

Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related  
potential investigation

Natasha Tokowicz

Learning Research and Development Center and The Center for the Neural Basis of Cognition,  
University of Pittsburgh

and

Brian MacWhinney

Carnegie Mellon University

Address for Correspondence:

Natasha Tokowicz  
642 Learning Research and Development Center  
University of Pittsburgh  
Pittsburgh, PA 15260 USA  
Phone: +1 412-624-7026  
Fax: +1 412-624-9149  
E-mail: [Tokowicz@alumni.umass.edu](mailto:Tokowicz@alumni.umass.edu)

## Abstract

We used event-related brain potentials (ERPs) to investigate the contributions of explicit and implicit processes during second language (L2) sentence comprehension. We tested 20 native English speakers enrolled in the first four semesters of Spanish classes, using an L2 grammaticality judgment task (GJT), while recording both accuracy and ERP data. We reasoned that any difference in the ERP between grammatical and ungrammatical sentences would reflect on-line, implicit processing, and that overt grammaticality judgments would reflect primarily explicit processing. We reasoned that because end-of-sentence grammaticality judgments are open to conscious inspection, they can be influenced by strategic processes that reflect on formal rules. On the other hand, because ERPs are a direct reflection of on-line processing, they reflect automatic, non-reflective, implicit responses to stimuli (Rugg et al, 1998; Schnyer et al., 1999; Tachibana et al., 1999). We used a version of the GJT that has been adapted for the ERP environment. The sentences were presented one word at a time; each word was presented for 300 ms with a blank between words of 350 ms. Grammaticality judgments were given after a brief delay following the last word of each sentence. Half of the sentences were grammatically acceptable. The critical sentences varied the form of three different syntactic constructions. First, we included sentences that were either acceptable or not in their tense-marking; this construction is formed *similarly* in L1 and L2. We also included sentences that were either acceptable or not in their determiner number agreement; this construction is formed *differently* in L1 and L2. Finally, we included sentences that were acceptable or not in their determiner gender agreement; this construction is *unique* to L2. 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. 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 brain responses that differed to grammatical and ungrammatical sentences) to violations in L2 for constructions that are formed similarly in L1 and L2, but were not sensitive to violations for constructions that differ in L1 and L2. Critically, a robust grammaticality effect was found in the ERP data for the construction that was unique to 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 L1 and L2. Furthermore, there is a divergence between explicit and implicit measures of L2 learning, which may be due to the behavioral task demands. We conclude that comparing ERP and behavioral data may provide a sensitive method for measuring implicit processing. In future research, we will attempt to improve the accuracy of grammaticality judgments using feedback, and then examine the consequences of these improvements for ERPs.

## Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation

Do adult second language 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 second language (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 may be exposed to explicit rules in classrooms and textbooks, they rely on implicit knowledge and implicit processing to comprehend sentences in L2.

Hulstijn (2002) suggested that one way to address this issue would be to measure readers' immediate on-line neuronal reactions to L2 sentences, using event-related potentials or ERPs. In the current study, we follow this suggestion by examining ERP data from beginning learners of Spanish as they are engaged in a grammaticality judgment task. Our findings indicate that beginning L2 learners show implicit sensitivity to violations of grammar in L2, but that the extent to which L2 is processed implicitly depends on the similarity between L1 and L2.

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 L1 to L2 is bad. When the two languages are similar, positive transfer will assist learning. However, cross-language mismatches may hinder the acquisition of L2 in two ways. First, cross-language mismatches can impede the process of learning by leading learners to entertain false hypotheses. For example, learners may 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 determining the exact range of L2 structures.

The second possible source of difficulty for adult L2 learners is on-line competition between the two language systems (Frenck-Mestre, in press; Kroll & Tokowicz, in press). When L1 and L2 provide contrasting interpretations of a given structure, the stronger L1 patterns will often be used. 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 an L1 syntactic “accent.”

Although transfer and competition pose similar challenges to the L2 learner, they have different consequences—transfer from L1 to L2 would cause an initial problem that should be resolved as L2 information is learned, whereas on-line 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 in order to effectively use L2.

The present study addressed two research questions. Are L2 learners at beginning stages able to process L2 implicitly on-line? And, to what extent do L1 transfer and competition effects modulate implicit processing? Specifically, we hypothesized that learners would show less implicit sensitivity to grammatical constructions that differ in the two languages than to constructions that are similar in the two languages. On the basis of the analysis of the Competition Model (MacWhinney & Bates, 1989; MacWhinney, in press), we also predicted that learners would show more implicit sensitivity to violations of constructions that are unique to L2 and provide valid cues to comprehension relative to constructions that differ in L1 and L2.

Tense-marking verbal auxiliaries are used and positioned similarly in English and Spanish. Therefore 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 translation of that sentence: “\*Su abuela cocinando muy bien”. We expect that both positive transfer and the absence of on-line 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 major challenge for beginning learners of Spanish. Violations of gender agreement in Spanish are not affected by either negative transfer from English or on-line competition, since English makes no use of gender in sentence processing. As a result, we would expect at least moderate sensitivity in response to the violation in a sentence such as “\*Ellos fueron a un fiesta.” [\*They went to a(masculine) party(feminine)].

In contrast, there is a mismatch between English and Spanish in the formation of determiner number agreement. In English, we use the same determiner with both singular and plural nouns, saying both “the boy” and “the boys.” In Spanish, on the other hand, the article takes different forms in “el niño” and “los niños.” Because English speakers have learned not to pay attention to the number of the noun in choosing the determiner, we would expect that they would also tend to ignore this information when processing Spanish. Thus, we expect little sensitivity in response to the violation in a sentence such as “\*El niños están jugando.” [\*The(singular) boys(plural) are playing.]. See Table 1 for the sample stimuli.

To examine *explicit* processing, we asked subjects to produce formal grammaticality judgments after the entire sentence was presented. This type of off-line 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 in rendering a judgment. However, this measure may not only reflect explicit processing, but rather a combination of implicit and explicit processing, because

learners could use their intuition about the sentences' grammaticality in making their judgments (e.g., R. Ellis, this volume). We return to this issue in the general discussion.

To examine *implicit* processing of L2 syntax, we used ERPs to measure comprehension as it unfolds over a very short period of time (less than 800 milliseconds). ERPs are electrophysiological brain responses to particular stimulus events (e.g., reading a word) that are recorded from electrodes placed on the scalp. Specific ERP components can be considered indices of specific cognitive events (Coles, Gratton, & Fabiani, 1990). In particular, an ERP component has been identified that corresponds to syntactic anomalies. In past research, ERPs have been used with great success to study the degree to which individuals are sensitive to such syntactic anomalies (e.g., Osterhout & Holcomb, 1992).

We focused our attention on a late positivity in the ERP waveform that peaks at approximately 600 ms post-stimulus and is centro-parietally distributed (the "P600"; see Figure 1), as an index of syntactic anomaly. For example, a P600 can be observed in response to the sentence "\*The cat won't eating" (e.g., Osterhout & Holcomb, 1992). This ERP reflects initial non-reflective 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 meta-cognition 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 off-line grammaticality judgments than in the on-line ERP measure which allows for explicit knowledge to be used.

