THE RELATIONSHIP BETWEEN IMPLICIT AND EXPLICIT SECOND-LANGUAGE PROCESSING: THE ROLE OF CROSS-LANGUAGE SIMILARITY

by

Leida Curado Tolentino

BS, University of Massachusetts, 2002

Submitted to the Graduate Faculty of the

University of Pittsburgh in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2008

UNIVERSITY OF PITTSBURGH

SCHOOL OF ARTS AND SCIENCES

This thesis was presented

by

Leida Curado Tolentino

It was defended on

April 14, 2008

and approved by

Julie Fiez, Professor, Department of Psychology

Tessa Warren, Assistant Professor, Department of Psychology

Thesis Advisor: Natasha Tokowicz, Assistant Professor, Department of Psychology

THE RELATIONSHIP BETWEEN IMPLICIT AND EXPLICIT SECOND-LANGUAGE PROCESSING: THE ROLE OF CROSS-LANGUAGE SIMILARITY

Leida C. Tolentino, M.S.

University of Pittsburgh, 2008

The present study investigated the relationship between implicit and explicit second-language (L2) processing in beginning L2 learners, and how cross-language similarity influences this relationship. The brain activity of native English speakers was recorded as they performed grammaticality judgments on Spanish sentences. The three types of agreement violations used were similar in the two languages ("Similar" type), different in the two ("Different" type), and unique to L2 ("Unique" type). After a baseline assessment, we improved participants' accuracy and then retested them on new and repeated items. Results showed that the explicit increase in accuracy was accompanied by a significant increase in brain sensitivity, as measured by the P600 ERP component. This effect was most pronounced for the Similar and Different types. Additionally, grammatical sensitivity was greater for repeated items in both measures of processing, and this was also modulated by cross-language similarity as well as the specific type of repetition. The obtained results are in line with an interface position on the relationship between implicit and explicit L2 processing and suggest a strong modulatory role of cross-language similarity in both types of processing.

Keywords: Implicit/Explicit Processing; Second Language Processing; Cross-Language Similarity; ERPs; P600.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

I wish to thank Chelsea Eddington and Michael Rose as well as other members of the PLUM lab at the University of Pittsburgh for their help with testing participants; Rob Simon of Neuroscan for extensive technical assistance; Gwen Frishkoff for informative consultations; and the Spanish professors at Pitt for allowing me to advertise the present study in their classrooms. I thank my academic advisor, Natasha Tokowicz, for comments on previous versions of this manuscript. Finally, I'd like to thank the members of my master's committee for their insightful comments.

1.0 INTRODUCTION

Most adults who attempt to learn a foreign language can relate to the fact that it usually constitutes a highly effortful and oftentimes quite lengthy process. The ease with which one's first language (L1) was acquired is conspicuously absent and the developing second language (L2) has to now overcome the often detrimental influence of a highly ingrained L1 (e.g., MacWhinney, 2005). This, however, does not imply that a high level of proficiency and automaticity in the processing of an adult-learned L2 cannot be achieved. Some research has been conducted in the field of L2 learning examining the different ways in which beginning adult learners process their L2. Specifically, various studies have investigated the nature of this processing in terms of its implicit and explicit components, drawing a distinction between them. According to some authors (e.g., R. Ellis, 2005; Hulstijn, 2005), explicit and implicit *knowledge* differ, among other things, in the extent to which one is consciously aware or intuitively aware of regular patterns in the information one possesses, and the extent to which one can or cannot verbalize such patterns, respectively. Furthermore, access to implicit knowledge is thought to occur automatically whereas access to explicit knowledge is thought to involve controlled processes (N. Ellis, 2005; R. Ellis, 2005). The term explicit knowledge is often used interchangeably with the term declarative knowledge. Thus, two important distinctions between implicit and explicit knowledge of an L2 are the level of awareness as well as the amount of effort involved in accessing relevant linguistic rules or patterns.

The goal of the present study was to investigate the relationship between implicit and explicit L2 processing and furthermore to examine the influence that similarity between L1 and L2 exerts on this relationship. To this end, we manipulated the level of explicit L2 processing in native English speakers who are beginning adult learners of Spanish by increasing their accuracy on grammaticality judgments. We then examined the effect of this intervention on their implicit processing as measured by event-related potentials (ERPs). Additionally, we manipulated the nature of the L2 stimuli such that they fell on a continuum of similarity between L1 and L2 with some grammatical violations exhibiting a high degree of cross-language correspondence and others practically none.

Determining the relationship between implicit and explicit knowledge has also been the subject of L2 learning research and has been framed in terms of the interface between them, resulting in three separate views (R. Ellis, 2005). At one extreme is the *noninterface position*, which posits a fundamental separability between implicit and explicit L2 knowledge such that they involve different learning mechanisms, are subserved by different brain regions, and are accessed by different, automatic or controlled, processes. According to this view, explicit knowledge cannot be directly transformed into implicit knowledge or vice versa. In contrast, the *strong interface position* maintains that not only can explicit knowledge be drawn from implicit knowledge, but the opposite is also true. Implicit knowledge can result from the practice of explicit rules or facts, possibly simultaneously maintaining its original explicit representation. Finally, the *weak interface position* posits that explicit knowledge can be converted into implicit knowledge but nevertheless establishes some limits as to the specific mechanisms or points in time that this could come about. For example, N. Ellis (2005) proposed that explicit knowledge can exert a large influence on implicit knowledge by shaping its content via top-down attentional processes as well as through rule practice and subsequent proceduralization. Nevertheless, the two types of processing remain dissociable and no actual *conversion* occurs according to this view. The current study can be viewed as a test of the noninterface position: if explicit accuracy improvement is accompanied by a parallel increase in brain sensitivity to violations then a strict noninterface view of the relationship between implicit and explicit processing is likely inaccurate. That is, if the two types of processing do not influence each other, then ERPs and accuracy scores should not systematically co-vary. On the other hand, if explicit processing exerts an influence on implicit processing, as demonstrated by corresponding changes in accuracy and ERPs, then some version of the interface position is likely to be true.

For the purposes of the present study, it is worth noting that the concepts of L2 *knowledge* and L2 *learning* are distinct from that of L2 *processing*, although they are all closely related (R. Ellis, 2004). The distinction lies mainly in the content of L2 memory (knowledge) versus the mechanisms underlying its formation (learning) versus its access during performance (processing), with the latter constituting the focus of the current study. Nevertheless, some studies that investigate these related concepts are reviewed here because they are relevant to the present study. Additionally, implicit and explicit types of processing will be viewed not as dichotomous concepts but rather as representing different segments along a continuum of explicitness. Thus, “explicit” processing is viewed as *relatively* more explicit, and the whole construct is considered to be continuous rather than discrete and absolute (see also Dienes & Perner, 1999).

In an attempt to establish valid and distinct operational measures of implicit and explicit L2 knowledge, R. Ellis (2005) conducted a psychometric study that employed a total of five tests specifically designed to promote access to implicit or explicit knowledge of 17 different English grammatical structures. The tests were administered to a group of adult L2 learners of English (as well as native English-speaker controls) of mixed proficiency and primarily of a Chinese background and consisted of an imitation test; an oral narrative test; a timed grammaticality judgment test (GJT); an untimed GJT; and a metalinguistic knowledge test. The latter two tests were designed as measures of explicit knowledge whereas the rest were designed to measure implicit knowledge. A two-factor solution was specified in a principal components analysis and the experimental results indeed conformed highly to the predictions: the untimed GJT and the metalinguistic knowledge test loaded heavily on one factor accounting for about 16% of the variance, whereas the remaining tests loaded on a different factor accounting for approximately 58% of the variance. These were interpreted as corresponding to separate measures of explicit and implicit knowledge, respectively. A similar earlier study (Han & R. Ellis, 1998) also examined the relationship between implicit and explicit measures of L2 knowledge (including a timed oral production test, a timed GJT, a delayed GJT, and a metalinguitstic knowledge interview) and their relationship to measures of general language proficiency. Results again yielded a two-factor solution with the timed and untimed tests loading on different factors and these results were subsequently interpreted as reflecting access to implicit and explicit knowledge, respectively. Because both factors also correlated highly with standard tests of English proficiency it was concluded that, though separable, both implicit and explicit knowledge contribute to general L2 proficiency (although only one of the measures of explicit knowledge, the delayed GJT, correlated significantly with tests of proficiency).

To the extent that the experiments described above did measure implicit and explicit knowledge, their results suggest that these components of L2 knowledge can be dissociable and thus lends validity to the distinction between the two types of knowledge. A different experiment that further supports this notion but in the context of L2 *learning* compared the effectiveness of implicit and explicit types of exposure in the learning of the “soft-mutation” grammatical feature of the Welsh language (N. Ellis, 1993). These mutations are triggered by various specific grammatical contexts and result in the change of the first consonant in a word (e.g., “Boston” becomes “Foston” after the preposition “o” meaning “from”). Participants were assigned to three experimental groups according to whether they were exposed only to examples in random order illustrating the various mutations (“Random” group); whether they were explicitly taught the 8 rules of soft mutations (“Rule” group); or whether they were first taught the rules and then saw them applied to specific instances of vocabulary (“Rules & Instances” group). The three groups corresponded to implicit, explicit, and “structured” types of exposure and their performance was assessed with a variety of measures including an implicit timed GJT, explicit knowledge of the rules, generalization to new words, and error and learning curve analyses when translating from Welsh to English. The three training regimens resulted in very different patterns of learning with implicit and explicit components often being dissociated: the “Random” (implicit) group, although competent on familiar mutations, showed little evidence of having acquired implicit (timed GJT) or explicit working knowledge of the soft mutation rules, whereas the “Rule” (explicit) learners acquired firm knowledge of the rules but failed to apply them in the identification of ungrammatical forms in the GJT, at least initially. The “Rules & Instances” (structured) learners were the most successful, and performed well on both implicit and explicit tests and generalized to new words. Interestingly, the “Rule” group was later able to transfer and apply their knowledge to new structures. Based on these results and the fact that this group also performed equivalently to the “Rules & Instances” group on the grammatical constructions in the GJT, the author concluded that the study provides support for an interface position which allows for the interaction between implicit and explicit knowledge, with the latter providing the attentional focus that the learner needs to abstract relevant structure from the language.

From a cognitive neuroscience perspective, other studies have examined the neural correlates of L2 processing in various languages. Many of these studies typically use ERPs to measure participants' brain activity while processing various types of linguistic information in their L2. Because specific components of the ERP record directly capture brain activity that occurs on the order of hundreds of milliseconds and is often dissociable from overt performance, they are thought to reflect relatively automatic cognitive (or sensory) processes that are free of conscious reflection (see Hulstijn, 2002; Osterhout, Bersick, & McLaughlin, 1997). Furthermore, ERPs have been shown to correlate with other measures of implicit memory (Rugg, Mark, Walla, Schloerscheidt, Birch, & Allan, 1998). Thus, the ERP record can be used as a measure of implicit performance because it reflects brain activity that is not subject to conscious reflection and other controlled processes. The validity of using specific ERP components to measure implicit processing shall be later addressed in more detail.

One specific component of the ERP record, the P600, a positivity that peaks at approximately 600 ms post stimulus and exhibits a centroparietal scalp distribution (Osterhout & Holcomb, 1992), has been widely used in studies of syntactic violations in the L2. The P600 is observed in response to syntactic anomalies (e.g., "The cat won't eating") reflecting its online processing from about 500 to 900 ms after the presentation of the violation. Similarly, the N400 component, a centroparietal negativity that peaks at approximately 400 ms post-stimulus presentation, is generally reflective of semantic integration processes and is sensitive to lexical variables such as word frequency (e.g., Carreiras, Vergara, & Barber, 2005). ERPs have been used in a variety of within-language and bilingual studies. For example, McLaughlin, Osterhout, and Kim (2004) recorded the ERPs of beginning adult learners of French while they made overt lexical decisions to target letter strings preceded by primes. Their results showed that learners' brain activity was indicative of a discrimination between L2 words and pseudowords after only 14 hours of classroom instruction and despite their behavioral performance at chance levels ($d' < 1$). Thus, these results suggest that there is a discrepancy between a learner's implicit and explicit processing of L2 linguistic violations; the learners appear to be sensitive to the unacceptability of stimuli at an implicit level and yet seem incapable of accessing, or perhaps reporting, this knowledge explicitly.

In another ERP study, Hahne (2001) found that Russian-German bilinguals exhibited a P600 in response to phrase structure violations in auditorily-presented German sentences in an untimed GJT (heneceforth simply "GJT" unless otherwise noted). Compared to the native German speakers, the P600 was slightly delayed in the bilinguals, peaking at about 950 ms. Interestingly, a similar study (Hahne & Friederici, 2001) examining Japanese-German bilinguals' ERPs found no evidence of a P600 in response to auditorily-presented phrase structure violations (in contrast to native speakers' responses). Hahne (2001) suggests that the discrepancy between the two sets of results may be explained by the behavioral performance (and thus level of proficiency) of the different groups of bilinguals such that the Russian L1 group was more proficient (8% errors) than the Japanese L1 group (20% errors). The implication is that implicit (ERP) sensitivity to L2 grammatical violations is correlated with L2 overt proficiency. Though an interesting suggestion lending support to the interface position of the relationship between implicit and explicit L2 processing, a different and perhaps complementary possibility is that the pairs of languages examined exerted an influence on the effects obtained. It is possible that the specific grammatical structures employed in the two studies are formed more similarly in Russian and German than in Japanese and German. This would then potentially lend a processing advantage to the Russian L1 bilinguals if similarity between languages is posited to play a role in implicit L2 processing.

1.1 THE ROLE OF CROSS-LANGUAGE SIMILARITY

Indeed, theories of language acquisition suggest that the degree of similarity between an adult-learned L2 and the L1 is of great importance to the learner mainly due to the transfer of various aspects of the L1 to the processing of the L2 (MacWhinney, 2005). According to interactive activation models such as the Unified Competition Model of language acquisition (MacWhinney, 2005) two outcomes occur when an L1 system is applied to L2 learning: transfer and competition. Positive transfer occurs when grammatical constructions, for example, are similar between the two languages and the learner can thus employ her native system effectively. On the other hand, when L1 and L2 linguistic structures do not directly map on to each other, competition arises and processes from the more dominant L1 can result in negative transfer and subsequent mistakes. Central to these processes is the notion of a linguistic *cue* that represents the mapping between surface form and underlying function of a given word, for example. The relative strength of cues associated with specific linguistic items is what determines the outcome of competition and the subsequent selection of the item associated with the strongest cue. Cue strength is in turn dependent on cue *availability*: if the correspondence between form and function is made available through both positive and negative evidence (i.e., in contrastive examples), cue strength will increase. Thus, cue strength and availability in the L2 input as well as cross-language similarity between various types of cues are seen as crucial to the L2 learner because they will largely determine how potentially conflicting cues in the two languages are resolved. Alternatively, according to this model, the processing of the L2 is subject to a vast influence from the L1 not only because of transfer processes *per se* but also due to the fact that, unlike the former, the latter is highly entrenched in an adult learner. Nevertheless, because the model is interactive, it also allows for some influence from the L2 on to the processing of the L1 albeit a weaker one.

