Are you sure?: Using the error-related negativity to examine adult L2 learning

by

Sarah Blackstone

Submitted to the University Honors College

of the University of Pittsburgh in partial fulfillment

of the requirements for the degree of

Bachelor of Philosophy

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH
University Honors College

This thesis was presented

By

Sarah Blackstone

It was defended on

April 9, 2012

and approved by

Charles Perfetti, University Professor at the University of Pittsburgh, Department of Psychology

Alan Juffs, Associate Professor at the University of Pittsburgh, Department of Linguistics

Carrie Jackson, Assistant Professor at The Pennsylvania State University, Department of Germanic and Slavic Studies

Thesis Director: Natasha Tokowicz, Associate Professor at the University of Pittsburgh, Department of Psychology

ABSTRACT

Are you sure?: Using the error-related negativity to examine adult L2 learning.

Sarah Blackstone, BPhil

University of Pittsburgh, 2012

Previous investigations of second language (L2) learners' proficiency have focused on explicit measures of overt responses. Recent data have shown discrepancies between L2 learners' overtly measured behaviors and covertly measured implicit processes (McLaughlin, Inoue, & Loveless, 2000). Event-related potentials (ERPs) have been used as a covert measure of implicit sensitivity. Prior studies have focused on the P600 component as a measure of sensitivity to syntactic violations in L2 (Tokowicz & MacWhinney, 2005; Tolentino & Tokowicz, 2012). Tokowicz and MacWhinney (2005) used the P600 to index cross-language similarity effects, and found that GJT accuracy scores were lowest in conditions with features unique to L2, however, the ERP responses were the strongest to unique features. This suggests that L2 learners possess implicit sensitivity to L2 violations that is not always indicated by their overt behaviors.

The present study looks at another ERP component, the error-related negativity (ERN), which is elicited in response to error processing (Sebastian-Gallés, Rodríguez-Fornells, Diego-Balaguer, & Díaz, 2006). We reprocessed previously collected ERP data from Tolentino and Tokowicz (2012; unpublished data) to see if an ERN is present, which would indicate that L2 learners are sensitive to L2 violations. The ERN will be investigated as a function of stimulus grammaticality, response accuracy, electrode site and laterality. We found a significant four way interaction between these variables, as well as significant contrasts in mean amplitudes at certain electrodes. Additionally, we found a positive component elicited in response to errors made in the judgment on ungrammatical stimuli, suggesting the context and the type of error influences how errors are processed. Overall, our data indicate that L2 learners are sensitive to L2 violations, and are at some level aware, not only of what is grammatical, but also what is ungrammatical.

Key Words: second language processing; second language learning; ERN; ERP

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

Thank you to my committee members: Dr. Chuck Perfetti, Dr. Alan Juffs, and Dr. Carrie Jackson, and to my committee chair Dr. Natasha Tokowicz. Also thank you to Leida Tolentino for permission to reprocess the data from Tolentino and Tokowicz (2012; unpublished data), as well as to members of the University of Pittsburgh PLUM lab for their feedback and support.

1.0 INTRODUCTION

Many studies have investigated the proficiency of adult second language (L2) learners using grammaticality judgment tasks, in which people are presented with a stimulus and asked to judge whether it is grammatical or ungrammatical. The results, unsurprisingly, reveal that adult L2 learners do not perform as well as native speakers. MacWhinney's (1997) competition model explains that L2 learners experience transfer from L1 to L2, and that even proficient L2 speakers use cues from their L1 to process L2 (Bates & MacWhinney, 1981). The grammaticality judgment tasks that are used to asses this simply ask the participant to make a yes or no judgment about grammaticality. However, recent evidence suggests that yes/no grammaticality judgments may not always be sensitive enough to reveal to what L2 learners know (e.g., Tokowicz & MacWhinney, 2005). Instead, measurements of brain responses such as *event-related brain potentials* (ERPs) may yield more sensitive results regarding syntactic knowledge. Here, we re-examine previously collected data (Tolentino & Tokowicz, 2012; unpublished data) to determine whether an ERP component associated with erroneous responses (the error-related negativity, or ERN) yields additional information regarding L2 learners' sensitivity to their errors.

1.1 EVENT RELATED POTENTIALS

ERPs are one of the convergent measures used to investigate the difference between L1 and L2 processing. ERPs are typically used to measure participants' brain activity while they perform a task, such as the grammaticality judgment task. ERP components such as the N400 and P600 have been used to better understand language processing. Here, we focus on the P600 component, which has been hypothesized to index syntactic reanalysis, and the difficulty of syntactic integration. This component is a positive-going wave that peaks at approximately 500-900ms after the presentation of a morphosyntactic error, and typically has a centro-parietal scalp distribution.

Kaan, Harris, Gibson, and Holcomb (2000) found a P600 component in L1 processing of correct garden path sentences, which are temporarily ambiguous because they contain a word group that appears to be compatible with more than one structural analysis (e.g. "the man who bought Jane flowers yesterday"). The P600 has been used to study processing in bilinguals as well. For example, Hahne (2001) found a P600 component in Russian-German bilinguals in response to phrase structure violations in German. In this sense, phrase structure violations refer to morphological and syntactic changes that render a phrase or sentence ungrammatical. Compared to native German speakers, the P600 in the bilinguals was delayed to 950 ms, as opposed to 650 ms in monolinguals. Similarly Hahne and Friederici (2001) found that Japanese-German bilinguals do not show a P600 in response to phrase structure violations, whereas native speakers do. It was suggested that the P600 was present in Hahne's Russian-German bilinguals because they were more proficient and have higher accuracy than Hahne and Friederici's Japanese-German speakers. This suggests that the P600 component gets larger with language proficiency and can be used to measure the degree at which one is able to identify and process syntactic errors. Hahne and Friederici also found that L2 learners showed little difference between the neural responses evoked in regards to sentences with and without a syntactic violation. Conversely, native Russian speakers showed a P600 modulation in response to German errors, whereas native Japanese speakers did not. This suggests that native Japanese speakers who are L2 learners of German were not as sensitive to grammatical structures present only in German. For example, Russian and German similarly form structures containing grammatical violations preceded by a preposition, whereas these structures are not possible in Japanese.