However, there is already an indication from a relevant study that this is not the case. Osterhout, McLaughlin, Inoue, and Loveless (2000) showed that brain responses may indicate better comprehension in L2 learners than would be suggested by overt responses obtained from accuracy to off-line grammaticality judgments. Overt grammaticality judgments obtained at the ends of sentences showed that, as early as during the fourth month 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. These results suggest that it is possible that the type of overt responses obtained in this study may reflect the results of integration or reflective processes that occur after the entire sentence has been read rather than only the on-line incremental comprehension processes of the reader.

ERPs have been used extensively to study implicit processing. For example, Tachibana et al. (1999) consider the N400 repetition effects they observed to be a measure of implicit memory processing. ERPs are also believed to reflect automatic processing (see Schnyer, Kaszniak & Forster, 1999), which is often assumed to be absent in L2 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 non-musicians. Finally, 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.

In the present study, native English speakers in the early stages of learning Spanish as a second language judged whether sentences were syntactically appropriate in Spanish (explicit measure) while the electrical activity of the brain was recorded non-invasively from the surface of the scalp (implicit measure). We included syntactic constructions that were similar or different in L1 and L2, and one that was unique to L2.

## Method

### *Participants*

The participants were 34 right-handed native English speakers who were learning Spanish as a second language at the University of Pittsburgh. Students were enrolled in one of the four semesters of beginning Spanish. There were five subjects in the first term, three in the second term, nine in the third term, and two in the fourth term. Although students in the more advanced classes were more proficient



than those in the less advanced classes, we found that L2 proficiency itself was not a predictor of any of the results in this study. We return to this issue in the results section. People who had been exposed to other languages before age 14 were not included because the present study was not designed to control for acceptability in languages other than English and Spanish.

### *Procedure*

The experiment was conducted in a dedicated ERP lab, with the participant seated comfortably in an isolated room. The participants 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 to 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. This greater risk is due to the drying of the sponges in which the electrodes are seated; to alleviate this problem, we re-wet the electrodes between the Spanish and English blocks. 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.<sup>1</sup>

Participants read sentences on a computer screen; half of these sentences were well-formed and the other half were not. The sentences were presented in a random order determined by the computer program (E-Prime, Psychological Software Tools, Pittsburgh, PA, USA) that also recorded the reaction times, and sent critical word onset information to the ERP acquisition software. The participants responded 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 (e.g., Osterhout et al., 2000); these timing parameters were used to maximize the likelihood of detecting sensitivity to grammatical violations without the post-violation word obscuring the effect. <sup>2</sup>

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 often having difficulty keeping up with the speed of presentation. 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 with a grammaticality judgment.

At the end of the on-line task, each participant completed a language history questionnaire that requested information regarding L1 and L2 language experiences. The questionnaire included open-ended questions and self-ratings of reading, writing, speaking, and speech comprehension abilities in L1 and L2 on a 10-point Likert-type scale.

### *Design*

We implemented a 3 (cross-language similarity: similar, different, unique to L2) X 2 (acceptability: acceptable, unacceptable) X 3 (“lobe”: frontal, central, parietal) X 3 (“hemisphere”: left, midline, right) within-participants design.

### *Stimuli*

The Spanish experimental stimuli came from 3 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 3 experimental syntactic constructions (subject verb agreement, tense omission, and reflexive agreement). The subject verb and reflexive agreement sentences were adapted from Osterhout and Mobley (1995) and the tense 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. 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 the participants should have been able to notice a violation (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 from 14 of the participants were removed for several reasons. 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 three hours. In addition, task difficulty 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).

### *ERP Measures*

*ERP recording and pre-processing details.* The data were recorded using 129-channel Electrical Geodesics Sensor Nets and associated NetStation acquisition software (Electrical Geodesics Incorporated, Oregon, USA). The electrodes used in these analyses correspond to these international 10-20 system (Jasper, 1958) electrode locations: F3, Fz, F4, C3, Cz, C4, P3, Pz, and P4 (see Figure 3). All impedances were kept below 40k $\Omega$  (Ferree, Luu, Russell, & Tucker, 2001). The vertex (Cz) electrode was used as the reference during recording; data were re-referenced off-line using the average of all electrodes (Lehmann & Skrandies, 1980). The sampling rate was 500 Hz. The hardware filter setting was between 0.1 and 200 Hz. The data were filtered off-line using a 30 Hz low-pass filter. Each recording file was subjected to artifact detection processing. This processing excluded trials on which an eye blink or movement obscured the data, as well as trials on which too few good electrodes were available. For the Spanish sentences, these procedures resulted in the removal of 31 % of trials on average; 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

ERPs were averaged within each acceptability and cross-language 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, cross-language similarity, lobe, and hemisphere as factors (2 X 3 X 3 X 3). The analysis focused on the mean amplitude of the waveform during a particular time window. The time windows of 500-700 ms and 700-900 ms after the onset of the critical word were examined, because these windows should include the P600 or syntactic anomaly response (the later time window is also used because Weber-Fox & Neville, 1996, and visual inspection of the waveforms showed that the onset of the language processing can be delayed in L2).<sup>3</sup>

*ERPs in Spanish.* The critical questions of the present study were whether learners show on-line implicit processing in L2 and whether this processing is sensitive to cross-language 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 that we would see evidence of syntactic anomaly sensitivity (i.e., a significantly more positive mean amplitude in the waveform between 500 and 900 ms post-stimulus for unacceptable versus acceptable stimuli) for the similar construction and the construction unique to L2.

To evaluate these predictions, we ran two analyses of variance, the first corresponding to the early P600 time window (500-700 ms post-stimulus; 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, 2003) that may be more typical of L2 processing. The grand average waveforms for acceptable and unacceptable sentences overall are shown in Figure 4. The grand average waveforms for the similar (tense) 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 did the acceptable constructions. This main effect indicates that learners were sensitive to syntactic violations in L2,  $F(1, 19) = 4.16, p = .06$ . In addition, the unique (determiner gender) sentences elicited

marginally more positive-going ERP responses than the similar (tense omission) sentences,  $F(2, 18) = 3.30, p = .06$ . 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 cross-language 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 a marginal sensitivity to the tense omissions (similar), no sensitivity to the determiner number violations (different), and significant sensitivity to the determiner gender violations (unique). These findings are consistent with our predictions that learners would be sensitive only to violations in constructions that are similar in L1 and L2 and unique to L2. Finally, cross-language similarity, lobe, and hemisphere interacted,  $F(8, 12) = 3.88, p < .05$ . This finding suggests that there may be multiple brain generators for the processing of the determiner gender violations, in that the same amount of activation is found over all three lobes along the left hemisphere.

In the mid P600 time window (700-900 ms post-stimulus), 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 cross-language similarity by 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 sensitive to the tense omissions (similar in L1 and L2) and to the violations of determiner gender agreement (unique to L2), but were not sensitive to the violations in determiner number agreement (different in L1 and L2).

In sum, the pattern of ERP responses to Spanish sentences 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, whether implicit processing of L2 occurs depends on the similarity between the new and the existing language. We had predicted that the learners would be moderately sensitive to violations of the construction that was unique to L2. However, the learners were highly

sensitive to these violations suggesting that they had already learned them to a sufficient degree. Finally, we would expect that learners of greater proficiency 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 tense sentences are shown in Figure 11, for reflexive sentences are shown in Figure 12, and 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 500-700 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 by lobe, type by hemisphere, and acceptability by hemisphere interactions. These effects are most likely due to the dipolar nature of the ERP generators. It is also notable that the unacceptable subject-verb agreement sentences elicited a more negative-going deflection in the N400 range (300-500 ms), as was observed by Osterhout and Mobley (1995). Thus, we have replicated past findings of sensitivity to violations in native-language syntax, which shows that our experimental procedures were sound.

