (2003) used ERP components to measure implicit social attitudes. This large body of evidence supports our use of ERPs as a measure of implicit processing. Osterhout, Bersick, and McLaughlin (1997) used ERPs—in particular, the P600—to examine implicit stereotypes by manipulating the match between a reflexive pronoun and the gender definition or bias of an occupation. For example, the sentence "The nervous actress prepared himself to face the crowd" elicits a gender definition violation, whereas the sentence "The adventurous nurse put himself on the list of volunteers" elicits a gender stereotype violation. Osterhout et al. found that overall acceptability judgments indicated acceptance of sentences with stereotype violations but rejection of sentences with definition violations. Despite this pattern of acceptability judgments, gender definition violations elicited a large P600 effect and gender stereotype violations elicited a large—although somewhat attenuated—P600 effect. This dissociation between acceptability judgments and ERPs is therefore not limited to processing in the L2.

In the present study, native speakers (NSs) of English in the early stages of learning Spanish as a L2 judged whether sentences were syntactically appropriate in Spanish (explicit measure) while the electrical activity of the brain was recorded noninvasively from the surface of the scalp (implicit measure).

We included syntactic constructions that were similar or different in the L1 and the L2, and one that was unique to the L2. Following the Spanish sentences, the participants judged the grammaticality of English sentences so that we could verify the soundness of our experimental procedures.

METHOD

Participants

The participants were 34 right-handed NSs of English who were learning Spanish as a L2 at the University of Pittsburgh. Students were enrolled in one of the four semesters of beginning Spanish. There were five participants in the first term, three in the second term, nine in the third term, and two in the fourth term. None of these students had any significant previous experience with Spanish or any other Romance languages (at most, they had taken a high-school Spanish class). Learners exposed to any language other than English before age 14 were not included because the present study was not designed to control for acceptability in languages other than English and Spanish. In addition to a prescreening process, each participant completed a language history questionnaire that collected information regarding L1 and L2 language experiences; participants listed all languages to which they had been exposed and the age at which exposure began. The questionnaire included open-ended questions (e.g., how long have you studied Spanish?) and self-ratings of L1 and L2 reading, writing, speaking, and speech comprehension abilities on a 10-point scale.

Procedure

The experiment was conducted in a dedicated ERP lab, with the participant seated comfortably in an isolated room. Each participant read the sentences from a computer monitor in the testing room while the experimenter monitored the ERP recording in the adjacent room.

Participants made grammaticality judgments for Spanish and English sentences. They were asked to indicate whether the sentences were acceptable in terms of grammar in the language of presentation. The language of presentation was blocked; the block of Spanish sentences was always presented first because of the greater risk of bad trials later in the recording session.[1] Following the Spanish block, the participants judged grammaticality of sentences in English so that we could validate our ERP setup; replicating the extensive past research showing P600s in response to syntactic anomalies in L1 demonstrates the soundness of our experimental setup.[2]

Participants read sentences on a computer screen; half the sentences were well formed and half were not. The sentences were presented in a random order determined by the computer program (E-Prime; Psychological Software Tools, 2000), which also recorded the reaction times and sent critical word onset information to the ERP acquisition software. The participants responded

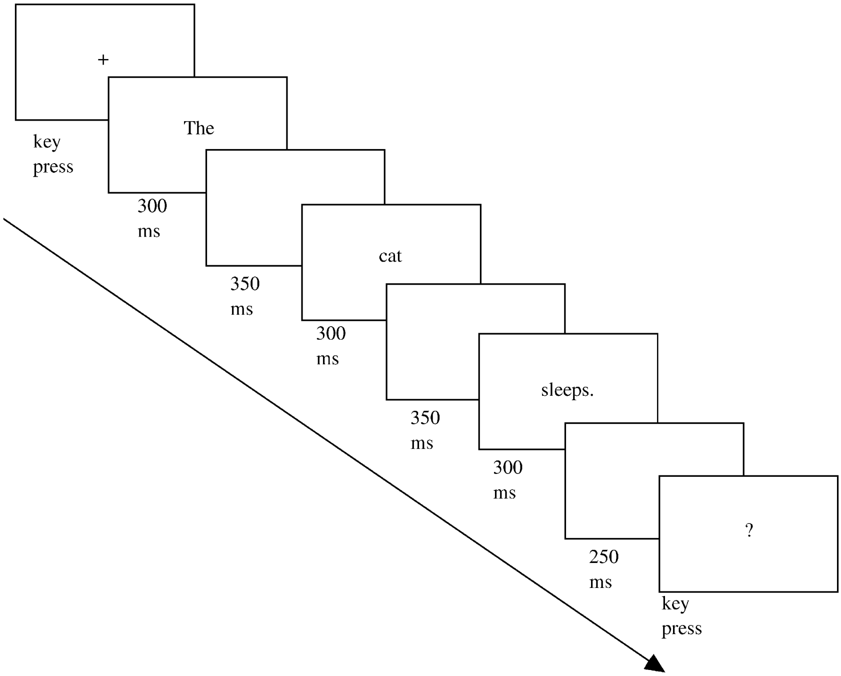

Figure 2. Time line of events during trials.

by pressing buttons on a computer keyboard; they pressed a button marked "Y" with their left hand to indicate if they thought the sentence was acceptable and a button marked "N" with their right hand if they thought the sentence was unacceptable.

Figure 2 provides an overview of the time line of events during a trial. Prior to each sentence, a fixation cross (+) appeared at the center of the computer screen. Participants were asked to blink when the fixation was on the screen. When they had finished blinking, they were to press the space bar to initiate the beginning of the trial. Sentences were presented one word at a time, at the center of the computer screen. Each word remained on the screen for 300 ms with a blank screen appearing for 350 ms between words (cf. Osterhout et al., 2000); these timing parameters were used to maximize the likelihood of detecting sensitivity to grammatical violations without the postviolation word obscuring the effect. After the offset of the final word of the sentence, a blank screen appeared for 200 ms, followed by a question mark (?) that served as a prompt. As soon as the prompt appeared, participants were supposed to respond by pressing either "Y" or "N".[3]

At the end of the online task, each participant completed the language history questionnaire.

Design

We examined four different variables: crosslanguage similarity (similar, different, unique to the L2); acceptability (acceptable, unacceptable); "lobe" (frontal, central, parietal); and "hemisphere" (left, midline, right). This entailed a $3 \times 2 \times 3 \times 3$ within-participants design.

Stimuli

The Spanish experimental stimuli came from three syntactic constructions. One is formed similarly in English and Spanish, one is formed differently in English and Spanish, and one is unique to Spanish (see Table 1). A total of 360 Spanish sentences were presented to each participant; 240 served as filler items to add variety to the constructions that appeared during the experiment. There were 40 items from each experimental construction. Nine different varieties of constructions were included in total; some varied in only two ways (acceptable or unacceptable) and others varied in four ways (acceptable in English only, acceptable in Spanish only, acceptable in both languages, acceptable in neither language). In total, there were 22 different syntactic patterns used in the experiment.