Research indeed suggests that L2 learners often do rely on the similarities between two linguistic systems to achieve their goal: Basnight-Brown, Chen, Hua, Kostic, and Feldman (2007) found that differences in participants' L1 (English, Serbian, or Chinese) accounted for their discrepant sensitivities to inflectional processing of regular and irregular English verbs, even when matched for overall L2 proficiency. Whereas all three experimental groups showed facilitation for regular verbs in an audio-visual primed lexical decision task, only the Serbian L1 and native English speakers showed facilitation for irregular verbs with a nested stem. Furthermore, only the native speakers showed facilitation for irregular verbs without a nested stem. The authors suggest that the pattern of results can be accounted for by the fact that Serbian is a highly inflected language that, like English, uses an alphabetic writing system whereas the use of characters in the logographic Chinese language does not allow for as consistent a mapping of form to phonology and vice versa. Thus, it is possible that the Chinese L1 speakers were attuned to only the more superficial *–ed* suffix in regular verbs but that the Serbian's L1 experience allowed them to extract more information from their L2 and thus become additionally sensitive to changes in some types of irregular verbs.

Another study examining the influence of cross-language similarity in L2 processing used ERPs to measure participants' sensitivity to subject-verb agreement in English (Chen, Shu, Liu, Zhao, & Li, 2007). Despite being highly accurate on a GJT (88% on average) Chinese learners of English showed no P600 in response to ungrammatical sentences, unlike native

English speakers. Because the Chinese language does not have a system for marking grammatical number, the authors interpreted the findings as suggestive of an instance of negative transfer from L1 to L2. It is also interesting to note once again the dissociation between ERP and behavioral measures (although a *d'* measure may have been more informative than pure accuracy scores because L2 learners tend to have response biases): above-chance performance is not accompanied by implicit sensitivity as was also the case in the Hahne and Friederici (2001) experiment. In a different study, Osterhout, McLaughlin, Pitkanen, Frenck-Mestre, and Molinaro (2006) found that native English speakers learning French showed a P600 in response to verbal-person agreement violations after four months of instruction but no sensitivity to noun phrase (NP) number agreement violations even after eight months of instruction. These results were interpreted in terms of cross-language similarity: although performed differently in the two languages, verbal-person agreement constitutes a feature of participants' L1 whereas NP number agreement does not and therefore takes longer to be incorporated into their online processing system. However, in this study cross-language similarity was confounded with phonological realization of French grammatical morphemes (only the verbal-person agreement condition entailed sounding out of the relevant word endings) so results cannot be clearly attributed to either factor.

Finally, Tokowicz and MacWhinney (2005) also used ERPs to investigate the extent to which transfer and competition mechanisms influence the processing of Spanish grammar by native English speaking adult learners. The authors directly manipulated cross-language similarity by using various types of grammatical constructions that differed in the degree of similarity when translated word for word between the two languages. Their results indicated that learners were implicitly sensitive to violations in the construction that is formed similarly in the

L1 and L2 (aspect-marking) and for the construction that is unique to the L2 (determiner gender agreement). Their ERPs did not show sensitivity to the construction that exists in both L1 and L2 but is implemented somewhat differently in the two languages (determiner number agreement). Additionally, in contrast to the ERP data, participants' behavioral judgments (*d'*) in a concurrent GJT reflected near chance performance for all constructions.

Of the studies reviewed above, the Tokowicz and MacWhinney (2005) is the most relevant to the present experiment, which can be viewed as an extension of it. In the present study we also investigated the influence of cross-language similarity in implicit and explicit processing. Here, however, we directly manipulated explicit processing and examined the effect on implicit processing.

1.2 THE PRESENT STUDY

The present study investigated the relationship between implicit and explicit L2 processing in beginning adult learners of an L2, and the influence that cross-language similarity exerts on this relationship. In this study we use ERPs, and specifically the P600 component, as our measure of implicit processing and grammaticality judgments as our measure of explicit processing. As stated earlier, the use of ERPs as a measure of implicit processing is based on the fact that this component occurs at a time scale usually attributed to implicit processes (see Hulstijn, 2002); is often dissociable from overt performance (e.g., McLaughlin et al., 2004; Osterhout et al., 1997; Tokowicz & MacWhinney, 2005); and has been shown to correlate with other measures of implicit memory (Rugg et al., 1998). However, this is not to say that the P600 cannot also reflect processes of repair and/or revision of ungrammatical stimuli as has been alternatively suggested (Hahne & Friederici, 1999; Kaan & Swaab, 2003). In our view, such processes can still occur relatively automatically and without conscious awareness. Hahne and Friederici (1999) claimed that the ERP component known as the early left anterior negativity (ELAN) represents automatic processing of syntactic violations whereas the P600 reflects processes that are under strategic control. This claim has been based partly on the observation of the ELAN in native speakers for auditorily-presented stimuli but not in L2 speakers thus allegedly reflecting early and automatic processes in the former but not the latter group. However, the basis of the argument can be problematic as a number of other studies have failed to find an ELAN in native speakers (e.g., Allen, Badecker, & Osterhout, 2003; Kim & Osterhout, 2005; Tokowicz & MacWhinney, 2005; see also Osterhout, McLaughlin, Kim, Greenwald, & Inoue, 2004 for a related discussion on LAN effects), possibly due to the fact that the stimuli were presented visually instead of auditorily or, alternatively, because of the specific constructions used. Furthermore, because we view implicit and explicit processing not as dichotomous concepts but rather as points along a continuum of explicitness, it is appropriate to compare a *relatively* more implicit measure to a less implicit (or more explicit) one. We therefore used accuracy (and *d'*) scores on the delayed GJT as our measure of explicit processing because presumably they reflect the fact that participants had enough time to access the relevant grammatical knowledge and also reflect on their response. Finally, we used participants' production accuracy scores on a brief English-Spanish translation task (which can be viewed as a measure of explicit processing at the extreme of our explicitness continuum) and qualitatively compared the pattern of results to those from the GJT. Compared to sentence comprehension, as in the GJT, learners are expected to engage in even longer reflection periods in L2 sentence production because they undertake an analytical approach to the selection of the appropriate L2 linguistic form (R. Ellis, 2004).

The various studies reviewed above attempted to measure implicit and explicit processing (or knowledge) and often also the relationship between the two measures. However, only the N. Ellis (1993) study directly attempted to experimentally manipulate either type of processing and thus systematically measure if (or how) they interact with each other within a given participant. Therefore, the goal of the present study was to measure the effect of an explicit accuracy-improving intervention on implicit processing. Specifically, we compared learners' brain sensitivity in response to L2 grammatical violations to their overt grammaticality judgments pre- and post-accuracy improvement. Furthermore, we concurrently manipulated the degree of similarity of the grammatical forms between the two languages to examine the effect of cross-language similarity on implicit and explicit L2 processing.

To examine if and how implicit L2 processing varies with explicit processing we manipulated explicit processing by improving participants' accuracy on the GJT after a baseline assessment. Accordingly, the current study required a pilot version in which the efficacy of various experimental interventions aimed at increasing explicit accuracy were compared. The goal of the pilot study was to identify the condition that proved to be the most effective in improving GJT accuracy so that this condition alone could be used in the primary experiment. In a different pilot study, Tokowicz (2004, unpublished data) showed that increasing the salience of the locus of grammatical violation by isolating the relevant word pairs from a sentence and providing feedback to adult learners of Spanish increased their accuracy and discrimination of grammatically acceptable and unacceptable sentences. Thus, our pilot experiment compared four experimental conditions in which the accuracy-increasing interventions varied according to whether presented items consisted of sentences or word pairs and/or the presence of feedback. As a result, depending on the specific condition, participants saw whole sentences in Spanish with or without accuracy feedback, or word pairs that were extracted from sentences with or without feedback. The pilot experiment is described in more detail below.

The role of cross-language similarity was examined by comparing implicit and explicit sensitivity to types of sentence agreement violations that can be placed along a continuum of L1/L2 similarity, ranging from very similar in English and Spanish to not very similar. The definition of cross-language similarity was based on correspondence of L1-L2 translations for each word in a sentence. For example, if all words in a given sentence could in principle be translated from Spanish into English and retain all of their grammatical features then such a sentence was classified as similarly-formed in the two languages. This resulted in three types of grammatical constructions (see Table 1): 1) constructions in which all of the L2 words directly map onto L1 words without conflict ("Similar" type); 2) constructions in which the relevant L2 word does not directly correspond to the L1 word and thus conflicts with it ("Different" type); and, 3) constructions in which the relevant L2 word simply cannot map onto a word in the L1 system ("Unique" type). We used the demonstrative determiner number agreement system as our Similar construction because both English and Spanish make use of this grammatical feature (e.g., "*this* cat" and "*these* cats" vs. "*este* gato" and "*estos* gatos") and thus predicted positive transfer from L1 to L2 and no competition. Definite determiner number agreement was used for the Different construction because even though the English grammar includes the concept of number marking, unlike Spanish this is usually restricted to the noun thus resulting in the use of the same determiner for both singular and plural nouns (e.g. "*the* cat" and "*the* cats" versus "*el* gato" and "*los* gatos"). This type of L2 construction was expected to elicit some competition between the L1 and L2 and possibly result in negative transfer. Finally, we used definite determiner gender agreement for the Unique construction because whereas the English language does not make use of grammatical gender, determiners in Spanish must agree in gender (and number) with nouns (e.g., "*the* cat" versus "*el* gato" or "*la* gata"). Because determiner gender agreement does not exist in the L1 it is not in direct conflict with the L2 form and thus was not expected to result in competition. Although it may pose some difficulty to learners because they need to build an entirely new grammatical system for processing this type of construction, relatively high cue availability in the Spanish determiner-noun gender agreement system may facilitate this process.

Table 1. Sample stimuli

Demonstrative determiner number agreement (**Similar**)		Definite determiner **number** agreement (**Different**)		Definite determiner **gender** agreement (**Unique**)	
Grammatical	**Ungrammatical**	**Grammatical**	**Ungrammatical**	**Grammatical**	**Ungrammatical**
Ese gato duerme.	***Esos*** gato duerme.	***El*** gato duerme.	***Los*** gato duerme.	***El*** gato duerme.	***La*** gato duerme.
(***That*** cat sleeps.)	(***Those*** cat sleeps.)	(***The***-SING cat sleeps.)	(***The***-PL cat sleeps.)	(***The***-MASC cat sleeps.)	(***The***-FEM cat sleeps.)

2.0 THE PILOT EXPERIMENT

The goal of the pilot experiment was to identify the experimental condition that would result in the greatest improvement in participants' grammaticality judgment accuracy so that it could be investigated further in the primary experiment. In particular, there were four between-participants experimental conditions that varied according to the degree of salience of the locus of grammatical violation and/or the presence of accuracy feedback. Therefore, in the specific phase of the experiment during which the interventions occurred, participants were exposed to: 1) whole sentences with no feedback; 2) whole sentences with accuracy feedback; 3) word pairs with no feedback; or 4) word pairs with accuracy feedback. Word pairs consisted of a determiner and a noun that could constitute an agreement violation or not. Previous studies have shown that providing feedback to adult ESL learners can be beneficial (R. Ellis, Loewen, & Erlam, 2006; Ferris & Roberts, 2001; Rosa & Leow, 2004; Tokowicz 2004, unpublished data) and that increased stimulus salience can help focus attention to relevant aspects of the L2 (N. Ellis, 2005). Thus, in light of these findings it was expected that the more salient violations (i.e., word pairs only) would draw attention to the relevant segments of the stimuli and that providing accuracy feedback to participants would increase their ability to discriminate grammatically correct and incorrect stimuli and therefore result in subsequent learning. As will be seen, our results are in line with this suggestion: the presentation of isolated word pairs as well as the inclusion of immediate accuracy feedback resulted in the greatest improvement in accuracy (i.e., word pairs with accuracy feedback condition).

2.1 METHOD

2.1.1 Participants

Seventeen native-English speaking adults participated in the pilot study. Participants were beginning learners of Spanish enrolled in semesters 1 through 4 at the University of Pittsburgh and Carnegie Mellon University. They each participated in a single session for which they were paid $10 per hour or received credit toward an Introductory Psychology course. All participants had normal or corrected-to-normal visual acuity, were right-handed (according to an abbreviated version of the Edinburgh Handedness Inventory, Coren, 1992), had no implanted brain devices, were not taking any psychoactive medications, and had not been exposed to any language other than English before age 13.

2.1.2 Design

The experiment was divided into three blocks containing Spanish stimuli and a fourth block conducted in English. For all participants, the first block consisted of whole sentences in Spanish and served as a baseline assessment of performance. The second block was the locus of the critical intervention consisting of whole sentences with no feedback as in Block 1 (hereafter referred to as SNF); whole sentences with accuracy feedback (SF); word pairs with no feedback (WPNF); or word pairs with accuracy feedback (WPF). The third and final Spanish block was similar to Block 1 consisting again of sentences with no feedback.

Thus, a 4 intervention-block condition (SNF, SF, WPNF, WPF) X 3 cross-language similarity (similar, different, unique to L2) X 2 grammaticality (grammatical, ungrammatical) mixed design was employed. The only between-subject factor was intervention-block condition. Participants were randomly assigned to one of the four intervention-block conditions.

2.1.3 Stimuli

The Spanish stimuli consisted of three different kinds of grammatical constructions that can be placed along a continuum of L1/L2 similarity as described above. A total of 288 original Spanish sentences and 96 Spanish word pairs were used in the pilot experiment. These were identical in all cross-language similarity conditions except for the determiner preceding the critical noun at the point of grammatical agreement. The stimuli were grouped into six versions combining the three different types of cross-language similarity and two levels of grammaticality. [1] This ensured that each sentence/word pair appeared in both its grammatical and ungrammatical form in each of the cross-language similarity conditions across all participants. Thus, within each block, half of the experimental stimuli were grammatical and the other half were not.

The 288 sentences were equally split among the three Spanish blocks (but Block 2 in the word-pair conditions consisted of only a determiner and a noun, e.g., "el gato") resulting in 96 items per block. The final Spanish block (Block 3) contained an additional 96 sentences (for a total of 192), which were repeated from Block 2 in that the same word pairs from Block 2 were incorporated into a sentence in Block 3. Of these, half were presented in their original form and, in half, the determiner preceding the critical noun was changed to reflect a switch in grammatical acceptability. Finally, 30 practice sentences were constructed. All stimuli were reviewed by two native Spanish speakers who verified their acceptability in the context of the experimental goals.

In addition, there were 38 sentences in English employing two grammatical agreements (demonstrative determiner number agreement and reflexive pronoun agreement) in the final block. Half of these were grammatical and two were practice sentences.

2.1.4 Procedure

Participants performed the Spanish part of the experiment first, which was divided into 3 blocks. After completing the Spanish phase of the experiment, participants proceeded to the final English block.

Participants made grammaticality judgments during all blocks. Figures 1 and 2 provide an overview of the timeline of events during sentence and word-pair trials. Sentences were presented one word at a time in the center of a computer screen in white, 36-point Arial font on a black background using the Neuroscan STIM software program (Neuroscan, El Paso, TX). The computer screen was situated inside a sound-attenuated, electrically-shielded booth (Industrial Acoustics, Inc.) that also contained a button box and the electroencephalographic amplifiers. Each trial began with a fixation cross that remained in the center of the screen until the participant pressed a key on the button box. Each word in a trial was presented for 300 ms followed by a 350-ms blank screen between words. The final word in the sentence contained a period and was followed by a shorter (250-ms) blank, after which a question mark probe appeared and remained on the screen until the participant made a response. The presentation of a single word at a time ensured that we averaged ERP responses from the point at which participants should have first noticed a violation (or not). These parameters were the same as those used by Tokowicz and MacWhinney (2005). Participants were instructed to indicate whether the trial was grammatically correct as quickly and accurately as possible following the probe. In the conditions in which feedback was included, a feedback screen displaying the words “Correct !” or “Incorrect” immediately followed the participants’ response and remained on the screen for 1000 ms. In all conditions and blocks, stimuli was presented in random order. ERPs were recorded continuously during the GJT but were analyzed only in conjunction with the data from the main experiment.