#### *Accuracy Data*

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

*Spanish accuracy.* Overall, individuals responded less accurately to the unique (determiner gender) constructions (57.88%) than to the other two types (70.13% for tense 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. In addition, 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 cross-language similarity and acceptability such that the difference between acceptable and unacceptable

sentences was greatest for the unique (determiner gender) sentences,  $F(2, 18) = 12.87, p < .01$  (see Figure 14).

To determine whether performance was at or above/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 (tense omission) sentences,  $t(39) = 5.45, p < .01$ , and the different (determiner number) sentences,  $t(39) = 6.83, p < .01$ . However, participants performed at chance on the unique (determiner gender) sentences,  $t(39) = 1.96, p = .06$  (but below chance for the gender unacceptable sentences). Performance was generally poorer for the unacceptable sentences, which reflects a bias for the participants to respond “yes” to most sentences. At an overt level, it appears that the learners are still confused about assigning gender to nouns, appearing willing to accept errors as possible forms.

*English accuracy.* Overall, individuals responded more accurately to the tense omission sentences (97.38%) than the other two kinds of sentences (93.13% for reflexive and 94.75% for subject verb),  $F(2, 18) = 4.85, p < .05$ . In addition, type of construction and acceptability interacted such that for the reflexive condition participants responded more accurately to the acceptable constructions than the unacceptable constructions. The reverse was true for tense, and accuracy was similar for acceptable and unacceptable subject verb constructions,  $F(2, 18) = 8.54, p < .01$  (see Figure 15). We believe this reflects the fact that assimilation to a correct formation is not possible for the tense omissions, whereas the subject can be assimilated for the reflexives. For example, after reading “boy kicked themselves” with word-by-word presentation, you may 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 may assume you saw “boys make...”.

*Effects of second language proficiency and experience.* To determine whether proficiency or experience with Spanish influenced the results of this study, we correlated the years of study of Spanish 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 (all  $ps > .05$ ). These findings suggest that similarity across languages accounts for more of the variance in on-line 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 did correlate 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 ( $r = .46, p < .05$ ) sentences. 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

The results obtained in this study provide good support for two key ideas regarding the early stages of L2 acquisition. First, we observed significantly more positive-going ERPs between 500 and 900 ms after a grammaticality violation relative to acceptable sentences for two of our three sentence types. This effect indicates that learners are able to detect grammaticality violations as they process sentences word by word. At the same time, learners did not demonstrate any clear ability to judge grammatical violations correctly at the end of the sentence. 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 be able to demonstrate their command of explicit grammatical rules in formal test situations that are very different from the context of this experiment.

Note that our grammaticality judgment task may have tested only explicit knowledge, or may have tested a combination of explicit and implicit knowledge. If our measure was not a pure measure of explicit knowledge, it is curious that learners did not demonstrate good use of the implicit knowledge that was observed using our ERP measure. It is certainly possible that they do not have reliable access to

that information off-line, and one of our future goals is to determine how second language learners can better use their implicit knowledge to make overt judgments (see “Creating improvements in performance” section below). In either case, we have demonstrated that learners at early stages are implicitly sensitive to some violations of L2 grammar.

The results also provided support for a second idea regarding the early stages of L2 acquisition. This is the prediction, derived from work in the framework of the Competition Model, that L1 syntactic processes will transfer to L2 and compete with these processes on-line. These predictions were borne out most clearly for sentence with determiner number violations that were different between English and Spanish. English speakers 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. They may detect the number of the article, since it is easy to decode, 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 indeed learn to make the article agree with the noun. 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 has no system of grammatical gender. It is not that learners think they can ignore gender on the pronoun. Rather they have no idea at first 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. The physical profile of the ERP results for sentences with gender agreement violations point toward multiple source generators, perhaps located in temporal and inferior frontal cortex. These areas

may be involved in the detection of the mismatch between the article which may be encoded more anteriorly and the noun, which may be encoded in temporal cortex. Interestingly enough, despite the clear cortical reactivity subjects present for gender agreement violations, their grammaticality judgments are at chance. Again, this suggests that they are developing effective implicit processing for L2 in the absence of ability to make use of explicit and/or implicit processes for grammaticality judgments (of gender, in particular).

### Implications

The techniques used in the research may assist in the development of adequate tools to isolate problem areas in second language learning that may inform second language teaching techniques. Indeed, part of the battle for teachers is to identify what students know and what they do not. Further, the proposed techniques may be used to identify ERP markers for learning milestones that can later be applied more broadly to studies of second language learning. If we can better understand the structures to which learners are sensitive, even though their overt behavior may not reflect such sensitivity, then we may be able to assist learners in harnessing this sensitivity such 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/acceptable constructions outside of the sentence context. For example, instead of reading the Spanish equivalent of “\*I walking to school.” they would read “\*I walking”. In addition, feedback is given as to 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 this pilot study show a vast improvement of responses during the interpolated block relative to the first block. In addition, accuracy during the third block is improved relative to that during 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 may 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.

## References

- Coles, M. G. H., Gratton, G., & Fabiani, M. (1990). Event-related brain potentials. In J. T. Cacioppo, & L. G. Tassinary (Eds.). *Principles of Psychophysiology: Physical, Social, and Inferential Elements* (pp. 413-455). New York:: Cambridge University Press.
- DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22, 499-533.
- Ellis, N. (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.
- 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? Chapter to appear in J. F. Kroll and A. M. B. De Groot, Eds., *Handbook of Bilingualism: Psycholinguistic Approaches*. New York: 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. (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. D. (1994). The input hypothesis and its rivals. In N. C. Ellis, Ed., *Implicit and explicit*

*learning of languages*. San Diego: Academic Press.

Kroll J. F., & Tokowicz, N. (in press). Models of bilingual representation and processing: Looking back and to the future. Chapter to appear in J. F. Kroll and A. M. B. De Groot, Eds., *Handbook of Bilingualism: Psycholinguistic Approaches*. New York: 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.

MacWhinney, B. (in press). A unified model of language acquisition. Chapter to appear in J. F. Kroll and A. M. B. De Groot, Eds., *Handbook of Bilingualism: Psycholinguistic Approaches*. New York: Oxford University Press.

MacWhinney, B., & Bates, E. (Eds.). (1989). The crosslinguistic study of sentence processing. New York: Cambridge University Press.

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

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

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

- Osterhout, L., McLaughlin, J., Inoue, K., & Loveless, J. (2000). *Brain responses in L2 learning*. Paper presented at the meeting: Bilingualism: From basic research to educational practice, University of Trento, Rovereto, Italy.
- Psychological Software Tools, Inc. (2000). E-Prime [Computer software]. Pittsburgh, PA.
- 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.
- Schnyer, D. M., Allen, J. J. B., Kaszniak, A. W., & Forster, K. I. (1999). An event-related potential examination of masked and unmasked repetition priming in Alzheimer's Disease: Implications for theories of implicit memory. *Neuropsychology*, *13*, 323-337.
- 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). Maturation 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.