Studies have also used the P600 to examine speakers' sensitivity to gender agreement violations. Sabourin and Stowe (2008) measured ERPs in response to verb inflection violation and gender agreement violations. The participants were native Dutch, German, and Romance language speakers. They were presented Dutch sentences containing grammaticality violations. Because gender agreement in Dutch and German works differently than in Romance languages, native speakers of Romance languages could not rely on cues from their L1 to help with Dutch gender agreement. They instead had to map gender individually to each lexical item. Despite similar levels of proficiency overall, only Dutch and German speakers showed a P600 component in response to gender violations, suggesting that cross-language similarity and language overlap improved ability to detect gender agreement.

Tokowicz and MacWhinney (2005) address the issues of whether L2 learners are able to carry out implicit online processing, and the extent to which L1 transfer and competition modulate differences in behavioral and electrophysiological responses. They hypothesized that L2 learners would show less implicit or automatic sensitivity to constructions that differ between the two languages as opposed to constructions that are similar in both languages, or are unique to L2. Similar constructions focused on auxiliary omission, which is ungrammatical in English and Spanish (*His grandmother cooking well; *Su abuela cocinando muy bien). Different constructions were focused on determiner-noun number agreement, with Spanish containing singular and plural determiners, and English only one (*the* book, *the* books; *el* libro, *los* libros). Unique constructions focused on determiner-noun gender agreement, which is not used in English, and is therefore unique to the L2, Spanish. Tokowicz and MacWhinney (2005) studied native English speakers in the beginning stages of Spanish L2 learning. Participants were asked to judge whether sentences were syntactically appropriate while electrical brain activity was recorded. Tokowicz and MacWhinney predicted that L2 learners would not show a P600 component in response to ungrammatical constructions that are different in English and Spanish. However, they predicted that learners would display sensitivity to violations of the similar and unique constructions because these, unlike the different constructions, would not suffer from competition between the two languages. The results indicated that there was sensitivity in the similar and unique conditions, with the unique condition eliciting a larger P600 component. The results supported the initial hypothesis that L2 learners do not display sensitivity to violations of constructions that are different between the two languages, and that L2 learners in the beginning stages are sensitive only to certain types of violations that depend on the similarity between the L1 and L2.

One interesting aspect of this study is the comparison of the behavioral accuracy results with the physiological data. Overall, participants responded least accurately to unique constructions, which showed the strongest implicit responses in the ERP data. Participants' overt accuracy did not display sensitivity to violations, and in fact was the lowest for the condition exhibiting the highest ERP sensitivity. It seems that L2 learners are sensitive to grammar violations during sentence processing, yet do not always demonstrate correct grammatical judgments. These data suggest that learners may have better access to implicit knowledge than explicit knowledge during sentence processing, and that beginning L2 learners indeed show online implicit processing to L2 morpho-syntactic violations. (Note that: the terms implicit and explicit processing and knowledge are controversial in the SLA literature; for discussion see Ellis, (2005).)

Tolentino and Tokowicz (2010) reviewed a series of language studies that used physiologically-based methods to examine how differences between L1 and L2 influence morphological or syntactic processing in L2, based both on the location of active neural tissue and the time course of processing. They focused on the reflection of cross-language similarity in electrophysiological studies of L2. Koltz et al. (2008) showed a comparable P600 effects in native and L2 speakers of English (with Spanish as L1) during processing of syntactic ambiguities, such as transitive-intransitive verb use (*The broker persuaded to sell the stock went to jail*). A P600 component was not found during processing of structures that are similar in English and Spanish (**The broker hoped to sell the stocks was sent to jail).* Another study done by Chen et al. (2007) compared Mandarin Chinese-English speakers' performance to native English speaking control subjects. They focused specifically on number agreement violations (**The price of the car were too high; *The price of the cars were too high),* which is a feature of English that does not exist in Mandarin. Native English speakers showed a P600 effect to ungrammatical sentences whereas L1 speakers of Mandarin did not. These data complement the results of Sabourin and Stowe (2008), in which native Dutch and German speakers showed P600 effects to gender agreement violations, whereas L1 Romance language speakers did not, despite being highly proficient in Dutch.

1.1.1 The Error Related Negativity

As mentioned above, one component of particular interest with regard to error processing and ERPs is the ERN. The ERN is a component of the ERP that has a fronto-central scalp distribution and peaks about 80 ms after an overt incorrect response (Bernstein, Scheffers, & Coles, 1995; Holroyd & Yeung, 2003; Scheffers, Coles, Bernstein, Gehring, & Donchin, 1996). The ERN was originally thought to be associated only with conscious errors. However, recent evidence has revealed that it is present in the cases of unperceived errors as well, making it a good candidate to examine for conscious and unconscious error processing. In a study examining phonological representations, Sebastián- Gallés, Rodríguez-Fornells, Diego-Balaguer, and Diaz (2006) used the ERN to uncover unconscious sensitivity to the difference between real and false L2 words. The aim of the study was to explore the difficulty in learning non-native phonemic contrasts. It is more difficult for adults to learn an L2 because it is difficult to discriminate and produce sounds as a native speaker would. ERPs were used to examine this unconscious processing of L2 words and non-words. Participants were asked to push one button, indicating that they heard a real word, and another button to indicate a non-word. Some of the non-words differed from a real words by two phonemes, both represented orthographically as ‘e,’ with one sounding like an ‘a.’ (real words: *gall**a**da, fin**e**stra;* non-words: *gall**e**da, fin**a**stra*). There were also control words for which the phoneme manipulation created a new word meaning (*llençol, llençal)*. For Catalan speakers, erroneous responses elicited an ERN. Correct responses given for the non-word condition also elicited an ERN because the stimulus itself was erroneous. This could indicate the conflict participants had while differentiating the phonemes. Sebastián- Gallés et al. (2006) argued that the ERN component is related to uncertainty when responding. After an error is made, the stimulus continues to be processed thereby activating the correct response, which competes with the incorrect response that has been given, thus yielding an ERN component.

