THE ROLE OF THE L1 AND INDIVIDUAL DIFFERENCES IN L2 SENSITIVITY TO MORPHOSYNTACTIC FEATURES: AN ERP INVESTIGATION

By

Kristi Bond

Submitted to the graduate degree program in Linguistics and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

____

Co-chair: Dr. Alison Gabriele

____

Co-chair: Dr. Robert Fiorentino

____

Member: Dr. Harold Torrence

____

Member: Dr. Amy Rossomondo

____

Member: Dr. Utako Minai

Date Defended: January 25, 2013

The Dissertation Committee for Kristi Bond
certifies that this is the approved version of the following dissertation:

THE ROLE OF THE L1 AND INDIVIDUAL DIFFERENCES
IN L2 SENSITIVITY TO MORPHOSYNTACTIC FEATURES:
AN ERP INVESTIGATION

Co-chair: Dr. Alison Gabriele

Co-chair: Dr. Robert Fiorentino

Member: Dr. Harold Torrence

Member: Dr. Amy Rossomondo

Member: Dr. Utako Minai

Date approved: January 25, 2013

ABSTRACT

This study used ERP (event-related potentials) to examine both the role of the L1 and the role of individual differences in the processing of agreement violations. Theories of L2 acquisition differ with regard to whether or not native-like acquisition of L2 features is possible (Schwartz and Sprouse, 1994, 1996; Tsimpli and Mastropavlou, 2007), and the results of previous ERP studies are inconsistent when it comes to whether or not native-like processing is observed in response to L2 agreement violations (e.g., Sabourin, 2003; Tokowicz & MacWhinney, 2005). Furthermore, studies of learners in early stages of L2 acquisition have found variability in the emergence of native-like responses (e.g., McLaughlin et al., 2010; Tanner et al., 2012), but sources of variability have not been investigated. The current study examines responses to gender and number agreement violations in English-speaking learners of Spanish (n=24). Stimuli targeted agreement in three conditions: subject-verb agreement (el barco…flota/*flotan), which is similar in Spanish and English; number agreement on adjectival predicates (la isla…rocosa/*rocosas), a context in which agreement is not instantiated in English; and gender agreement on adjectival predicates (la isla…rocosa/*rocoso), which is unique to Spanish. Grammaticality judgments and ERP responses were also tested for correlations with aptitude scores on the Modern Languages Aptitude Test (MLAT; Carroll and Sapon, 1959) and LLAMA tests (Meara, 2005) and the Raven Advanced Progressive Matrices (Raven, 1965). Results are in line with theories that claim native-like processing is acquirable even for features that are not present in the L1 or are instantiated differently in the L1 and L2. Learners demonstrated similar ERP responses to a control group of native Spanish-speakers (n=12) with regard to all three agreement types, although the response to gender violations was more limited for learners, who also exhibited low proficiency. Additionally, the MLAT and LLAMA were significantly correlated with sensitivity to agreement violations, both in terms of grammaticality judgments and ERP amplitudes, indicating a role for verbal aptitude in L2 processing. No correlations were found for nonverbal aptitude.

ACKNOWLEDGMENTS

The past six years have taught me a lot about linguistics, but even more about life and the surprises that God can bring our way. I am grateful for His continued guidance, and for the strength that He has provided in tough times. “I have learned the secret of being content in any and every situation…I can do all things through him who gives me strength” (Philippians 4:12-13). He has supplied much of that strength through my husband, whose support has been constant. Eric, thank you for your frequent sacrifices over these last few years. This work would not have been completed without your support. My parents Les and Vicki Jones and my in-laws Dean and Judy Bond also deserve many thanks for their support and encouragement for so long.

My advisors Dr. Alison Gabriele and Dr. Robert Fiorentino have provided expert guidance with great patience, for which I am very grateful. Their valuable and timely feedback has moved the project along at every point. Special thanks also to my other committee members Harold Torrence, Utako Minai, and Amy Rossomondo, and to colleagues in the Second Language Acquisition Lab, the Neurolinguistics and Language Processing Lab, and the Department of Linguistics at the University of Kansas. This project was also supported by a National Science Foundation Graduate Research Fellowship, with additional funding provided by the University of Kansas General Research Fund (awarded to Alison Gabriele).

Many others deserve thanks for their roles in this work. Sincere thanks goes to José Alemán Bañon for his collaboration on every aspect of the project, as well as others who contributed to recruiting participants and/or data collection: Rosalea Postma-Carttar and Leigh Rysko of the KU Department of Spanish and Portuguese, Alonso Canales, Jamie Bost, Ella Fund-Reznicek, Yuka Naito-Billen, Wendy Herd, Lamar Hunt, Stephen Politzer-Ahles, Jeff Shouse, Jon Coffee, and Sally Nelson.

DEDICATION

This dissertation is dedicated

to those from whom I continue to learn

the meaning of sacrifice -

to my husband, my parents, and my God.

TABLE OF CONTENTS

CHAPTER 1:

INTRODUCTION

Adult second-language learners display wide variability in the extent to which they are able to acquire native-like representations and/or processing of features of their second language (L2). The current study measures event-related potentials (ERPs) using electroencephalography (EEG) to examine factors that may account for such variability. ERPs are a measurement of electrical activity in the brain time-locked to a specific stimulus. When averaged across multiple participants and multiple tokens of the same type of stimuli, ERPs can reveal very small changes in the brain's electrical signals related specifically to complex stimuli like syntactic violations or semantic anomalies. Here, learners' ERP responses to violations of morphosyntactic agreement are investigated with regard to the role of differences between specific features of the L2 and the native language (L1). A second aim of the study is to investigate the role of individual differences between learners in L2 processing, particularly differences in aptitude. Thus, the study combines two disparate fields of study that share one common characteristic: knowledge of an L1, coupled with the learner's own aptitudes for learning, necessarily form the basis of what the learner brings to L2 acquisition. The use of ERP methodology makes it possible to explore the neural correlates of these two factors in L2 processing.

With regard to native speakers, ERP studies have demonstrated an important distinction between lexical/semantic processing and syntactic processing. Lexical/semantic anomalies (e.g., *He spread the warm bread with socks...*) frequently produce negative-going EEG waveforms across central and posterior electrodes occurring between 200 and 600ms post-stimulus, a response that is referred to as the N400 (e.g., Kutas and Hillyard, 1980; Osterhout and Nicol, 1999). In contrast, the processing of morphosyntactic features (e.g., *The elected officials*

**hopes...*) is usually associated with a late positivity in the posterior electrodes that peaks around 600ms post-stimulus (the P600) and is sometimes preceded by a negative-going waveform (the LAN) whose left anterior distribution differs from the often central/posterior distribution of the N400 (e.g., Osterhout and Mobley, 1995; Hagoort et al., 1993; Friederici, 2002; Barber and Carreiras, 2005).

The results of ERP studies targeting adult L2 learners vary greatly. In particular, L2 ERP studies differ as to the presence or type of ERP components observed in response to morphosyntactic processes like number or gender agreement. In response to gender agreement violations, for example, Tokowicz and MacWhinney (2005) find the late syntactic response (P600) in learners whose L1 is genderless, but Sabourin (2003) does not. Additionally, some studies have found that native-like processing is not present even for features that are shared between the L1 and L2 in cases where those features are instantiated differently (McLaughlin et al., 2010; Foucart and Frenck-Mestre, 2010). In order to further investigate the role of features in adult L2 acquisition, the study reported here investigates ERP responses during the L2 processing of number and gender agreement in Spanish by English-speaking learners.

The Spanish language offers a set of number and gender features that trigger agreement on a variety of other sentential elements, e.g., verbs, adjectives, and determiners. Although the syntactic locus of grammatical gender features in Spanish is debated, gender is argued to be a property of noun phrases that is lexically determined (Corbett, 1991). The lexical gender of animate nouns, especially those referring to humans, often reflects biological gender while the lexical gender of inanimate nouns is arbitrary. Harris (1991) argues that the prototypical gender markers in Spanish are not in fact gender markers but rather word markers, since they appear on words with no gender, but he still categorizes the gender of Spanish nouns based on their conformity to the prototypical markers of *–o* for masculine nouns (1a) and *–a* for feminine nouns (1b).

(1) a. espejo$_{masc}$
'mirror'

b. espada$_{fem}$
'sword'

Non-canonical gender marking may also be present, wherein some masculine nouns end in *–a* some feminine nouns end in *–o*, and many nouns end in consonants, *–e*, or *–u*. It is important to note that L2 Spanish learners must assign the correct gender to nouns as they are acquired, in addition to acquiring the gender feature itself.

In addition to gender, Spanish nouns uniformly bear number. Singular morphology is null or phonetically unmarked; plural morphology depends on the phonetic properties of the final segment in the root. If the root ends in an unstressed vowel (2a), the plural marker is [-s] (2b). The marker [-es] is used if the final segment is a stressed vowel or consonant (2c-d)[1].

(2) a. vestido$_{sg}$
'dress'

b. vestidos$_{pl}$
'dresses'

c. nación$_{sg}$
'nation'

d. naciones$_{pl}$
'nations'

Critically, in Spanish, agreement marking surfaces on elements other than nouns. An example is provided in (3), where the determiner and adjective must agree in both gender and number with the masculine singular noun *vestido* 'dress' in (3a) and the feminine, plural noun

[1] A null (unmarked) plural is possible only in the case of roots of more than one syllable ending in [-s, -x], such as *tesis*, which is both the singular and plural form of "thesis".

faldas ‘skirts’ in (3b). Note that the auxiliary verb, as is also true of all main verbs in Spanish, must agree with the subject in number, but not gender.

(3) a. el vestido es *blanco/*a/*os/*as*
the$_{masc.sg}$ dress$_{masc.sg}$ is$_{sg}$ white$_{masc.sg/*fem.sg/*masc.pl/*fem.pl}$
‘the dress is white’

b. las faldas son *blancas/*os/*a/*o*
the$_{fem.pl}$ skirts$_{fem.pl}$ are$_{pl}$ white$_{fem.pl/*masc.pl/*fem.sg/*masc.sg}$
‘the dresses are white’

Gender and number on adjectives is marked in the same way as was described above for nouns.

Spanish also instantiates across-phrase number agreement on verbs, which display a number-marking system that is conflated with person-marking. Since person agreement is not under investigation here, only 3rd-person verb forms were used in the current study. In (4a), the verb appears with singular marking because it must agree with the singular subject NP. In contrast, the plural subject NP requires plural marking on the verb in (4b):

(4) a. el muchacho *estudia/*an* mucho
the$_{sg}$ boy$_{3sg}$ study$_{3sg/*3pl}$ much
‘the boy studies a lot’

b. los muchachos *estudian/*a* mucho
the$_{pl}$ boy$_{3pl}$ study$_{3pl/*3sg}$ much
‘the boys study a lot’

Under the current Minimalist Program (Chomsky, 1995, 2000), an important distinction is made between interpretable and uninterpretable instantiations of features like number and gender at two different levels of interpretation, Logical Form (LF) and Phonetic Form (PF). Features that are interpretable at LF are those that contribute semantic information, e.g. number and gender features on nouns. Uninterpretable number and gender features may occur on other elements in the sentence, but at LF they do not represent an inherent semantic feature of the element on which they occur. Features that are uninterpretable at LF may, however, be interpretable at PF if they display a phonetic realization common to that feature in the given language. Critically, number and gender features in Spanish are only interpretable at LF on the noun even though canonical markers may be interpretable at PF on other elements (Carstens, 2000).

The question of the adult L2 acquisition of LF-uninterpretable features like number and gender on elements other than nouns is under debate, particularly with regard to L1/L2 similarities. Theories of L2 acquisition like the *Interpretability Hypothesis* (Tsimpli and Mastropavlou, 2007; Tsimpli and Dimitrakopoulou, 2007) and the *Full Transfer/Full Access* theory (Schwartz and Sprouse, 1994, 1996) differ with regard to claims about whether or not native-like acquisition of uninterpretable features not present in the L1 is even possible for adult L2 learners, a debate that centers on distinct claims regarding the role of a proposed universal set of features and parameters for language, or Universal Grammar (UG) (Chomsky, 1965; 1980; 1981). The *Interpretability Hypothesis* claims that adult learners have indirect access to UG only through their L1 and so are unable to incorporate features of the L2 that are not instantiated in the L1 if those features are uninterpretable. The *Full Transfer/Full Access* approach, on the other hand, claims that L1 features and parameter settings are influential in the early stages of L2 acquisition, but that L1 properties can be abandoned in favor of other UG-constrained settings as needed in order to accommodate L2 input. These two theories are tested here by contrasting number agreement, which occurs in both English and Spanish, with gender agreement that is unique to Spanish. Additionally, since English instantiates number on verbs but not adjectives, the current study examines number agreement on both verbs and adjectives in order to further investigate recent claims that native-like processing is only possible when features are instantiated similarly in the L1 and L2 (McLaughlin et al., 2010; Foucart and Frenck-Mestre, 2010).