### Author Note

Natasha Tokowicz, Learning Research and Development Center and the Center for the Neural Basis of Cognition, University of Pittsburgh; Brian MacWhinney, Department of Psychology, Carnegie Mellon University.

This research was supported by a National Institutes of Health Individual National Research Service Award (NIMH HD42948-01) awarded to NT and a National Institutes of Health Institutional National Research Service Award (T32 MH19102) awarded to BM. We thank Dayne Grove, Kwan Hansongkitpong, Katie Keil, Lee Osterhout, Chuck Perfetti, Kelley Sacco, Alex Waid, and Eddie Wlotko for their assistance with this project. A portion of these results were presented at the Forty-Third Annual Meeting of the Psychonomic Society (2002, November).

Correspondence concerning this article should be addressed to Natasha Tokowicz, 642 Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA 15260. Electronic mail may be sent to NT at [tokowicz@alumni.umass.edu](mailto:tokowicz@alumni.umass.edu) or to BM at [macw@cmu.edu](mailto:macw@cmu.edu)



## Footnotes

1. It would be ideal to include a native Spanish-speaking control group that could demonstrate that native speakers are equally sensitive to all of the violations. However, the sample we of native speakers 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 cross-language similarity which would bias the results.

2. In future research, it would be advantageous to consider including an additional grammaticality judgment task with more traditional timing, run after the ERP recording session. However, care should be taken to not repeat the items as repetitions may alter processing. This procedure would allow the comparison of the two judgment tasks.

3. We also conducted our analyses using an adaptive mean which 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.

Table 1. Sample stimuli.

Construction	Similarity	Example/Translation
Spanish Sentences		
Tense	Similar	Su abuela * <u>cocinando/cocina</u> muy bien. <i>His grandmother *cooking/cooks very well.</i>
Determiner gender	Unique	Ellos fueron a *un/una <u>fiesta</u> . <i>They went to *a (m.)/a (f.) party.</i>
Determiner number	Different	*El/Los <u>niños</u> están jugando. <i>*The (s.)/the (pl.) boys are playing.</i>
English Sentences		
Subject-verb	N/A	The boys *makes/make excellent ice cream.
Reflexive	N/A	The children enjoyed *himself/themselves.
Tense	N/A	His grandmother *cooking/cooks very well.

Note: Critical words are underlined. English translations of the Spanish sentences are shown in italics.

## Figure Captions

Figure 1. Event-related brain potentials elicited by syntactic (e.g., The cat won't *eating*.) anomalies (adapted from Osterhout & Nicol, 1999) recorded from an electrode at the surface of the scalp over parietal lobe at the vertex (Cz). Onset of the critical words is indicated by the vertical bar.

Figure 2. Time line of events during trials. The sequence begins with a fixation cross which disappears when the participant presses the space bar. It ends when the participant makes a judgment in response to the question-mark probe.

Figure 3. Electrodes used in the ERP analyses.

Figure 4. Grand average waveforms for the Spanish acceptable and unacceptable sentences overall. 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 tense omission (similar) acceptable and unacceptable sentences.

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

Figure 7. Grand average waveforms for the Spanish determiner gender (unique to L2) acceptable and unacceptable sentences.

Figure 8. Mean amplitudes for the cross-language similarity by acceptability interaction in the early P600 (500-700 ms post-stimulus) time window.

Figure 9. Mean amplitudes for the cross-language similarity by acceptability interaction in the mid P600 (700-900 ms post-stimulus) time window.

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

Figure 11. Grand average waveforms for the English tense 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.

Figure 15. Accuracy by condition in English.