Participants used the left thumb to press the 1 key on the button box to indicate a positive response, and the right thumb to press the 4 key to indicate a negative response. Participants were asked to blink only during the fixation-cross screen and to sit still and not move their eyes during word presentations to reduce movement artifact in the ERP analysis windows.

Participants completed 30 practice trials prior to the Spanish phase and two prior to the English phase. After completing both the Spanish and English grammaticality judgment phases of the experiment (which took approximately 1 to 1.5 hours) participants were asked to complete a brief translation task by typing into an Excel sheet the Spanish translations of six sentences in English on a different computer outside the sound-attenuated booth. Following this, participants answered a general questionnaire containing questions about how they approached the experimental task as well as about handedness and medication or drug use. Finally, participants filled out a language history questionnaire (adapted from Tokowicz, Michael, & Kroll, 2004). The entire experiment lasted approximately three hours.

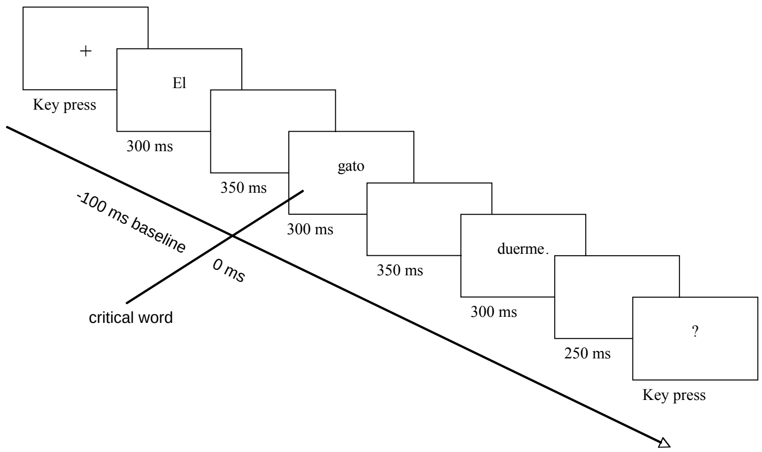


Figure 1. Timeline of events during sentence trials.

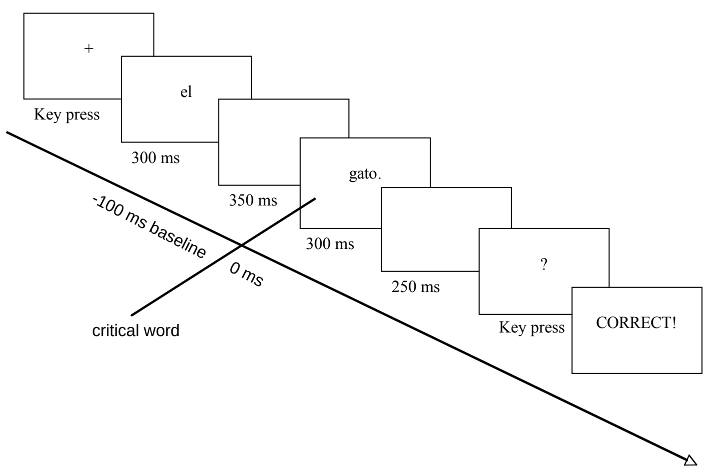


Figure 2. Timeline of events during word-pair trials.

2.1.5 Data Analysis

2.1.5.1 Behavioral data. Data from one participant were excluded due to computer error and data from four other participants were excluded from a subset of analyses due to below-standard (95%) English accuracy performance. Thus, the final analyses were conducted on both a set of 16 participants and a set of 12 participants due to the English accuracy cutoff. The results were unaffected by the removal of data from the four participants.

Accuracy performance in Block 3 relative to Block 1 served to determine the effectiveness of the intervention in each intervention-block condition. Furthermore, items in Block 3 were analyzed separately depending on whether they constituted new or repeated items. Thus, a 4 intervention-block condition (SNF, SF, WPNF, and WPF) X 3 "block" (Block 1, Block 3 old, Block 3 new) mixed analysis was employed.

Accuracy rates were calculated for each participant in each condition during each experimental block. It was then verified that the accuracy-improving intervention was successful in most intervention-block conditions, with the vast majority of participants (81%) showing a numerical increase in accuracy in Block 2 relative to Block 1. An ANCOVA was then conducted that analyzed accuracy scores in Block 3 adjusted for Block 1 accuracy (as a covariate). This served to evaluate the effectiveness of the intervention block in the four different pilot conditions and allowed us to choose the one that elicited the highest improvement in accuracy scores to use in the primary experiment. Due to the small number of participants in each condition and the general goal of the pilot experiment we relied on numerical differences in accuracy measures across conditions rather than standard levels of statistical significance.

2.1.5.2 ERP data. ERPs obtained in the pilot experiment were only analyzed in conjunction with data obtained in the primary experiment.

2.2 RESULTS

The analysis including all participants (*N*=16) showed that the WPF condition elicited the highest absolute accuracy scores in Block 3 and also the greatest accuracy improvement relative to Block 1. The SF condition also resulted in similar improvements, constituting a close second. The pattern of results was the same when Block 3 items were split into “old” and “new” and also when four participants were excluded due to English accuracy scores that were below the cutoff.

2.3 DISCUSSION

In light of the stated goal of the pilot experiment, the WPF condition was thus chosen as the one to be used in the within-participants design employed in the primary experiment. This outcome had been predicted based on previous research demonstrating the influence of increased stimulus salience and the role of feedback. Interestingly, the SF condition resulted in a very similar accuracy pattern thus emphasizing the role of accuracy feedback in L2 learning. There is some controversy in the L2 literature regarding what *type* of feedback is most useful to learners, with some studies showing that more explicit forms of feedback (e.g., correction + metalinguistic explanation) are most effective (R. Ellis et al., 2006; Rosa & Leow, 2004) whereas other studies suggest that any type of feedback or just task-essential practice is sufficient (Ferris & Roberts, 2001; Sanz & Morgan-Short, 2004). It is nevertheless reasonable to conclude that *some* type of feedback is better than none (but see R. Ellis et al., 2006). Indeed, a study by Mackey and Silver (2005) found that native Chinese-speaking children in Singapore had higher rates of improvement in English question formation when they received corrective feedback (i.e., recasts) than when they did not. In a different study, Ferris and Roberts (2001) also found that adult ESL students (mostly from Southeast Asian and Chinese backgrounds) who received either of two types of feedback significantly outperformed a control group that received no feedback on a selfediting writing task. One of the types of feedback that students received in that study was simple underlining of their written errors (as opposed to underlining + coding of errors), which Chandler (2003) has also shown to be as effective as corrective feedback and superior to describing the type of error, even with underlining, again in an ESL writing task. Simple underlining of errors can be roughly equated to the procedure in WPF condition in the present pilot experiment. Underlining of errors informs students that they made a mistake as we also did in this pilot by telling students whether their responses were correct or incorrect. Furthermore, it serves to point out the specific location of the error as is also implied in the WPF condition because there is only one possible location of violation when only a determiner and a noun are presented.

It is important to note, however, that the above studies involved mainly production tasks unlike our comprehension task. R. Ellis et al. (2006) employed a GJT when comparing implicit and explicit forms of feedback and showed no clear advantage of feedback *itself* (though explicit metalinguistic feedback was superior to implicit recasts and/or no feedback on several analyses). However, the results in that study could have been affected by the fact that the control group received no practice with the target grammatical structure in addition to receiving no feedback. In the present pilot study, it is also difficult to determine whether the increased salience of the isolated word pairs had an effect *per se* or if it was due to the fact that there were fewer words to process in general. Because this was not the primary goal of the pilot experiment, it will thus be interesting for future studies to directly compare the effect of different types of feedback and increased stimulus salience/focus on both production and comprehension tasks with adequate control groups and thus help determine the separate contributions of each factor on different L2 tasks.

3.0 THE PRIMARY EXPERIMENT

One goal of the primary experiment was to investigate the effect of explicit accuracy improvement in the detection of L2 grammatical violations on implicit L2 processing as measured by ERPs. If implicit processing is not influenced by explicit processing, as posited by the noninterface position, then we should not expect brain sensitivity to vary with an increase in explicit performance. On the other hand, if implicit and explicit processing interact we should see some change in brain sensitivity with an improvement in behavioral performance. As mentioned above, we also asked participants to complete a brief translation task from English to Spanish and qualitatively compared their explicit performance on this task to their performance on the GJT and their ERP sensitivity.

Another objective of the primary experiment was to examine how similarity between the two languages modulates the implicit-explicit relationship. Based on past research and cross-linguistic transfer theories that were reviewed above, we predicted that grammatical constructions that are formed similarly in the two languages (Similar type) would show a processing advantage both implicitly and explicitly as compared to constructions that conflict in the two languages (Different type). The construction that neither benefits from transfer nor suffers from competition (Unique type) could show implicit and explicit sensitivity depending on whether the learner has acquired the foreign grammatical system. According to the Unified Competition Model (MacWhinney, 2005) described above, such an outcome would, in this case, largely depend on cue strength and availability for grammatical gender agreement in the Spanish language. Because this type of cue is widespread in the language and usually quite reliable (though to a lesser extent than its availability), cue strength should be relatively high. Thus, it is feasible that L2 learners would be sensitive to this, though possibly to differing degrees in implicit and explicit processing.

We therefore compared ERPs and accuracy scores for the three types of agreement violations pre- and post-intervention. Furthermore, topographical maps of ERP scalp distributions were examined to explore the possibility that grammatical constructions that are formed similarly in L1 and L2 are subserved by more overlapping brain areas in L1 and L2 than construction types that are formed differently. This would support the notion that linguistic processes that are similar in two languages also rely on analogous neural processes from a spatial and temporal point of view (see Abutalebi, Cappa, & Perani, 2005 for analogous fMRI results in relation to proficiency and age of acquisition).

3.1 METHOD

The experimental method employed in the primary experiment was identical to that of the pilot experiment except where indicated below.

3.1.1 Participants

Forty-nine native English speakers participated in the primary experiment. They shared the characteristics of the pilot experiment population.

3.1.2. Design

We employed a 3 cross-language similarity type (similar, different, unique to L2) X 2 grammaticality (grammatical, ungrammatical) X 3 electrode laterality (left, midline, right) X 3 electrode site (frontal, central, parietal) mixed design. In addition, different levels of the experimental block variable (block 1, block 2, block 3 new, and block 3 old) were included depending on the specific comparison.

3.1.3 Stimuli

Stimuli were fully counterbalanced across participants for grammaticality and cross-language similarity but not experimental block because this would add unreasonable complexity to the study. Nevertheless, blocks 1 and 3 new were swapped for six out of 24 participants to allow for an examination of possible effects of the nature of the stimuli in each of the two blocks.

3.1.4 EEG recording and pre-processing procedures

Digital images of individual head shapes were obtained prior to the beginning of the experimental task using a head digitizer device (Polhemus for Neuroscan, Inc). These measurements were intended for subsequent source localization of the ERP signals obtained at the scalp. After a complete image was obtained, each electrode was filled with conducting gel using a syringe, and then the participant was comfortably seated inside the sound-attenuated booth to begin the experiment following a brief verbal explanation. Participants read the stimuli from the computer screen while the experimenter monitored the ERP recording from outside the booth in the main room.

Electrophysiological activity was recorded continuously at a sampling rate of 1000 Hz and the EEG signal was amplified with Neuroscan SynAmps2 amplifiers with 24-bit analog-to-digital conversion (Compumedics, NeuroScan Inc., El Paso, TX). Participants wore a high-density electrode cap fitted with 128 Ag/AgCl electrodes (QuikCap, Compumedics, NeuroScan Inc., El Paso, TX). In addition to the cap electrodes spanning the entire scalp, others were placed over the right and left mastoid bones, below and above the left eye, and in the outer canthi of the left and right eyes.

EEG data were processed off-line using Neuroscan Edit 4.3 software (Compumedics, NeuroScan Inc., El Paso, TX). All electrodes were re-referenced offline to linked mastoids and low-pass filtered at 30 Hz. Channels that contained large artifacts in a recording session (fewer than 3% of electrodes), typically due to poor contact with the scalp, were excluded from the averages. This corresponded to a maximum of three channels in each of the 19 affected participants. For a participant's data to be included, a minimum of 10 trials (63%) of each grammaticality and cross-language similarity stimulus type had to be artifact-free. The ERP epoch ranged from 100 ms pre-stimulus (baseline) to 1000 ms post-stimulus. The data were quantified by calculating the mean amplitude (relative to the 100 ms pre-stimulus baseline) for each stimulus type in two main latency windows: 500 to 625 ms and 625 to 750 ms. In addition to these windows that are representative of the latency of the P600 component, further visual inspection of the waveforms resulted in secondary analyses of two additional windows: 75 to 125 ms and 300 to 400 ms. Average ERPs were formed from trials that were corrected for ocular and movement artifacts. Ocular artifact reduction was based on estimates of average eye blink duration; trials on which the EEG epoch was contaminated by clipping, movement artifact, etc. were rejected.

3.1.5 Data analysis

The data obtained from 10 participants had to be replaced due to recording equipment failure; data from seven participants were replaced due to high levels of artifact; data from six participants were replaced due to below-cutoff (90%) English accuracy performance; and, finally, data from four participants were replaced because they did not meet experiment recruitment criteria. Data from a total of 24 participants were included in the final analyses (including 2 from the pilot experiment).

The nine relevant electrodes used in the analyses correspond to the International 10-20 Electrode System (Jasper, 1958) 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. Statistical analysis consisted of repeated measures analyses of variance (ANOVAs) that were conducted on electrodes that included data from trials on which the participants made both correct and incorrect responses (see Tokowicz & MacWhinney, 2005). The decision to include both types of response trials was based on the fact that the use of correct trials only would result in a very low number of valid ERP trials, especially in the case of ungrammatical stimuli. Thus, only including participants who reached an acceptable number of correct ERP trials would likely have required a different population than the one of interest in the present study (e.g., more advanced students). Also, because beginning learners are yes-biased, it is unclear that a correct response reflects correct processing instead of guessing. Furthermore, because this experiment is predicated on the idea that the participants are not making behavioral responses in a way that is consistent with their implicit knowledge, we are interested in data from all trials.

In the case of significant interactions, these ANOVAs were followed-up using the Duncan's multiple-range test (with $p < .05$ cutoffs) to identify the source of the effect. All analyses of the Spanish data included three levels of cross-language similarity type (similar, different, and unique), two levels of grammaticality (grammatical, ungrammatical), three levels of laterality (left, midline, right), and three levels of electrode site (frontal, central, parietal). In addition, different levels of the experimental block variable (block 1, block 2, block 3 new, and block 3 old) were included depending on the specific comparison. All variables were manipulated within-subjects with the exception of the between-participants variable ("exchange") that was included in an additional analysis comparing blocks 1 and 3 new to ascertain that the results obtained in the relevant comparison were not influenced by the fact that stimuli were not fully counterbalanced across experimental blocks.