This leads to an alternative explanation: the ERN component may indicate the amount of conflict that exists when making a response. Yeung, Botvinick and Cohen, (2004) administered a version of the flanker task in which participants responded using key presses to indicate the direction of the central arrow in a group of five arrows. There were four types of stimuli: congruent left, congruent right, incongruent left, incongruent right. In the congruent conditions, the surrounding arrows faced the same direction as the center arrow (right for congruent right, left for congruent left). In the incongruent conditions the surrounding arrows faced the opposite direction of the center arrow. They found an ERN that was most prominent for the incongruent conditions. This suggests that the ERN reflects the conflict that develops after an incorrect response as the information continues to be processed. Their theories are consistent with previous accounts that error detection relies on the stimulus to continue being processed, and that this tends to produce a reliable representation of the correct response within the task processing where the initial error occurred.

fMRI studies have found that the caudal region of the anterior cingulate cortex is active in the correct and incorrect trials in which a response conflict is present (Carter et al.,1998). The conflict appears around the time of the response and peaks approximately 80 ms later. Yeung et al., (2004) describe this conflict in terms of two response units, the incorrect response and the target response. The stimulus continues to be processed after the error which causes the target response to dominate the competition units because the participant realizes that it is correct. Activity in this response unit increases which leads to a brief period in which both incorrect and correct responses are activated, producing a larger conflict signal. In summary, the ERN is explained in terms of the conflict that results in the period following errors due to continued processing of the stimulus, leading to later activation of the correct response. The ERN itself is not explicitly associated with the occurrence of an error, but is a signal of the level of conflict involved in responding.

Another account of the ERN is the idea that it indicates a mismatch signal. In this view the ERN detects errors as mismatches between the actual response and the knowledge of the correct response, which is thought to be derived from continued processing (Coles et al. 2001). This notion is that errors occur when the execution of a response is impulsive and takes place before the stimulus has been fully processed. Therefore, the response selection system has not been able to decide on the appropriate response. The conflict emerges when the stimulus continues to be processed and is matched to the correct response after an error has been produced. One of the ideas of this theory is that error responses do not always indicate a lack of knowledge, but could rather reflect impulsivity in responding.

Bernstein, Scheffers, and Coles (1995) reported that a larger ERN is associated with a greater degree of discrepancy between the actual error response and the correct response. That is, errors that more closely resemble the correct response will elicit a smaller ERN than errors that do not closely resemble the correct response. Scheffers and Coles (2000) continued this investigation by evaluating the relationship between perceived accuracy and the ERN amplitude. Participants performed the flanker task, and at the end of each trial participants were asked if they thought they responded correctly and how confident they were with their self -assessment. Previous research had indicated that errors could be classified into two sub-groups: errors made due to premature responses, and errors made due to data-limited processing. In premature error conditions, the response is made before enough information has been gathered from the target to guide response selection. A large ERN is expected in these conditions because the response was activated before the stimulus presentation, and released before the target had been completely identified. The comparison of the correct representation with the actual response would be unambiguous and should yield a clear error signal. Converging evidence was provided for this idea by comparing that data for the incorrect trials that were judged as 'correct' by the participants, to the correct trials that were accurately identified as correct. In data-limited error conditions, a smaller ERN is predicted because the participant could not acquire enough information from the target to clearly represent that appropriate response. Thus the actual response and the compromised representation of the appropriate response would result in a partial mismatch and a medium amplitude ERN. Scheffers and Coles found that on average, participants' beliefs about the accuracy of their responses were correct. Using a subset of participants from this study Scheffers and Coles analyzed the ERN amplitude to see if their perception of the accuracy of their answer influenced the ERN. The results indicated that on incorrect trials, if participants had a stronger belief that they answered incorrectly, the ERN amplitude was larger in comparison to other trials. This supports the idea that there is an association between loss of confidence and an increased ERN. The data revealed a systematic relationship for correct and incorrect trials between the accuracy and the ERN; the smallest ERN was associated with trials that the participants rated as correct, and larger ERNs were found when participants believed more strongly that they had answered incorrectly. This analysis supported the idea that the ERN is associated with error detection during performance.

Although several theories behind the ERN have been proposed, many studies have referred to the ERN in association with error processing in action monitoring, a control mechanism used to inhibit and correct an error (Desmurget & Grafton, 2000; Rodríguez-Fornells et al., 2002). When the wrong selection of a motor command is generated, a copy of the response is produced and compared to the correct representation. An error signal is generated when there is a mismatch between the response and the representation (Coles, Scheffer, & Holroyd, 2001). Other studies have shown that the ERN appears in response to verbal monitoring errors (Sebastián-Gallés et al., 2006). Ganushchack and Schiller (2009) examined the relationship between the ERN and verbal monitoring in a nonnative language, by studying the difference in the amplitude of the ERN in response to phonemic errors made under time pressure and errors made in non- pressure situations. All participants were German-Dutch bilinguals. Participants were given target phonemes to monitor and then shown a series of pictures in which they indicated the presence of one of the target phonemes. Participants were tested individually and asked to carry out a learning phase, in which they were familiarized with the pictures and their corresponding names, a practice block, a picture naming task, and then either the time pressure or no-time pressure condition. Prior to the practice and experimental blocks they heard a sample of the phoneme they were required to monitor. Participants were then instructed to press a button indicating whether the target phoneme was present in the picture task. Ganushchack and Schiller observed an increase in the ERN amplitude under time pressure. This finding was contrary to previous reports that found decreased ERN amplitudes in action and verbal monitoring tasks. Presumably, this is because under time pressure there may not be enough time to make an optimal comparison between the intended and actual response, which lowers the ERN amplitude.

However, this decreased ERN was seen in monolinguals, and Ganushchack and Schiller tested bilinguals. The increased ERN found in this study could be due to interference from the dominant language in L2 processing conditions. In summary, this study showed that the ERN can be elicited by errors of verbal monitoring and is sensitive to verbal manipulation.

1.1.2 The Pe

Another ERP component that is relevant to response and error processing in the Pe, which is a positive component associated with incorrect responses (Herrmann, Rommler, Ehlis, Heidrich, & Fallgatter, 2004). The Pe shows a parietal distribution that peaks significantly between 200 and 500 ms at Fz and FCz (Carp 2009) after an incorrect response, with significant activation in the rostral anterior cingulate cortex (ACC) and superior parietal cortex. Unlike the ERN, the Pe is more pronounced for perceived than unperceived errors, which may indicate conscious processing of the error context. For the Pe, higher activation was found in the anterior cingulate cortex for incorrect responses as opposed to correct responses. Although the two components shared overlapping regions, they clearly originate from two different sources, with the ERN originating in the medial and middle frontal gyrus, and the Pe in the rostral part of the ACC. The data suggest that these two components have different neural generators and therefore may represent two different aspects of error processing (Herrmann, et al. 2004).