Figure 1

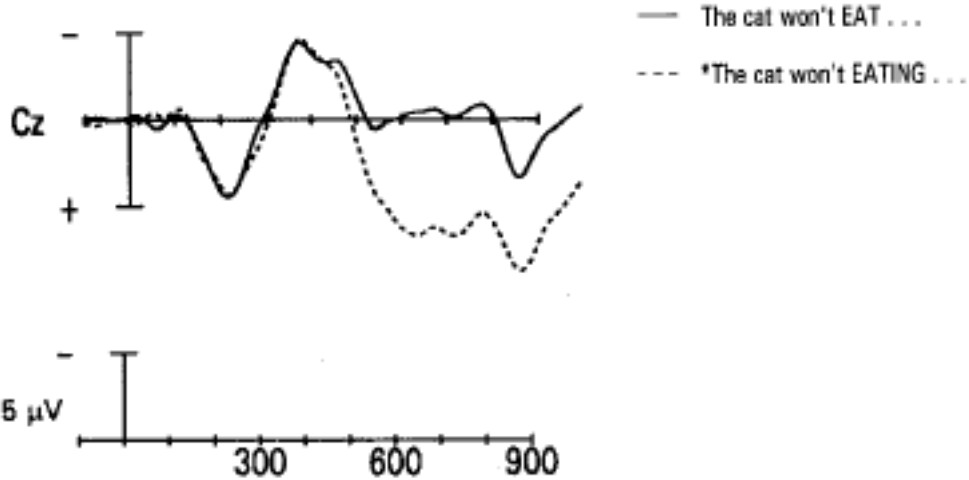


Figure 2

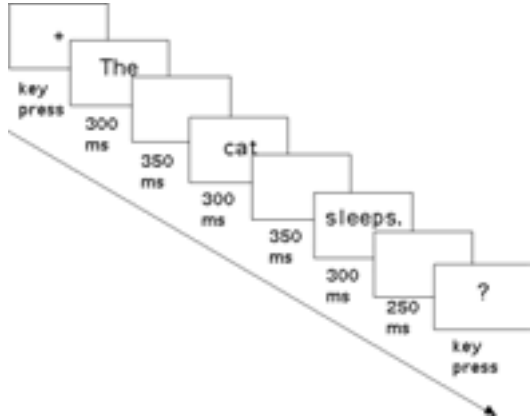


Figure 3

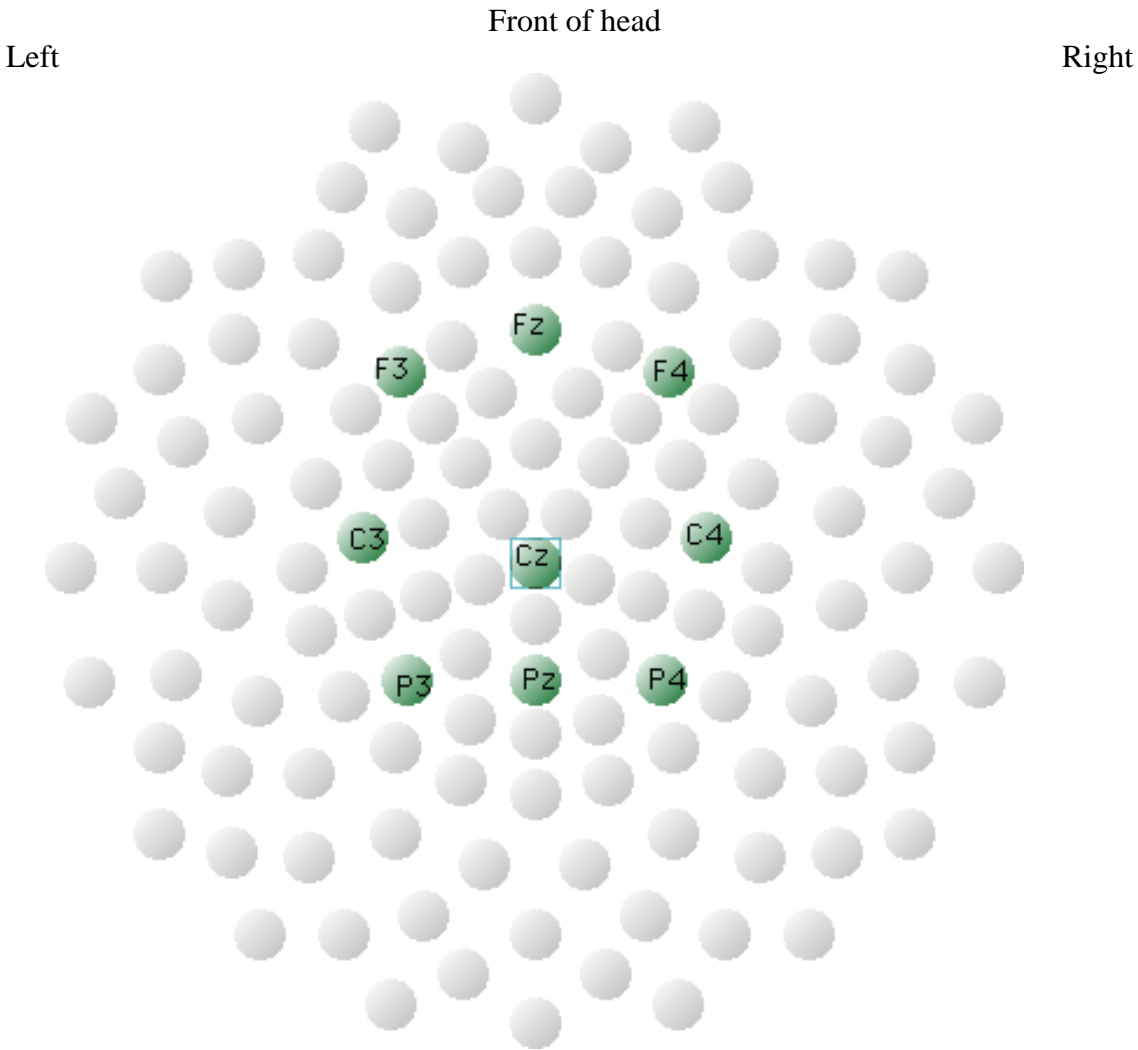


Figure 4

■ acceptable  
■ unacceptable

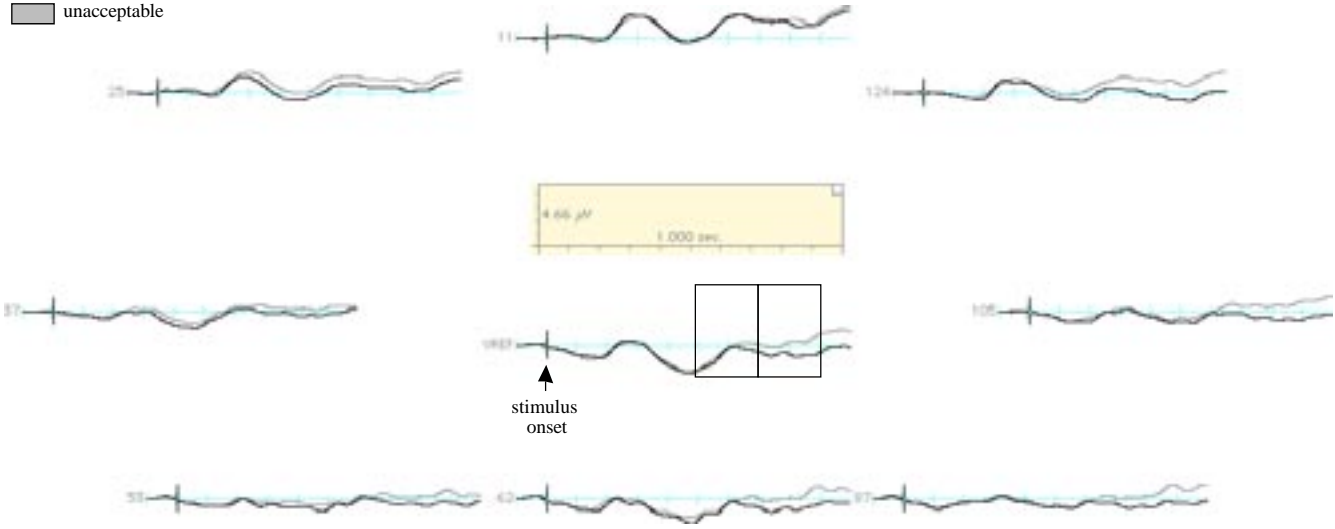