A third theory related to access to Universal Grammar in adult L2 acquisition, the *Fundamental Difference Hypothesis* (Bley-Vroman, 1989, 1990), emphasizes not just L1/L2 similarities, but also the role of individual differences between learners. Greater variability in adult L2 acquisition, as compared to child L1 learners, led Bley-Vroman to propose that adult L2 acquisition is dependent on domain-general cognitive processes that are susceptible to individual differences. DeKeyser (2000) argues in support of these claims based on evidence from verbal aptitude scores, but does not specifically investigate measures reflecting domain-general cognitive processes. The current study, therefore, examines the role of both verbal and nonverbal aptitude in L2 processing.

The role of individual differences in predicting brain responses to morphosyntactic violations has not been thoroughly explored in previous literature, but there is recent evidence that learners at the same level of exposure to an L2 can differ in the extent to which they demonstrate native-like ERP responses (e.g., McLaughlin et al., 2010; Tanner et al., 2012). The beginning learners in these studies demonstrate a possible development from an N400 early in the language-learning process, indicative of lexical processing, to a more native-like syntactic P600 in later stages, at least for features that are similar in the L1 and L2. These studies found that at least one stage of learning revealed differences between learners with regard to whether a P600 or an N400 was present, indicating an as-yet-uninvestigated role for individual differences between learners in ERP studies. Accordingly, the current study begins to examine these differences by investigating correlations between the amplitudes of ERP responses and various measures of individual differences. Critically, measures of both verbal and nonverbal aptitude are included in order to test Bley-Vroman’s claims regarding L2 reliance on domain-general capacities.

The dissertation is organized as follows. Chapter 2 presents in more detail the contrasting theories of L2 acquisition being tested here, with attention to the role of uninterpretable morphosyntactic features. Literature on the impact of both verbal and nonverbal aptitude on adult L2 acquisition is presented in Chapter 3. Chapter 4 describes ERP methodology in more detail and then presents processing patterns that have been observed in native speakers, while ERP studies of L2 learners are presented in Chapter 5. Research questions and predictions are presented in Chapter 6, and Chapter 7 presents the methods employed in the study. Chapter 8 presents behavioral and ERP results for native speakers, followed by behavioral and ERP results for learners in Chapter 9. Chapter 10 presents the relationships of ERP responses to measures of individual differences. Finally, Chapter 11 concludes the study, along with a discussion of the results and future directions for research.

CHAPTER 2:

THE ROLE OF FEATURES IN ADULT L2 ACQUISITION

Theoretical considerations regarding the role that uninterpretable features play in adult L2 acquisition will be considered in this chapter, which begins with a brief description of important concepts related to adult L2 acquisition, followed by a more detailed description of the three primary theories under investigation in the current study.

2.1 The Critical Period and Access to Universal Grammar in Adult L2 Acquisition

Second-language learners who acquired their L2 as children have been found in a number of studies to have an apparent advantage over those who learned as adults (for reviews see Long (1990), Birdsong (1999), and Marinova-Todd et al. (2000)). One theory that attempts to account for the effect of age of acquisition in L2 learning is adapted from Penfield and Roberts' (1959) and Lenneberg's (1967) proposals of a Critical Period (CP) for first language acquisition. In general, the *Critical Period Hypothesis* claims that language acquisition is impaired after a certain age due to maturation of cognitive abilities and/or changes in the neuroanatomy of the brain. Lenneberg's (1967) *Critical Period Hypothesis* was a brain-based theory, claiming that Critical Period effects began at two years of age when lateralization of language to the left hemisphere began to be observed. While not all such theories claim neurobiological causes for the Critical Period, many do include changes in cognitive processing and/or the brain, e.g. development of localized responses (Seliger, 1978), decrease in plasticity (Penfield and Roberts, 1959), maturation of different types of neurons (Diller, 1981), or myelination (Pulvermüller and Schumann, 1994).

While the effects of these processes on language acquisition are as yet unclear, child language acquisition is believed by many to be guided by an innate learning mechanism, or Language Acquisition Device, that facilitates the selection of features in a particular language, along with their parameter settings (often assumed to be binary options), from a universal set of features and parameters, or Universal Grammar (UG), which was first proposed by Chomsky (1965; 1980; 1981). In second language research, a major question is whether or not that mechanism is available to post-Critical-Period learners[2]. This question also provides the framework for distinctive theoretical predictions regarding linguistic features involved in morphosyntactic processes.

In this regard, one of the primary differences between contrasting theories of L2 acquisition, many of which directly address morphosyntactic features, is their approach to Universal Grammar. Most theories based on an assumption of Universal Grammar can be categorized as either restricted-access or full-access theories. Theories of restricted access are of two varieties: some claim that UG is not available at all (no access) past the Critical Period, while others claim that adult L2 learners have partial access to UG, usually through the L1. Examples of partial-access theories include the *No Parameter Setting Hypothesis* (e.g., Clahsen and Muysken, 1986) and the *Failed Functional Features Hypothesis* (Hawkins and Chan, 1997), along with its more recent version, the *Interpretability Hypothesis* (Tsimpli and Mastropavlou, 2007; Tsimpli and Dimitrakopoulou, 2007). On the other hand, the *Fundamental Difference Hypothesis* (Bley-Vroman, 1989, 1990) claims that adult learners have no access to Universal Grammar, but instead rely on their knowledge of the L1 and general problem-solving skills.

---

[2] The question remains as to whether the results of studies on age of L2 acquisition support a distinct Critical Period or a more linear progression of age effects (see Bialystok and Hakuta, 1999; Birdsong, 2005), but age effects continue to be observed in L2 studies, whatever shape the function of age of acquisition takes.

These theories have in common the claims that L1 and L2 acquisition are *qualitatively* different and that the adult learner's mental representation of the L2 will therefore be incomplete.

Accounts based on theories of partial or no access to UG contrast with theories that claim full access to UG for all learners, such as the *Strong Continuity Hypothesis* (Epstein, Flynn, and Martohardjono, 1996), the *Full Transfer/Full Access Hypothesis* (FTFA) (Schwartz and Sprouse, 1994; 1996), and the *Missing Surface Inflection Hypothesis* (Prèvost and White, 2000). Under these theories, native-like grammatical representations can exist even in the absence of consistent accuracy in morpheme usage. The acquisition of native-like representations is theoretically possible for all learners; however, some of these theories – and in particular FTFA – argue that the transfer of L1 features may be a primary source of difficulty in L2 acquisition. Additionally, some accounts, including the *Missing Surface Inflection Hypothesis*, have recently been discussed in terms of the processing constraints experienced by L2 learners. In other words, complete L2 representations may be present in the knowledge of the learner but may be underused due to heavy processing loads.

The next sections discuss the claims of the *Full Transfer/Full Access Hypothesis* (Schwartz and Sprouse, 1994; 1996), the *Interpretability Hypothesis* (Tsimpli and Mastropavlou, 2007; Tsimpli and Dimitrakopoulou, 2007), and the *Fundamental Difference Hypothesis* (Bley-Vroman, 1989, 1990) as representative theories of full, partial, and no access approaches to UG, respectively.

2.2 Full Access: The Full Transfer/Full Access Theory

Claiming a role for both UG and L1-specific representations in adult L2 acquisition, Schwartz and Sprouse (1994, 1996) have proposed the *Full Transfer/Full Access* (FTFA) hypothesis. Under this theory, the initial state of adult L2 learning is considered to be the L1 grammar, but L1 parameter settings can be abandoned in favor of UG-constrained settings as needed, in order to accommodate L2 input. Restructuring of the grammar occurs in cases where an L1-based analysis is not sufficient, allowing learners to handle features that are not present in their L1. It is important to note that restructuring may not always result in native-like L2 representations during development, since any combination of UG-constrained settings can be selected as long as the L2 input is accounted for. Furthermore, since the L2 input may be limited and L1 or other UG-constrained representations can continue to interfere with acquisition of native-like representations of the L2, "convergence on a grammar identical to that of a native speaker is not guaranteed" (White, 2003: 68). However, a number of studies have found that learners converge both qualitatively and quantitatively on native-like behavioral responses as their level of proficiency increases (e.g., White and Genesee, 1996; White et al., 2004; Sagarra and Herschensohn, 2010), and in some cases, even lower-proficiency learners have exhibited native-like patterns in online measures (e.g., Renaud, 2010), providing support for theories claiming Full Access to UG.

As an example, White, Valenzuela, Kozlowska-MacGregor and Leung (2004) take advantage of the nominal agreement properties of Spanish to investigate the acquisition of L2 features. Recall that Spanish determiners and adjectives must agree in number and gender with their head noun. It is also possible in certain contexts in Spanish to refer to an object using only the determiner and adjective, a phenomenon known as N(oun)-drop:

(5)	¿Dónde	puse	*el*	*nuevo*	que	compré?
	where	put_{1sg}	$\text{the}_{masc.sg}$	$\text{new}_{masc.sg}$	that	bought_{1sg}?

"Where did I put the new one that I bought?"

(Adapted from White et al., 2004: 126)

Along with a control group of native Spanish speakers, White et al. (2004) tested adult learners of Spanish in Canada whose first language (L1) was either English or French and who were divided into three proficiency levels based on results of the vocabulary portion of the MLA Cooperative Foreign Language test (Educational Testing Service, Princeton, N.J.) and a cloze section from the Diploma de Español como Lengua Extranjera (DELE) test (Spanish Embassy, Washington D.C.). Due to Full Transfer, it was predicted that at low proficiency, the L1 French group would outperform the L1 English group on gender agreement, since French but not English encodes gender in much the same way as Spanish. It was also predicted that the L1 English group would have less difficulty with number agreement than gender agreement since some determiners in English are marked for number. In line with Full Access, it was predicted that at higher proficiency levels, the L1 English group would perform as well on gender agreement as on number agreement and that differences between the L1 English and L1 French groups would disappear, perhaps converging on native speaker performance.

Tasks included production and comprehension tests. In one production task, the experimenter and participant were given the same cards depicting people, and the participant had to ask questions to figure out which person card the experimenter had chosen. Another production task involved describing what was going on in three different pictures. For the comprehension task, participants were asked to identify the referent of test sentences like the one in (6) below, where the referent is "*la roja*". Pictures of three items were provided for each test sentence, only one of which matched the gender or number of the reference.

(6) Ponlas ahí cerca de *la*$_{fem.sg}$ *roja*$_{fem.sg}$.
"Put them there next to *the red (one)*."

(Adapted from White et al., 2004: 115)

Results showed that, in general, the intermediate and advanced groups of both L1s performed at above 90% accuracy on gender across all production and comprehension tasks. Furthermore, neither the English- nor the French-speaking advanced learners differed from the Spanish native speakers in comprehension or production. White et al. argued that L2 gender features are acquirable due to full access to UG, regardless of the fact that these were adult learners. A problem for Full Transfer emerges here, however, in that there were no significant differences between the English- and French-speaking groups at low proficiency, both of which performed less well on gender than on number. The authors discuss differences between Spanish and French with regard to gender at the lexical level which may have contributed to the results for the French L1 group. They also propose that perhaps their learners were not absolute beginners, a level of proficiency that is difficult to test but would provide more solid evidence for transfer. However, White (2003) points to research indicating the presence of L1 properties in the developing grammar (e.g., Haznedar, 1997), as well as the behavior of different L1 groups with regard to properties of the same L2 (White, 1985, 1986; Yuan, 1998; Slabakova, 2000), as evidence in favor of Full Transfer.