We employed Greenhouse-Geiser (1959) non-sphericity correction for effects with more than one degree of freedom in the numerator. Following convention (Picton et al., 2000), we report uncorrected degrees of freedom, the corrected *p*-value, and the Greenhouse-Geiser epsilon value (ε). Mean square error values reported are those corresponding to the Greenhouse-Geiser correction.

3.2 RESULTS AND DISCUSSION

To better guide the reader through the various experimental findings and therefore facilitate the understanding of how these results relate to the specific research questions that were investigated in this study, we organize the reported accuracy and ERP results according to three general questions: 1) What is the relationship between implicit and explicit L2 processing?; 2) What is the role of cross-language similarity in L2 processing in general and how does it influence the implicit-explicit relationship?; and 3) What is the effect of item repetition on the implicit-explicit relationship and the role of cross-language similarity? At the end of this section, we also briefly address the question of whether the spatial distributions of scalp ERPs (and therefore likely brain sources) differ extensively across similarity types and in relation to ERPs in response to the English sentences and between block 1 and block 3 new.

Because the results for the accuracy data followed the same pattern as the *d'* analyses, we report *d'* data only (*d'* = 0 indicates no sensitivity; *d'* = 4 indicates close to absolute sensitivity; Green & Swets, 1974). Also, as mentioned earlier, we analyzed the ERP data in two main latency windows (500 to 625 ms and 625 to 750 ms) and two additional secondary windows (75 to 125 ms and 300 to 400 ms). We report only significant effects that involve the grammaticality variable (suggesting brain sensitivity to a difference between grammatical and ungrammatical stimuli) or its interaction with another relevant variable in each specific comparison. In addition, "grammatical sensitivity" is assumed to reflect more positive ERPs in response to ungrammatical stimuli than to grammatical stimuli in the two main (P600) time windows unless otherwise noted. Furthermore, we do not report or interpret main effects of laterality or electrode site, or interactions of these two factors because they reflect the dipolar nature of ERPs and are not theoretically relevant when they do not interact with manipulated variables.

3.2.1. 1) What is the relationship between implicit and explicit L2 processing?

Figure 3 illustrates the mean *d'* scores for blocks 1 (B1), 2 (B2), 3 new (B3 new) and 3 old (B3 old) for the Spanish stimuli. The grand average ERPs for B1 overall and B3 new overall are plotted in Figure 4.

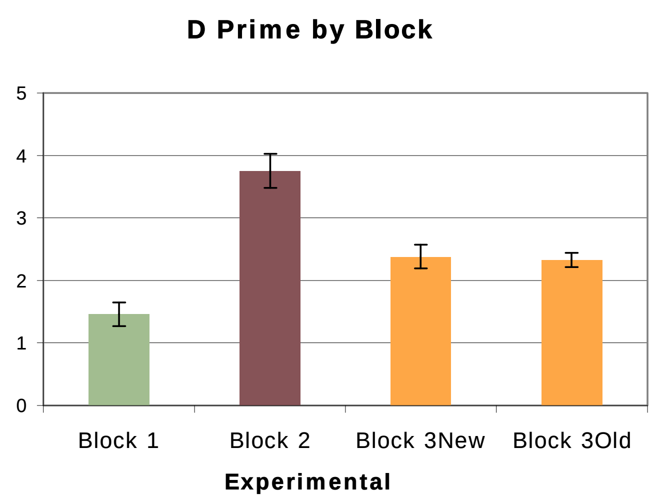
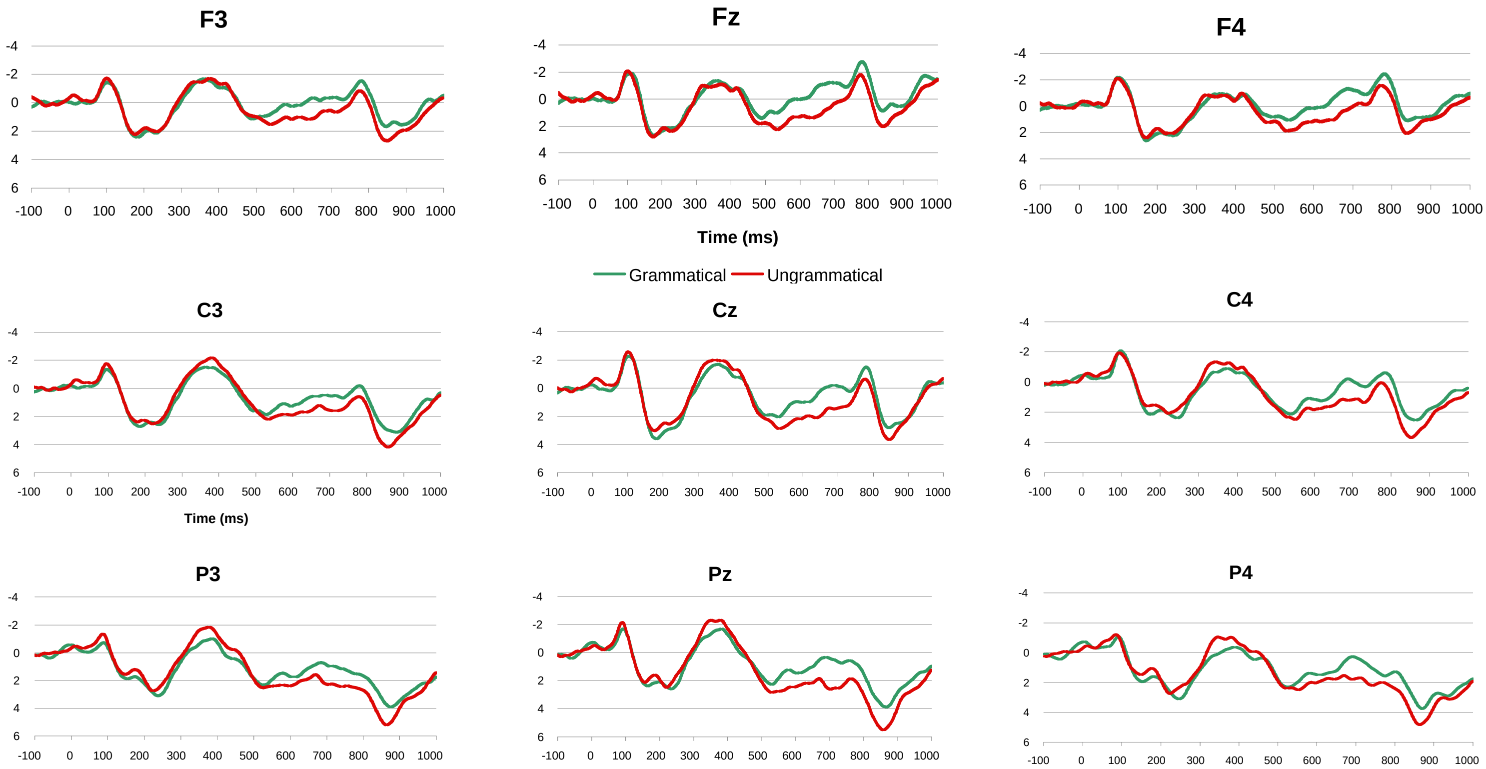


Figure 3. Mean *d'* scores in each experimental block.

Block 1:


F3
Fz
F4
Time (ms)
Grammatical
Ungrammatical
C3
Cz
C4
Time (ms)
P3
Pz
P4
-4
-2
0
2
4
6
-100
0
100
200
300
400
500
600
700
800
900
1000

Block 3 new:

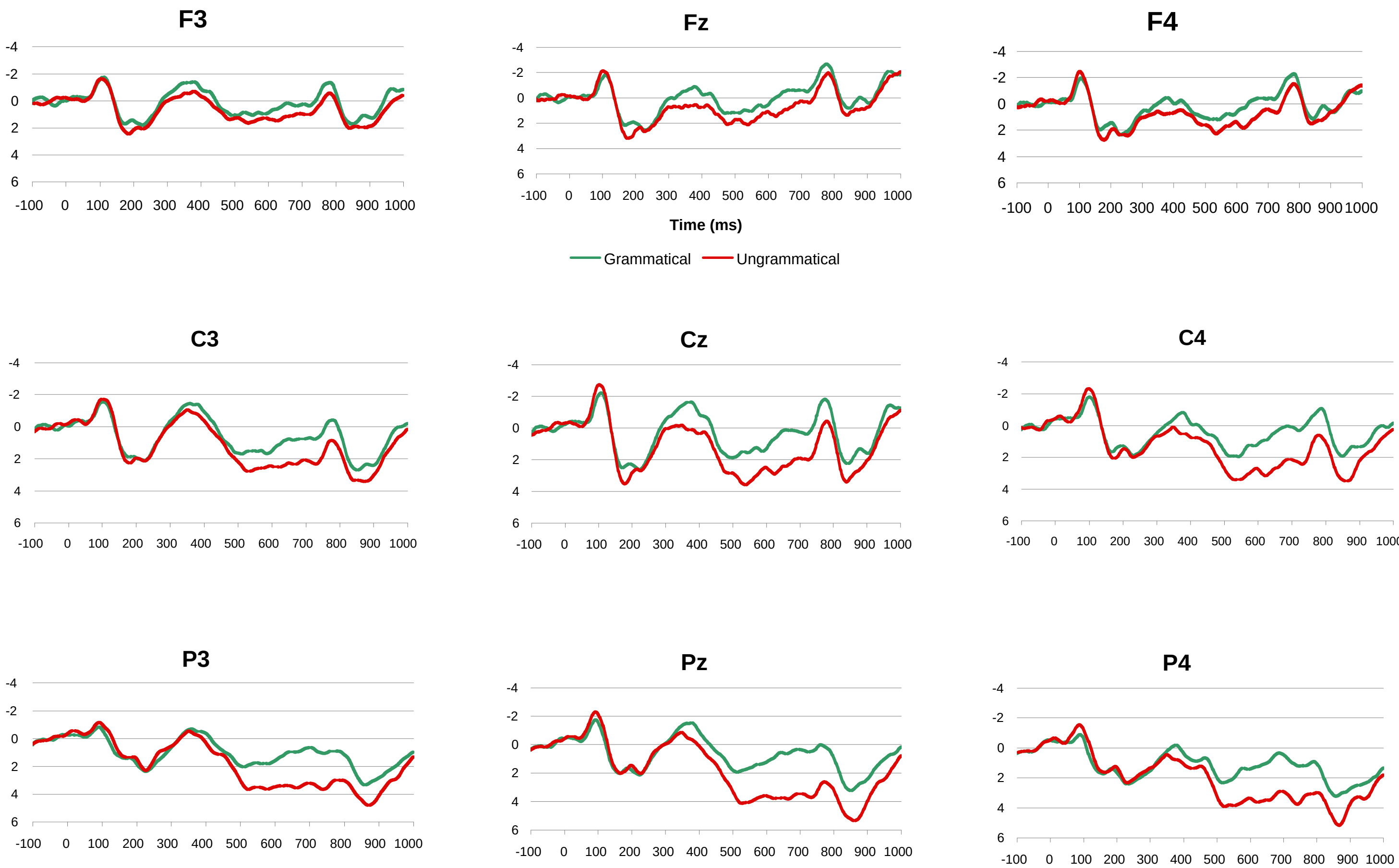


Figure 4. Grand average ERPs for block 1 overall and block 3 new overall at nine electrode sites.

3.2.1.1 Behavioral data. In the analysis comparing *d'* scores in all experimental blocks there was a significant effect of block, F (3, 69) = 54.46, MSE = 1.79, p < .00. Specific contrasts showed that scores in B1 were reliably lower than those in B2, B3 new, and B3 old, t (92) = -8.1, p < .001; t (92) = -3.56, p <.01; and t (92) = -3.07, p <.01, respectively. This result demonstrates an explicit accuracy increase not only in the intervention B2 but also in subsequent whole-sentence blocks. Thus, participants' processing became more explicitly accurate between B1 and B3 new, F (1, 23) = 20.37, MSE = 2.62, p < .001.

3.2.1.2 ERP Data.

3.2.1.2.1 500 to 625 ms. In the analysis comparing ERPs in response to grammatical and ungrammatical stimuli in B1 and B3 new there was a significant interaction between grammaticality, block, and electrode site, F (2, 46) = 4.65, MSE = 14.61, p < .05, ε = .55. Raw scores indicated a trend toward increased grammatical sensitivity in B3 new in centroparietal sites in comparison to B1 but this trend was not reliable in follow-up Duncan's tests. The interaction qualifies a main effect of grammaticality, F (1, 23) = 10.78, MSE = 62.3, p < .05.

3.2.1.2.2 625 to 750 ms. The analysis comparing ERPs in B1 and B3 new in this window also revealed a significant interaction between grammaticality, block, and site, F (2, 46) = 7.35, MSE = 11.36, p < .01, ε = .573. Follow-up tests indicated a significant grammaticality effect in parietal sites in B3 new only[2]. This interaction qualified a main effect of grammaticality, F (1, 23) = 23.69, MSE = 67.09, p < .001.

Thus, in summary, results indicated both an increase in behavioral GJT performance and brain sensitivity (across the two ERP time windows) between B1 and B3 new. This shows that an increase in explicit processing in B3 new was accompanied by an increase in implicit processing, suggesting a close (interface) relationship between the two types of processing.

3.2.2 2) What is the role of cross-language similarity in L2 processing in general and how does it influence the implicit-explicit relationship?

3.2.2.1 Behavioral data. Figure 5 illustrates the mean *d'* scores for similar, different, and unique types of stimuli collapsed across experimental blocks; Figure 6 shows the same data organized by block. In the analysis comparing *d'* scores for the three types of cross-language similarity in each block, we found a marginally significant interaction between type and block, F (6, 138) = 2.09, MSE = 1.54, p = .058. This interaction indicated that scores for the Similar and Different cross-language similarity types were not reliably different from each other and were both higher than the scores for the Unique type in all blocks except B3 old in which there was no type effect. The interaction qualifies a main effect of type, F (2, 46) = 31.58, MSE = 1.37, p < .001.

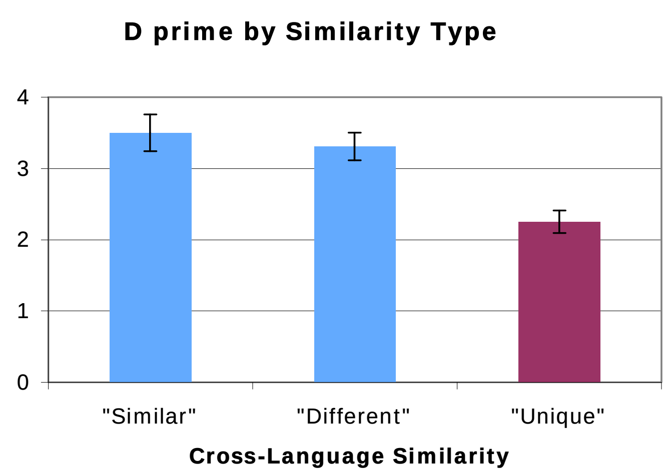


Figure 5. Mean *d'* scores for each cross-language similarity type overall.