Figure 5

■ acceptable  
■ unacceptable

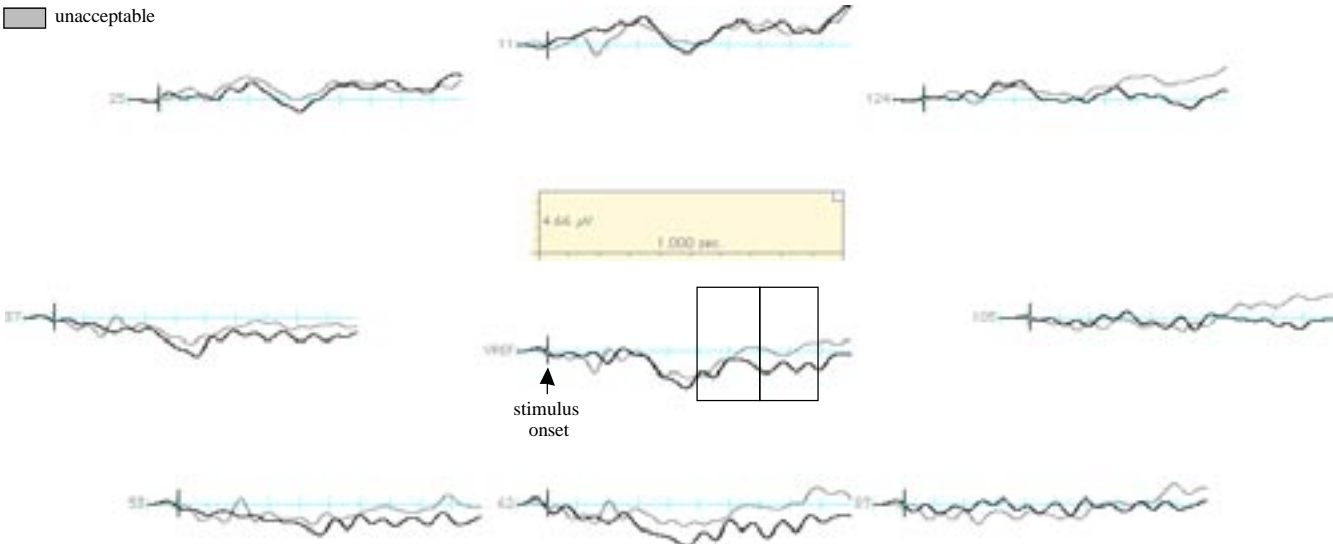


Figure 6

■ acceptable  
■ unacceptable

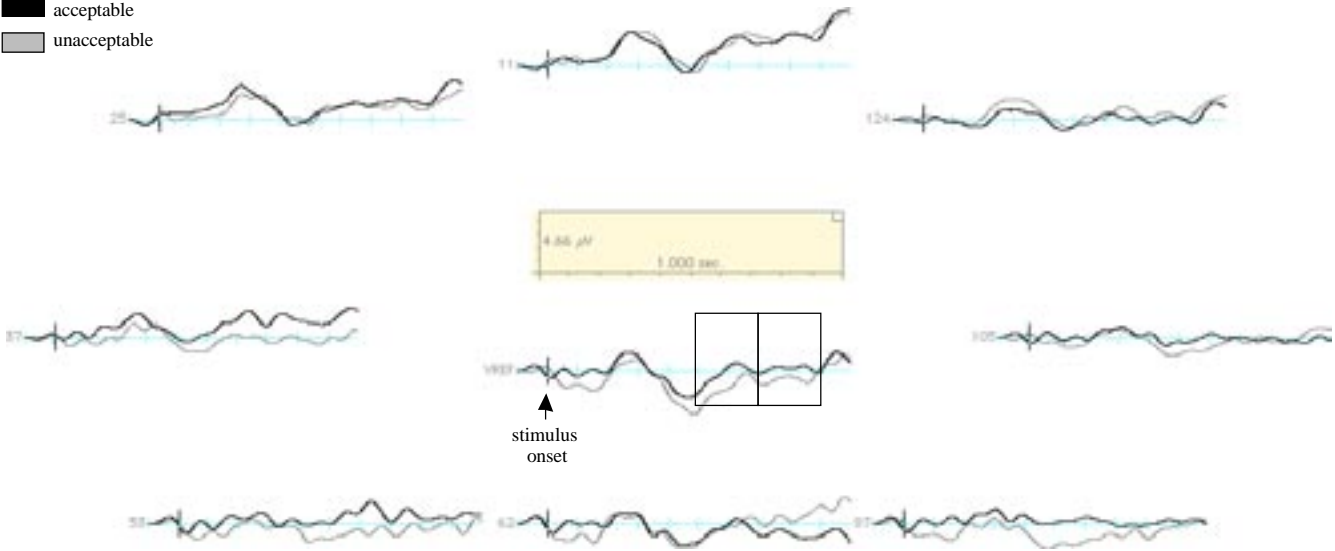


Figure 7

■ acceptable  
■ unacceptable