2.3 Partial Access: The Interpretability Hypothesis

One of the most current formulations of the partial-access approach is the *Interpretability Hypothesis* of Tsimpli and Mastropavlou (2007; also Tsimpli and Dimitrakopoulou, 2007). This theory is a refinement of the *Failed Functional Features Hypothesis* (FFFH) (Hawkins and Chan, 1997), which claims that learners have indirect access to UG only through their L1 and so are unable to incorporate features of the L2 that are not instantiated in the L1. As an example, Hawkins and Chan (1997) provide evidence that Chinese-speaking learners of English are unable to acquire the (±*wh*) feature on complementizers[3]. Gender features have been argued to provide the same difficulty for learners whose L1s do not instantiate grammatical gender. Hawkins (1998) investigated gender agreement in English-speaking learners of French. Participants were asked to describe a short animated film. Overall, the learners were relatively accurate with regard to gender agreement, but displayed errors that suggested the overgeneralization of one form or the other, leading to the conclusion that gender was not accurately represented in the learners' L2 grammar. Franceschina (2001) also analyzed production data, but in a case study of Martin, a single high-proficiency L2 Spanish learner. Gender errors were minimal (8.3% on determiners, 7% on adjectives), but significantly higher than number errors (0.5% on determiners, 2% on adjectives). Given that Martin makes no mistakes in gender assignment to nouns, his errors seem indicative of persistent problems with gender features on determiners and adjectives.

Franceschina (2002) provides data from both production and comprehension tasks in L2 Spanish that are put forth as support for FFFH. Subjects for each of the two experiments were native Spanish speakers and two groups of advanced learners, one group whose various native languages include a gender feature and another group whose native English language does not. In the production experiment, participants had to provide the missing pronoun in sentences where either a masculine or feminine accusative-marked pronoun or a neutral dative-marked pronoun should have occurred. An example sentence is provided in (7), where the blank should be filled with the masculine pronoun *los* 'them', referring to the masculine word in italics:

(7) Los dos *enchufes*$_{\text{masc}}$ que compré estaban fallados. ¿Será posible cambiar___ por unos nuevos?
'The two plugs I bought were faulty. Could I change (them) for new ones?'

(Adapted from Franceschina, 2002: 79)

[3] But see White and Juffs (1998), who found evidence that Chinese-speaking learners can acquire this feature.

Participants in the comprehension experiment were presented with sentences containing gender-bearing pronouns or adjectives and had to choose the referents from among three choices where only one of the choices was of the appropriate gender, as in the following practice item:

(8) Los trajo Martín y dijo que son para usted
them$_{masc}$ brought Martín and said that are for you
'Martín brought them and said that they were for you'

a. flores b. joyas c. chocolates
flowers$_{fem}$ jewels$_{fem}$ chocolates$_{masc}$

(Adapted from Franceschina, 2002: 81)

In this experiment, there was a significant difference between the native Spanish speakers and the genderless L1 group, but not the gendered L1 group. There was a marginally significant difference between the two L1 groups. Despite the fact that the learners in the genderless L1 groups in these experiments were fairly accurate[4], Franceschina takes these results to indicate that they have not been able to acquire the gender feature of the L2.

As a further refinement of the FFFH, Tsimpli and Mastropavlou (2007) propose the *Interpretability Hypothesis*, under which it is only uninterpretable features that are constrained by Critical Period effects. Accordingly, those features which are unique to the L2 *and* uninterpretable are expected to be the ones that pose a problem for learners even at high levels of proficiency. Following this line of inquiry, Tsimpli and Dimitrakopoulou (2007) investigated Greek speakers' acceptance of illicit resumptive pronouns after *wh*-extraction in English, as in (9-10). Resumptive pronouns are allowed following *wh*-extraction in Greek, except in the case of *ti* 'what' in object position, which is argued to be a reflex of its lack of (uninterpretable) *phi*-features.

[4] In the production experiment, the genderless L1 group produced only 11 gender errors and demonstrated an overall mean accuracy rate of 85% compared to 89% for native speakers. In the comprehension experiment, the mean score for the genderless L1 group was 12.20/18 points, while the native speakers were only at 14.69/18 points.

(9) ***Grammatical and ungrammatical object extraction***
a. Which student/Who do you think that Jane likes *t*/**him*?
b. Which book/What do you remember that Peter read *t/*it* carefully?

(10) ***Grammatical and ungrammatical subject extraction (±that)***
a. Which politician/Who have you suggested *t/*he/*that-he* should not resign?
b. Which party/What does John think *t*/**it/*that-it* was very boring?

(Adapted from Tsimpli and Dimitrakopoulou, 2007: 227)

In line with the *Interpretability Hypothesis*, results revealed an L1-like acceptance of illicit resumptive pronouns, except in cases where interpretable features like animacy and d-linking provided independent constraints on the resumptive strategy. The researchers propose that interpretable features are available even in the early stages of L2 acquisition and "can improve L2 performance so that it approximates target output" (Tsimpli and Dimitrakopoulou, 2007: 236), but that the uninterpretable features necessary for the rejection of the resumptive strategy had not been acquired by the learners in their study.

Importantly, the *Interpretability Hypothesis* makes a distinction between features which are interpretable in the underlying Logical Form (LF) and those which may be interpretable when spelled out in Phonetic Form (PF) but are not interpretable at LF. Only the former are considered to be acquirable. This distinction is an important one in the investigation of agreement, since gender and number marking on adjectives and determiners in Spanish are realized phonetically and thus are argued to be PF-interpretable, at least when canonical (Carstens, 2000). However, unacquirable features are argued to be those whose role at LF is restricted to syntactic derivations, regardless of whether or not they are realized phonetically (Tsimpli and Mastropavlou, 2007; Tsimpli and Dimitrakopoulou, 2007).

2.4 No Access: The Fundamental Difference Hypothesis

While it is possible to make clear predictions regarding specific roles for uninterpretable features under both the *Interpretability Hypothesis* (Tsimpli and Mastropavlou, 2007; Tsimpli and Dimitrakopoulou, 2007) and the *Full Transfer/Full Access Hypothesis* (Schwartz and Sprouse, 1994, 1996), the *Fundamental Difference Hypothesis* (FDH) (Bley-Vroman, 1989, 1990), which can be classified as a no-access approach, makes more general claims about the success of adult L2 learners. Bley-Vroman (1990) points to such factors as L2 fossilization, the need for negative evidence during L2 development, greater variation in adult L2 learners' success (ultimate attainment), varied L2 learning paths and strategies, and affective aspects of L2 learning as evidence for a fundamental difference between adult L2 acquisition and first-language (L1) acquisition in children. The original formulation of the FDH claimed that post-Critical-Period learners, having no access to UG, succeed in L2 acquisition only to the extent that they are able to rely on general problem-solving skills and their knowledge of the L1. Under this theory, learners with near-native abilities in the L2 are "pathologically" successful learners who exhibit high non-linguistic (domain-general) cognitive abilities that allow them to overcome Critical Period limitations on acquisition (Bley-Vroman, 1990: 7). These cognitive abilities would matter for adult L2 acquisition, as opposed to child L2 acquisition, precisely because a different mechanism would be at work.

Although the cognitive abilities in question were originally posited to be non-specific to language, subsequent studies have taken this to mean that language aptitude plays a significant role in L2 acquisition. DeKeyser (2000) was among the first to test the hypothesis that the lack of sensitivity to L2 grammaticality attested in post-Critical-Period learners (in particular, in Johnson and Newport, 1989) might be modulated by language aptitude. Adapting Johnson and Newport's study, which required L2 learners to make grammaticality judgments (GJ) on sentences involving English morphosyntactic features, he tested Hungarian-speaking learners who had acquired English at an early age versus those who had acquired it as adults. In order to test the prediction of the FDH that individual cognitive abilities might modulate performance, DeKeyser also included a test of verbal analytical ability in the native language of the participants, specifically a Hungarian version of the Words in Sentences portion of the Modern Language Aptitude Test (or the MLAT4). DeKeyser hypothesized that the results would show an overall significant negative correlation between age of acquisition and performance, and that any overlap between the early and late learners' performance would provide evidence for the role of aptitude. Additionally, he predicted that not all structures would be subject to Critical Period effects. In general, these predictions were borne out. A negative correlation between age of acquisition and GJ performance was present (r=-.63, p<.001), but in the late learners this correlation was not significant, in line with the *Critical Period Hypothesis* (but see Bialystok (2002) for arguments that DeKeyser's results indicate a general decline over the lifespan rather than Critical Period effects). Very few late acquirers performed near the level of the early learners, but of those who did so, all but one also had above-average scores on the aptitude test. Additionally, aptitude scores were significantly correlated with overall GJ performance in late learners (r=.33, p<.05) but aptitude was not a factor for the early learners (r=.07, p>.05). Finally, the late learners were only successful on structures that DeKeyser argued to be quite salient in L2 input. DeKeyser concluded that aptitude, i.e., language analytical ability, plays a role in adult L2 acquisition but not in early L1 or L2 acquisition.

In a more recent version of the *Fundamental Difference Hypothesis*, Bley-Vroman (2009) proposes that the fundamental difference between child L1 and adult L2 acquisition is a heavier reliance on patches and shallow processing after the Critical Period, a theory which is heavily influenced by processing accounts in general, and for L2, the *Shallow Structure Hypothesis* of Clahsen and Felser (2006). The reformulated FDH predicts that for L2 properties not present in the L1, learners will only be successful in the case of those properties that can be acquired through the use of such shallow processing, resulting in nonconvergent (non-native-like) processing and/or representations.

The *Shallow Structure Hypothesis* (Clahsen & Felser, 2006; Clahsen et al., 2010) itself argues that adult L2 learners do not build detailed abstract syntactic representations of the L2 during processing, but instead rely on lexical/semantic processing and shallow syntactic representations. Number and gender agreement, therefore, could possibly be constructed when the agreeing elements are in a local relationship within the same phrase, but not when agreement occurs across phrases (as in the examples in (3,4) above). Interestingly, the *Shallow Structure Hypothesis* (SSH) can be contrasted with the previous theories discussed here in that it does not predict a role for the L1 at all. It is important to note that the shallow processing routines used by the L2 parser under the SSH are also available to native speakers and may therefore at least include linguistic processing, as opposed to the FDH (Bley-Vroman, 1990) which claims that domain-general capacities are employed for the L2.

2.5 Summary

At issue here are the different approaches taken by L2 acquisition theories with regard to the availability of Universal Grammar and therefore to uninterpretable features not present in the L1. While each of the no-, partial- and full-access approaches presented here assumes a role for the L1 grammar, each makes a different prediction with regard to features that are unique to the L2. Under the *Interpretability Hypothesis* (Tsimpli and Mastropavlou, 2007; Tsimpli and

Dimitrakopoulou, 2007), features that are both uninterpretable and unique to the L2 are unacquirable. Under a full-access approach like *Full Transfer/Full Access* (Schwartz and Sprouse, 1994, 1996), such L2 features are theoretically acquirable despite the challenges they may present in initial stages of learning, and performance will improve as proficiency increases. For proponents of the *Fundamental Difference Hypothesis* (Bley-Vroman, 1989, 1990, 2009), there is also a role for the L1, but processing strategies and/or individual differences are argued to be highly relevant. Due to the lack of access to UG, unique L2 features may only be acquirable given the right combination of salient input and learner characteristics, particularly with regard to domain-general processing. Additionally, both the FDH and the *Shallow Structures Hypothesis* (Clahsen and Felser, 2006; Clahsen et al., 2010) claim that L2 processing may rely on shallow parsing routines rather than abstract syntactic representations.