The English stimuli came from three experimental syntactic constructions (subject-verb agreement, auxiliary omission, and reflexive agreement). The subject-verb and reflexive agreement sentences were adapted from Osterhout and Mobley (1995) and the auxiliary omissions were adapted from Osterhout and Nicol (1999). A total of 120 English sentences were presented; all were experimental items. There were 40 instances of each construction type. In both English and Spanish, the sentences were randomly assigned to four versions of the stimuli. These multiple versions were created so that the sentences that one set of participants saw in their acceptable form were seen in their unacceptable form by another set of participants.

The critical word in each sentence was at the violation point. In unacceptable sentences, the critical word was defined as the word at which a violation was noticeable (e.g., the word "cooking" in "*His grandmother cooking very well."). In acceptable sentences, the critical word was in the same position as the critical word in the corresponding unacceptable sentence (e.g., the word "cooks" in "His grandmother cooks very well.").

DATA ANALYSES

Data Trimming

Data from 14 of the participants were removed for several reasons; the analyses were conducted on data from the remaining 20 participants. Data from six participants were lost due to equipment failures. Data from six participants were lost either because there were too many eye movements or blinks during recording or because there were too many high-impedance measurements. This relatively high level of data loss was a result of the fact that the experimental session lasted nearly 3 hours. This total includes about 45 minutes of preparation time and any additional optional breaks taken by participants during the trials. Additionally, task difficultly can increase movement artifact (e.g., brow scrunching, eye blinking), which also leads to bad trials. Finally, data from two participants were excluded to maintain a full counterbalancing of the stimuli (five participants in each of four rotations of the stimuli). Note that because more L2 than L1 sentences were rejected overall, the analyses we report are based on fewer trials in the L2 than in the L1.

ERP Measures

The data were recorded using 129-channel Electrical Geodesics Sensor Nets and associated NetStation acquisition software (Electrical Geodesics Incorporated). The electrodes used in these analyses correspond to these international 10-20 system electrode locations: F3, Fz, F4, C3, Cz, C4, P3, Pz, and P4 (Jasper, 1958; see Figure 3). These particular electrodes were selected for the analysis

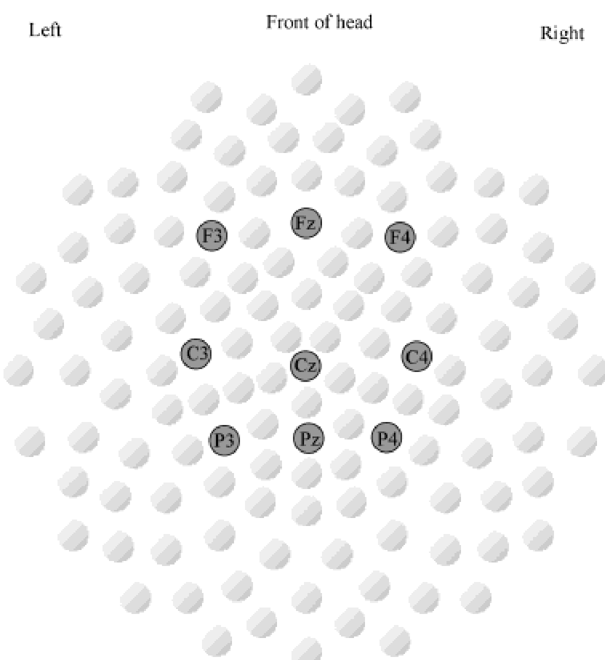

Figure 3. Electrodes used in the ERP analyses.

to maximize comparability with previous results. All impedances were kept below 40 kΩ, as recommended for use with EGI equipment (Ferree, Luu, Russell, & Tucker, 2001). The vertex (Cz) electrode was used as the reference during recording; data were rereferenced offline using the average of all electrodes (Lehmann & Skrandies, 1980). The sampling rate was 500 hertz (Hz). The hardware filter setting was between 0.1 and 200 Hz. The data were filtered offline using a 30-Hz low-pass filter. Each recording file was subjected to artifact detection processing. This processing excluded trials for which an eye blink or movement obscured the data as well as trials for which too few good electrodes were available. For the Spanish sentences, these procedures resulted in the removal of an average of 31% of trials; thus, on average, 249 of the 360 trials remained. A participant was excluded if more than half of the data consisted of bad trials. In English, these procedures resulted in the removal of 13% of the trials, leaving 105 of the 120 trials on average. Eye movements and blinks were monitored using two horizontal and four vertical eye channels. When possible, data from bad channels were replaced using data from the surrounding electrodes. The 100 ms prior to the critical word was used as the baseline for each trial.

RESULTS AND DISCUSSION

ERP Data

Event-related potentials were averaged within each acceptability and crosslanguage similarity condition for each participant. Our analysis of the ERP data included both correct and incorrect trials because past studies (e.g., Osterhout et al., 2000) have shown that the ERPs produced by beginning L2 learners show sensitivity to grammaticality, even when formal grammaticality judgments are near chance in terms of accuracy. The grand average across participants for each condition was then calculated. These grand-average ERPs were analyzed using repeated measures analyses of variance with acceptability, crosslanguage similarity, "lobe," and "hemisphere" as factors ($2 \times 3 \times 3 \times 3$). The analysis focused on the mean amplitude of the waveform during a particular time window.[4] The time windows of 500 to 700 ms and 700 to 900 ms after the onset of the critical word were examined because these windows should include the P600 or syntactic anomaly response. The inclusion of the later time window was further supported by findings reported in Weber-Fox and Neville (1996) and visual inspection of the waveforms, both of which indicated that the onset of language processing can be delayed in the L2. Note that even in the L1, the P600 response typically lasts for more than 200 ms and therefore extends into the later time window. Comparison of the patterns of results in English and Spanish is possible for the 500- to 700-ms time window. Following each repeated-measures ANOVA, we probed significant interactions that corresponded to our a priori predictions by examining the 95% confidence intervals generated by the ANOVA. Examination of these confidence intervals allows us to determine whether or not the means from various conditions were significantly different (nonoverlapping confidence intervals) without running additional statistical analyses and thereby inflating our experimentwide error rate.

ERPs in Spanish. The critical questions of the present study were whether learners show online implicit processing in L2 and whether this processing is sensitive to crosslanguage similarity. Our prediction was that there would be no observable P600 (i.e., syntactic anomaly) response to sentences containing the construction that differs between the two languages. However, we predicted evidence of syntactic anomaly sensitivity (i.e., a significantly more positive mean amplitude in the waveform between 500 and 900 ms poststimulus for unacceptable vs. acceptable stimuli) for the construction similar in the L1 and the L2 and the construction unique to the L2.

To evaluate these predictions, we ran two ANOVAs, the first corresponding to the early P600 time window (500 to 700 ms poststimulus; e.g., Kaan & Swaab, 2003) and the second corresponding to a delayed-onset P600 (hereafter referred to as the mid-P600; e.g., Kaan & Swaab). It is the mid-P600 that might be more typical of L2 processing. It is interesting to note that even though the same rate of presentation was used for the two languages, participants reported believing that the Spanish sentences were presented more quickly than the English sentences. Although this presentation rate is not as fast as that of fluent speech or even rapid self-paced reading, participants in this experiment reported difficulty with the speed of presentation during the practice trials. This supports the idea that the participants found L2 processing more difficult (and slower) than L1 processing.