1.2 THE PRESENT STUDY

The present study re-examines previously collected data from Tolentino and Tokowicz (2012; unpublished data) that were re-analyzed to examine the ERN in relation to L2 processing. Tolentino and Tokowicz investigated the relationship between implicit and explicit L2 processing in native English speakers learning Spanish by measuring the brain activity of participants as they made grammaticality judgments of Spanish sentences. For the purposes of this study, explicit and implicit are differentiated, respectively, by the extent to which someone is consciously aware of the regular patterns in information processing, and the extent to which one is unable, or able, to verbalize such patterns. Additionally, it is thought that explicit knowledge requires more controlled processing, whereas implicit knowledge occurs automatically. The most important distinctions are the level of conscious awareness and the amount of effort put into the processing of information. The goal of this study was to investigate whether similarities between L1 and L2 influenced implicit and explicit processing. Influences of cross-language similarity and the effects on processing were examined based on a continuum of the similarity in sentence violation agreements from similar to English to not similar. Three types of stimuli were constructed from this; words that mapped from the L1 to L2 without conflict were placed in the 'Similar' category. If the L1 and L2 translations did not directly correspond the category of the stimulus was 'Different.' Finally, if the L2 word could not be mapped to an L1 translation the stimulus was 'Unique.'

Similar constructions were tested using the demonstrative determiner because it is used similarly in both English and Spanish (e.g., "this" shirt, "these" shirts vs. "esta" camisa, "estas" camisas). Different constructions were tested with definite determiners, which are the same in

English regardless of number, but change in Spanish based on whether the object is singular or plural (e.g., “the” shirt, “the” shirts vs. “la” camisa, “las”camisas). Unique constructions consisted of the gender determiner which is not present in English and is present in Spanish. (e.g., “the” book vs. “el” libro, “la” mesa). Participants were shown sentences containing constructions from the categories previously listed and were asked to make grammaticality judgments of the sentences that were presented.

Tolentino and Tokowicz (2012; unpublished data) used a three block design so that performance before and after an accuracy-increasing intervention (Block 2) could be examined. In Blocks 1 and 3, all participants saw Spanish sentences and received no accuracy feedback for their grammaticality judgments. The data analyzed here came from two condition representing different types of stimuli seen during Block 2: Spanish sentences with no accuracy feedback (SNF), and word pairs with accuracy feedback (WPF). However, here we only analyze Block 1 data during all participants saw comparable stimuli: sentences without feedback.

Tolentino and Tokowicz (2012) compared sensitivity of the grammaticality judgments and P600 effects to violations before and after accuracy improvement, to investigate the relationship between behavioral and ERP L2 processing. The results of this study revealed that accuracy improvement was accompanied by an increased P600, which suggests the two processes share similar underlying mechanisms. Further analysis revealed that both Block1 and Block 3 accuracy were higher for the ‘Similar’ and ‘Different’ conditions than for the ‘Unique’ condition. Sensitivity was high in the ‘Similar’ condition and slightly lower in the ‘Different’ condition, and was not present in the ‘Unique’ condition. The consistently lower performance in the gender agreement (unique) condition suggests that L2 learners have not yet acquired grammatical features unique to L2. The participants did not show a P600 effect in the unique condition, which contrasts from the similar and different conditions, and suggests an influence of L1 on L2 processing. The behavioral accuracy increase and P600 increase were significant only in the unique construction, suggesting that the relationship between accuracy and the P600 component is, in part, triggered by unique constructions. This may have been related to more variability in the responses for the unique condition, due to the grammatical feature not being fully acquired. Additionally, because the unique constructions do not rely on L1 modulation, it is possible that experimental inputs caused the changes in behavior results and ERP components.

Because Blocks 2 and 3 of the Tolentino and Tokowicz (2012; unpublished data) studies had higher accuracy than the initial block, there were only a sufficient number of errors for analysis in Block 1. Furthermore, although examining ERNs for the three cross-language similarity conditions would be very interesting, because the similar condition yielded very few errors, we could not examine these differences and instead collapsed over cross-language similarity condition.

Based on the previous theories of the ERN, such as the conflict hypothesis, it is possible that the presence of an ERN on error trials in grammaticality judgments of L2 learners could indicate that they are aware of their mistakes and have a better knowledge of the language than other data suggest. Using the data from Tolentino and Tokowicz (2012; unpublished data) we can examine this more closely. Participants' responses stating whether a sentence in the L2 was grammatically correct or incorrect can yield data supporting the conflict theory of the ERN, or the idea that some of the errors may have resulted from impulsive execution. Regardless of whether an ERN is caused by conflict or impulsivity, previous evidence has indicated an ERN should be present for both conscious and unconscious errors. Using the Block 1 data we can examine whether the grammaticality of the initial stimulus is reflected in the elicited neural component.

1.2.1 Goals of the Present Study

The goal of this study was to reanalyze the ERP data collected by Tolentino and Tokowicz (2012; unpublished data) for an ERN to see if L2 learners are more sensitive to grammatical violations that their behaviors indicate. If learners in the study are implicitly sensitive to L2 grammar, we would expect to see an ERN based on the conflict hypothesis because the learner will continue to process the information after responding, and the incorrect response will be reevaluated. Additionally, under the representational mismatch theory, we would expect to see an ERN because of incongruencies between the internal representation of the correct response, and the erroneous response.

This research uses a novel application of the ERN that is the beginning of a larger body of work focusing on L2 acquisition. Revisiting the ERP data from Tolentino and Tokowicz (2012; unpublished data) could reveal that simply asking if a sentence is grammatical or not is only one way to estimate how well someone is acquiring a language, and we may gain more sensitivity by also observing brain activity that follows such judgments. The presence of an ERN component could show that adults do indeed recognize that the sentences are ungrammatical but pushed the incorrect button in a split-second decision, giving the impression that their knowledge of the language is worse than it actually is. And, although Tolentino and Tokowicz examined the P600 responses in their study, those measurements were taken earlier in the sentence, so the ERN provides a measure of knowledge at the point of response execution.

2.0 METHODS

2.1 PARTICIPANTS

Tolentino and Tokowicz (2012) collected data from thirty (7 male) Spanish learners enrolled in semesters 1-4 at the University of Pittsburgh and Carnegie Mellon University. For the present study, data from sixteen participants were used in the final analyses. Each participant completed one session and received $10 compensation, or credit for an Introduction to Psychology course. All participants had normal, or corrected to normal vision, were right handed, had no implanted brain devices, and were not taking psychoactive drugs. Additionally, they had not been exposed to a language other than English before the age of 13. See Table 1 for participants' background information.

Table 1. Participants' Background Information. N = 16.