CHAPTER 3:

THE ROLE OF INDIVIDUAL DIFFERENCES IN ADULT L2 ACQUISITION

In a discussion of arguments for the *Fundamental Difference Hypothesis* (Bley-Vroman, 1989, 1990) in the previous section, individual differences in cognitive abilities were argued to play a substantial role in the late acquisition of an L2. Indeed, measures of verbal aptitude typically yield predictive validity coefficients in the .40 to .60 range when correlated with L2 proficiency tests and measures of classroom learner success (Carroll, 1981). With regard to the predictions of the FDH, however, two gaps in the literature can be noted. First, the relationship between individual differences like aptitude and acquisition of *specific* structures within the L2 has not been adequately investigated. Second, the original formulation of the FDH has been largely supported by evidence from studies claiming that verbal aptitude can explain apparent exceptions to the *Critical Period Hypothesis*. Bley-Vroman's (1990) original hypothesis, however, gave preference to domain-general problem-solving capacities that may be better tested outside the realm of language. These two issues are considered in the current study, which investigates whether individual learner characteristics related to either verbal or nonverbal aptitude explain a percentage of the variability observed in measures of sensitivity to specific L2 features that differ in their similarity to the L1. Accordingly, this section presents background on both verbal and nonverbal aptitude. Here, verbal aptitude and language aptitude are used synonymously. Nonverbal aptitude is used to refer to domain-general cognitive abilities, which have most often been tested in conjunction with verbal measures in studies of the relationship of intelligence to L2 outcomes. Therefore, nonverbal aptitude will necessarily be linked to the concept of general intelligence.

3.1 Verbal Aptitude

Verbal aptitude is claimed to be the single most effective predictor of L2 success (Carroll, 1981; Skehan, 1989; R. Ellis, 1994; Gardner and MacIntyre, 1992). It is normally-distributed within the population and is believed to be a fixed characteristic of the learner[5] and distinct from but subsumed under a general intelligence factor (Carroll, 1981; Skehan, 1986b, Wesche et al., 1982). Carroll (1965) defined aptitude in terms of phonemic coding ability, associative memory (ability to match L1 to L2 words), grammatical sensitivity, and inductive language learning ability. Carroll and Sapon’s (1959) Modern Languages Aptitude Test was designed to test for these four skill sets and is still used widely today. Skehan (1986a) adds that the ability to handle de-contextualized language is also crucial, but it is not clear whether this skill is directly related to the stable characteristic of aptitude, since it appears to improve with instruction even in adults.

In a resurgence of aptitude research, several recent studies have demonstrated the role of verbal aptitude in successful adult L2 acquisition. Abrahamsson and Hyltenstam (2008) used a grammaticality judgment task to test learners of Swedish whose onset of acquisition began either in childhood or adulthood. All were judged by native speakers of Swedish to be virtually indistinguishable from native speakers.Results showed that 7 out of 11 late (adult) learners and 13 out of 31 early (child) learners fell outside of the native speaker range on the grammaticality judgment task, which involved very difficult Swedish structures.In order to determine the role of verbal aptitude in child versus adult L2 acquisition, these learners were also tested using an adapted version of the Swansea Language Aptitude Test (Meara, Milton, and Lorenzo-Dus,

[5] But see Sternberg (2002) for arguments that language aptitude can fluctuate based on the experiences of the learner and Thompson (2009), who finds higher language aptitude in multilinguals than in bilinguals. She leaves open the question of whether these results stem from greater desire on the part of high-aptitude learners to acquire more languages, or whether the experience of learning more than one language influences aptitude.

2003). Early learners patterned with native speakers with regard to a normal distribution of aptitude scores, and there was a significant correlation between aptitude and GJ scores for early learners (r=.70, p<.001).Aptitude scores for all of the late learners were higher than the average of the early learners even though there was no significant correlation between GJ scores and aptitude in the late learners. The researchers conclude that while aptitude is helpful in child L2 acquisition, it may be absolutely necessary (although not entirely sufficient) for native-like attainment in adult L2 acquisition. Thus Abrahamsson and Hyltenstam (2008) claim support for a Critical Period in general and for DeKeyser's (2000) conclusions regarding the role of aptitude in adult L2 acquisition.

Harley and Hart (1997) also make substantial contributions to the investigation of aptitude and L2 acquisition. Participants in their study were L2 French learners in partial immersion programs in Canadian schools. One group of participants began receiving 50% of their instruction in French in Grade 1, and the other in Grade 7. Testing included a variety of L2 tasks targeting vocabulary, comprehension, and production, along with three aptitude tests designed to assess both memory and analytical skills related to language aptitude. Harley and Hart found significant correlations between proficiency and memory skills for the early immersion group, but even greater significant correlations between proficiency and analytical skills in the late immersion group, suggesting that language analytical ability in particular may play a stronger role in late L2 acquisition. Unfortunately, it is not clear whether the reported differences between memory and analytic ability can be distinguished on the basis of age of acquisition rather than learning environment. Additionally, Harley and Hart report that early learners do not have a higher aptitude resulting from early exposure to a second language. In other words, aptitude seems to be fixed by the time second language acquisition begins, even for relatively early learners, supporting the stability of the construct over time. Clearly, language aptitude is related to the success (or lack of success) of adult L2 acquisition, but whether or not that relationship constitutes evidence for the FDH is not so clear, since the role of nonverbal factors must also be investigated.

Another question that has not been sufficiently tested is whether the correlation of aptitude and learner success in general is observable for specific structures, and whether L1/L2 similarities have any bearing on that correlation. If individual differences like aptitude can overcome the effects of age of acquisition, as is claimed under the *Fundamental Difference Hypothesis* (Bley-Vroman, 1989, 1990), then the effects of aptitude should be observable even, and maybe especially, for features that are unique to the L2. Work in this area is sparse. Recall that DeKeyser (2000) predicted that not all structures would reveal age-related effects. He tested this prediction by analyzing correlations between age of arrival in the U.S. and performance on particular test items. Test structures for which there was a significant correlation ($p<.01$) included auxiliaries, determiners, *wh*-questions, plurals, subcategorization, and adverb placement, which were not compared to the L1 Hungarian. The structures that demonstrated no significant correlation with age of arrival ($p>.05$) were argued to be highly salient either because the faulty use of these structures in production results in salient errors (word order), in particular on lexical verbs in sentence-initial position (subject-verb inversion, do-support in yes-no questions), or with regard to the likelihood that learners would be corrected by native speakers (gender on pronouns). It should be noted that DeKeyser did not test any direct correlation between performance on test items and aptitude test scores. However, overall performance was correlated with aptitude for late learners. Interestingly, with regard to the question of L1/L2 differences, the Hungarian L1 of these participants differs from the L2 English with respect to the structures for which learners were generally successful. Based on DeKeyser's overall results, then, it could be argued that both aptitude and salience (which is not explicitly defined by DeKeyser) can overcome L1/L2 differences even for post-Critical-Period learners. There is reason to believe, then, that aptitude effects could be found in adult learners for at least less salient structures. The current study employs number and gender agreement paradigms that do not appear to be salient in the input. Regardless, any attempt to test for aptitude effects on particular grammatical structures, as is undertaken here, should be considered exploratory in nature. In general, most studies of aptitude effects usually employ broad tests of L2 proficiency, without regard to the effects of aptitude for specific structures, either similar to the L1 or unique to the L2.

3.2 Nonverbal Aptitude

The relationships of verbal and nonverbal aptitude to L2 acquisition have rarely been tested as separable constructs. In many cases where verbal aptitude and intelligence have both been tested for correlations with L2 outcomes, researchers often rely on tests of intelligence that themselves include verbal measures without a factor analysis of the resulting correlations. For example, Robinson (2002) tested the relationship of intelligence, aptitude, and working memory (WM) to the L2 performance of native speakers of Japanese (n=160) under three conditions: implicit and explicit instruction in an artificial grammar, and incidental learning of aspects of Samoan grammar. Among other predictions, he hypothesized that individual differences in intelligence and aptitude would be most influential in explicit learning conditions. Aptitude was measured by Sasaki's (1996) Language Aptitude Battery for the Japanese (LABJ), and intelligence by the Short Form of the Wechsler Adult Intelligence Scale-Revised (WAIS-R) in Japanese, which includes verbal, arithmetic, and block design subtests. While there were differences between aptitude and intelligence when correlated with learning under Implicit and Incidental conditions, both aptitude and intelligence were moderately correlated with a measure of Explicit learning[6]. The use of the WAIS-R test that included verbal tasks leaves room for the possibility that verbal intelligence was the underlying factor in the correlation between L2 performance in the Explicit group and general intelligence. Other investigations of general intelligence and verbal aptitude in L2 studies show the same trend of including verbal measures of intelligence. For example, Sasaki's (1996) construct of a general cognitive factor contributing to L2 proficiency lumps together the three factors of language aptitude, verbal intelligence and reasoning abilities, which were found to be correlated but not identical to an L2 proficiency construct in the best-fit model uncovered by Sasaki's structural equation modeling.

Research into other cognitive factors points to the fact that it might be reasonable to separate nonverbal and verbal factors in the investigation of L2 performance. Consider working memory (WM), which is argued to play an important role in cognition generally and also in language processing (e.g., Miyake and Friedman, 1998; Juffs, 2004; McDonald, 2006). It is not clear that the WM resources recruited for language processing and those involved in more domain-general processes are the same because studies investigating the relationship of WM to successful L2 acquisition have largely used verbal measures of WM, again, not necessarily being concerned with the question of domain-specificity (e.g., McDonald, 2006; Sagarra, 2007; Sagarra and Herschensohn, 2010; Juffs, 2004; also see a description of WM measures in L2 studies presented in Ortega, 2009). For example, Sagarra (2007) used a reading span test to investigate the role of WM in the L2 processing of gender agreement. She found that beginning

[6] Aptitude scores demonstrated no relationship to either Implicit or Incidental learning in immediate post-tests, but a 6-month post-test using guided production revealed a significant correlation with LABJ scores for participants in the Incidental group (r=.56, p<.01). The WAIS-R measure of intelligence demonstrated a negative correlation with results for the Implicit group and no correlation in the Incidental group.

learners of Spanish (n=209) whose L1 English does not incorporate a gender feature were generally insensitive to gender agreement violations on adjectives, but taken together, the reading times and accuracy rates of the high-span individuals demonstrated a developing sensitivity to violations. Additionally, there is evidence that even L1-based measures of working memory may not completely capture individual differences affecting L2 processing. Service et al. (2002) reported that learners in their study had lower WM capacity in their L2 than in their L1, but that L2 but not L1 WM capacity correlated with L2 comprehension.

In order to test the relationship between verbal aptitude and intelligence, Wesche et al. (1982) conducted a factor analysis of the scores of 793 participants on the Modern Language Aptitude Test (MLAT; Carroll and Sapon, 1959) and the Primary Mental Abilities Test (PMA; Thurstone and Thurstone, 1965), a test of cognitive abilities that includes the following subtests: Verbal Meaning, Number Facility, Reasoning, and Spatial Relations. Unsurprisingly, they found evidence for a hierarchical model in which specific language-related skills are subsumed under a more general ability. They conclude that while verbal aptitude and general intelligence are not completely distinct in that they rely on a general ability for abstract reasoning using symbols, verbal aptitude measures tap abilities that are specific to language learning and distinct from the cognitive abilities assessed by the PMA. They do not, however, test the relationship of verbal aptitude and intelligence to L2 outcomes. The only study where subtests of the MLAT and the PMA were examined in a factor analysis along with L2 outcomes was a study of 96 teen learners by Gardner and Lambert (1965), who examined 14 measures of French achievement, the MLAT, and an earlier version of the PMA including an additional subtest of Word Fluency (Thurstone and Thurstone, 1941). Like Wesche et al. (1982), they had also found high degrees of correlation between the MLAT and the PMA. However, in their study the MLAT subtests and the PMA subtests collapsed into separate factors. Indeed, four of the PMA measures loaded onto a factor that Gardner and Lambert identified as intelligence, and the researchers concluded that the four measures themselves "share little in common with measures of French achievement or language aptitude" (1965: 198-199).

Only one study of the relationship between proficiency and intelligence can be found that involves a nonverbal measure of intelligence. Flahive (1980) tested 20 learners of English from seven different language backgrounds to determine whether intelligence or proficiency correlated more precisely with scores on three reading comprehension tests: a multiple choice test, a paraphrase recognition test, and a cloze test. The measure of intelligence used in the study was the Raven's Standard Progressive Matrices (Raven, 1965), which was chosen for its nonverbal format due to the varied language backgrounds of the participants. Proficiency was measured by scores on the Test of English as a Foreign Language (TOEFL). The RAVEN and the TOEFL measures were intercorrelated (r=.61, p<.01). With regard to the reading comprehension tasks, the Raven was found to correlate most strongly with the multiple choice test (r=.84, p<.01), but also with the other two tests (paraphrase: r=.68, p<.01; cloze: r=.61, p<.01). The opposite pattern was true of the TOEFL even though it was also correlated with all three measures (multiple choice: r=.59, p<.01; paraphrase: r=.84, p<.01; cloze: r=.75, p<.01). Flahive concludes that reading comprehension tests, especially the multiple choice test used in his study, may test more than just reading comprehension. The main point here, though, is that a nonverbal measure of intelligence was found to be correlated with general measures of L2 performance. It is unclear whether nonverbal intelligence would correlate with less complex measures of L2 proficiency. Another factor that may limit the interpretation of these results is that the learners in Flahive's study were not normally-distributed in terms of intelligence since their mean score on the Raven placed them at the 75th percentile. It is unknown whether the RAVEN would correlate with reading comprehension measures in a more normalized group of learners.