The grand-average waveforms for acceptable and unacceptable sentences overall are shown in Figure 4. The grand-average waveforms for the similar (auxiliary omission) condition are shown in Figure 5, for different (determiner number) in Figure 6, and for unique (determiner gender) in Figure 7.

Overall, unacceptable constructions elicited marginally more positive-going ERP responses than the acceptable constructions. This marginally significant main effect indicates that learners were sensitive to syntactic violations in the L2, $F(1, 19) = 4.16$, $p = .06$. Additionally, the unique L2 construction sentences (determiner gender) elicited marginally more positive-going ERP responses than the similar L1-L2 construction sentences (auxiliary omission), $F(2, 18) = 3.30$, $p = .06$. This reflects the fact that participants had a more positive-going initial response to gender agreement sentences. However, these two main effects were qualified by an interaction between crosslanguage similarity and acceptability, $F(2, 18) = 4.06$, $p < .05$. Examination of the 95% confidence intervals for the means (see Figure 8) demonstrates that there was marginal sensitivity to the auxiliary omissions (similar construction), no sensitivity to the determiner number violations (different construction—in fact, the means are in the opposite direction of the pre-

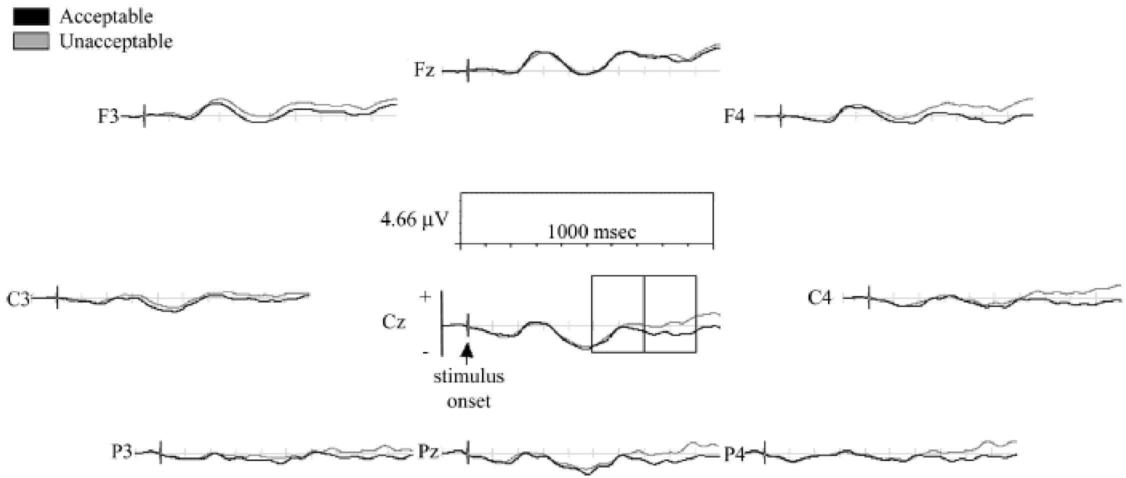

Figure 4. Grand average waveforms for the Spanish acceptable and unacceptable sentences. In all waveform plots, positive amplitude is plotted up, and an additional 15 Hz lowpass filter was applied for graphical purposes only. The boxes indicate the two critical time windows for statistical testing.

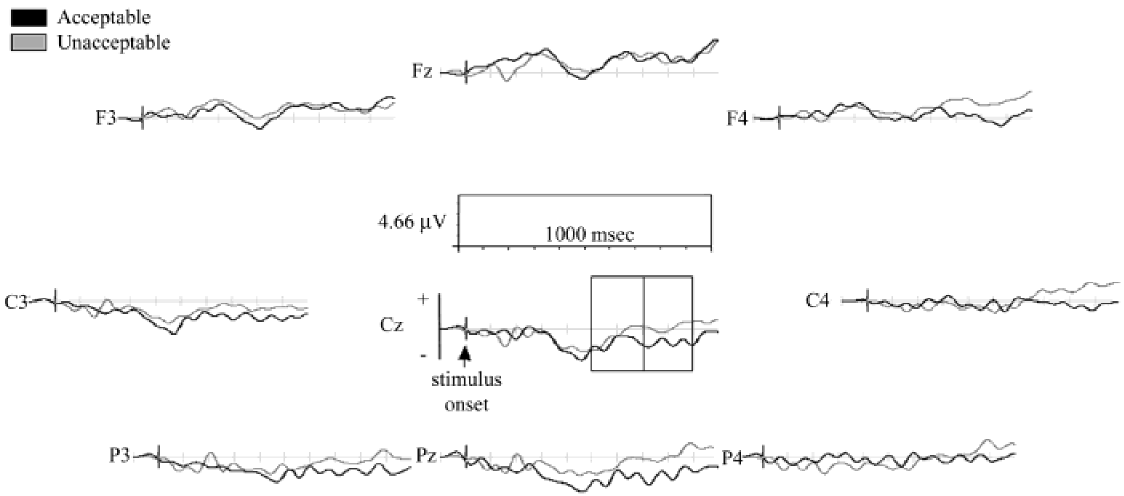

Figure 5. Grand average waveforms for the Spanish auxiliary omission (similar construction) acceptable and unacceptable sentences.

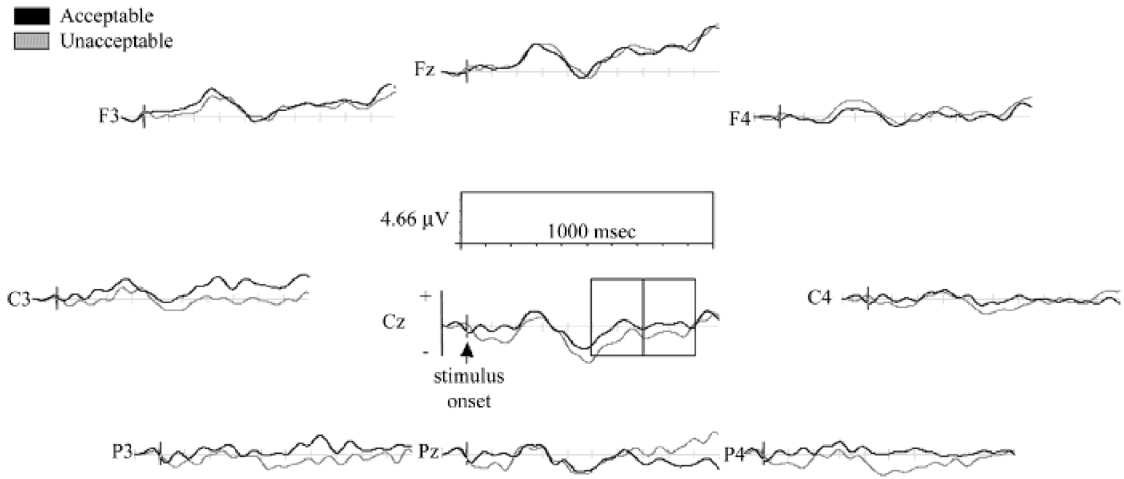

Figure 6. Grand average waveforms for the Spanish determiner number (different construction) acceptable and unacceptable sentence.