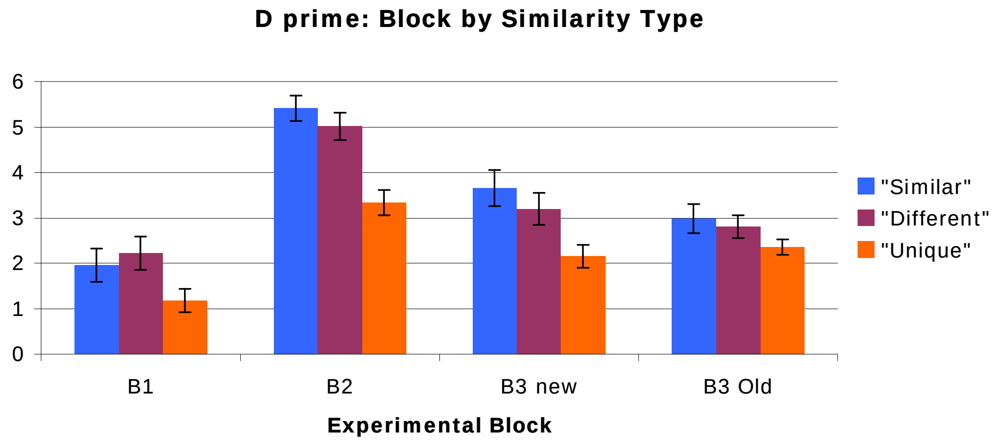
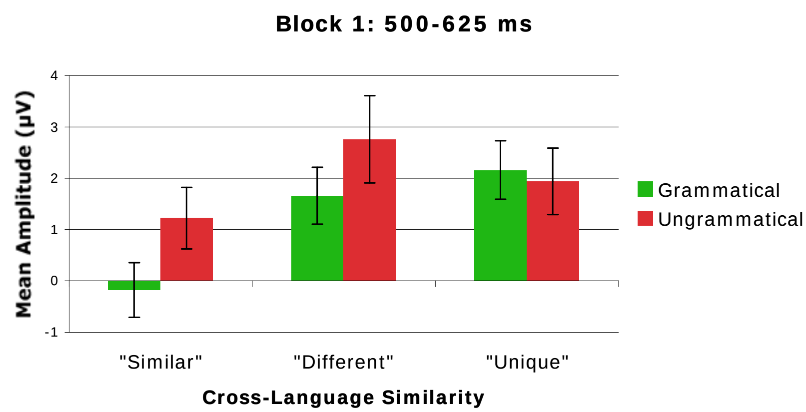


Figure 6. Mean *d'* scores for each cross-language similarity type in each experimental block.

In a specific comparison between B1 and B3 new there were main effects of type, F (2, 46) = 15.17, MSE = 1.27, $p < .001$, ε = .889, and block, F (1, 23) = 20.37, MSE = 2.62, $p < .001$. This again indicated higher scores for the Similar and Different types than the Unique type, and also higher scores in B3 new overall. The two variables did not interact.

3.2.2.2 ERP data.

3.2.2.2.1 500 to 625 ms. Figure 7 illustrates the mean amplitude of ERPs for grammatical and ungrammatical stimuli for each cross-language similarity type in B1 and B3 new in the 500 to 625 ms latency window. The overall ANOVA comparing B1 and B3 new including cross-language similarity type did not show an interaction with this variable. However, because the raw data showed a clear trend for an effect of type (see Figure 7), additional ANOVAs were conducted comparing B1 and B3 new *within* each similarity type. The results showed a significant interaction between grammaticality, block, and site for the Similar type, F (2, 46) = 7.29, MSE = 7.6, $p < .01$, ε = .639. A Duncan's follow-up test indicated significant grammatical sensitivity in B3 new at parietal sites only. An equivalent marginally-significant interaction between grammaticality, block, and site was also found for the Different type, F (2, 46) = 3.49, MSE = 9.62, p = .069, ε = .564 with follow-up tests indicating significant grammatical sensitivity in B3 new at parietal sites only, as in the Similar type. These interactions qualified main effects of grammaticality for both types, F (1, 23) = 6.01, MSE = 84.98, $p < .05$, and F (1, 23) = 6.2, MSE = 63.73, $p < .05$, respectively. There were no effects of grammaticality for the Unique type.



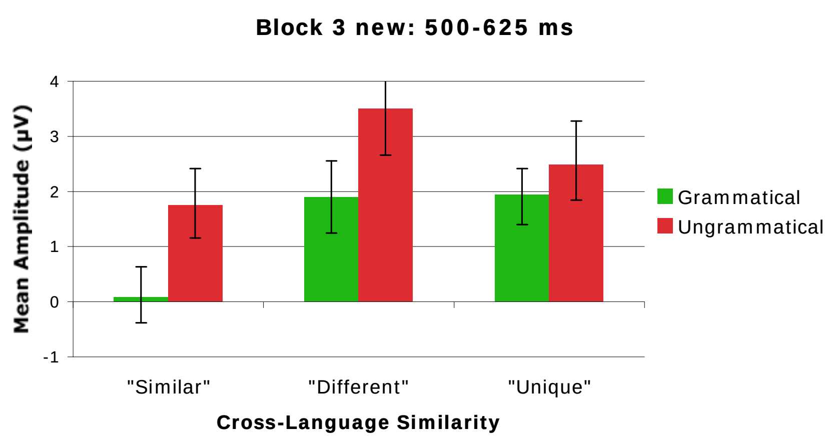


Figure 7. Mean amplitude of ERPs for each cross-language similarity type in B1 and B3 new in the 500 to 625 ms latency window.

Finally, an analysis of B2 ERPs revealed an interaction between type, grammaticality, and electrode site, F (4, 92) = 4.15, MSE = 7.12, $p < .05$, $\varepsilon = .532$. Follow-up analyses showed that the grammaticality effect was only significant for ERPs from central sites in the Similar condition. Figure 8 illustrates the mean amplitude of ERPs for grammatical and ungrammatical stimuli for each cross-language similarity type in B2.

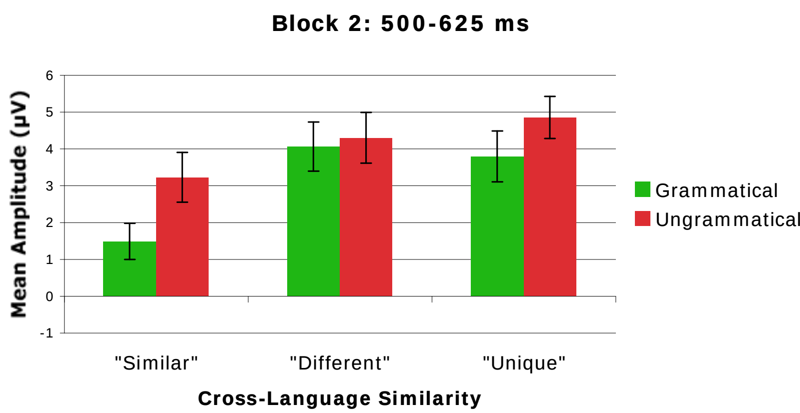
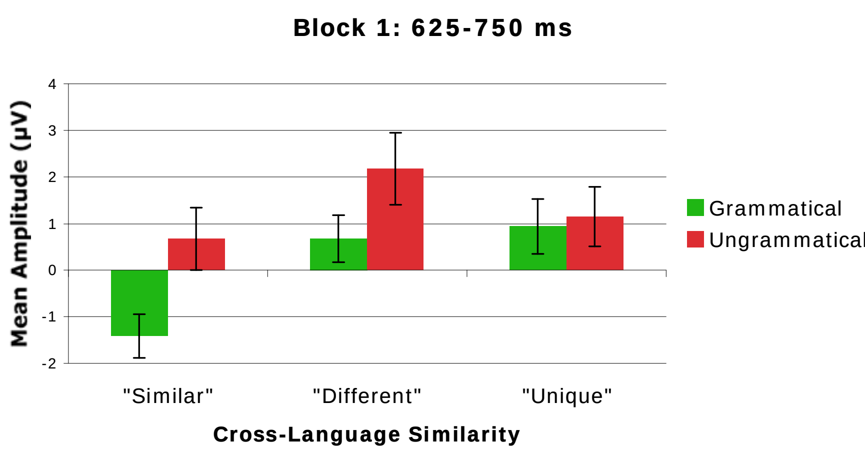


Figure 8. Mean amplitude of ERPs for each cross-language similarity type in B2 in the 500 to 625 ms latency window.

3.2.2.2.2 625 to 750 ms. Figure 9 illustrates the mean amplitude of ERPs for grammatical and ungrammatical stimuli for each cross-language similarity type in B1 and B3 new in the 625 to 750 ms latency window. As in the previous time window, the overall ANOVA comparing B1 and B3 new including cross-language similarity type did not show an interaction with this variable. However, analyses within each similarity type again showed an effect of block. Specifically, there was a significant interaction between grammaticality, block, and site (as well as between grammaticality and site) for the Similar type, F (2, 46) = 5.86, MSE = 7.08, $p < .05$, ε

= .665. Follow-up tests indicated grammatical sensitivity for the Similar type in B1 across all sites and in B3 new in centro-parietal sites only. The latter was also of greater magnitude than the former. The Different type again showed a marginally-significant interaction between grammaticality, block, and site, F (2, 46) = 3.69, MSE = 8.77, p = .058, ε = .603. Follow-up tests indicated grammatical sensitivity in B3 new in parietal sites only, as in the previous time window. These interactions again qualified main effects of grammaticality for both types, F (1, 23) = 12.58, MSE = 98.67, p < .01, and F (1, 23) = 11.84, MSE = 49.78, p < .01. Again, there were no effects of grammaticality for the Unique type, even though graphical display of the data seem to suggest an emerging trend toward grammatical sensitivity in B3 new (see Figure 9).



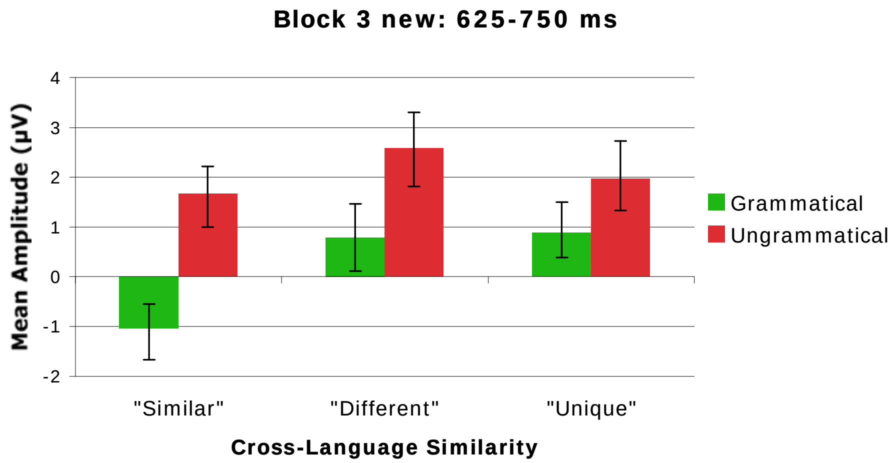


Figure 9. Mean amplitude of ERPs for each cross-language similarity type in B1 and B3 new in the 625 to 750 ms latency window.

Finally, analyses of B2 responses showed a significant interaction between grammaticality, laterality, and site, F (4, 92) = 4.89, MSE = 1.32, p < .01, ε = .752. Follow-up tests showed significant grammaticality effects in all electrode sites but of greater magnitude in left parietal sites. There was no interaction with type, unlike in the previous time window. Figure 10 illustrates the mean amplitude of ERPs for grammatical and ungrammatical stimuli for each cross-language similarity type in B2.

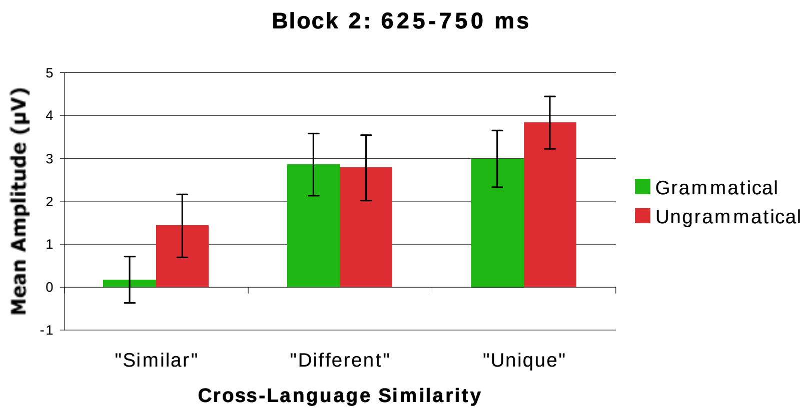


Figure 10. Mean amplitude of ERPs for each cross-language similarity type in B2 in the 625 to 750 ms latency window.

In summary, the data strongly suggest an influence of cross-language similarity in the processing of L2 stimuli as reflected by both behavioral and ERP results. *D'* scores were higher for the Similar and Different types than for the Unique type both pre- and post- accuracy increase. Furthermore, even though this was not apparent in the overall statistical analysis that included all variables, subsequent targeted analyses strongly suggested an increase in grammatical sensitivity between B1 and B3 new for the Similar and, less so, for the Different type (across ERP time windows) but not for the Unique type. Therefore, results at this level of analysis indicate a modulatory influence of cross-language similarity in explicit L2 processing and likely in implicit processing as well. Furthermore, cross-language similarity seemed to slightly modulate the *relationship* between the two types of processing because although increases in accuracy between B1 and B3 new were accompanied by some level of increased brain sensitivity in all types, the latter was of greater magnitude for the most cross-linguistically similar types.

3.2.3 3) What is the effect of item repetition on the implicit-explicit relationship and the role of cross-language similarity?

3.2.3.1 Behavioral data. Figure 11 illustrates *d'* scores for each similarity type in B3 new and B3 old. The analysis comparing *d'* scores in B3 new and B3 old showed no effect of block, suggesting that participants were explicitly equally sensitive to items that had been previously encountered and to those that were not. There was, however, a main effect of type, F (2, 46) = 9.08, MSE = 1.47, $p < .001$, ε = .948. This showed that scores for the Similar and Different types were not different from each other and were both higher than those for the Unique type across both blocks. There was no interaction between block and type.

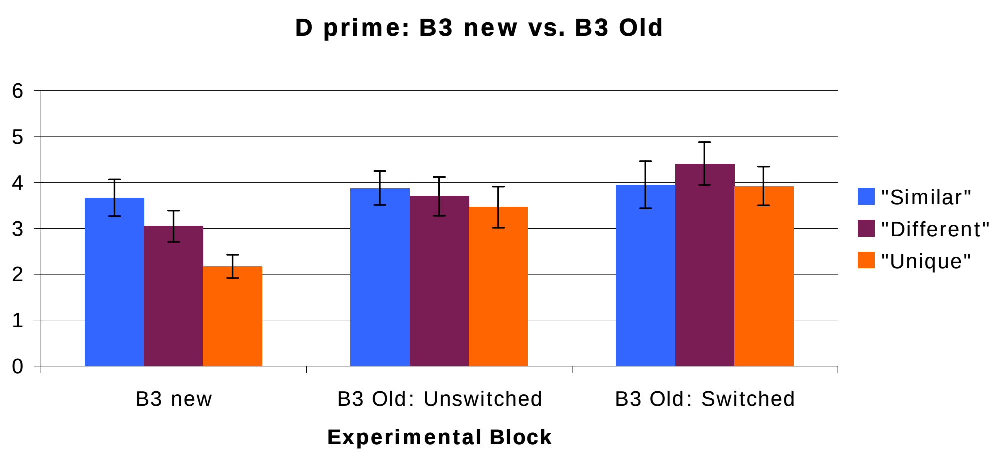


Figure 11. Mean *d'* scores for each cross-language similarity type in B3 new and B3 old unswitched and switched items.