As pointed out by Skehan (1998), case studies of exceptional post-puberty learners may also provide evidence that intelligence is not a key factor in successful language learning. Obler (1989) and Novoa et al. (1988) present a case study of a learner who had rapidly acquired five foreign languages in informal or mixed formal/informal environments and was judged to be near-native by native speakers. Schneiderman and Desmarais (1988a, b) reported on two learners who were at native-like proficiency in three foreign languages, as judged by native speakers and grammaticality judgment tests. All of these learners demonstrated slightly above-average but not exceptional intelligence quotients (IQ). Additionally, Smith and Tsimpli (1995) studied a man who was gifted in his knowledge of ten languages, but was mentally retarded due to brain damage and did not score well on nonverbal measures in standard IQ tests.

In sum, there is no consensus on whether or not domain-general capacities are key to successful L2 acquisition. The only study to clearly demonstrate a relationship between L2 performance and a nonverbal measure of intelligence is Flahive (1980), whose participants were clearly of above average intelligence to begin with. This does not appear to be the case with learners in a variety of case studies cited in Skehan (1998). The sophisticated analysis of Wesche et al. (1982) found that a general intelligence factor subsumes more language-specific aptitude measures, but those aptitude measures may correlate with L2 achievement without a corresponding relationship between L2 achievement and measures of intelligence (Gardner and Lambert, 1965). In other research on cognitive abilities, whether broadly defined (e.g., intelligence) or specific (e.g., working memory), the construct of domain-general capacities has not always been so clearly distinguished from domain-specific or verbal capacities. The question remains, then, as to whether there is independent evidence that general cognitive factors are correlated with L2 performance *outside the realm of verbal measures*. A secondary question to be explored in the current study is whether L1/L2 differences modulate the role of domain-general and domain-specific cognitive factors in L2 performance on particular structures.

3.3 Summary

Individual differences appear to play a role in L2 acquisition. More specifically, it is possible that adult learners with strong verbal aptitude can acquire a high level of proficiency in the L2, although such results are not guaranteed. Bley-Vroman's (1990) proposal regarding domain-general capacities should predict that nonverbal aptitude would be a strong predictor of success in L2 acquisition, but this prediction has not been adequately tested due to the difficulty of teasing apart verbal and nonverbal aptitude in the results of previous studies. Furthermore, the results of case studies and other research into the relationship of intelligence and aptitude suggest that domain-general capacities may not play as strong a role in language learning as domain-specific aptitude. It is also unclear whether either verbal or nonverbal aptitude differentially impacts the acquisition of different structures in the L2.

CHAPTER 4:

EVENT-RELATED POTENTIALS IN NATIVE LANGUAGE PROCESSING

The extent to which learners exhibit native-like acquisition of agreement features has largely been investigated using behavioral measures (e.g., White and Genesee, 1996; White et al., 2004; McCarthy, 2008; Franceschina, 2002; Montrul et al., 2008), but an emerging area of research in second-language acquisition includes the study of event-related potentials (ERPs) obtained by electroencephalography (EEG) in response to syntactic and semantic anomalies. EEG records electrical signals at the scalp that reflect the activity of groups of neurons. The precise location of the neural generators of these signals cannot be straightforwardly determined by the use of EEG alone, in part since the path of electrical signals generated in the brain will be determined by the conductivity of the anatomical structures surrounding their source. However, EEG provides very high temporal resolution, on the order of milliseconds, making it an excellent tool for investigating the temporal dynamics of language processing.

Event-related potentials (ERPs) are essentially a measure of voltage changes that are time-locked to the presentation of a stimulus. Generally, ERP data from multiple tokens of the same type of stimulus must be averaged in order to reduce noise in the electrical signal. Differences between ERPs for different conditions can then be compared on a number of dimensions, including latency, amplitude, and topography. The latency of an ERP component measures the timing of voltage changes and therefore provides a temporal measure of processing associated with the stimulus. Amplitude measures the intensity of the voltage change and therefore may index the amount of resources allocated to processing specific properties of the stimulus. Topography refers to the location of electrodes on the scalp where the voltage changes were recorded. Differences in the topography of ERP components may at least indicate differences in neural generators. In this chapter, a summary of research findings on ERP responses to linguistic stimuli in native speakers is presented, along with specific results of studies of agreement processing.

4.1 Lexical/Semantic Processing

As mentioned in the introduction, ERP responses involved in lexical/semantic processes can be distinguished from those that index syntactic processing. The primary component involved in lexical/semantic processing is the N400, a negative-going waveform in comparison with baseline conditions, occurring between 200 and 600ms post-stimulus and typically captured by central and parietal electrodes, sometimes stronger in the right hemisphere (e.g., Kutas and Hillyard, 1980; Osterhout and Nicol, 1999). The amplitude of the N400 response to a word is modulated by its context. Kutas and Hillyard (1980) were the first to report a negative deflection in waveforms during this time window for semantically anomalous words. For example, a greater N400 amplitude was observed for the final word in (11b) as compared to (11a):

(11) a. He spread the warm bread with *butter*.
b. He spread the warm bread with *socks*.

(Adapted from Kutas and Hillyard, 1980: 203)

Although words in sentential contexts have been found to exhibit longer-lasting N400 deflections (Van Petten, 1993), the N400 amplitude is also modulated by context in word lists. Specifically, a priming effect is observed wherein the N400 amplitude decreases for a word presented following a semantically-related word (e.g., Rugg, 1985; Holcomb and Neville, 1990). The N400 is also sensitive to lexical properties like word frequency (Van Petten and Kutas, 1990) and phonotactic probablility (Kutas and Federmeier, 2000), leading Kutas and Federmeier (2000) to argue that it indexes the cost of lexical access rather than the cost of integrating the stimulus into its context. This view of the N400 is also supported by Lau et al. (2008), based on evidence from imaging studies regarding the location of neural generators of the N400.

4.2 Early Syntactic Processing

Syntactic processing in native speakers may occur in two phases. First, some studies report a Left Anterior Negativity (LAN), a negative-going waveform occurring between 300 and 500 ms post-stimulus. This ERP component, when it is found, typically has a distribution in the left anterior electrodes, distinguishing it from the N400. The LAN has been observed in response to a variety of morphosyntactic anomalies, including – but not limited to – agreement violations (e.g., Osterhout & Mobley, 1995; Coulson et al., 1998; Gunter et al., 2000; Barber & Carreiras, 2005). Due to its earliness in the response waveform, at least one account of sentence processing claims that the LAN reflects first-pass, automatic processing (Friederici, 2002).

Although the LAN has long been held as proof of native-like morphosyntactic processing and is critical to some theories regarding adult L2 acquisition, including the *Shallow Structure Hypothesis* discussed earlier (see Clahsen and Felser, 2006), both Osterhout et al. (2004) and Alemán Bañon et al. (2012) point out that a substantial number of studies have not reported the LAN for agreement violations in native speakers of various languages. Molinaro et al. (2011) point out that in many studies, the absence of LAN effects may be related to the lack of overt morphophonological cues on the trigger of agreement that would set up expectations for agreement on the target. They also suggest that averaging using the left mastoid subtracts out enough signal in the left electrodes to prevent detection of any LAN effects.

Osterhout et al. (2004) offer an alternative explanation for the *presence* of the LAN in some studies. They report a Japanese study of agreement that at first glance resulted in a LAN/P600 biphasic response. However, upon further analysis, the LAN was a product of grand averaging across participants who exhibited an N400-like response and those who exhibited a P600 response, indicating that one of two distinct processes may have been engaged in each participant. Indeed, a number of theories of L1 processing concur that both structural and lexical/semantic information may be used by native speakers to process relative clauses and other ambiguous structures, and recent studies have indicated that the selection of parsing strategies may be related to individual differences in working memory (for discussion, see Gibson & Pearlmutter, 1998; Felser et al., 2003). Indeed, a relatively unresolved question in L1 ERP studies is whether dual syntactic and semantic processes are serial (e.g., Friederici, 2002) or parallel (e.g. Hagoort, 2005), with the faster process potentially cancelling out the other. The answer to this debate will be critical to understanding the nature of dual processing in the L1, but it at least seems possible that native speakers might have two options available to them for processing at least local agreement dependencies – both a shallow approach that involves direct association and results in an N400 response, and a deeper syntactic operation producing a P600. If Osterhout et al. are correct in their analysis of the LAN as a by-product of averaging two different processes, then the presence of the LAN may be heavily influenced by properties of the stimuli or by individual differences between participants.

It should also be noted that anterior negativities with a similar onset have also been observed for syntactic processes that involve the integration of displaced elements over a distance, such as those required for filler-gap dependencies (Kluender & Kutas, 1993; Fiebach et al., 2002) or object relative clauses (King & Kutas, 1995). This ERP component is sensitive to working memory load (Vos et al., 2001) and has been argued to reflect working memory costs. It is typically bilateral and of longer duration than the LAN. Additionally, an early left anterior negativity (ELAN) emerges in some studies between 150-200ms post-stimulus that is believed to reflect phrase-structure building (Friederici, 2002).

4.3 Late Syntactic Processing

The P600, also known as the Syntactic Positive Shift, is a late ERP component elicited during syntactic processing that is reported much more reliably than the LAN. The P600 is a positive-going deflection of the waveform for anomalous or ungrammatical stimuli starting around 500ms post-stimulus, typically peaking around 600ms and with a duration of about 400ms. The P600 has been observed in response to a variety of morphosyntactic violations (e.g., Hagoort et al., 1993; Osterhout and Mobley, 1995; Barber and Carreiras, 2005; Nevins et al., 2007), subcategorization violations (e.g., Osterhout and Holcomb, 1992), constructions involving syntactic dependencies (e.g., Gouvea et al., 2010), and garden path sentences where stimuli are unexpected but grammatical (Kaan and Swaab, 2003; Gouvea et al., 2010), leading researchers to conclude that this component indexes not just repair but reanalysis of sentence structure.

The P600 has been argued to be composed of two stages (Barber and Carreiras, 2005; Carreiras et al., 2004; Hagoort and Brown, 2000; Kaan and Swaab, 2003). The early P600 (generally from 500-750ms) exhibits a broad scalp distribution including anterior electrodes and is believed to index integration of a new constituent with the previous material (Kaan et al., 2000; Phillips et al., 2005), while the later stage (from roughly 750-1000ms) is typically captured only in the posterior electrodes and may reflect reanalysis and repair processes (Hagoort et al., 1993; Hagoort and Brown, 2000; Osterhout & Mobley, 1995; Gunter et al., 2000; Barber & Carreiras, 2005). More frontal P600s have also been reported that are sensitive to syntactic complexity (Friederici et al., 2002; Gouvea et al., 2010; Kaan & Swaab, 2003). For example, Kaan and Swaab (2003) found a frontal complexity effect for the P600 in response to verbs when two possible subjects (versus only one) were present. Gouvea et al. (2010) also found a more frontal P600 for the integration of moved wh-elements in grammatical sentences (12a) than for length-matched control sentences (12b).

(12) a. The patient met the doctor to whom the nurse with the white dress showed the chart during the meeting.
b. The patient met the doctor while the nurse with the white dress showed the chart during the meeting.