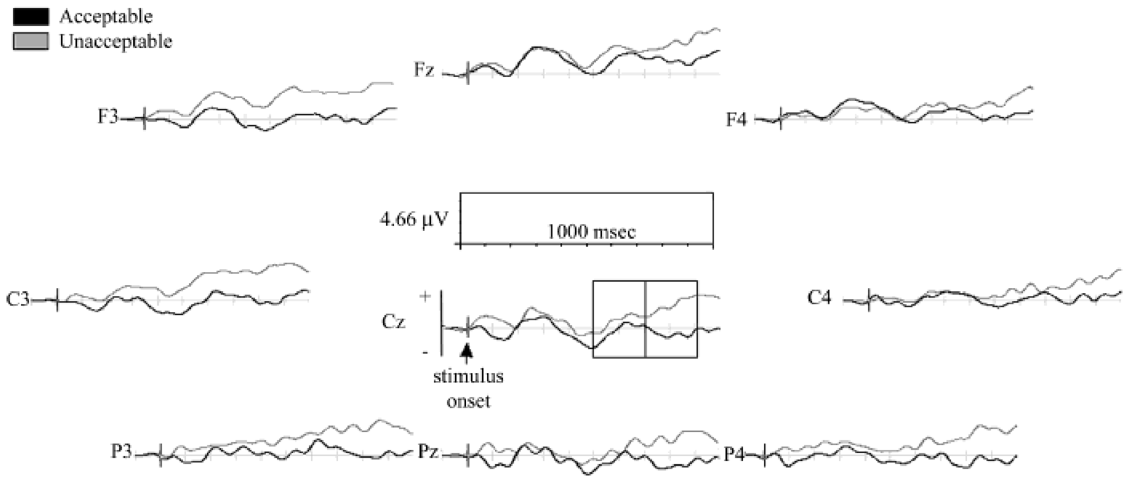

Figure 7. Grand average waveforms for the Spanish determiner gender (unique construction) acceptable and unacceptable sentences.

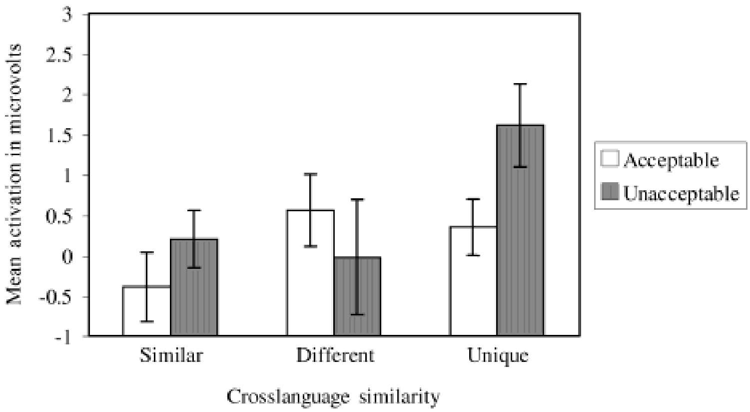

Figure 8. Mean amplitudes for the crosslanguage similarity by acceptability interaction in the early P600 (500–700 ms poststimulus) time window.

dicted effect), and significant sensitivity to the determiner gender violations (unique construction). The mean amplitudes are as follows: -0.39 ± 0.89 for similar construction acceptable sentences (with 95% confidence intervals), 0.21 ± 0.74 for similar construction unacceptable sentences, 0.56 ± 0.92 for different construction acceptable sentences, -0.02 ± 1.48 for different construction unacceptable sentences, 0.35 ± 0.72 for unique construction acceptable sentences, and 1.61 ± 1.08 for unique construction unacceptable sentences.

These findings are consistent with our predictions that learners would not be sensitive to violations in constructions that are different in the L1 and the L2. Finally, crosslanguage similarity, "lobe," and "hemisphere" interacted, $F(8, 12) = 3.88$, $p < .05$. This finding suggests that there might be multiple brain generators for the processing of the determiner gender violations, in that the same amount of activation was found over all three lobes along the left hemisphere.

In the mid-P600 time window (700 to 900 ms poststimulus), unacceptable sentences were responded to more positively than the acceptable sentences, showing that, overall, there was sensitivity to the violations, $F(1, 19) = 11.96$, $p < .01$. The Crosslanguage similarity × Acceptability interaction only approached significance in this time window, $F(2, 18) = 3.00$, $p = .075$. However, examination of the 95% confidence intervals for the means (see Figure 9) confirms our predictions: Individuals were marginally sensitive to the auxiliary omissions (similar in L1 and L2) and to the violations of determiner gender agreement (unique to the L2) but were not sensitive to the violations in determiner number agreement (different in L1 and L2; the data again suggest an effect in the opposite direction from the predictions in this condition). The mean amplitudes are as follows: -0.40 ± 0.92 for similar construction

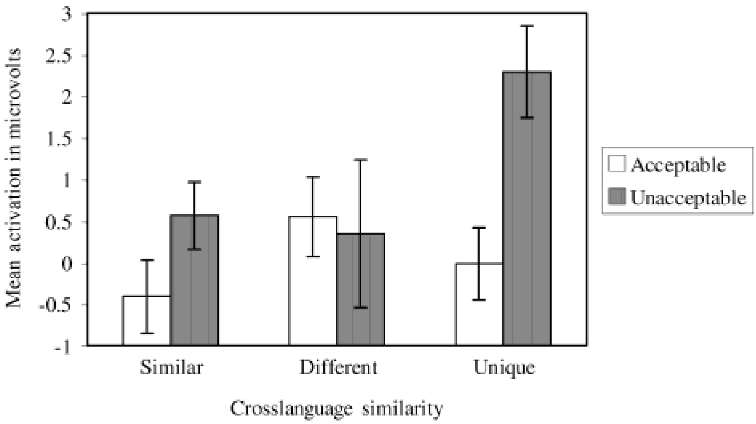

Figure 9. Mean amplitudes for the crosslanguage similarity by acceptability interaction in the mid P600 (700–900 ms poststimulus) time window.

acceptable sentences (with 95% confidence intervals), 0.57 ± 0.85 for similar construction unacceptable sentences, 0.55 ± 0.99 for different construction acceptable sentences, 0.35 ± 1.86 for different construction unacceptable sentences, −0.01 ± 0.91 for unique construction acceptable sentences, and 2.30 ± 1.16 for unique construction unacceptable sentences.