Furthermore, a more fine-grained analysis comparing *d'* scores in B3 new and B3 old restricted to items that were identically repeated (no switch in grammaticality) yielded main effects of type, F (2, 46) = 3.65, MSE = 3.03, $p < .05$, $\varepsilon = .989$, and block, F (1, 23) = 10.3, MSE = 1.83, $p < .01$. These effects indicated that scores for the Similar type were higher than those for the Different type which, in turn, were higher than those for the Unique type. Scores for all types were higher for repeated unswitched items than for new items. Along the same lines, the analysis comparing *d'* scores in B3 new and B3 old restricted to repeated items that incurred a switch in grammaticality showed a marginally significant main effect of type, F (2, 46) = 2.99, MSE = 2.83, $p = .06$, $\varepsilon = .987$, and a significant effect of block, F (1, 23) = 27.73, MSE = 1.68, $p < .001$. These results again indicated higher scores for Similar and Different than Unique types, and higher scores in response to repeated switched items than to new items. Finally, the direct comparison between switched and unswitched B3 old items showed no significant effects.

Thus, accuracy results showed no difference between old and new items overall. However, scores in response to old items were consistently higher than those to new items (for both switched and unswitched old items) for all similarity types. Therefore, magnitude differences in *d'* scores for new and old items between types are likely responsible for the cancellation of the effect at the overall block comparison level.

3.2.3.2 ERP data.

3.2.3.2.1 500 to 625 ms. Figure 12 illustrates the mean ERP responses in B3 new and B3 old switched and unswitched items for each similarity type in the 500 to 625 ms time window. A comparison of B3 new and B3 old overall ERPs showed marginally-significant interactions between type, block, grammaticality, and site, F (4, 92) = 2.89, MSE = 4.51, $p = .052$, $\varepsilon = .629$; and between block, grammaticality, laterality, and site, F (4, 92) = 2.72, MSE = 1.23, $p = .055$, ε

= .706. Follow-up tests indicated grammatical sensitivity to Similar and Different types in bilateral centroparietal sites in B3 new as opposed to sensitivity to all cross-language similarity types in B3 old. B3 old sensitivity was also apparent in left and midline frontal sites, although greater in bilateral centroparietal sites.

Finally, the analysis comparing ERPs in B3 new and B3 old restricted to items that were identically repeated revealed various significant interactions between block, type, grammaticality, and site, F (4, 92) = 4.96, MSE = 9.73, $p < .01$, $\varepsilon = .602$; block, grammaticality, laterality, and site, F (4, 92) = 3.96, MSE = .992, $p < .01$, $\varepsilon = .790$; and between block, type, grammaticality, laterality, and site, F (8, 184) = 2.78, MSE = .808, $p < .01$. Probing of the latter interaction with follow-up tests indicated grammatical sensitivity of greater magnitude in B3 old compared to B3 new in all cross-language similarity types with a centroparietal distribution in the Different type and a left frontoparietal distribution in the Unique type. The distribution in the Similar type was widespread in this block, and the overall distribution in B3 new was mostly centroparietal. There was no effect of block in the comparison between B3 new and B3 old restricted to repeated items that incurred a switch in grammaticality. Also, a direct comparison between B3 old switched and unswitched items showed no significant effect of this variable.

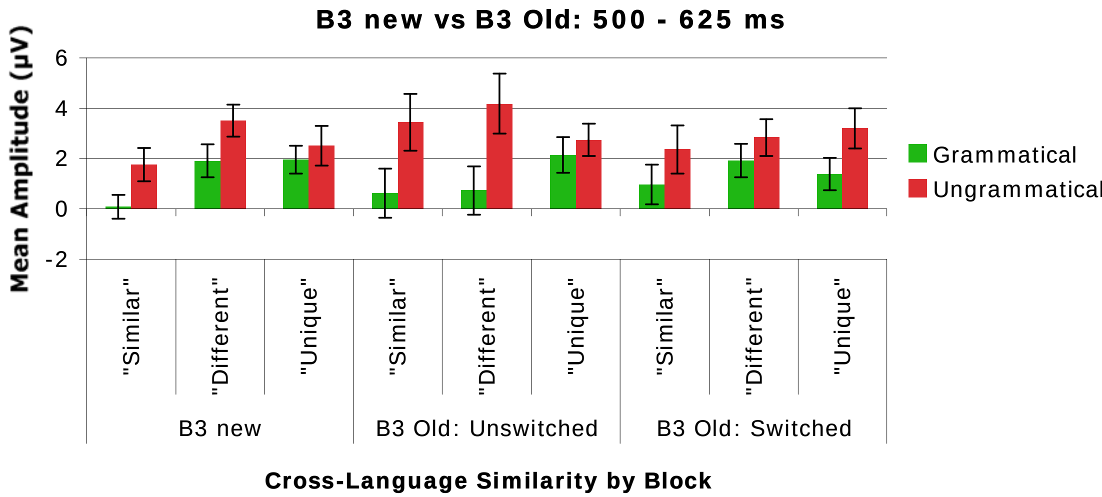


Figure 12. Mean ERPs in B3 new and B3 old switched and unswitched items for each cross-language similarity type in the 500 to 625 ms latency window.

3.2.3.2.2 *625 to 750 ms*. Figure 13 illustrates the mean ERP responses in B3 new and B3 old for switched and unswitched items for each similarity type in the 625 to 750 ms time window. In the analysis comparing ERPs in B3 new and B3 old overall there was a significant interaction between type, block, grammaticality, laterality, and site, F (8, 184) = 2.28, MSE = 1.05, $p < .05$, ε = .648. Follow-up tests revealed grammatical sensitivity in all cross-language similarity types in B3 old as well as B3 new, unlike in the previous time window. Furthermore, ERPs in B3 new were bilateral and mostly centroparietally distributed for all types but for the Similar type in B3 old they were of similar magnitude across sites and right-lateralized in frontal sites.

The analysis comparing ERPs in B3 new and B3 old restricted to items that were identically repeated revealed significant interactions between block, type, grammaticality, and site, F (4, 92) = 4.84, MSE = 10.99, $p < .01$, ε = .595; and between block, type, grammaticality, laterality, and site, F (8, 184) = 3.43, MSE = .893, p = .001. Follow-up tests to the latter interaction indicated that ERPs in B3 new were mostly centroparietally distributed in all cross-language similarity types but more bilateral in the Different type. ERPs in B3 old were more widespread across the scalp, though more right-lateralized in the Unique type. The magnitude of B3 old responses were greater than those in B3 new for the Different type only. As in the previous time window, there was no relevant effect of block when comparing B3 new and B3 old restricted to repeated items that incurred a switch in grammaticality.

Finally, and unlike in the previous time window, the direct comparison between between B3 old switched and unswitched items showed a significant interaction between switch, type, grammaticality, laterality, and site, F (8, 184) = 2.07, MSE = 1.07, p < .05. Follow-up tests indicated mainly a centroparietal distribution of the grammaticality effect in response to switched items for all types. ERPs in response to unswitched items were centroparietally-distributed mainly for the Different type; the Similar type showed equivalent widely distributed responses and responses to the Unique type were more right-lateralized in central sites. Unswitched items showed greater magnitude of responses but only for the Similar and Different types. Responses to the Unique type actually showed an opposite trend: greater grammatical sensitivity to the switched items.

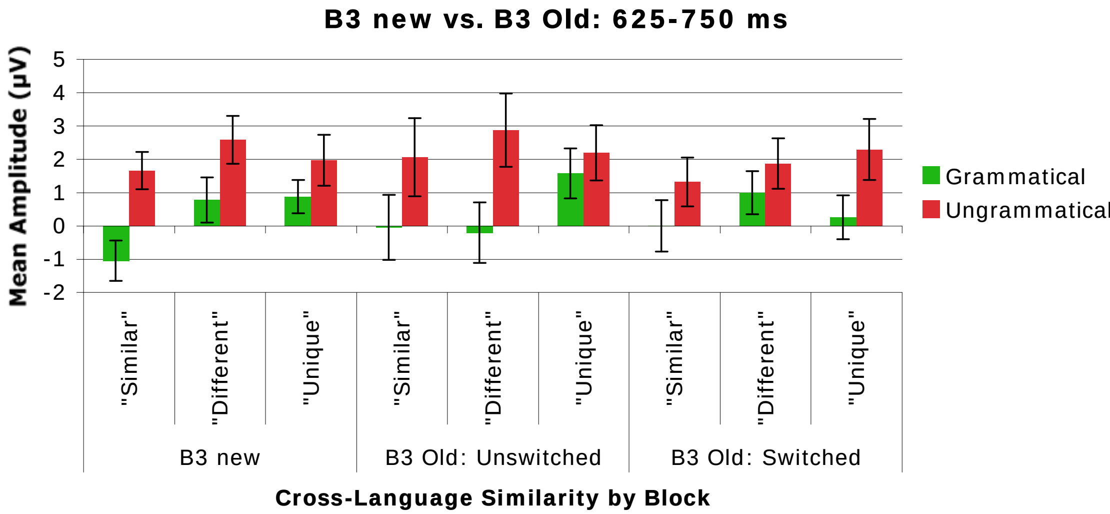


Figure 13. Mean ERPs in B3 new and B3 old for switched and unswitched items for each cross-language similarity type in the 625 to 750 ms latency window.

In summary, the learning processes underlying the increase in accuracy after the B2 intervention appear to be more robust for items that were previously encountered because *d'* scores for both switched and unswitched repeated items were higher than those for new items. Thus, even though learning generalized to new items, it was greater for repeated items. The ERP data showed a similar advantage in magnitude for repeated unswitched items, but for the Different type only. On the other hand, ERPs in response to repeated switched items were no different than those to new items, suggesting that participants were implicitly more sensitive to items that were identically repeated than to items that incurred a grammatical switch. Indeed this was true in the ERP analysis comparing switched and unswitched items, but only for the Similar and Different types. This pattern, however, was not apparent in the accuracy data (no difference between switched and unswitched). Thus, the ERP and accuracy data diverged for responses to switched items: there was no difference in explicit processing between the two item types but participants were more implicitly sensitive to the unswitched items. This pattern was furthermore modulated by cross-language similarity, as these differences were most apparent for the Similar and Different types.

3.2.4 Additional time windows. In this section we report ERP results from the analyses of the secondary time windows as well as of the English stimuli. Additionally, we present topographical map data for the various similarity types and blocks.

3.2.4.1 75 to 125 ms. Figure 14 illustrates the mean ERP responses for each similarity type in B1 and B3 new in the 75 to 125 ms time window. The analysis comparing B1 and B3 new revealed significant interactions between block, type, and grammaticality, F (2, 46) = 3.8, MSE = 19.37, $p < .05$, ε = .979, and between block, grammaticality, and laterality, F (2, 46) = 6.05, MSE = 1.21, $p < .01$, ε = .963. Follow-up tests indicated that ERP responses to grammatical items were significantly more negative than ungrammatical items in B3 new for the Similar type only. Additionally, marginally-significant interactions between block, type, grammaticality, and site, F (4, 92) = 2.78, MSE = 4.29, p = .066, ε = .557, and between block, type, grammaticality, and laterality, F (4, 92) = 2.34, MSE = 1.17, p = .061, ε = .826 suggested that ungrammatical items showed more negative responses than grammatical items in midline and right centroparietal sites for B3 new for the Different type only.

Efforts to determine which ERP components might be at play in this time window were limited by the fact that ERPs for the Similar and Different types showed opposite patterns of grammaticality effects (see Figure 14). Furthermore, to our knowledge, the effect in neither type closely matches any known component in this latency window. We return to this issue in the general discussion.

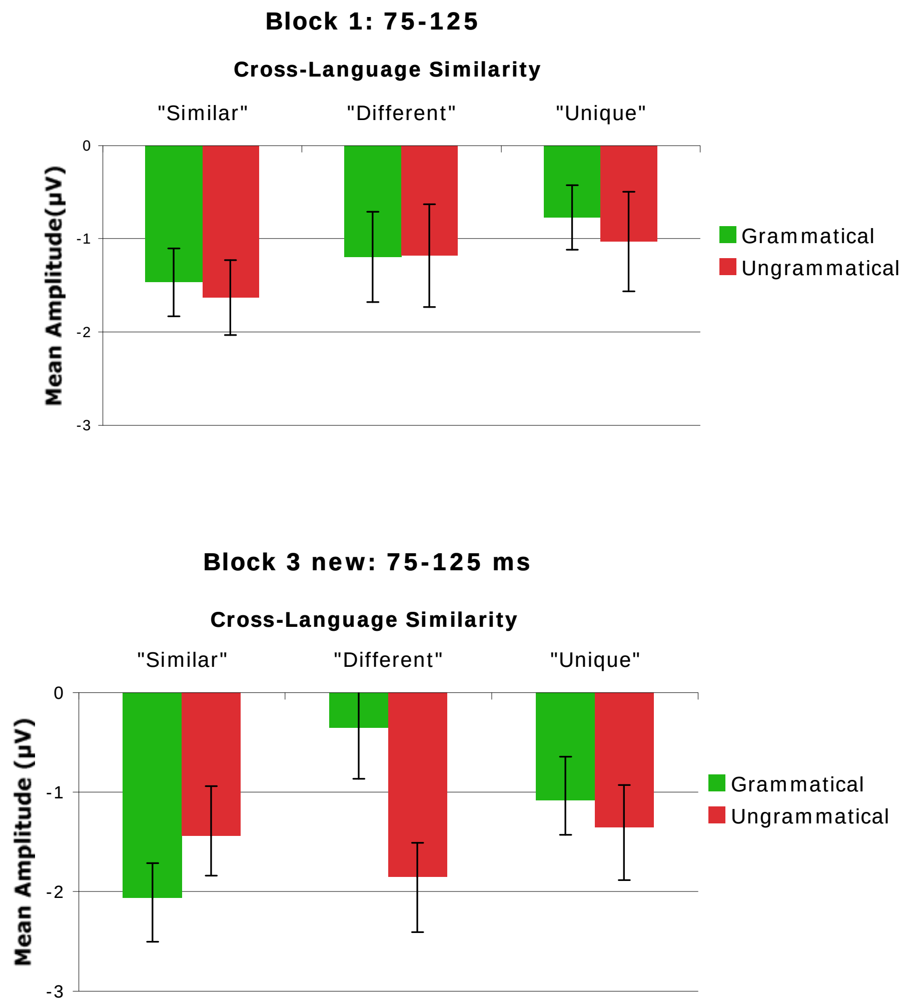
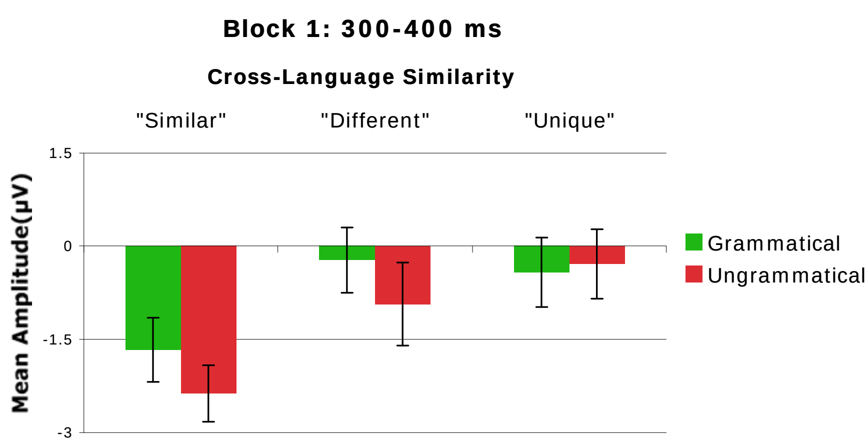


Figure 14. Mean ERPs for each similarity type in B1 and B3 new in the 75 to 125 ms latency window.