Measure	Mean (SD)
Age (years)	18.63 (.81)
Age at first L2 exposure (years)	14.84 (1.36)
L2 learning period (years)	4.25 (1.18)
L2 self-ratings (10-point scale)	
Reading	5.81 (1.64)

Writing	5.00 (1.71)
Speaking	4.63 (2.24)
Listening	5.31 (2.39)
Immersed	
Yes	4
No	12
Class Level	
Spanish 1	5
Spanish 2	6
Spanish 3	3
Spanish 4	2

NOTE: Self-ratings were made on a scale on which 1 indicated the lowest level of performance and 10 indicated the highest level of performance

2.2 DESIGN

Tolentino and Tokowicz (2012; unpublished data) used a within-subjects design with 3 cross language similarity types (similar, different, unique) X 2 grammaticality conditions (grammatical, ungrammatical) X 3 electrode laterality (left, midline, right) X 3 electrode site (frontal, central, parietal). The participants were presented whole sentences is Spanish and asked to make a grammaticality judgment. Each participant completed three blocks of this experiment, but data reported here are only from Block 1. See Table 2 for sample stimuli from Block 1.

Table 2. Sample Stimuli from Block 1

Condition	Stimulus	Gloss
Similar	Este/*Estos lago es vasto	This/*These lake is vast.
Different	El/*Los lago es vasto.	The(sing)/*The(plur) lake is vast.
Unique	El/*La lago es vasto.	The(masc)/ *The(fem) lake is vast.

For the present study we used a 2 grammaticality (grammatical, ungrammatical) x 2 response accuracy (correct, incorrect) x 3 electrode laterality (left, midline, right), x 3 electrode site (frontal, central, parietal) within-subjects design.

There were four different conditions that we observed, from: correct grammatical trials (gramCORRECT); erroneous grammatical trials (gramERROR), correct ungrammatical trials (ungramCORRECT); and erroneous ungrammatical trials (ungramERROR). Here, the grammaticality refers to the grammaticality of the stimulus and the accuracy refers to the accuracy of the grammaticality judgment.

2.3 DATA REPROCESSING

ERP data were reprocessed using Neuroscan Edit 4.3 software. All electrodes were rereferenced to the mastoids, and filtered using a 30 Hz low-pass filter. Trials that contained large artifact were excluded from analysis. The ERP epoch ranged from -100ms -1000ms (100ms before the response until 1000ms after the response). This would allow the brain activity before the presentation of the stimulus and response to be compared to the later activity after the response, while the stimulus is still being processed. Data were baseline corrected to the pre-response baseline.

The average for a sequence of single sweeps was then computed for each participant. In total four averages were taken for each participant: gramCORRECT, gramERROR, ungramCORRECT, and ungramERROR. Correct and error responses were averaged in grammatical and ungrammatical sentences presented in Block 1. These were averages of all the correct and error responses made across every condition.

2.4 DATA ANALYSIS

2.4.1 ERP Data Analysis

A subset of nine electrodes was used, sufficiently covering the frontal, central and parietal regions of scalp distribution involved in ERP components related to linguistic error processing. The subset corresponded to the 10-20 Electrode System (Jasper, 1958), with locations of Fz, Cz, and Pz along the midline, F3, C3, and P3 over the left hemisphere, and F4, C4, and P4 over the right hemisphere. Data from participants who had trials in all four conditions (gramCORRECT, gramERROR, ungramCORRECT, ungramERROR) were used. Data from a given participant were excluded if they had fewer than four trials in any condition. The final analysis included subjects from the WPF and SNF condition, totaling sixteen participants.

2.4.2 Statistical Analysis

Analyses of variance (ANOVAs) were used to determine whether changes to the grammaticality of the stimulus and the accuracy of the response were significantly related to the brain responses at each electrode. The amplitudes of the ERN waveforms were compared using ANOVAs, with accuracy (correct, incorrect), grammaticality (grammatical, ungrammatical), electrode laterality (left, midline, right), and electrode site (frontal, central, parietal) as factors. We also observed the interactions between the dependent variables to see if the laterality and site of the electrode location was related the amplitude of the ERN. The analysis involved a four way comparison between accuracy, grammaticality, laterality and site.

3.0 RESULTS

3.1 BEHAVIORAL DATA

An ANOVA was done to analyze response accuracy and latency, with grammaticality as the independent variable. There was a significant effect of grammaticality on performance accuracy, F (1, 15) = 38.823, MSE =.369 p < .001, with higher accuracy for grammatical conditions, (M = .82; SD = .077) than for ungrammatical conditions, (M = 0.60; SD = .184). Grammaticality did not have significant effects on response latency, F (1, 15) = 2.382, p > .1. However, this was expected because responses were delayed due to a 250 ms blank screen before the prompt.

3.2 ELECTROPHYSIOLOGICAL DATA

An ANOVA was conducted with grammaticality, accuracy, laterality, and site as the independent variables, and mean amplitude within the 60-100 ms post-response time window as the dependent variable. See Figure 1 for the waveforms from each electrode by condition, and Figure 2 for the mean amplitudes as a function of condition and electrode A significant four way interaction was found between grammaticality, accuracy, hemisphere, and lobe, F (4, 60) = 2.618, MSE = .551, $p < .05$. We also conducted post-hoc Duncan's multiple range tests on the four-way interaction between grammaticality, accuracy, laterality, and site. Significant contrasts in mean difference amplitude were found in the following trials: grammatical Cz (Mean Difference = .68, $p < .05$), ungrammatical F3 (Mean Difference =.32, $p < .01$), ungrammatical P3 (Mean Difference = 1.01, $p < .01$), ungrammatical Pz (Mean Difference = .67, $p < .05$), ungrammatical F4 (Mean Difference = .93, $p < .01$). The post-hoc tests revealed that the effects for grammatical trial errors were more negative and for ungrammatical trial errors were more positive.