Note that there appears to be an early separation between the responses to grammatically acceptable and unacceptable gender sentences (see Figure 7). This early difference most likely reflects carryover processing from the previous word and is not based on the processing of the stimulus, because ERPs up to 100 ms are thought to reflect sensory—not cognitive—processing (see Fabiani et al., 2000). To verify that these early differences are not related to our later effects, we ran an analysis on the gender sentences for the half of the subjects who showed the smallest difference during the initial 0 to 100-ms time window (the same division of subjects would be obtained if the first 200 ms were used; the mean difference for these subjects was −0.03 μV, a *very small* difference in the opposite direction of the visual effect). In this analysis, we included the acceptability of the sentences, “hemisphere,” and “lobe” as within-subjects factors. The acceptability effect was still significant, $F(1, 9) = 5.61$, $p < .05$, and in the predicted direction; the mean amplitude (with 95% confidence intervals) for the acceptable sentences was −.41 ± 1.69, and the mean amplitude for unacceptable sentences was 1.39 ± 1.74. Therefore, our main finding remains the same regardless of whether there is an early difference.

In sum, the pattern of ERP responses to Spanish sentences generally supports our predictions. At the beginning of L2 learning, participants are only sensitive to violations of particular types, depending on the match between

L1 and L2. Thus, the occurrence of implicit processing of L2 appears to depend on the similarity between the L2 and the L1. We predicted that the learners would be sensitive to violations of the construction that was unique to the L2. The results support our prediction; learners were highly sensitive to these violations, which suggests that they had already learned them to a sufficient degree. We also predicted that the learners would be sensitive to violations that were similar in the L1 and the L2. The results showed marginal sensitivity to these violations. We also predicted that learners would not be sensitive to violations in the construction that differs in the L1 and the L2; this prediction was confirmed by the data. If anything, the different construction tended toward an opposite effect, although this reversal was not statistically significant. Finally, we would expect that learners of greater proficiency than those we tested would be sensitive to violations for constructions that differ in L1 and L2.

ERPs in English. The grand-average waveforms for the acceptable and unacceptable sentences are shown in Figure 10. The grand-average waveforms for auxiliary omission sentences are shown in Figure 11, data for reflexive sentences are shown in Figure 12, and data for subject-verb sentences are shown in Figure 13. The critical words in the unacceptable constructions elicited more positive-going ERPs than the acceptable constructions in the 500 to 700 ms following the onset of the critical word, $F(1, 19) = 13.66$, $p < .01$. The distribution of the effect varied as a function of type and acceptability of construction, as evidenced by Type × "Lobe," $F(4, 16) = 3.91$, $p < .05$, Type × "Hemisphere," $F(4, 16) = 3.35$, $p < .05$, and Acceptability × "Hemisphere" $F(2, 18) = 7.02$, $p < .01$, interactions. These effects are most likely due to the dipolar nature of the ERP generators. Thus, we have replicated past findings of sensitivity to violations in L1 syntax, which demonstrates that our experimental procedures were sound.

Accuracy Data

Accuracy for each condition was calculated for each participant. These data were analyzed with ANOVA using acceptability and type of construction as factors.

Spanish Accuracy. Overall, individuals responded less accurately to the unique construction (57.88% for determiner gender) than to the other two types (70.13% for auxiliary omission and 70.38% for determiner number), $F(2, 18) = 11.99$, $p < .01$. This result is interesting in light of the ERP effects that showed that the implicit responses to the determiner gender violations were the strongest. Additionally, individuals responded more accurately to the acceptable than the unacceptable constructions (80.33% vs. 51.92%, respectively), $F(1, 19) = 66.51$, $p < .01$. These two main effects are qualified by an interaction between crosslanguage similarity and acceptability, such that the dif-

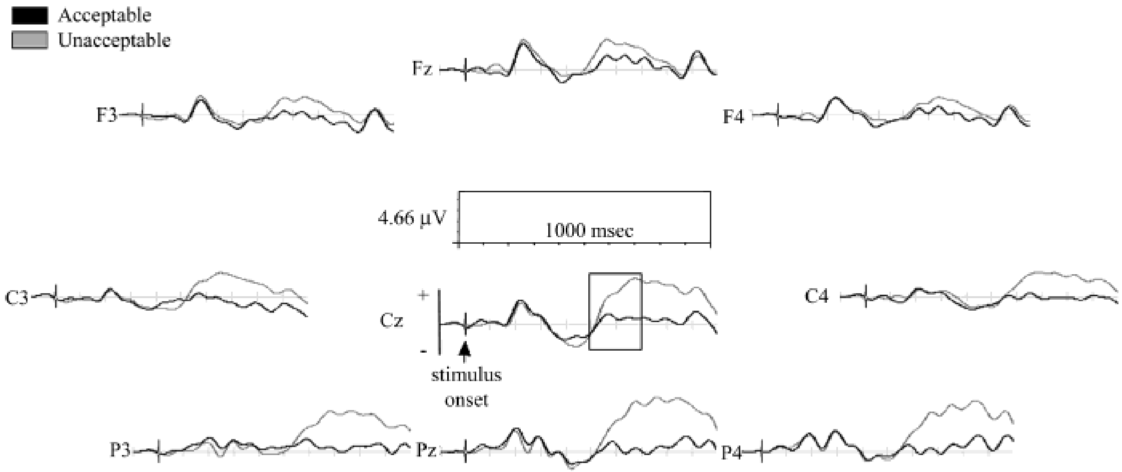

Figure 10. Grand average waveforms for the English acceptable and unacceptable sentences.

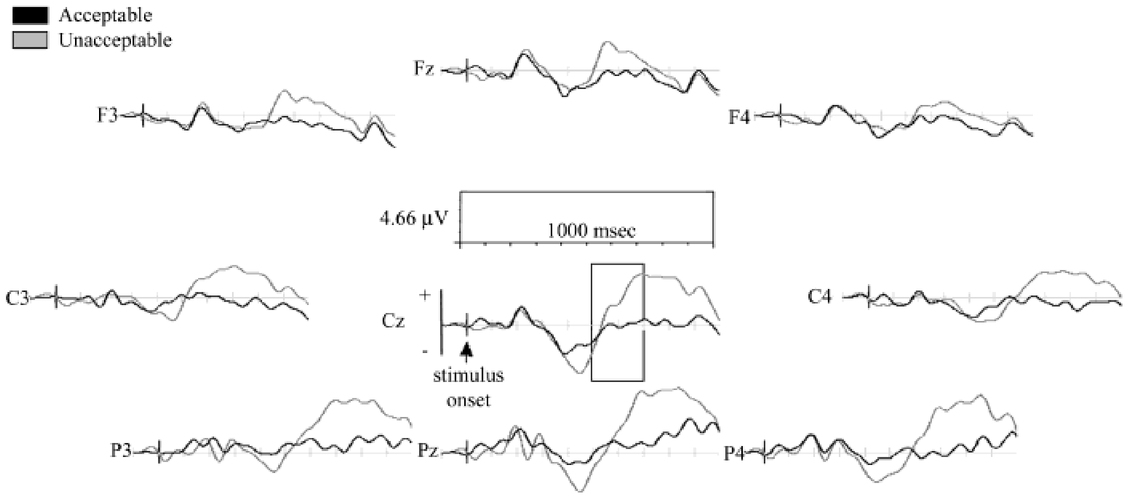

Figure 11. Grand average waveforms for the English auxiliary omission acceptable and unacceptable sentences.