(Adapted from Gouvea et al., 2009: 157)

Coulson et al. (1998) have proposed that the P600 does not reflect syntactic repair or reanalysis, but is one of a family of responses related to the P300, a component that is sensitive to the probability and saliency of the stimulus itself and is not domain-specific to language. In their study, stimuli included grammatical and ungrammatical sentences involving subject-verb agreement or case marking. The researchers manipulated the percentages of grammatical and ungrammatical items in two different blocks of stimuli in an effort to compare the sensitivity of P600 responses to grammaticality versus probability. The P600 was found to be sensitive to probability whether the larger percentage of stimuli in a block was grammatical or ungrammatical, although the probability effect was greater for ungrammatical stimuli. No significant differences in P600 distribution were reported for the grammaticality effects versus the improbability effects. Critically, the study did not include any conditions that tested probability apart from grammaticality effects. Osterhout (1999) responds with arguments based on an earlier study (Osterhout et al., 1996) that employed both a violation of subject-verb agreement and an unexpected non-syntactic anomaly (words in all capital letters). Osterhout and colleagues found that the anomaly elicited three positivities related to the P300 (an early positivity in frontal and central electrodes, a larger-amplitude positivity in the centroparietal electrodes, and a late positivity with longer duration) while ungrammatical subject-verb agreement elicited a canonical P600 effect. Osterhout argues that while the P600 in Coulson et al. (1998) may be sensitive to probability, there are other non-P300 components that are also sensitive to probability, and that when effects of probability and grammaticality are separated out, they have different topographies, amplitudes, and latencies.

4.4 Native Processing of Agreement Violations

A biphasic response including the LAN and the P600 is often reported for native speakers in response to violations of number or gender agreement. For example, Osterhout and Mobley (1995) reported a LAN/P600 combination in a study of verbal and pronominal agreement in English. Participants were asked to make grammaticality judgments on stimuli that included sentences containing number agreement violations on verbs as in (13), where the noun in the subject NP triggers agreement and the verb (in italics) contributes either a grammatical continuation or a violation of agreement:

(13) The elected officials$_{\mathrm{pl}}$ *hope*$_{pl}$/**hopes*$_{sg}$ to succeed.

(Adapted from Osterhout and Mobley, 1995: 742)

Compared to grammatical sentences, agreement violations exhibited both a greater negative amplitude in the 300-500ms time window in anterior and temporal electrodes in the left hemisphere (the LAN), and a widely-distributed positivity (P600) beginning around 500ms.

Barber and Carreiras (2005) found the same general ERP response patterns to gender and number agreement violations in Spanish when target words were presented in sentences. In the second of two experiments with Spanish monolinguals, their stimuli target agreement on determiners and adjectives, both of which must agree with the noun they modify. Violations were realized either on the noun in the subject NP (mismatched to the incorrect preceding article, as in the ungrammatical versions of (14)) or on an adjective in the predicate occurring in the middle of the sentence (15). In the following examples, the noun triggering agreement is underlined, while the critical word at which the violation is realized will be in italics.

(14) El/*La/*Los *piano* estaba viejo y desafinado.
The$_{masc.sg/*fem.sg/*masc.pl}$ piano$_{masc.sg}$ was old and off-key.
'The piano was old and off-key.'

(15) El faro es *alto/*alta/*altos* y luminoso.
The lighthouse$_{masc.sg}$ is high$_{masc.sg/*fem.sg/*masc.pl}$ and bright.
'The lighthouse is high and bright.'

(Adapted from Barber and Carreiras, 2005: 151)

In comparison to grammatical sentences, number/gender violations in both contexts produced a LAN effect in anterior left-hemisphere electrodes between 300 and 450ms. Additional effects for grammaticality were observed in the P600 time window, beginning around 500ms after presentation of the critical word and lasting throughout at least the subsequent 400-ms window analyzed by the researchers. The P600 effects were greater on electrodes in the posterior regions of the right hemisphere. Barber and Carreiras also report differences in the amplitude of the P600 between within-phrase agreement (14) and across-phrase agreement (15), and between number and gender agreement in the late stage of the P600, from 700-900ms, suggesting that these types of agreement are not processed similarly.

Interestingly, in their first experiment where agreement violations were presented on word pairs, Barber and Carreiras did not find a P600 effect for agreement violations targeting either article-noun agreement (*el* / *piano* 'the / piano') or noun-adjective agreement (*faro* / *alto* 'lighthouse / high'), but instead found more negative-going waveforms following violations. According to the researchers, this apparent N400 effect may be related to direct association of the two agreeing words without regard to syntactic processing. It should be noted that the N400 effect was present for noun-adjective agreement regardless of whether the nouns and adjectives displayed markers that were canonical or noncanonical (*poesía* / *triste* ‘poetry / sad’), indicating that lexical associations might have been at play rather than morphology.

In a study designed to investigate the effects of structural distance on native-speaker sensitivity to morphosytactic features, Alemán Bañón, Fiorentino, and Gabriele (2012) tested number and gender agreement on adjectives in Spanish both within (16) and across (17) phrases. In response to Barber and Carreiras (2005), stimuli were controlled for both linear distance and realization of the violation on the target of agreement (in italics), which followed the trigger (underlined).

(16) ***Experiment 1: Within-phrase agreement on adjectives***
El banco es un edificio muy _seguro/*-os/*-a_ y el juzgado también.
the bank is a building$_{masc.sg}$ very safe$_{masc.sg/*masc.pl/*fem.sg}$ and the courthouse too.
‘The bank is a very safe building and so is the courthouse.’

(Adapted from Alemán Bañón et al., 2012: 53)

(17) ***Experiment 2: Across-phrase agreement on adjectives***
El cuento es _anónimo/*-os/*-a_ y el manuscrito también.
the story$_{masc.sg}$ is anonymous$_{masc.sg/*masc.pl/*fem.sg}$ and the manuscript also
‘The story is anonymous and so is the manuscript.’

(Adapted from Alemán Bañón et al., 2012: 53)

No LAN was present in response to agreement violations in either experiment. In both experiments, however, number and gender agreement violations yielded canonical P600 responses between 400 and 900ms, with a peak around 600ms. Violations within phrases yielded more positive-going waveforms than violations across phrases, but no differences in P600 amplitude were observed between number and gender violations, in contrast with Barber and Carreiras (2005).

4.5 Summary

To summarize results of previous ERP studies for native speakers, then, at least two distinct types of processing have been identified. Lexical/semantic processing elicits an N400 response while syntactic processing consistently elicits a P600 response, sometimes preceded by a LAN component. At least in the case of agreement, the P600 is always present and may in fact be very similar for different types of agreement. However, the LAN may be absent potentially due to either properties of the stimuli or to EEG data processing. Alternatively, it has been proposed that the LAN may be present in many studies as the result of grand averaging across participants who variably exhibit either P600 or N400 responses in cases where the properties of the stimuli allow either syntactic processing or direct associations between trigger and target.

CHAPTER 5:

EVENT-RELATED POTENTIALS IN L2 PROCESSING

Due to a focus on the Critical Period as discussed previously, early L2 ERP studies largely investigated age of acquisition as a factor in L2 processing, but other factors that have recently emerged as critical to the development of native-like processing include proficiency and L1/L2 similarities. A brief review of the literature regarding these factors is presented in this chapter, along with recent developments that impact the current study.

5.1 Age of Acquisition and Proficiency as Factors in L2 Processing

Early L2 ERP studies that were designed to test for effects of age of acquisition (e.g., Weber-Fox and Neville, 1996; Hahne and Friederici, 2001) found no differences between native speakers and learners in terms of the types of ERP responses found for lexical/semantic processing. Syntactic processing, on the other hand, was argued to be deficient due to the absence of an early component often seen in native speakers in the same studies in response to phrase-structure violations, the Early Left Anterior Negativity (ELAN). For example, Weber-Fox and Neville (1996) tested groups of L1 Chinese participants that varied according to age of acquisition: acquisition at 1-3 years old, 4-6 years, 7-10 years, 11-13 years, and 16 or older. Native English speakers in their study elicited the ELAN, LAN, and P600 in response to phrase structure violations such as those in (18), but most L2 participants demonstrated only a LAN and P600. The group who had acquired English at 16 years of age or older did not exhibit a P600 at all, and their LAN was not native-like.

(18) a. The scientist criticized Max's *proof* of the theorem.
b. The scientist criticized Max's **of* proof the theorem.

(Adapted from Weber-Fox and Neville, 1996: 233)

A problem with most early studies like Weber-Fox and Neville (1996) is that effects of age of acquisition could not be tested separately from proficiency. Steinhauer et al. (2009) offer an overview that demonstrates the role of proficiency in L2 processing, specifically with regard to morphosyntactic development. These researchers cite studies of novice learners that find either no grammaticality effects at all (e.g., Hahne, 2001), or N400 lexical/semantic responses, which are purportedly not native-like in response to syntactic errors (but see discussion of Osterhout et al., 2004 above). However, they note that higher-proficiency learners can demonstrate native-like P600 responses (e.g., Bowden et al., 2007; Steinhauer et al., 2006; Rossi et al., 2006). The view that L2 processing becomes more native-like with proficiency is also supported by research on monolingual children, whose P600 responses are late and reduced before the age of 7 (Hahne, 2004), and also by research on monolingual adults, where lower-proficiency native speakers exhibit P600 effects that are reduced in comparison with those of high-proficiency speakers in terms of both distribution and amplitude (Pakulak and Neville, 2004). The same high-proficiency monolinguals were then taught Jabberwocky, a nonsense language based on English grammar rules – they exhibited reduced P600s to Jabberwocky at low proficiency levels.

A number of studies have shown that high-proficiency adult L2 learners can acquire native-like processing (e.g., Bowden et al., 2007; Steinhauer et al., 2006; Rossi et al., 2006). For example, Rossi et al. (2006) found sensitivity to grammaticality in both behavioral and ERP results for their L1 Italian/L2 German and L1 German/L2 Italian participants, who were grouped by proficiency in each L2. Proficiency levels were determined by self-ratings, L2 background, and translation tests. An example of German stimuli for their verbal agreement condition can be seen in (19):

(19) Der Junge im Kindergarten *singt/*singst* ein Lied.
The boy in-the kindergarten sings/*sing a song.
"The boy in the kindergarten sings a song."

(Adapted from Rossi et al., 2006: 2034)

Participants were also tested on word category violations and a combination of category violations/agreement violations. High-proficiency learners in both languages combined demonstrated an accuracy rate of 87.9% on the grammaticality judgment task, as well as both LAN and P600 effects for grammaticality. There were only amplitude differences in these responses when compared with native speakers tested on the same materials in a previous experiment (Rossi et al. 2005). Low-proficiency participants, on the other hand, did not demonstrate a LAN effect, and their P600 response was delayed in comparison with native speakers even though their accuracy rate at 81.4% was not much lower than the high-proficiency learners.

Unfortunately, results of studies looking at proficiency may not take into account similarities and differences between the L1 of the learners and the L2 being targeted. The fact that high-proficiency learners of both languages in Rossi et al. (2006, cited above) were able to acquire native-like processing of verb agreement errors may not be surprising since the features involved in verb agreement are similar in both languages. This was not the case in a study that looked at Japanese learners of English (Ojima et al., 2005) who were tested on subject-verb agreement as in (20).

(20) Some scientists *find*/**finds* solutions by chance.

(Adapted from Ojima et al., 2005: 1226)

While the English control group in this study exhibited a LAN/P600 pattern, high-proficiency learners exhibited a LAN but not a P600. Low-proficiency learners, who had acquired English at around the same age as the high-proficiency learners, did not exhibit either a LAN or a P600.

Ojima et al. conclude that it is not so much the proficiency of the learners that results in the lack of a P600, but rather the lack of subject-verb agreement in the L1 Japanese. It should be noted that Ojima et al. were able to find native-like N400 effects for semantically anomalous words in sentences, further suggesting that semantic processing is similar in the L1 and L2.

5.2 The Role of Features in L2 Processing

Recall that both the *Interpretability Hypothesis* (Tsimpli and Mastropavlou, 2007) and *Full Transfer/Full Access* (Schwartz and Sprouse, 1994, 1996) assume a role for L1 features since they take the L1 to be the initial state of L2 acquisition, an assumption shared by Bley-Vroman (1990). These theories, however, make different predictions regarding unique L2 features. Another open question is whether differences can be predicted even for features that are shared between the L1 and L2 if they are instantiated differently in the L2. Recent ERP studies to be described below seem to suggest that this is the case, but as Kotz (2009) points out, only a few ERP studies so far have directly tested L1 transfer. Three of these studies will be described below.