Figure 1. Grand average ERPs for Block 1 at nine electrode sites as a function of grammaticality and accuracy

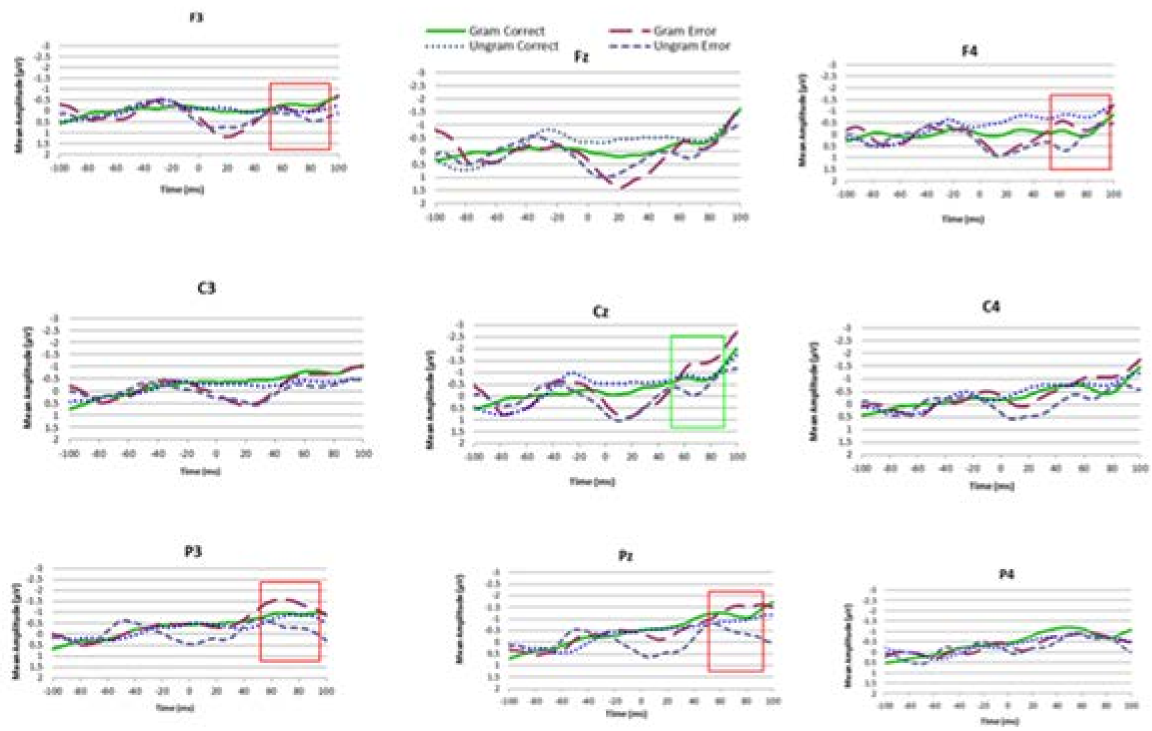


Figure 2: Mean ERP amplitudes for grammatical trials as a function of accuracy

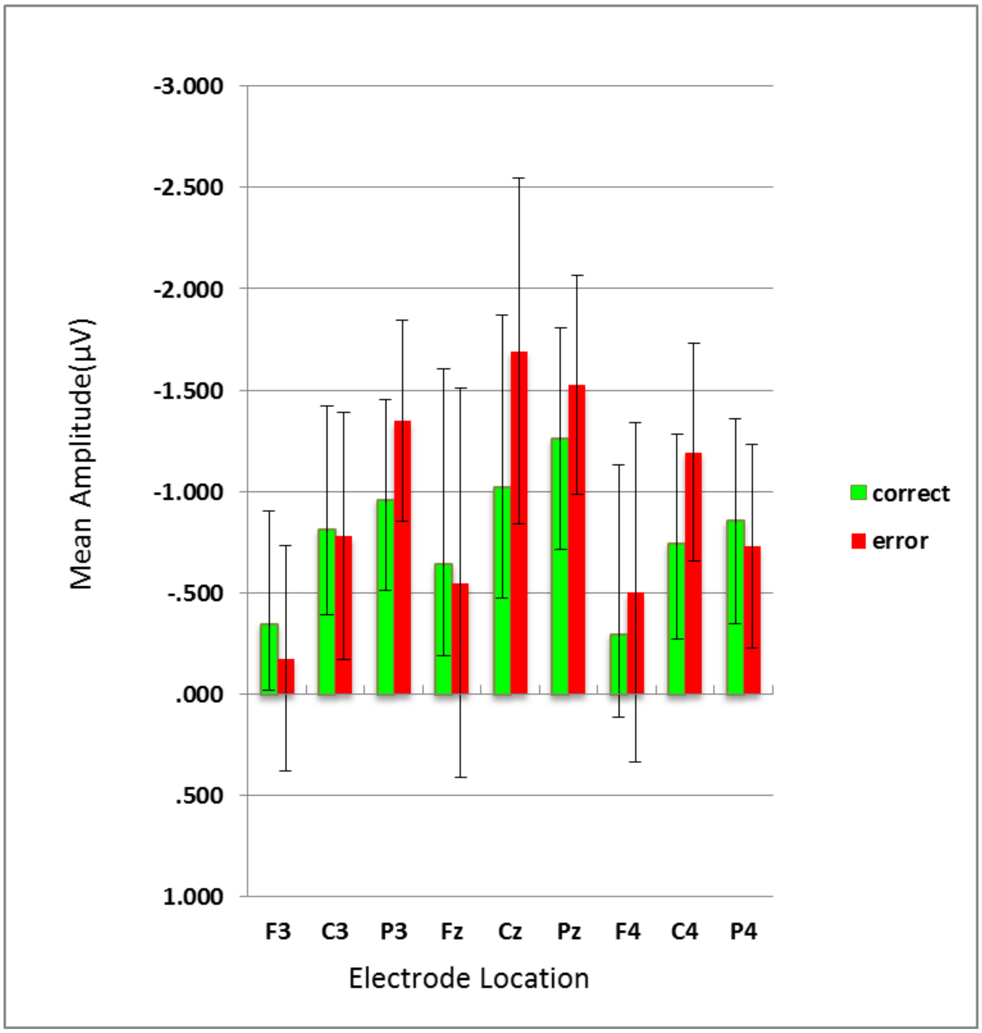
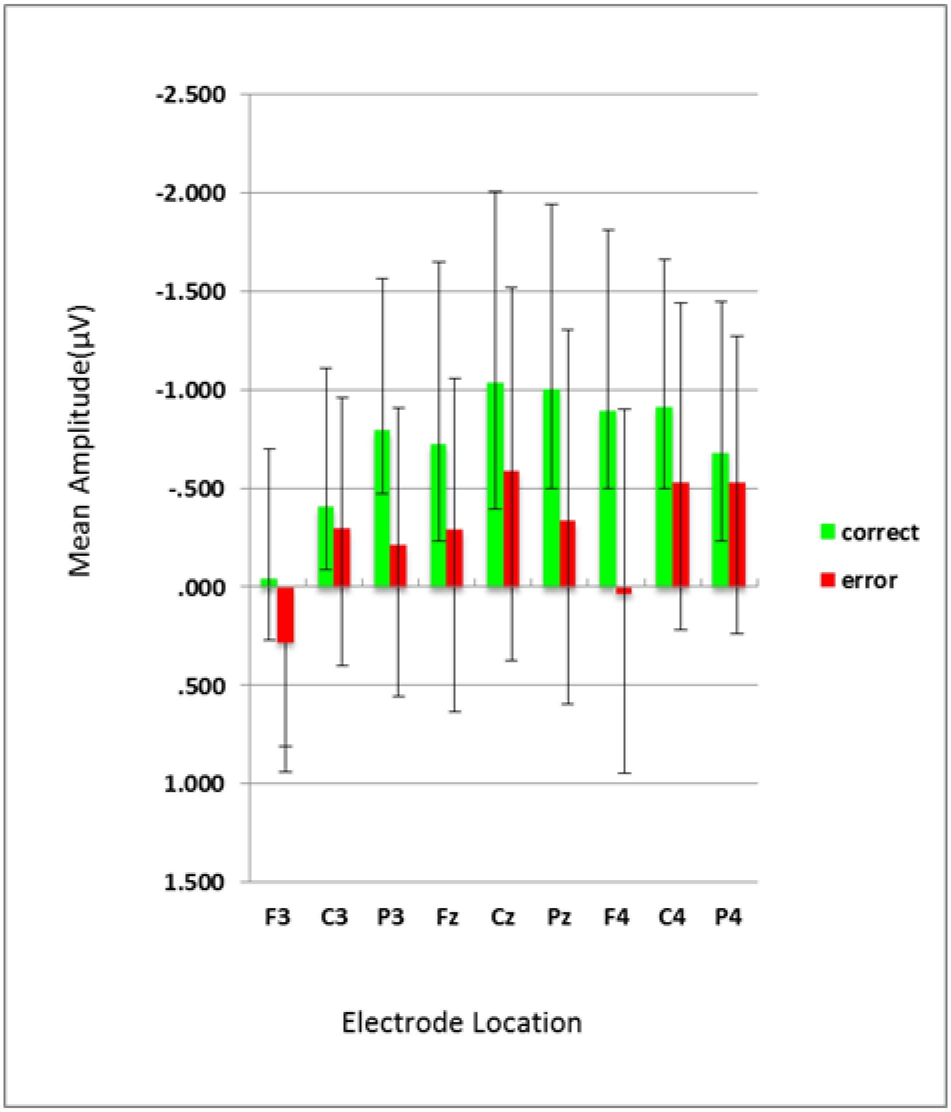