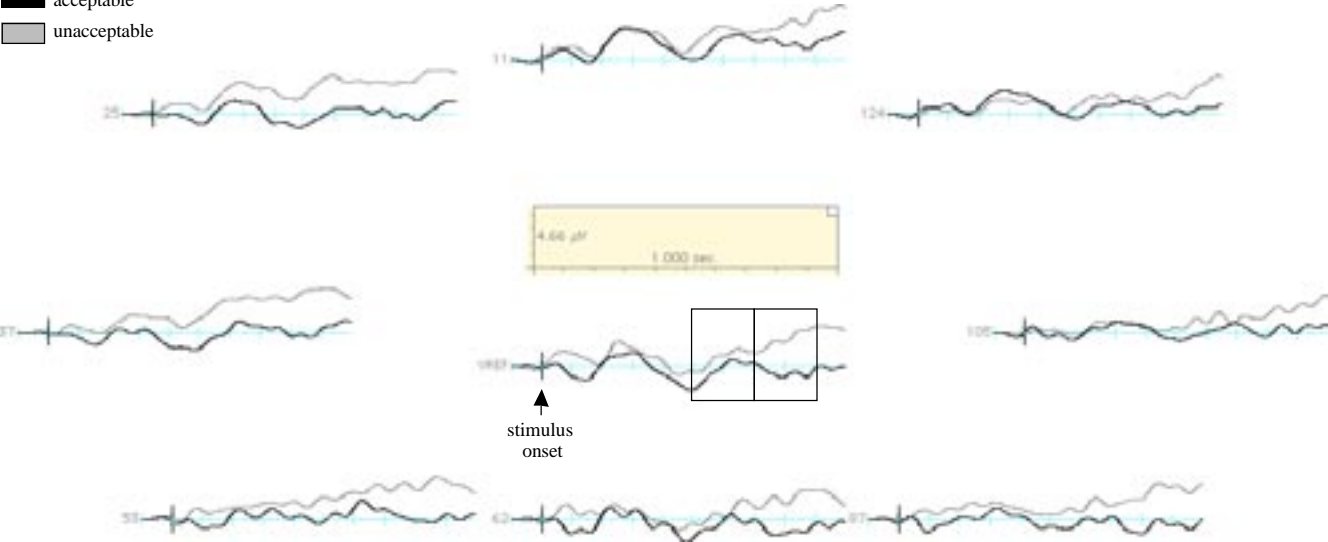


Figure 8

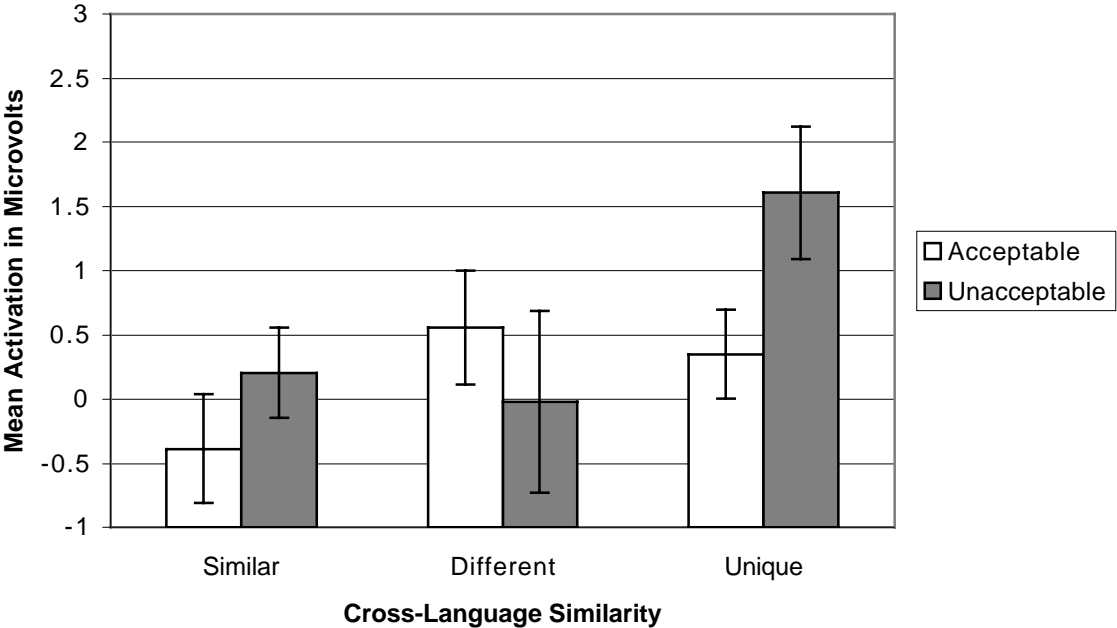


Figure 9

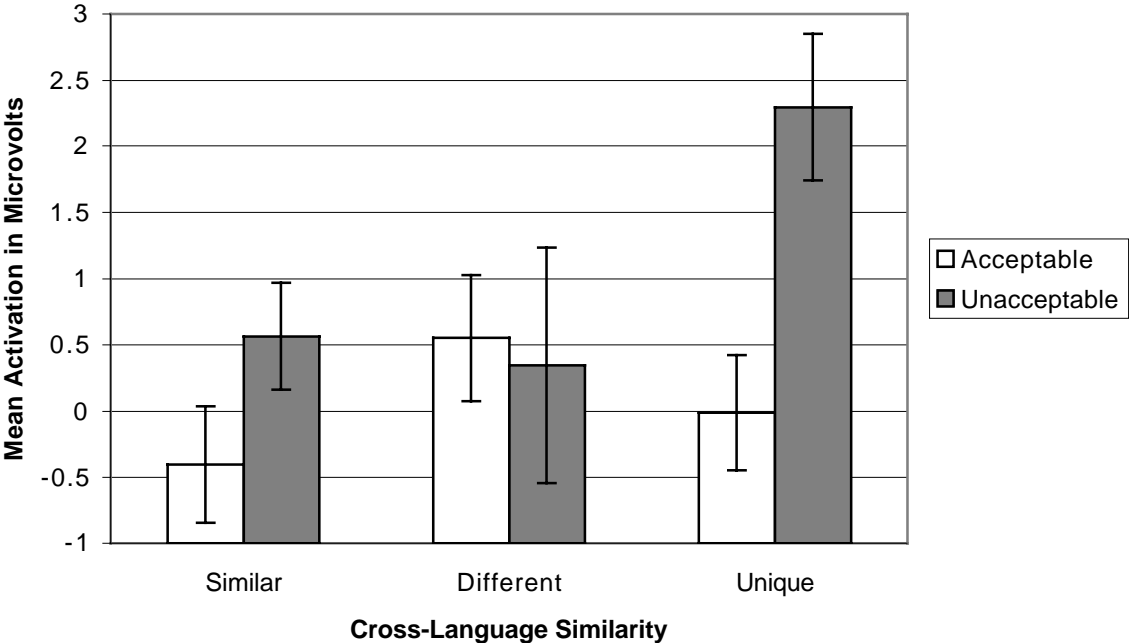


Figure 10

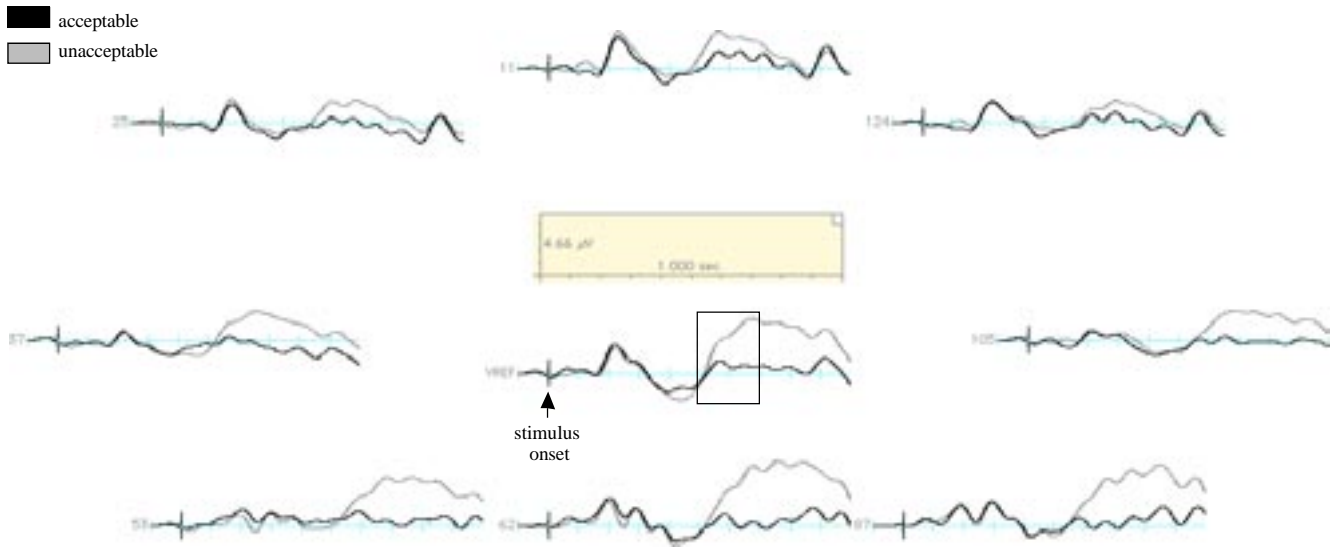


Figure 11