First, Sabourin (2003) compared groups whose native languages vary in similarity to the L2. Sabourin tested gender agreement in four contexts in L2 Dutch. Learners composed three groups, depending on the similarity of their L1 to Dutch: L1 German (surface similarity with Dutch), Romance (abstract similarity but surface differences), and English (no grammatical gender agreement). Gender agreement between determiners and nouns (21), as well as between nouns and relative pronouns (22), were tested both in definite (a) and indefinite (b) phrases. Here, nouns triggering agreement are underlined, while the critical word at which the violation would be realized (sometimes the noun itself) are in italics:

(21) a. *Noun Phrase, Definite*

Het/*De kleine *kind* probeerde voor het eerst te lopen.
The$_{neut/*com}$ small child$_{neut}$ tried for the first to walk.
'The small child tried to walk for the first time.'

b. *Noun Phrase, Indefinite*

Hij loopt op een gekke/*gek *manier*.
He walks in a funny$_{com/*neut}$ way$_{com}$.
'He walks in a funny way.'

(22) a. *Relative Pronouns, Definite*

De baron *die/*dat* in het kasteel woonde, is overleden.
The baron$_{com}$ that$_{com/*neut}$ in that castle lived, has died.
'The baron that lived in that castle has died.'

b. *Relative Pronouns, Indefinite*

Een lichaam *dat/*die* slap is, heeft training nodig.
A body$_{neut}$ that$_{neut/*com}$ flabby is, has training necessary.
'A body that is flabby needs training.'

(Adapted from Sabourin, 2003: 77)

Effects of grammaticality in native Dutch speakers were reported as a biphasic LAN/P600 response in all conditions. Only the L1 German group showed native-like grammaticality effects in conditions that demonstrated the most surface similarity to their L1 with regard to gender assignment, consisting of late or more limited P600 effects, but no LAN. The L1 Romance (largely native speakers of French) and L1 English groups demonstrated behavior at chance, and neither group produced P600 responses to gender agreement violations in any condition. Sabourin concludes that L1 transfer occurs only where the L1 and L2 demonstrate surface similarities. Learners whose L1 does not instantiate gender agreement, as well as those whose L1 gender system is not congruous with that of the L2, do not seem to have acquired the necessary L2 properties. It is important to note that the relative proficiency of Sabourin's L1 German, Romance, and English groups is unknown, since proficiency was measured using a different experimental condition targeting subject-verb agreement, which occurs in all three languages.

Another issue is that the L1 Romance and L1 English groups performed slightly more poorly on a gender assignment task.

In a different approach to L1 transfer, Tokowicz and MacWhinney (2005) compared L2 responses to three different kinds of syntactic violations: missing auxiliary verbs (23), violations of number agreement between nouns and determiners (24), and gender agreement violations on indefinite articles (25):

(23) Su abuela *cocina/*cocinando* muy bien.
His grandmother cooks/*cooking very well.

(24) Los/*El *niños* están jugando.
The$_{\text{masc.pl/*masc.sg}}$ boys$_{\text{masc.pl}}$ are playing.

(25) Ellos fueron a una/*un *fiesta*.
They went to a$_{\text{fem.sg/*masc.sg}}$ party$_{\text{fem.sg}}$.

(Adapted from Tokowicz and MacWhinney, 2005: 178)

These conditions were categorized, respectively, as similar to the L1, different but present in the L1, and unique to the L2. Participants were native speakers of English in their first four semesters of L2 Spanish study. Tokowicz and MacWhinney predicted difficulty for learners due to competition only in the condition where the rule for the number feature differed in the L1 and L2, but P600 responses in the conditions that were similar to the L1 and unique to the L2. As predicted, grammaticality effects for the P600 were observed in response to missing auxiliaries and gender agreement violations – the similar and unique conditions, but not to number agreement, in contrast with behavioral results. In the behavioral task, participants scored at around 70% accuracy in the missing auxiliary and number agreement conditions but only at chance for gender agreement violations. The researchers conclude that the ERP results show greater sensitivity to grammaticality than can be observed in explicit measures like acceptability judgments, and that native-like processing is possible where L1-based processing does not interfere in the L2. However, it could be argued that the low accuracy rates may have been due to the very rapid speed of presentation, 300 ms/word, which is faster than in many other studies that use RSVP (Rapid Serial Visual Presentation), and would have been especially difficult for these low-proficiency learners. Additionally, no comparison to native-speaker P600 responses is available for these stimuli since Tokowicz and MacWhinney did not test an L1 Spanish control group. It is interesting, though, that Tokowicz and MacWhinney find a P600 in response to gender agreement violations even in an L1 English group whose native language does not instantiate gender agreement, a finding that is not consistent with Sabourin (2003).

In one of very few studies designed to directly test whether incongruous L1/L2 systems impact processing of similar features in the L2, Foucart and Frenck-Mestre (2010) target gender agreement but in German-speaking learners of French. The study compared stimuli across two experiments. The stimuli in the first experiment tested gender agreement between determiners and nouns (26), which is a similar rule in both languages:

(26) Hier la/*le <u>*chaise*</u> était dans le salon.
Yesterday the$_{fem/*masc}$ chair$_{fem}$ was in the living.room.
'Yesterday the chair was in the living room.'

Learners demonstrated native-like processing in this experiment, producing a P600 to agreement violations. In Experiment 2, the stimuli targeted agreement between nouns and post-nominal plural adjectives (27), a context in which agreement is not established in the L1 German:

(27) En été, les <u>chaises</u> *blanches/*blancs* sont dans le jardin.
In summer, the chairs$_{fem}$ white$_{fem/*masc}$ are in the garden.
'In summer, the white chairs are in the garden.'

(Adapted from Foucart and Frenck-Mestre, 2010: 12)

In contrast with learner responses to violations in Experiment 1, the group of L2 participants in Experiment 2 demonstrates only a very early negativity. Foucart and Frenck-Mestre interpret the lack of a P600 in terms of differences between featural instantiations in the L1 and L2, but in a third experiment, they also tested agreement on adjectives in a pre-nominal position where gender would be instantiated in German. Results were similar – no grammaticality effects were present for learners.

Overall, these studies indicate a need for further testing of morphosyntactic features that are different in the L1 and L2. With regard to unique L2 features, Sabourin's L1 English group demonstrated no grammaticality effects in response to the unique gender feature in L2 Dutch, but Tokowicz and MacWhinney's English-speaking learners of Spanish exhibited a P600 in response to gender violations. Additionally, Sabourin's results for the L1 Romance group whose native languages instantiate gender agreement differently than in L2 Dutch, along with Tokowicz and MacWhinney's (2005) results for Spanish number agreement in L1 English participants and Foucart and Frenck-Mestre's (2010) results for French gender agreement in German-speaking learners, suggest that even features that are shared between the L1 and L2 can be problematic if they are not instantiated in the same way in the L2.

One challenge in comparing stimuli across experiments and sometimes even within experiments has to do with differences in the design of the stimuli itself. First, the position of the critical word in the sentence sometimes varies from one experimental condition to another, a factor that Barber and Carreiras (2005) found to impact the robustness of P600 responses to agreement violations in native speakers. In Tokowicz and MacWhinney (2005), for example, number agreement in sentence-initial clauses was compared to gender agreement in sentence-final position, where wrap-up effects are more likely to have impacted the ERP response. Another factor, particularly in studies of agreement, has to do with whether the violation is realized on triggers of agreement or on targets. Sabourin's study includes at least two sets of stimuli where the violation is not apparent until reaching the triggering noun itself, while in other experimental conditions, the violation is realized on relative pronouns that follow the trigger. Foucart and Frenck-Mestre directly compare two conditions involving gender agreement, but in one of them the violation is not apparent until the trigger noun itself is presented, while in the other condition, the trigger precedes the target of agreement where the violation is realized (the adjective). The current study will address these issues by presenting three different types of agreement that are realized only on targets, with all critical targets appearing in the same sentential position.

5.3 Individual Differences in L2 ERP Responses

The current study also attempts to address individual differences between learners that could be obscured in ERP studies and are generally not reported. Observation of the effects of individual differences on ERP responses within groups of L2 learners is made difficult by a number of factors. First, the possibility should be noted that individual differences might not be observable on all response components (LAN, N400, P600). Secondly, EEG methodology is not optimal for observing individual differences in localization of responses. However, the primary issue in investigating individual differences in ERP responses has to do with methodology: due to the signal-to-noise ratio present in EEG data, the primary means of observing and reporting results is to average ERP data across not only multiple tokens of a given type of stimulus, but also across groups of participants. In some cases, participants may vary widely with regard to factors other than individual differences, as in the case of the L2 participants in Tokowicz and MacWhinney (2005), who were recruited from across four different semesters of Spanish. More homogeneous group selection can also inhibit observation of individual differences. If advanced language learners are “self-selecting”, then high proficiency groups will be more uniform than average when it comes to motivation, aptitude, and the like, making it difficult to observe the variability necessary for the analysis of individual differences. Low proficiency groups will often demonstrate behavioral results at chance, making conclusions tenuous in regard to any ERP responses observed.

However, it may be possible to investigate individual differences related to variability between learners, if learner development in regard to brain responses to L2 stimuli is characterized by qualitative or quantitative changes over time. Critically, it is just such evidence of developmental changes that is observed in two pioneering studies of ERP responses in participants at early levels of L2 study (Tanner et al., 2012; McLaughlin et al., 2010). The beginning learners in these studies demonstrate that development from an N400 early in the language-learning process to a more native-like P600 is possible, at least for features that are similar in the L1 and L2. Furthermore, these studies reveal differences between learners in early stages of acquisition with regard to whether a P600 or an N400 was present, demonstrating that a) averaging across participants in ERP studies can mask P600 development in a subset of learners and b) individual differences between learners may play a role in second-language development that has yet to be investigated.

McLaughlin et al. (2010; also Osterhout et al., 2006) conducted a longitudinal ERP study of 14 novice L2 French learners whose native language was English. Testing took place after 4 weeks, 16 weeks, and 26 weeks of instruction. In addition to a semantic condition, participants were asked to perform acceptability judgments on sentences involving Subject-Verb agreement and number agreement between Determiners and Nouns, as seen in (28) and (29), respectively:

(28) Tu *adores/∗adorez* le français.
You$_{2sg}$ love$_{2sg/*2pl}$ (the) French.

(29) Tu manges des *hamburgers/*hamburger* pour diner.
You eat (the)$_{pl}$ hamburgers$_{pl}$/*hamburger$_{sg}$ for dinner.

(Adapted from Osterhout et al., 2006: 217)

Results differ between the two morphosyntactic conditions, indicating development from non-native to native-like processing for subject-verb agreement but not for determiner-noun agreement, a context in which English does not instantiate agreement. Interestingly, for the subject-verb stimuli, these beginning learners exhibited a negativity to ungrammatical stimuli in the first session – a response peaking between 300 and 500ms after presentation of the violation and resembling the N400 response elicited for semantically anomalous words in the semantic condition. Based on previous research with pseudowords and beginning learners, these researchers interpret the N400 response here as indexing learners' sensitivity to probabilistic dependencies between morphemes, i.e., "the probability of occurrence of particular pronoun-verb ending combinations" (McLaughlin et al., 2010: 141). By the third session, however, the N400 effect had been replaced by a native-like P600, indicating the grammaticalization of a generalized rule regarding morphosyntactic dependencies.

Critically for the purposes of the present study, the qualitative change from an N400 to a P600 response over time that was uncovered by McLaughlin et al. (2010) indicates an area where individual differences may make an impact. The most interesting of their results for subject-verb agreement were found in Session 2, where the learners who had previously exhibited an N400 suddenly demonstrated no grammaticality effects. However, further analysis discovered evidence not only of grammaticality effects, but also of differences between sets of learners that had been obscured by averaging across the whole group, similar to the results discussed above for native speakers of Japanese in Osterhout et al. (2004). One set exhibited a clear N400 effect and the other a P600 effect, indicating qualitative individual differences with regard to how learners were processing the ungrammaticality.