Figure 3. Mean ERP amplitudes for ungrammatical as a function of accuracy



4.0 DISCUSSION

Our analysis of Block 1 data from Tolentino and Tokowicz (2012; unpublished data) revealed sensitivity to errors in the ERN time window. Grammaticality had a significant effect on the ERN amplitude. When the stimulus was grammatical, and the participant made an error, an ERN was present. When the stimulus was ungrammatical and the participant made an error, an effect opposite of the ERN was observed Around 80 ms after the response, a positive component was seen. A similar positive response was seen by Herrmann et al. (2004), called the Pe. They found this to be more pronounced for perceived than unperceived errors. These data are interesting because they reveal a Pe like response in the ungrammatical conditions that is more frontally distributed that the ERN in the grammatical conditions. The ERN was seen when the sentence was grammatical and judged as unacceptable by the participant, whereas the positive component was seen when the sentence was ungrammatical, and misjudged as being correct. A significant neural component was seen in both instances of error judgment, but the direction depended on the grammaticality of the stimulus, suggesting that the context of processing and the nature of the error could have an impact on the type of brain response that is evoked. More research needs to be done on the factors affecting whether and ERN or a positive component is elicited.

Our results are compatible with both the conflict-monitoring and mismatch accounts of the ERN. To accurately judge the grammaticality of a phrase, one must have an internal representation of the correct structure. Under either the conflict-monitoring or mismatch theory, the degree of overlap between the internal representation and the stimulus being are coded are related to degree of activity of the internal representation. Holding to the conflict-monitoring theory, the degree of overlap can be reevaluated after the response has been made. The ERN appears after a decision has been made and evidence has accumulated indicating that the degree of overlap between the stimulus and the internal representation is much greater or much less than was initially judged. Under the representational mismatch hypothesis, reevaluation of the degree of overlap between the representation and the stimulus does not occur. Instead, the ERN results from direct incongruency between a *Yes* response in a condition with low overlap, or a *No* response in a condition with high overlap. Our results also complement previous ERP data (Hulstijn, 2002; Tokowicz & MacWhinney, 2005) indicating that L2 learners are implicitly sensitive to L2 violations.

Though we refer to implicit processing as the extent to which one is unable to verbalize patterns of information processing, and explicit processing as one's awareness of such processes, the topic is controversial in SLA literature. Although several differences have been noted between explicit and implicit processing, such as awareness, learnability, and accessibility, however, the question of how researchers can operationalize these constructs, to design testing methods, does not have a straightforward answer. Although GJTs are behavioral measures, it is not certain that they exclusively reflect explicit processing. Loewen (2003) found that grammatical and ungrammatical sentences seem to measure different constructs: grammatical sentences relying on implicit knowledge, and ungrammatical sentences relying of explicit knowledge (see Ellis, 2005). Similarly, it cannot be said with complete certainty that ERPs exclusively measure implicit knowledge. However, for the purposes of this study we used GJT scores and ERP responses to measure discrepancies between behavioral data and electrophysiological data. This study makes no attempts to define implicit or explicit processing and knowledge, rather the goal is to compare behavioral and brain potentials in response to grammatical and ungrammatical stimuli, so that more can be learned about the knowledge L2 speakers possess.

The present study expands on the previous ERN literature, specifically focusing on its implications for L2 acquisition. An ERP study, unlike behavioral assessments of L2 knowledge, can provide real-time information about L2 grammatical processing. An ERN recorded during these grammaticality judgments reveals that L2 learners have stronger internal representations than what is reflected in their overt behavior.