3.2.4.2 300 to 400 ms. Figure 15 illustrates the mean ERP responses for each similarity type in B1 and B3 new in the 300 to 400 ms time window. Analyses in this time window yielded no statistically significant effects. However, there was a trend toward an overall grammaticality effect mainly in frontal sites as suggested by a marginally-significant interaction between grammaticality and site in this time window, F (2, 46) = 3.64, MSE = 6.9, p = .059, ε = .610. Because of the trend toward frontal sites, this effect is likely more indicative of an earlier ERP component that is sensitive to syntactic anomalies such as the left anterior negativity (LAN) rather than an N400, which has a more centroparietal distribution and is usually seen in response to semantic anomalies. However, though anterior, the scalp distributions were not left-lateralized, as in a typical LAN effect. Furthermore, ERPs to ungrammatical items tended to be more negative than those to grammatical items in B1 only; in B3 new they tended to be more positive. We return to this issue in the general discussion.



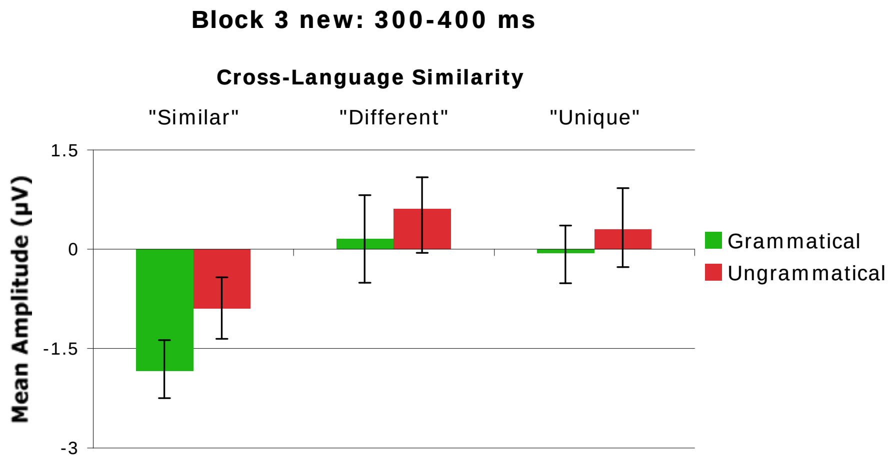


Figure 15. Mean ERPs for each cross-language similarity type in B1 and B3 new in the 300 to 400 ms latency window.

3.2.5 English Stimuli. Figure 16 illustrates the mean ERP responses for the English stimuli in each time window. There were no grammaticality effects in the earlier time windows.

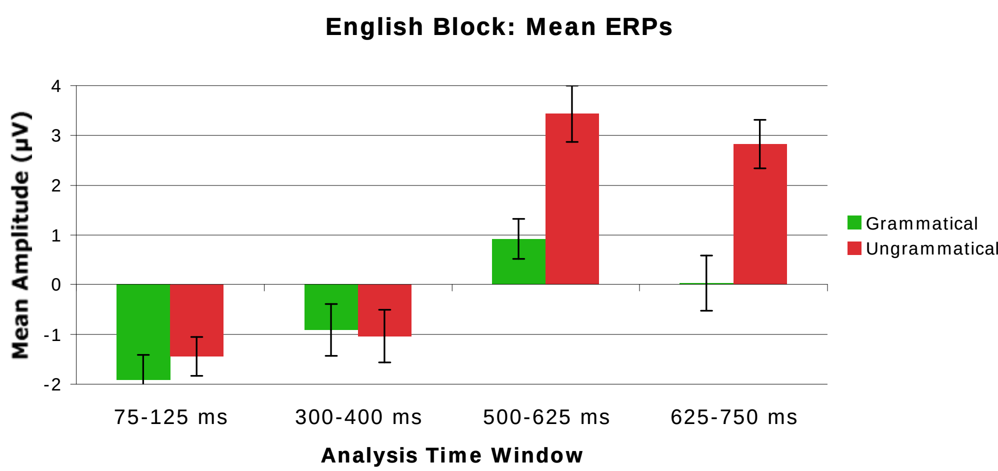


Figure 16. Mean ERPs for the English stimuli in each time window.

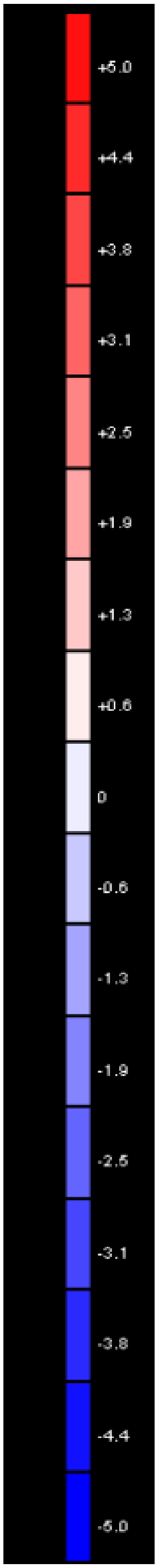
3.2.5.1 500 to 625 ms. In this time window there was a significant interaction between grammaticality and laterality, F (2, 46) = 3.79, MSE = 2.37, $p < .05$. Follow-up tests indicated greater sensitivity to grammaticality in left and midline sites. This interaction qualifies a main effect of grammaticality, F (1, 23) = 14.86, MSE = 46.06, p = .001.

3.2.5.2 625 to 750 ms. Grammaticality and site interacted in this time window, F (2, 46) = 10.05, MSE = 6.72, $p < .01$, ε = .615, and follow-up tests indicated greater sensitivity to grammaticality in centroparietal sites. This interaction qualifies a main effect of grammaticality, F (1, 23) = 24, MSE = 35.12, $p < .001$.

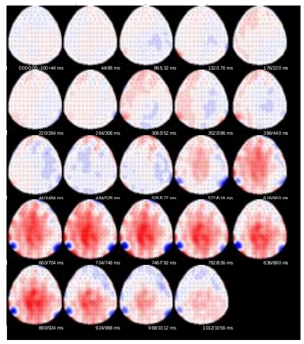
Topographical maps representing scalp distributions of ERPs in the English block were qualitatively compared to those of the Similar, Different, and Unique types with the goal of examining whether more similar types would show greater correspondence in scalp distribution to the English maps and, furthermore, whether this changed with increased explicit and implicit processing. If obvious differences in topography are present then it is reasonable to assume that these reflect underlying differences in neural generators because it would be nearly impossible for the same neural source to result in two very different scalp distributions (Michel, Murray, Lantz, Gonzalez, Spinelli, & Grave de Peralta, 2004).

Figure 17 illustrates the scalp topographies of difference waves (ungrammatical – grammatical) in B1 and B3 new for each similarity type as well as the English block. Whereas no striking differences in scalp distribution were observed between similarity types and in relation to English maps, suggesting similar neural sources in each case, some differences in magnitude were observed. Specifically, the magnitude of the difference waves in the P600 window (approximately 500 to 900 ms) exhibited a centroparietal distribution and was clearly greater for the Similar type, followed by the Different type, and was barely present in B3 new for the

Unique type. This pattern is consistent with the main findings from the accuracy and ERP analyses. Furthermore, there appeared to be a shift in topography for the Similar type from frontal sites in B1 to centroparietal sites in B3 new. This pattern closely matches that of the ERP data analysis within this specific similarity type, which indicated widely distributed and mostly frontal distributions in B1 but only centroparietal sites in B3 new.

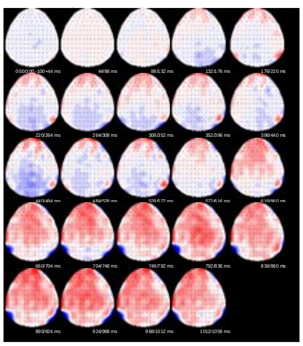


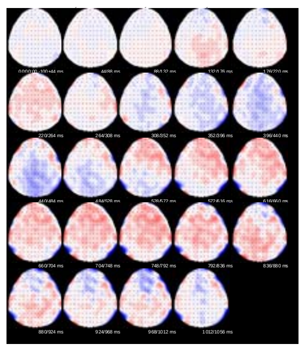
English:

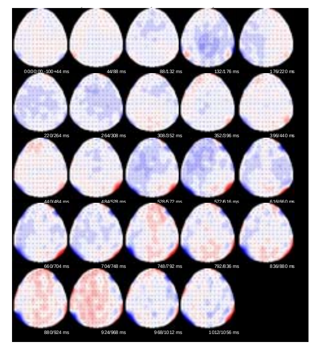


Similar Different Unique

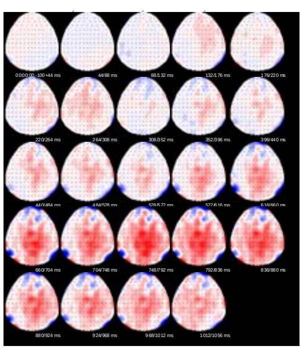
Block 1:

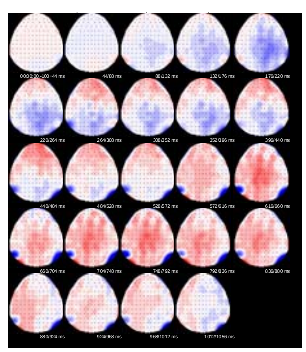






Block 3 new:





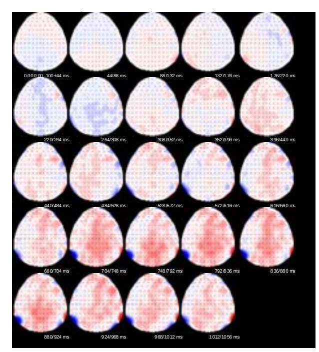


Figure 17. Topographical maps of difference waves from -100 ms to 1000 ms for English stimuli and B1 and B3 new for each cross-language similarity type.

3.3 GENERAL DISCUSSION

The present study investigated the relationship between implicit and explicit L2 processing by increasing participants' level of explicit accuracy in a GJT and simultaneously measuring their brain activity before and after accuracy increase. Results indicated that the explicit accuracy increase was accompanied by a concurrent increase in implicit sensitivity, thus suggesting an interaction between the two processing types. As participants became more accurate in their grammaticality judgments this was reflected in the pattern of their brain waves indicating an influence of explicit processing on implicit processing. An analogous increase in the amplitude of the P600 has been observed with increased L2 proficiency levels. Rossi, Gugler, Friederici, & Hahne (2006) showed that groups of highly proficient Italian learners of German and German learners of Italian exhibited a P600 in response to subject-verb agreement violations that was of greater magnitude than the P600 exhibited by moderately-proficient analogous groups of learners. Behavioral accuracy scores were also correlated with level of proficiency in this study such that high-proficiency learners were significantly more accurate than low-proficiency learners. Thus, although the present study was conducted on a shorter time scale (and within-participants), the two sets of findings suggest that increased L2 proficiency is accompanied by a parallel increase in implicit processing.

The present results conflict with a strict noninterface position of the relationship between implicit and processing and are instead more in line with an interface position. According to N. Ellis (2005), the learning, representation, and processing of language constitute interacting components of a larger dynamic network. Explicit knowledge can modify implicit knowledge via top-down attentional mechanisms that shape the input to our implicit learning systems. That is, what we explicitly pay attention to becomes the basis of patterns that are subsequently formed automatically and without awareness. Ellis also suggests that practice of linguistic rules promotes implicit learning and proceduralization. In the context of the present experiment, this would provide a plausible mechanism for the observed parallel changes in explicit and implicit processing. As participants improved at recalling patterns and/or practiced rules of grammatical agreement, their implicit systems became more tuned to these allowing them to recognize violations faster, more automatically, and without conscious awareness. However, it is unclear that the explicit increase in accuracy exerted a direct influence on implicit processing. An alternative mechanism would be that the intervention in block 2 simultaneously increased both explicit and implicit processing. Another alternative is that the B2 intervention actually increased implicit sensitivity which, in turn, contributed to an increase in explicit accuracy. Nevertheless, regardless of the exact mechanism, which remains to be elucidated, it can be concluded from the present results that implicit and explicit types of L2 processing do interact.

The results of the present study are similar to those of the N. Ellis (1993) study in which the (explicit) “Rule” group was later able to transfer and apply its knowledge to new structures. Participants in our study demonstrated an increase in accuracy not only in response to items that they had seen before but also to new items. It is worth pointing out that an argument could be made for the role of practice alone in increasing brain sensitivity to violations. Although possible, it is unlikely that in this experiment practice alone was responsible for the observed changes in implicit processing and that these were not instead reflective of the explicit accuracy increase due to the B2 intervention. The main reason for this is that an analysis of ERPs from the pilot experiment indicated that the SNF condition (whole sentences with no feedback) did not result in an increase in brain sensitivity in B3 new as compared to B1. Though from a limited set of participants, this preliminary result suggests that the increase in implicit sensitivity observed in the primary experiment was likely due to the accuracy-increasing intervention and not to overall practice effects. Furthermore, there were also no practice effects in accuracy as indicated by statistical comparisons between scores in the first and last thirds of B1 in the present experiment, and between B1 and B3 new for the SNF condition in the pilot experiment.

The present study also investigated the role of cross-language similarity in implicit and explicit L2 processing by employing violations that are formed similarly in the L1 and L2, differently in the L1 and L2, and that are unique to the L2. Results indicated that cross-language similarity exerted a strong modulatory influence on both types of processing. Specifically, accuracy scores were consistently higher for the Similar (demonstrative determiner number agreement) and Different types (determiner number agreement) than for the Unique type (determiner gender agreement), and similar patterns were observed in brain sensitivity to grammaticality. Similarity type modulated performance on new versus old items and exerted an influence on which kind of old items were most successfully processed. When explicit accuracy and ERP patterns diverged (e.g., equivalent accuracy scores for switched and unswitched B3 old items but greater brain sensitivity for unswitched items), this was usually due to the modulatory influence of cross-language similarity. Nevertheless, despite the fact that in the present experiment we attempted to control for any superfluous differences in the constructions used for each cross-language similarity type, it is possible that these varied in other ways. For example, it could be the case that the rules of agreement for the Similar and Different types are easier to apply in general than those for the Unique type, thus rendering participants relatively more sensitive to the former. It will be interesting for further studies to shed some light on this issue.