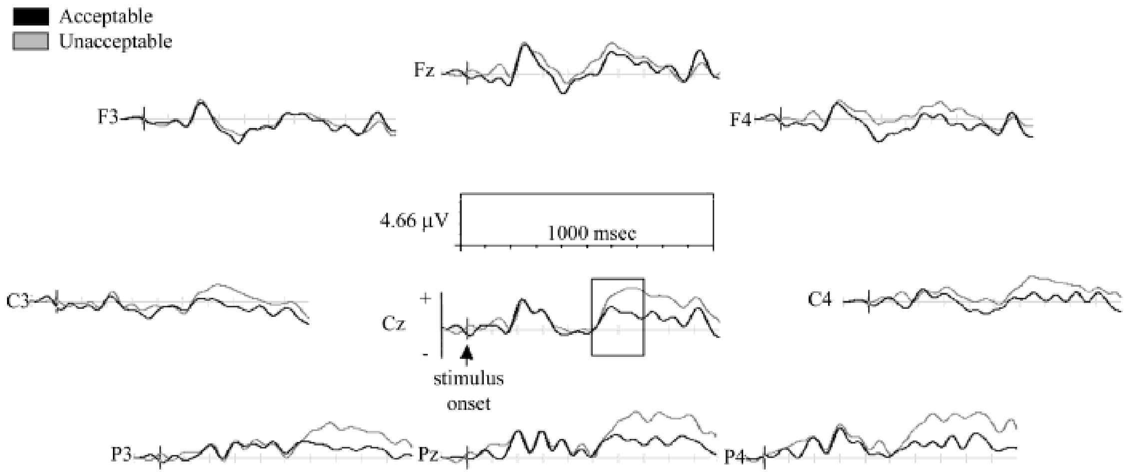

Figure 12. Grand average waveforms for the English reflexive agreement acceptable and unacceptable sentences.

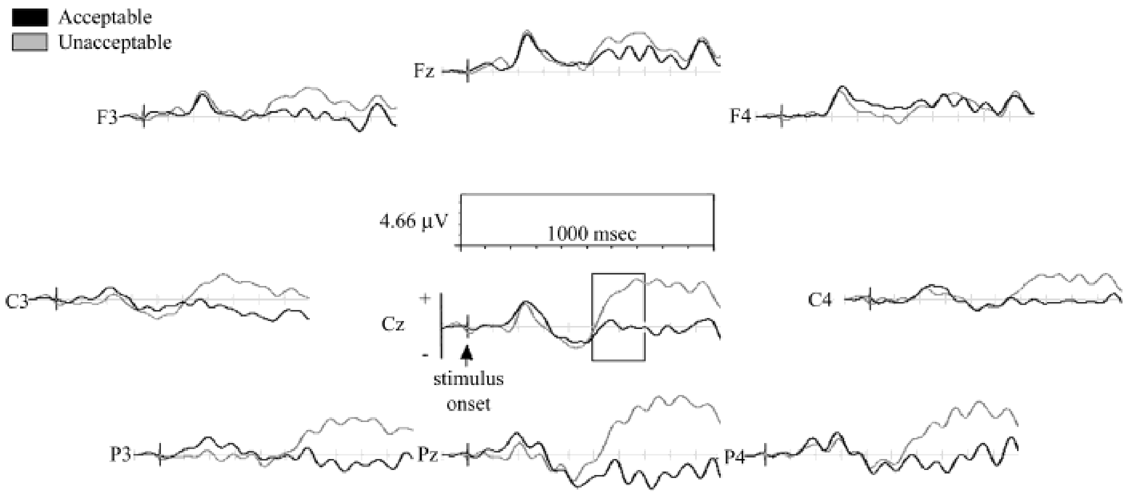

Figure 13. Grand average waveforms for the English subject-verb agreement acceptable and unacceptable sentences.

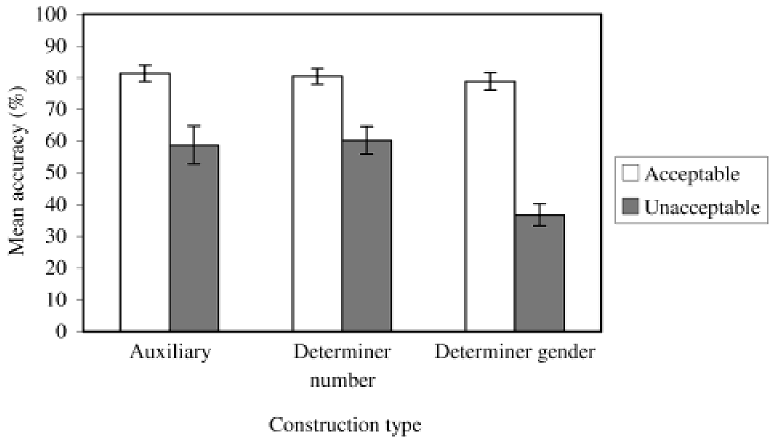

Figure 14. Accuracy by condition in Spanish.

ference between acceptable and unacceptable sentences was greatest for the unique construction sentences (determiner gender), $F(2, 18) = 12.87$, $p < .01$ (see Figure 14). The mean accuracy for similar construction acceptable sentences (with 95% confidence intervals) is 81.50 ± 5.38, 58.50 ± 12.77 for similar construction unacceptable sentences, 80.50 ± 5.37 for different construction acceptable sentences, 60.25 ± 9.16 for different construction unacceptable sentences, 79.00 ± 5.81 for unique construction acceptable sentences, and 36.75 ± 7.18 for unique construction unacceptable sentences.

To determine whether performance was at, above, or below chance (50%), we tested each mean individually against 50% in one-sample *t*-tests. We found that performance exceeded chance for two of the three syntactic constructions. Participants performed above chance on the similar construction sentences (auxiliary omission), $t(39) = 5.45$, $p < .01$, and the different construction sentences (determiner number), $t(39) = 6.83$, $p < .01$. However, participants performed at chance on the unique construction sentences (determiner gender), $t(39) = 1.96$, $p = .06$ (but below chance for the unique construction unacceptable sentences). Performance was generally poorer for the unacceptable sentences, which reflects a bias for the participants to respond "yes" to most sentences. Due to the yes bias, d' scores were examined as a measure of sensitivity to violations; a d' score of 0 indicates no sensitivity, whereas a d' score of 4 indicates perfect sensitivity. In this study, $d' \cong 1.2$ for similar and different constructions and $d' \cong .5$ for unique to the L2 constructions. Thus, the participants' overt accuracy did not display sensitivity to violations. Moreover, overt accuracy (and d') was lowest for the very condition that exhibited highest ERP sensitivity (constructions unique to the L2). At an overt level, it appears that the learners are still uncertain about assigning gender to nouns, appearing willing to accept errors as possible forms.