Our results add to the ERN literature by providing a new context in which the component can be studied. Previous ERN research has indicated that the ERN results from incongruencies between the internal representation and the stimulus, or the reevaluation of the degree of overlap between the stimulus being processed and the representation (Bernstein et al. 1995; Coles et al., 2001; Yeung et al., 2004). Our data complement both of these hypotheses, and provide additional information that the way L2 learners process errors is affected by the grammaticality of the stimulus, that in turn, yields a different pattern of data. On trials with grammatical stimulus in which participants made an error, and ERN was elicited approximately 80 ms after the erroneous response, similar to what previous data have shown. However, when the initial stimulus was ungrammatical and participants made and erroneous judgment, a positive component was elicited approximately 80 ms after the response, with a similar distribution to the ERN. Under the representational mismatch, and conflict-monitoring theories of the ERN, these data suggest that

L2 learners do have internal representations about correct and incorrect morpho-syntactic structures. However, it appears that L2 learners' knowledge of incorrect constructions elicits a different neural component.

An interesting follow up to this study would be an investigation of ERN amplitude differences between native and non-native speakers in response to these same violations. Sabourin and Stowe (2008) showed this with the P600 component: L2 learners did exhibit a P600 effect in response to errors, but the effect was smaller in amplitude than the P600 effect of native speakers. It is possible that this proficiency dependent grading of responses indicates a continuum of grammatical sensitivity. Under this assumption, the ERN would not be as prominent in less proficient L2 speakers in comparison to L2 learners of high proficiency of native speakers. However, the reflected ERP sensitivity despite different levels of proficiency, suggests that a common neural mechanisms underlies this process. It is possible that lower level L2 learners are implicitly sensitive to sentence violations but are not able to access or report the information (Tolentino & Tokowicz, 2010). Tokowicz and MacWhinney (2005) showed that learners are still hesitant to make judgments about unique constructions, and may be willing to judge ungrammatical stimuli as correct. This complements the idea that learners are unable to express or access the information regarding the grammaticality of some constructions. As the ERP data indicated, learners were most sensitive to violations of these unique constructions, but had a more difficult time expressing it as an accuracy judgment. The positive component that we observed in the unacceptable errors trials could be reflective of this uncertainty about judging sentences as ungrammatical. Further research about cross-language similarity and the ERN may further contribute the understanding of the brain areas involved in L2 processing and factors that influence their level of activity. Broadly, if we can better understand the structures and the influencing factors, we may be able to assist L2 learners who do not display overt sensitivity, to use their implicit sensitivity to make more accurate L2 judgments.

Overall, our results support the hypothesis that that when one makes an incorrect assumption about the grammaticality of a sentence, the brain elicits a response, showing that on some level, one knows that an error has been made. This could be a result of accidentally pressing the wrong response button, or a figurative slip of the tongue experience. Although more research needs to be conducted as to determine which factor is the most prominent in error causation, our results provide a foundation on which this research can develop. Additionally, these data can assist with the formation of new methods of assessing L2 proficiency which will sensitively analyze speakers' skills.

BIBLIOGRAPHY

Bates, E. & MacWhinney, B. (1981) Second language acquisition from a functionalist perspective: Pragmatic, semantic and perceptual strategies. In H. Winitz (Ed), *Annals of the New York Academy of Sciences conference on native and foreign language acquisition* (pp. 190-214). New York: New York Academy of Sciences.

Bernstein, P. S., Scheffers, M. K., & Coles, M. G. H. (1995). 'Where did I go wrong?' A psychophysiological analysis of error detection. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 1312-1322.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Nool, D., & Cohen, J. D. (1998). Anterior cingulate cortex error detection, and the online monitoring of performance. *Science*, 280, 747-749.

Carp, J., Compton, R. J. (2009) Alpha power is influenced by performance errors. *Psychophysiology,* 46 (2), 336-342.

Coles, M. G. H., Scheffers, M. K., & Holroyd, C. B. (2001). Why is there and ERN/Ne on correct trials? Response representations, stimulus-related components, and the theory of error processing. *Brain Research,* 1147, 192-208.

Desmurget, M., & Grafton, S. T. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Science,* 4, 423-431.

Ellis, R., Measuring implicit and explicit knowledge of a second language: a psychometric study., *SSLA*, 127, 141-172.

Ganushchack, L. Y., Schiller, N. O. (2009). Speaking one's second language under time pressure: An ERP study on verbal self-monitoring in German-Dutch bilinguals. *Psychophysiology*, 46, 410-419.

Hahne, A. (2001). What's different in second language processing: Evidence from event-related brain potentials. *Journal of Psycholinguistic Research.* 30 (3), 251-266.

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

Hermann, M. J., Römmler, J., Ehlis, A. C., Heirdrich, A., & Fallgatter, A. J. (2004). Source localixation (LORETA) of the error-related negativity (ERN/Ne) and positivity (Pe). *Cognitive Brain Research*, 20, 294-299.

Holroyd, C. B., Yeung, N. (2003). Alcohol and error processing. *Trends in Neuroscience,* 26, 402-404.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). *Language and Cognitive Processes*, 15

(2), 159-201.
Loewen, S., (2003), *Grammaticality judgment tests: What do they really measure?* Paper presented at the Second Language Research Forum, Tucson, AZ.
MacWhinney, B. (1997). Second language acquisition and the Competition Model. In J. Kroll & A. De Groot (Eds.), *Tutorials in bilingualism*. Mahwah, NJ: Lawrence Erlbaum.
Rodríguez-Fornells, A., Kurzbauch, A. R., & Münte, T. F. (2002). Time course of error detection and correction in humans: Neurophysiological evidence. *Journal of Neuroscience,* 22, 9990-9996.
Sebastián- Gallés, N., Rodríguez-Fornells, A., Diego-Balaguer, R., & Díaz, B. (2006). First- and second-language phonological representations in the mental lexicon. *Journal of Cognitive Neuroscience*, 18 (8), 1277-1291.
Sabourin, L., & Stowe, L. A., (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research,* 24 (3), 397-430.
Scheffers, M. K., Coles, M. G. H., Bernstein, P. S., Gehring, W. J., & Donchin, E. (1996). Event-related brain potential and error-related processing: An analysis of incorrect responses to go and non-go stimuli. *Psychophysiology,* 33, 42-53.
Scheffers, M. K., & Coles, M. G. (2000). Performance monitoring in a confusing world: Error-related brain activity judgments of response accuracy, and types of errors. *Journal of Experimental Psychology: Human Perception and Performance,* 26, 141-151.
Tokowicz, N. & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition*, 27, 173-204.
Tolentino, L. C., & Tokowicz, N. Cross-languages similarity modulates the relationship between brain activity & behavioral responses in L2 processing. *Manuscript under review.*
Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the Error-Related Negativity. *Psychological Review,* 111, 931-959.