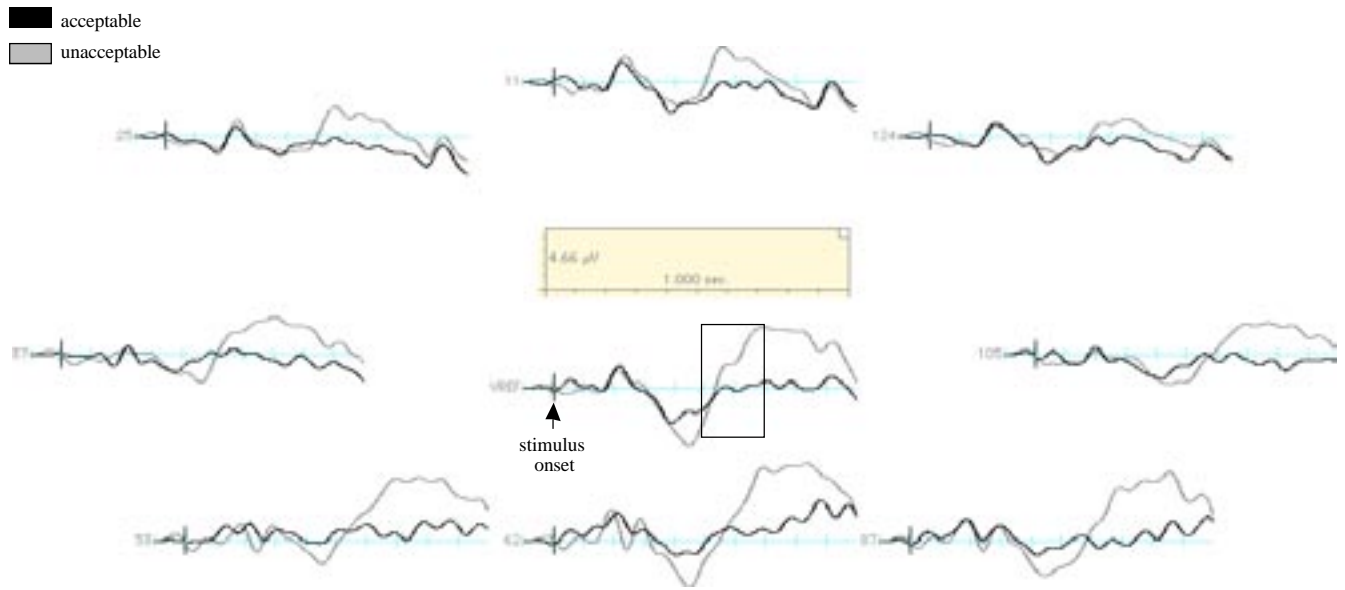


Figure 12

■ acceptable  
■ unacceptable

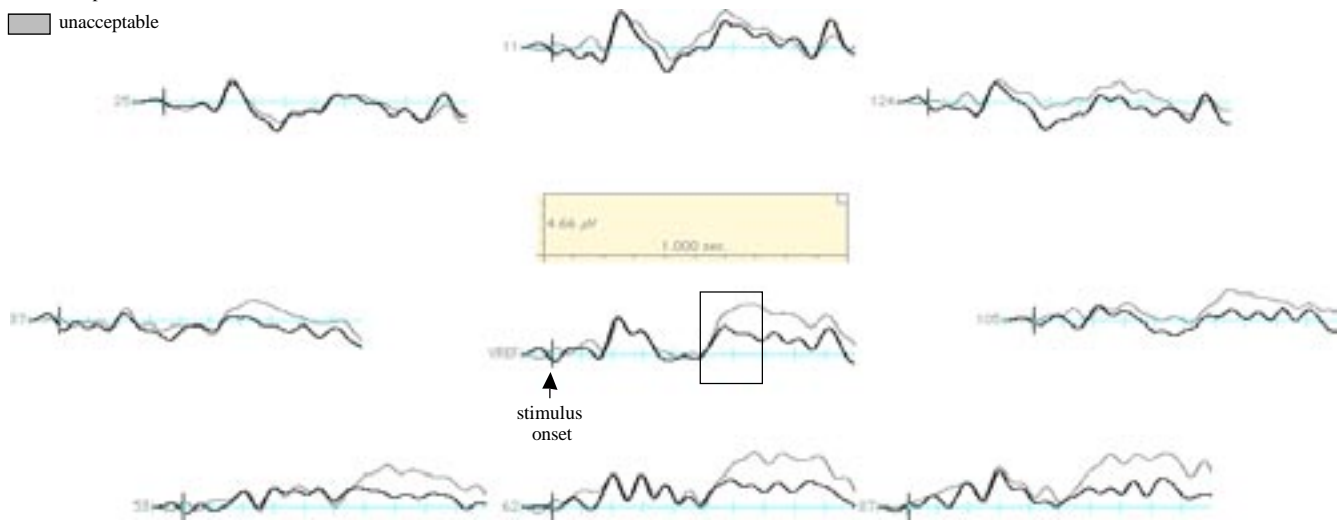




Figure 13

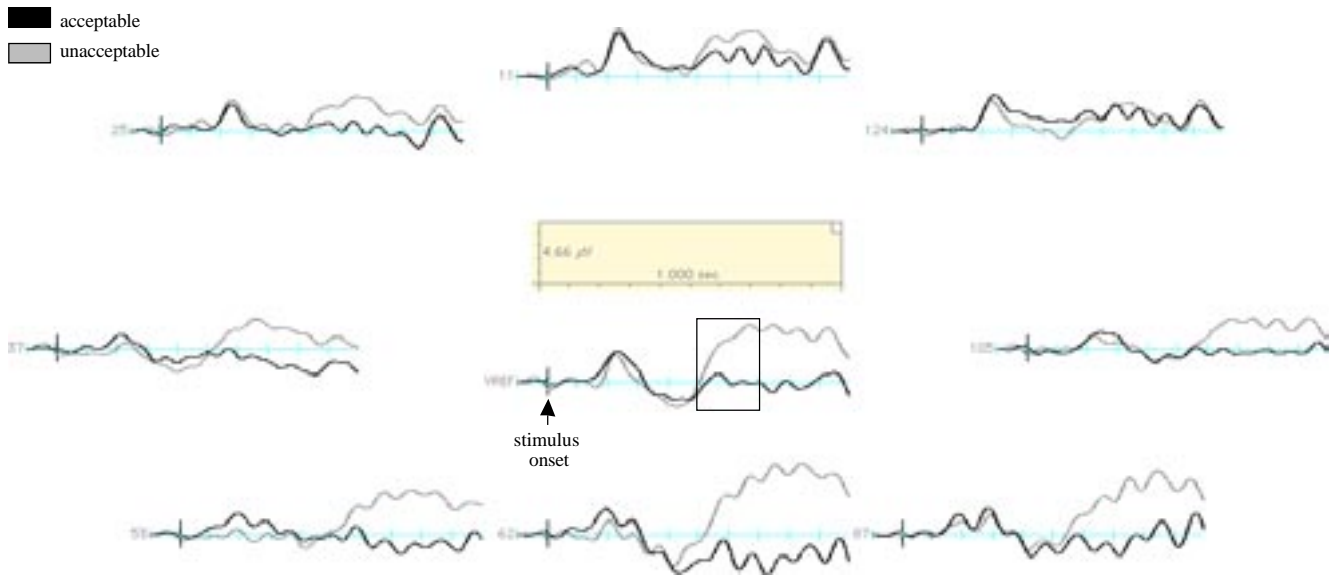


Figure 14

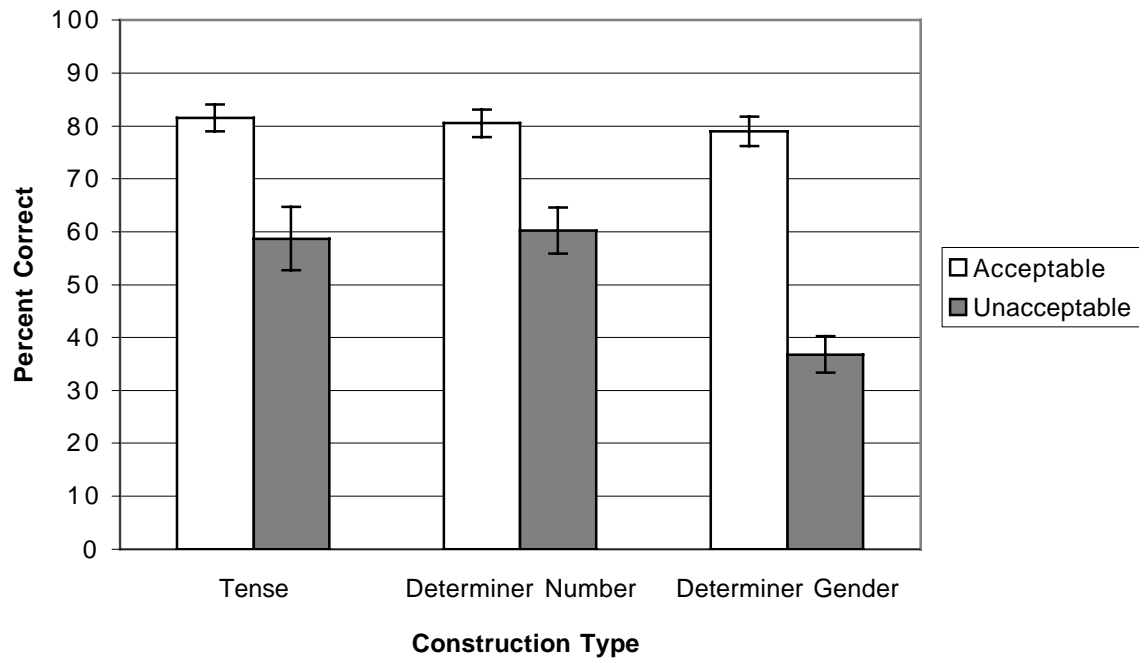


Figure 15