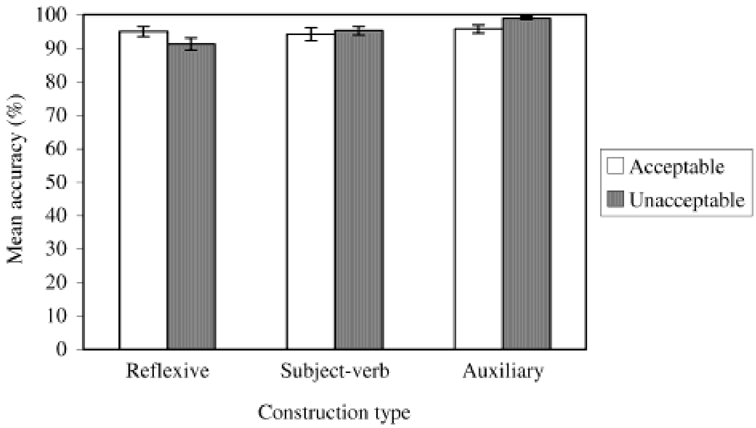

Figure 15. Accuracy by condition in English.

English Accuracy. Overall, individuals responded more accurately to the auxiliary omission sentences (97.38%) than the other two kinds of sentence (93.13% for reflexive and 94.75% for subject-verb), $F(2, 18) = 4.85$, $p < .05$. Additionally, type of construction and acceptability interacted such that participants responded more accurately to the acceptable constructions than the unacceptable constructions in the reflexive condition. The reverse was true for auxiliary omission, and accuracy was similar for acceptable and unacceptable subject-verb constructions, $F(2, 18) = 8.54$, $p < .01$ (see Figure 15). The mean accuracy for reflexive agreement acceptable sentences (with 95% confidence intervals) is 95.00 ± 3.31, 91.25 ± 3.71 for reflexive agreement unacceptable sentences, 94.25 ± 4.11 for subject-verb agreement acceptable sentences, 95.25 ± 2.79 for subject-verb agreement unacceptable sentences, 95.75 ± 2.55 for auxiliary omission acceptable sentences, and 99.00 ± 0.96 for auxiliary omission unacceptable sentences.

We believe this reflects the fact that assimilation to a correct formation is not possible for the auxiliary omissions, whereas the subject can be assimilated for the reflexives. For example, after reading “boy kicked themselves” with word-by-word presentation, you might imagine that the first word was actually “boys” rather than “boy,” thereby producing “boys kicked themselves.” Similarly for the subject-verb agreement sentences, if you saw “boy make,” you might assume you saw “boys make.”

Effects of L2 Proficiency and Experience. Although students in the more advanced classes rated themselves as more proficient than those in the less advanced classes, we found that L2 proficiency self-ratings themselves were not a predictor of any of the results in this study. By contrast, Osterhout et al. (2000) found a difference in the manner in which sensitivity to grammatical violations was borne out in the ERP record as a function of L2 proficiency. However, even our learners in the first semester of study were at the same point as the more proficient learners in the Osterhout et al. study because they were tested at the end of the semester. Therefore, we anticipated that our learners were already past the point at which grammatical sensitivity would be evidenced by an N400.

To determine whether proficiency or experience with Spanish influenced the results of this study, we correlated the years of study of Spanish (total number of years of study, which might include some prior exposure) and the self-ratings of Spanish proficiency with the accuracy of judgments in all critical conditions and the mean amplitude for the Cz electrode (which was representative of the results) for each condition. Neither experience with Spanish nor Spanish self-ratings correlated significantly with any measure of performance ($p > .05$ in all cases). These findings suggest that similarity across languages accounts for more of the variance in online sensitivity than experience with the language for a relatively homogeneous sample such as ours. We would expect that the results would be correlated with experience had we included a more heterogeneous sample. We also ran the same correlations with the semester of study. Semester of study also did not correlate with our ERP measures, but correlated with accuracy for two of the conditions; individuals in later semesters were more likely to correctly reject unacceptable determiner number, $r = .64$, $p < .01$, and gender agreement sentences, $r = .46$, $p < .05$. This finding is consistent with the idea that our judgment task measured explicit knowledge, because such knowledge should be greater for individuals in later semesters of L2 study.

GENERAL DISCUSSION

Second language processing and particularly the beginning stages of L2 learning have recently come to the attention of researchers who use electrophysiological methods to study language processing. We hope that the present study will provide preliminary evidence to help lay the foundation for future research in this domain. We view electrophysiological methods as an additional way in which researchers in this area can attempt to answer the many questions of interest about L2 learning.

The results obtained in this study and other related studies provide support for two key ideas regarding the early stages of L2 learning. First, as compared to grammatically acceptable sentences, we observed more positive-going ERPs between 500 and 900 ms after a grammaticality violation for two of our three sentence types in the L2. This effect suggests that learners are responsive to L2 grammaticality violations as they process sentences word by word (see also Osterhout et al., 2000, Experiment 2). At the same time, learners did not demonstrate any clear ability to judge grammatical violations correctly at the end of the sentence (e.g., Osterhout et al., 2000, Experiment 2). The comparison of these two effects suggests that learners have better access to implicit knowledge than explicit knowledge during sentence processing. Of course, it could be that learners would demonstrate their command of explicit grammatical rules in formal test situations that are very different from the context of this experiment.

Our offline GJT involves use of a combination of explicit offline processes with implicit knowledge acquired during comprehension. However, we believe that these beginning learners were unable to reliably integrate their online grammaticality reactions with the offline task. As a result of this task analysis, one of our future goals is to determine how L2 learners can better use their implicit knowledge to make overt judgments (see the Creating Improvements in Performance section). Regardless of the outcome of this additional exploration, the results of the current study suggest that beginning learners show implicit online sensitivity to some violations of L2 grammar.

The results also provided support for a second idea regarding the early stages of L2 acquisition. Derived from work in the framework of the competition model, we predicted that L1 syntactic processes would transfer to L2 and compete with these processes online. These predictions were borne out most clearly for sentences with determiner number violations that were different between English and Spanish. English NSs have learned that there is no agreement in number between the article and the following noun. When they read or hear an article, they know that they can move on to the following noun without storing any information regarding number on the article. Because learners tend to think that Spanish works the same way, they are simply insensitive to grammaticality violations for determiner number agreement. Given that it is easy to decode, they might detect the number of the article, but this information does not influence their processing of the following noun. In effect, their L1 processor is telling them to simply throw away some important information in the new L2. Note that this analysis is only relevant to comprehension. In production, learners must choose between forms on the basis of noun-article agreement. However, this marking is easy and regular in Spanish, requiring only minimal attention. As a result, this learning has little secondary impact on comprehension.