A role for cross-language similarity has been reported in several studies reviewed above. Chen et al. (2007) reported that Chinese learners of English showed no P600 in response to subject-verb agreement violations despite high accuracy scores. Because this type of agreement does not constitute a part of the Chinese language, this result can be viewed as akin to the obtained results for the Unique type in the present experiment: despite clearly above-chance overt performance ($d' > 2$), participants in the present experiment showed very little sensitivity to Spanish determiner gender agreement violations, except when these were repeated. The present results are also consistent with those of Osterhout et al. (2006) and Tokowicz and MacWhinney (2005) when analyzing ERP responses in English-speaking learners to violations of the Similar type. The former study employed verbal-person agreement in French and the latter used auxiliary omission agreement in Spanish. Both reported a P600 in participants' responses to violations, as in the present study. However, the pattern of results across studies differed for the Different and Unique types: whereas neither of the two studies found a P600 in response to determiner number agreement violations, this was observed in the present study, albeit to a lesser degree than that found in response to demonstrative determiner number agreement (Similar type). Although it is possible that the discrepancy with the Osterhout et al. (2006) study is due to the length of L2 instruction in the two groups of participants, this is unlikely because participants in that study had been learning the L2 for 8 months and participants in our study had a similar amount of instruction on average (though a broader range from only a few months to approximately 16 months). A different possibility is that participants in that study were not sensitive to determiner number agreement because this feature is not phonologically realized in French. However, Tokowicz and MacWhinney also did not find a P600 effect for this type of violation in Spanish, a language in which it is phonologically realized. Another discrepancy with this study is that, whereas in the present experiment participants were mostly insensitive to grammaticality for the Unique type as assessed by ERPs, Tokowicz and MacWhinney found that participants were implicitly sensitive to grammaticality in an identical stimulus condition. It is possible that differences in stimuli could explain the observed disparities. Specifically, Tokowicz and MacWhinney employed 9 different syntactic patterns (most served as fillers) whereas the present study only used three. Furthermore, the present study employed two different types of number-agreement (demonstrative determiner and determiner number agreement) whereas the former only used one. It has been shown that the structure of the input is highly relevant to L2 processing, resulting in higher rates of learning and generalization (N. Ellis, 1993). However, it remains unclear why more structure in the stimulus list would result in a benefit to Different but not to Unique violation types. One possibility is that increased structure in the stimulus list results in an increase in cue strength thus helping to resolve likely competition for the Different type. The Unique type would not necessarily benefit to the same extent because there may be no competition and, in the present case, no gender-system representation in the first place. Furthermore, it is also possible that the existence of two kinds of number-agreement in the present study but only one kind of gender agreement helped participants in the processing of that particular type of cue. Another major difference between the Tokowicz and MacWhinney study and the present study is the level of stimulus controls employed. Specifically, although the same constructions were used for the Different and Unique types in the two studies, sentences in the present study were identical across similarity types, differing only in the determiner preceding the locus of violation. Furthermore, the number of masculine and feminine critical words, as well as singular and plural words, were balanced across sentences. This was not the case in the former study, in which sentences varied across similarity types. This difference could have been responsible for the observed discrepancies in the two studies, possibly coincidentally rendering Unique type sentences relatively easier to process in the former study and Different type sentences more difficult.

Furthermore, participants' judgment accuracy in the Tokowicz and MacWhinney (2005) study was near chance levels for all constructions. It is possible, though unlikely, that differences in proficiency could account for these discrepancies because participants were drawn from similar college-student pools of Spanish learners. A more likely possibility is that there is a difference in the two studies in the number of relevant words that come after the violation point in a sentence, and thus constitute additional pieces of information that aids participants in their grammaticality judgment. Indeed, McClain (2007, unpublished honors thesis) found that judgment accuracy was positively correlated with the amount of additional information in the sentence following the critical word. If sentences in the present study exhibited a greater amount of relevant information, this could account for the higher accuracy scores overall.

In the present study we also investigated the generalizability of the increases in explicit and implicit processing. Results suggest that learning did generalize to new items (B3 new) but, in general, was more robust for repeated items. More specifically, items that were repeated in an identical way showed an advantage over repeated items that incurred a switch in grammaticality for both explicit and implicit measures. This pattern was further modulated by cross-language similarity and was most evident for the Different type, perhaps reflecting a greater "boost" to this type when items are repeated as compared to the easier Similar type. It is reasonable to assume that these effects of repetition were due to learning and subsequent application of the correct agreement rules because they are manifested in the P600 time window and not in earlier windows where more immediate episodic memory effects would be expected. Such rote memorization effects would more likely manifest themselves as a P300 effect instead, as suggested by studies in which rote memorizers were shown to exhibit larger P300s for words subsequently recalled than participants who employed more elaborative memorization strategies (Fabiani, Gratton, & Federmeier, 2007). There was no indication of early positivities in the present study.

ERP results from earlier time windows, specifically between 75 and 125 ms and between 300 and 400 ms, yielded unclear results, thus making it difficult to map any of the obtained grammaticality effects to any known ERP components. Nevertheless, possible components underlying the grammaticality effect peaking at about 100 ms post-stimulus with a centroparietal distribution for the Different type are the N100 (or N1) and/or the ELAN. However, neither constitutes a very likely match because the former has been observed mainly in response to elemental perceptual feature analysis and the latter, though indicative of syntactic violations, has a typical left and anterior distribution (Fabiani, Gratton, & Federmeier, 2007). Because in the present study grammatical and ungrammatical sentences were counterbalanced across participants and sentences across similarity types were identical except for the determiner preceding the critical word, it is highly unlikely that a basic perceptual feature difference would explain the obtained results. On the other hand, a possible component reflecting the opposite pattern for the Similar type (ERPs in response to ungrammatical items were more *positive* than those to grammatical items) in the same latency window is the P150. However, this component has been observed mainly in response to visual perception categorization and selective attention, rendering it an unlikely component in this case, again because stimulus features were carefully controlled for in this study. Finally, it is possible that the general trend toward more negative ERPs in response to ungrammatical than to grammatical items in the 300 to 400 ms time window reflects a LAN-type effect. This component is often (but not always) elicited by syntactic violations and has also been shown to exhibit a bilateral, and not just a left-lateralized, anterior distribution (Kutas, Federmeier, Staab, & Kluender, 2007). Nevertheless, such attempts to classify apparent early grammaticality effects should be interpreted with caution as a number of the analyses only yielded marginally-significant results and incidental effects always remain a possibility in any statistical analysis.

Qualitative analyses of the topographical maps as well as translation data showed similar patterns to those observed in the main data set. Visual inspection of the scalp distributions of the difference-wave ERPs in the various cross-language similarity types showed no striking differences in distribution to each other and to the English maps, suggesting similar neural sources in each case. However, a few observations are worth mentioning. Firstly, the magnitude of the difference waves in the P600 window exhibited a centroparietal distribution and was clearly greater for the Similar type. This pattern is consistent with the main findings from the accuracy and ERP analyses. Secondly, the only relevant shift in scalp topography appears to be for the Similar type going from frontal sites in B1 to centroparietal sites in B3 new. This pattern closely matches that of the ERP data analysis within this specific similarity type, which indicated widely distributed and mostly frontal distributions in B1 but only centroparietal sites in B3 new. This was also accompanied by an increase in the magnitude of the response. Decreases in activity in frontal parts of the brain have been reported in fMRI studies examining the neural correlates of skill acquisition and the development of automaticity. Specifically, decreases in activity in frontal and parietal parts of the brain that are associated with working memory and attentional processes, respectively, have been reported after just 60 minutes of practicing a simple visual tracking task (Hill & Schneider, 2006). Although the nature of the task in the present study is very different and the location of the neural source could be different from that observed at the scalp, it is possible that the shift in topography accompanying an increase in accuracy in the Similar type reflects increased automaticity and less effortful processing. The reason why this is only apparent for the Similar type may be because it was the only type that showed a robust grammaticality effect in B1 at the outset. Thus, it is possible that if the other similarity types had also shown grammatical sensitivity in B1 this could have exhibited a frontal distribution followed by a centroparietal shift.

Finally, qualitative analyses of the translation task data showed that participants made more mistakes overall in the Similar sentences than the other similarity types. However, the quality of the errors was superior in this case, as participants usually just substituted a letter (e.g. “Estes” instead of “Estos”) without also incurring a change in gender and/or number, as for the other types. Thus, when accounting for error quality, a different scenario surfaces in which participants made more prejudicial mistakes in the Unique type. It is interesting to note here that, though exhibiting a high degree of cross-linguistic correspondence, the grammatical constructions used in the Similar type are not prevalent in the input in the Spanish language. Thus, the Unique type, for example, benefits from a much higher degree of cue availability. Nevertheless, the present results seem to indicate that the processing of available cues is subject to the influence of cross-language similarity, with the latter factor contributing more to cue strength beyond availability.

In summary, the present experiment showed that implicit and explicit types of L2 processing can and do interact; specifically, an increase in explicit processing was closely coupled with an increase in implicit processing, as measured by accuracy judgments and ERPs, respectively. Furthermore, the present study demonstrated that similarity between the L1 and L2 plays a large role in the two types of processing, generally favoring constructions that are most similar to the L1. The translation data suggest that it would be interesting to examine these relationships and the role of cross-language similarity more closely in the context of a production task.

The findings from the present study have practical implications for several fields of current L2 research, from the development of classroom language-instruction curricula to treatment approaches for recovering bilingual aphasic patients as well as the issue of diglossia in former European colonies throughout the world. The finding that, despite possible divergence, explicit L2 processing exerts an influence on implicit processing highlights the importance of an integrated framework that asseses and capitalizes on both types of processing in language instruction and therapy. Furthermore, the current study suggests that feedback, repetition, and cross-language similarity may have important consequences for L2 learning and recovery.

BIBLIOGRAPHY

Abutalebi, J., Cappa, S. F., & Perani, D. (2005). What can functional neuroimaging tell us about the bilingual brain? In J. Kroll & A. de Groot (Eds.). *Handbook of Bilingualism: Psycholinguistic approaches*. New York, NY: Oxford University Press.

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

Basnight-Brown, D. M., Chen, L., Shu, H., Kostic, A. & Feldman, L. B. (2007). Monolingual and bilingual recognition of regular and irregular English verbs: Sensitivity to form similarity varies with first language experience. *Journal of Memory and Language, 57*, 65-80.

Carreiras, M., Vergara, M., & Horacio, B. (2005). Early event-related potential effects of syllabic processing during visual word recognition. *Journal of Cognitive Neuroscience, 17*, 1803-1817.

Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing, 12*, 267-296.

Chen, L., Shu, H., Liu, Y., Zhao, J. & Li, P. (2007). ERP signatures of subject–verb agreement in L2 learning. *Bilingualism: Language and Cognition, 10*, 161-174.

Coren, S. (1992). The Left-Hander Syndrome: The Causes and Consequences of Left-Handedness, Free Press, New York.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences, 22*, 735–808.

Ellis, N. (1993). Rules and instances in foreign language learning: Interactions of explicit and implicit knowledge. *European Journal of Cognitive Psychology, 5*, 289 – 318.

Ellis, N. (2005). At the interface: dynamic interactions of explicit and implicit language knowledge. *SSLA, 27*, 305-352.

Ellis, R. (2004). The Definition and Measurement of L2 Explicit Knowledge. *Language Learning, 54*, 227-275.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language. *SSLA*, *27*, 141-172.

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *SSLA, 28*, 339-368.

Fabiani, M. Gratton, G, & Federmeier, K. D. (2007). Event-related brain potentials: Methods, theory, and application. In J. T. Cacioppo, L. Tassinary, and G. Berntson (Eds.). *Handbook of Psychophysiology*, 3rd edition. Cambridge: Cambridge University Press.

Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing, 10*, 161-184.

Green, D.M. & Swets, J.A. (1974). Signal Detection Theory and Psychophysics. New York: Wiley.

Greenhouse, S. W. & Geiser, S. (1959). On methods in the analysis of profile data. *Psychometrica*, 24, 95-112.

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related potentials. *Journal of Psycholinguistics Research*, *30*, 251-265.

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

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

Han, Y., & Eliis, R. (1998). Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research, 2,* 1-23.

Hill, N.M. & Schneider, W. (2006). Brain changes in the Development of Expertise: Neuroanatomical and Neurophysiological Evidence about Skill-based Adaptations. In K. A. Ericsson, N. Charness, P. Feltovich, and R.Hoffman (Eds.), *Cambridge Handbook of Expertise and Expert Performance*. New York: Cambridge University Press.

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

Hulstijn, J. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning. *SSLA*, 27, 129-140.

Jasper, H.H. (1958). The Ten Twenty Electrode System of the International Federation. *Clinical Neurophysiology, 10*, 371-375.

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

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

Kutas, M., Federmeier, K. D., Staab, J., & Kluender, R. (2007). Language. In J. T. Cacioppo, L. Tassinary, and G. Berntson (Eds.). *Handbook of Psychophysiology*, 3rd edition. Cambridge: Cambridge University Press.

Mackey, A., & Silver, R. E. (2005). Interactional tasks and English L2 learning by immigrant children in Singapore. *System*, *33*, 239-260.

MacWhinney, B. (2005). A unified model of language acquisition. In J. Kroll & A. de Groot (Eds.). *Handbook of Bilingualism: Psycholinguistic approaches*. New York, NY: Oxford University Press.

McClain, R.M. (2007). *Piecing together the puzzle: the effects of agreement information availability on second language learners' grammaticality judgment, confidence and accuracy*. Unpublished manuscript.

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

Michel, C.M., Murray, M.M., Lantz, G., Gonzalez, S., Spinelli, L., & Grave de Peralta, R. (2004). EEG source imaging. *Clinical Neurophysiology, 115*, 2195-2222.

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

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

Osterhout, L., McLaughlin, J., Kim, A., Greenwald, R., & Inoue, K. (2004). Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. In M. Carreiras & C. Clifton, Jr. (eds.), *The on-line study of sentence comprehension: Eyetracking, ERP, and beyond*. Psychology Press.

Osterhout, L., McLaughlin, J., Pitkanen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A paradigm for exploring the neurocognition of second-language processing. *Language Learning, 56*, 199-230.

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, Jr., R., Miller, G. A., Ritter, W., Ruchkin, D. S., Rugg, M. D., & Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology, 37*,127–152.

Rosa, E. M., & Leow, R. P. (2004). Computerized Task-Based Exposure, Explicitness, Type of Feedback, and Spanish L2 Development. *Modern Language Journal, 88*, 192-216.

Rossi, S., Gugler, M.F., Friederici, A.D., & Hahne, A. (2006). The Impact of Proficiency on Syntactic Second-language Processing of German and Italian: Evidence from Event-related Potentials. *Journal of Cognitive Neuroscience,18*, 2030-2048.

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.

Sanz, C., & Morgan-Short, K. (2004). Positive evidence vs. explicit rule presentation and explicit negative feedback: A computer-assisted study. *Language Learning, 54*, 35-78.

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar. *SSLA*, *27*, 173-204.

Tokowicz, N., Michael, E. B., & Kroll, J. F. (2004). The roles of study-abroad experience and working-memory capacity in the types of errors made during translation. Bilingualism: Language and Cognition, 7, 255-272.

Footnotes

1. Because of the way the stimuli were created, the Different and Unique similarity conditions resulted in identical sentences/word pairs in their grammatical form but were nevertheless considered to be separate experimental versions for practical reasons.

2. Analyses of ERP data that included the variable "exchange" showed significant interactions with block and grammaticality in both analysis time windows. However, the nature of these interactions bore no relevance to our hypothesis as they failed to suggest that increased accuracy and brain sensitivity in B3 new was due to differences in the stimuli between B1 and B3 new.