The results for the gender agreement violations indicate a very different developmental pathway. There is no transfer of determiner gender-marking from L1 to L2 because English lacks a system of grammatical gender. It is not that learners think they can ignore gender on the pronoun. Rather, at first they have no idea how to use gender during processing. As they acquire this L2 system, learners begin to set up relations between the various forms of the article and endings on adjectives and the nominal lexicon. Interestingly enough, despite the clear cortical reactivity participants present for gender agreement violations, their grammaticality judgments are at-chance levels. Again, this suggests that they are developing effective implicit processing for L2 in the absence of ability to make use of explicit processes, implicit processes, or both for grammaticality judgments (of gender, in particular).

Implications

The techniques used in this research might assist in the development of adequate tools to isolate problem areas in L2 learning that could inform L2 teaching techniques. Indeed, part of the challenge for teachers is to identify what students know and what they do not know. Furthermore, the proposed techniques might be used to identify ERP markers for learning milestones that can later be applied more broadly to studies of L2 learning. If we can better understand the structures to which learners are sensitive—even though their overt behavior might not reflect such sensitivity—then we might be able to assist learners in harnessing this sensitivity so that they could use the L2 more accurately.

Creating Improvements in Performance

We are in the process of conducting a follow-up study to determine how to obtain behavioral measurements that better reflect the capabilities of the participants; that is, if their brain responses suggest that they are sensitive to violations in syntax in L2, can we improve their acceptability judgments? In this pilot study, participants process sentences during an initial block, very similar to the present study. They are then given an interpolated block in which they are shown word pairs with the violations or acceptable constructions outside of the sentence context. This procedure was selected to allow the participants to focus their attention on the problematic aspects of the sentences. For example, instead of reading the Spanish equivalent of "*I walking to school," they would read "*I walking." Additionally, feedback is given about the accuracy of the responses. This interpolated block is followed by another block of sentences without feedback, some of which duplicate concepts seen during the interpolated block and others that were not previously seen.

The accuracy data from the interpolated block of this pilot study show a vast improvement relative to the first block (Phase 1: $d' = .24$; Phase 2: $d' = 2.36$). Additionally, accuracy during the third block (Phase 3: overall $d' = 1.28$) showed improvement relative to the first block of sentences, both for repeated concepts and for new concepts. It seems unlikely that these results can be attributed simply to practice, because we were not able to observe changes of this sort in the main experiment we have reported here. These new results suggest that we might be able to improve behavioral performance by manipulating both feedback and by decontextualizing the errors. We are in the process of determining whether both facets are needed to improve performance and whether the improved overt performance is accompanied by enhanced sensitivity to violations during the third block relative to the first block of sentences.

NOTES

1. This greater risk is due to the drying of the sponges in which the electrodes are seated; to alleviate this problem, we rewet the electrodes between the Spanish and English blocks.

2. It would be ideal to include a Spanish NS control group to demonstrate that native NSs are equally sensitive to all of the Spanish violations. However, the sample of NSs we found consisted of individuals who were highly proficient in English and had begun learning it during childhood. Thus, these bilinguals were likely to be influenced by crosslanguage similarity that would bias the results.

3. In future research, it would be advantageous to consider including an additional GJT that is not adapted to the ERP environment (e.g., with full-sentence presentation), administered after the ERP recording session. However, care should be taken to not repeat the items, as repetitions might alter processing. This procedure would allow the comparison of the two judgment tasks.

4. We also conducted our analyses using an adaptive mean that corrects for variability across trials (*latency smearing*; e.g., Hoffman, Simons, & Houck, 1983). Using this procedure, a peak is identified during a particular time window. Then, the peak becomes the center of the newly defined 200-ms time window. The mean for the new window is then calculated (the *adaptive mean*). These analyses showed the same statistical pattern of results as the original analyses and are, therefore, not reported.

REFERENCES

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

Electrical Geodesics, Inc. (2003). *NetStation* [computer software]. Eugene, OR: Electrical Geodesics, Inc.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*, 143–188.

Fabiani, M., Gratton, G., & Coles, M. G. H. (2000). Event-related brain potentials: Methods, theory, and applications. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 53–84). New York: Cambridge University Press.

Ferree, T. C., Luu, P., Russell, G. S., & Tucker, D. M. (2001). Scalp electrode impedance, infection risk, and EEG data quality. *Journal of Clinical Neurophysiology*, *112*, 536–544.

Frenck-Mestre, C. (in press). Ambiguities and anomalies: What can eye-movements and event-related potentials reveal about second language sentence processing? In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches*. Oxford: Oxford University Press.

Hoffman, J. E., Simons, R. F., & Houck, M. R. (1983). Event-related potentials during controlled and automatic target detection. *Psychophysiology*, *20*, 625–632.

Hulstijn, J. H. (2002). Towards a unified account of the representation, processing, and acquisition of second language knowledge. *Second Language Research*, *18*, 193–223.

Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, *48*, 113–126.

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

Koelsch, S., Gunter, T., Schröger, E., & Friederici, A. D. (2003). Processing tonal modulations: An ERP study. *Journal of Cognitive Neuroscience*, *15*, 1149–1159.

Krashen, S. (1994). The input hypothesis and its rivals. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages*. San Diego, CA: Academic Press.

Kroll, J. F., & Tokowicz, N. (in press). Models of bilingual representation and processing: Looking back and to the future. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches*. Oxford: Oxford University Press.

Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multi-channels potential fields. *Electroencephalography and Clinical Neurophysiology*, *48*, 609–621.

Luce, D. (1959). *Individual choice behavior*. New York: Wiley.

MacWhinney, B. (1987). The competition model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249–308). Mahwah, NJ: Erlbaum.

MacWhinney, B. (in press). A unified model of language acquisition. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches*. Oxford: Oxford University Press.

McDonald, J. L. (1987). Sentence interpretation in bilingual speakers of English and Dutch. *Applied Psycholinguistics*, *8*, 379–414.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience*, *7*, 703–704.

Morris, J. P., Squires, N. K., Taber, C. S., & Lodge, M. (2003). Activation of political attitudes: A psychophysiological examination of the hot cognition hypothesis. *Political Psychology*, *24*, 727–745.

Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition*, *25*, 273–285.

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

Osterhout, L., McLaughlin, J., Inoue, K., & Loveless, J. (2000, November). *Brain responses in L2 learning*. Paper presented at the meeting Bilingualism: From basic research to educational practice, Rovereto, Italy.

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 response to syntactic and semantic anomalies. *Language and Cognitive Processes*, *14*, 283–317.

Psychological Software Tools, Inc. (2000). *E-Prime* [Computer software]. Pittsburgh, PA: Psychological Software Tools, Inc.

Rugg, M. D., Mark, R. E., Walla, P., Schloerscheidt, A. M., Birch, C. S., & Allan, K. (1998). Dissociations of the neural correlates of implicit and explicit memory. *Nature*, *392*, 595–598.

Tachibana, H., Miyata, Y., Takeda, M., Minamoto, H., Sugita, M., & Okita, T. (1999). Memory in patients with subcortical infarction: An auditory event-related potential study. *Cognitive Brain Research*, *8*, 87–94.

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

Zhang, S. Q. (1995). Semantic differentiation in the acquisition of English as a second language. *Language Learning*, *45*, 225–249.