Another study has found similar differences between groups of learners. In a cross-sectional study, Tanner et al. (2012) recruited L2 learners in their first and third years of German, with a control group of native German speakers. The first-year learners in this study were tested after roughly the same amount of instruction as learners in Session 3 of McLaughlin et al. (2010). Stimuli were similar to the French stimuli in McLaughlin et al. (2010), including grammatical and ungrammatical versions of sentences containing subject-verb agreement (30) as well as number agreement between determiners and nouns (31):

(30) <u>Ich</u> *wohne/*wohnt* in Berlin.
I live/*lives in Berlin.

(31) Viele/*ein <u>*Bücher*</u> liegen auf dem Tisch.
Many/*a books lie on the table.

(Adapted from Tanner et al., 2012: 4)

Results are reported for subject-verb agreement, where the third-year students demonstrated a native-like P600, although with a broader distribution than that of the native speakers. First-year learners exhibited a biphasic response, including an N400-like component followed by a trend toward a P600. Critically, further investigation of individual patterns of response revealed differences between first-year learners, with one group exhibiting an N400 response, while another group demonstrated a developing P600.

Taken together, McLaughlin et al. (2010) and Tanner et al. (2012) indicate that L2 acquisition of at least some L2 morphosyntactic features may involve development from an N400 that indexes lexical processing to a more syntactic P600. Interestingly, the N400/P600 split in both studies did not seem to be a dichotomous phenomenon, in that the effect sizes of the N400 and P600 were inversely related. In other words, across learners, as the N400 effect decreased, the P600 effect increased, creating a continuum along which individual learners could be identified and which may allow testing of the correlation of a measure like the P600 effect size to individual factors like aptitude. The current study will address the relationship of individual differences like aptitude to specific brain responses to linguistic stimuli, a question that has not been previously investigated.

5.4 Summary

While adult L2 learners engaged in lexical/semantic processing generally demonstrate native-like N400 responses, the results of studies looking at syntactic processing are varied. Just as in behavioral L2 studies, various factors have been proposed to impact L2 processing, including age of acquisition, proficiency, and L1/L2 similarities. The results of early studies testing for effects of age of acquisition are often confounded with proficiency, and indeed a compelling case for proficiency effects is made in recent reviews of L2 ERP studies (Kotz, 2009; Steinhauer et al., 2009). However, in many cases where higher-proficiency learners have demonstrated native-like responses, particularly the P600 for syntactic processing, the structures tested were also instantiated in the L1. Results of studies that test for effects of L1 transfer are inconsistent, with some studies claiming a P600 only for structures similar to the L1 (Foucart and Frenck-Mestre, 2010; McLaughlin et al., 2010) or only in L1 groups where features are similar to the L2 (Sabourin, 2003), while others find a P600 for structures unique to the L2 (Tokowicz and MacWhinney, 2005). Finally, little is known about the neural correlates of individual differences between learners. Not much has changed since Ioup et al., in their well-known study of two near-native L2 learners with high capacity for languages, concluded that aptitude was a factor in their success but that "how the talented brain acquires language in comparison with the normal brain remains a mystery" (1994:93). However, recent studies indicate that learners, at least at lower levels of proficiency, vary in the types of responses they demonstrate to syntactic violations (McLaughlin et al., 2010; Tanner et al., 2012). It is quite possible that ERP responses during syntactic processing may be sensitive not only to proficiency, L1 similarity, and the like, but also to factors like aptitude that vary from learner to learner.

CHAPTER 6:

RESEARCH DESIGN AND PREDICTIONS

The current study investigates adult learners' sensitivity to violations of number and gender agreement in Spanish, as measured by grammaticality judgments and event-related potentials. Three different types of agreement will be examined that differ in their similarity to the L1 English of the learners. The theories of second-language acquisition investigated here include the *Full Transfer/Full Access Hypothesis* (Schwartz and Sprouse, 1994, 1996), the *Interpretability Hypothesis* (Tsimpli and Mastropavlou, 2007; Tsimpli and Dimitrakopoulou, 2007), and the *Fundamental Difference Hypothesis* (Bley-Vroman, 1989, 1990, 2009). All three theories assume a role for the L1 at least in initial stages of L2 acquisition, but differ in their predictions regarding unique and uninterpretable L2 features. Additionally, regardless of theoretical approach, individual differences may be expected to play a role in adult L2 acquisition, but the nature of those individual differences, as well as their relationship to brain functions, have yet to be determined. The aim of the current study is to investigate the impact of both of these factors – L1/L2 differences with regard to uninterpretable features, and individual differences in verbal and nonverbal aptitude - on the processing of specific L2 structures involving morphosyntactic agreement, as measured by both grammaticality judgments and online measurements of event-related potentials.

6.1 Learner Sensitivity to Agreement Violations during L2 Processing

The role of L1/L2 differences with regard to uninterpretable features is addressed here by investigating three types of agreement in Spanish. As described in the introduction, Spanish requires both number and gender agreement between nouns and their modifiers (including adjectives), as well as number agreement between subject nouns and verbs. Critically, English instantiates number agreement on verbs, but not on adjectives, and gender agreement triggered by nouns is not present at all in English. By examining number agreement on both verbs and adjectives, as well as gender agreement on adjectives, the current study investigates differences in the processing of features that are present (number) and absent (gender) in the learners' L1, as well as differences in the processing of a shared feature (number) in contexts where agreement is instantiated in the L1 (on verbs) and where it is not (on adjectives). The experimental stimuli can be seen in Table 1 below and will be described further in the next chapter.

Table 1. Sample sentences for the grammatical and ungrammatical conditions in the three types of agreement tested (similar to L1, instantiated differently in the L1, and unique to the L2).

Similar to L1: Number Agreement on Verbs (Subject-Verb)	
Grammatical (S-V)	La <u>viajera</u> agotada *descansa* en el hotel. The$_{3sg}$ traveler$_{3sg}$ exhausted$_{3sg}$ rest$_{3sg}$ in the hotel. 'The exhausted traveler rests in the hotel.'
Ungrammatical (*S-V)	La <u>viajera</u> agotada **descansan* en el hotel. The$_{3sg}$ traveler$_{3sg}$ exhausted$_{3sg}$ rest$_{3pl}$ in the hotel.
Different between L1/L2: Number Agreement on Adjectives (Noun-Adj)	
Grammatical (N-Adj)	La <u>isla</u> es *rocosa* y la península también. The$_{fem.sg}$ island$_{fem.sg}$ is$_{3sg}$ rocky$_{fem.sg}$ and the peninsula too. 'The island is rocky and the peninsula too.'
Ungrammatical (*N-Adj NUM)	La <u>isla</u> es **rocosas* y la península también. The$_{fem.sg}$ island$_{fem.sg}$ is$_{3sg}$ rocky$_{fem.pl}$ and the peninsula too.
Unique to L2: Gender Agreement on Adjectives	
Ungrammatical (*N-Adj GEN)	La <u>isla</u> es **rocoso* y la península también. The$_{fem.sg}$ island$_{fem.sg}$ is$_{3sg}$ rocky$_{masc.sg}$ and the peninsula too.

6.1.1 Number Agreement

The first research question has to do with differences in the instantiation of shared features in the L1 and L2. This question will be addressed by comparing the conditions involving number agreement on verbs (similar) and adjectives (different).

Research Question 1:

> For number agreement, do low-proficiency learners demonstrate equally native-like sensitivity to violations realized on verbs and on adjectives, despite differences in instantiation of number agreement on these categories in the L1?

Two of the theories that are under investigation here, the *Interpretability Hypothesis* (Tsimpli and Mastropavlou, 2007) and *Full Transfer/Full Access* (Schwartz and Sprouse, 1994, 1996), both predict that learners will be sensitive to number agreement violations in L2 Spanish since the number feature is present in their L1. If these theories are on the right track, then, learners should pattern with native speakers with regard to number agreement on both verbs and adjectives, despite the fact that number agreement does not occur on adjectival predicates in English. It should be noted that the *Fundamental Difference Hypothesis* (Bley-Vroman, 1989, 1990, 2009) does not make predictions regarding specific features, although knowledge of the L1 is predicted to have an impact on L2 acquisition. Therefore, both native speakers and learners are predicted to exhibit behavioral and ERP sensitivity to violations of number agreement on both verbs and adjectives, with no significant differences between the two agreement types. ERP results are predicted to include at least P600 effects. However, the results of the ERP studies reviewed above (Tokowicz and MacWhinney, 2005; Sabourin, 2003; Foucart and Frenck-Mestre, 2010; McLaughlin et al., 2010) seem to suggest that no P600 will be found for number agreement on adjectives, where the L1 and the L2 differ. Even more extreme is the prediction of the *Shallow Structure Hypothesis* (Clahsen and Felser, 2006) with regard to the particular stimuli used in the current study: the P600 should be present on neither verbs nor adjectives since the agreement dependencies presented are not local, that is, agreement between noun and verb and noun and adjective must be computed across phrases since the verb or adjective is not found within the Noun Phrase itself. Finally, it should be noted that no LAN is predicted for learners due to their low proficiency level, and it is unclear whether native speakers will demonstrate a LAN (Alemán Bañón et al., 2010; Molinaro et al., 2011; Osterhout et al., 2004).

6.1.2 Gender Agreement

Research Question 2 addresses whether or not learners can exhibit native-like processing of a feature not present in the L1.

Research Question 2:

Can low-proficiency learners demonstrate developing native-like sensitivity to violations of the gender feature that is unique to the L2?

The theories under investigation make different predictions regarding gender agreement violations. The *Interpretability Hypothesis* (Tsimpli and Mastropavlou, 2007) predicts that learners whose L1 English does not instantiate gender agreement will not be able to acquire the uninterpretable gender feature on Spanish adjectives. Thus, learners are expected to diverge from native speakers in that learners will not demonstrate native-like P600 responses to gender agreement violations. It should be noted that if learners attempt to establish associations between nouns and adjectives based on orthographic regularities in the input (e.g., *vestido blanco* or *muchachos...estudian*), they may demonstrate sensitivity in ERP responses resulting in an N400 effect, as argued for initial stages of learning by McLaughlin et al. (2010) and similar to what was found by Barber and Carreiras (2005) for native speakers in word-pair conditions where associations were tested without reference to a syntactic context. The *Fundamental Difference Hypothesis* (Bley-Vroman, 1989, 1990, 2009) predicts that only isolated cases of native-like responses to gender agreement should be observed, and that this success would be attributable to exceptional ability.

On the other hand, the *Full Transfer/Full Access* (Schwartz and Sprouse, 1994, 1996) allows that learners could demonstrate sensitivity to gender agreement. Under this hypothesis, gender should ultimately be acquirable, and sensitivity to gender agreement should be largely a function of proficiency. In this case, given the low proficiency of the learners being tested, non-native ERP responses are possible. If a P600 effect is observed for noun-adjective gender agreement at low proficiency, it could be expected to reflect lower amplitude differences than in the number agreement conditions.

6.2 The Relationship of Individual Differences to Sensitivity to Agreement Violations

Predictions are given below for research questions 3 and 4, which investigate the relationship of verbal and nonverbal aptitude to L2 sensitivity to agreement violations. While neither FTFA (Schwartz and Sprouse, 1994, 1996) nor the *Interpretability Hypothesis* (Tsimpli and Mastropavlou, 2007) make specific predictions regarding correlations between measures of sensitivity to agreement violations and either verbal or nonverbal aptitude, such correlations might at least be expected for verbal aptitude, given these hypotheses' claims regarding the role of the L1. Under the *Fundamental Difference Hypothesis* (Bley-Vroman, 1989, 1990, 2009), both verbal and nonverbal aptitude might be expected to correlate positively with L2 outcomes.

6.2.1 Verbal Aptitude

First, predictions regarding verbal aptitude will be considered, as per Research Question 3 below:

Research Question 3:

Does verbal aptitude modulate sensitivity to agreement violations and/or to number and gender agreement differentially?

The results of previous aptitude studies (DeKeyser, 2000; Abrahamsson and Hyltenstam, 2008; Harley and Hart, 1997) show clear effects for verbal aptitude when it comes to L2 performance, at least for adult learners. The overall responses of late learners in these studies demonstrated a significant correlation with aptitude scores, particularly with regard to language analytic ability (Harley and Hart, 1997; DeKeyser, 2000). Given these relationships, it is predicted here that in late learners, verbal aptitude (at least as it measures language analytic ability) might be correlated with measures of sensitivity to agreement violations. Since Bley-Vroman argues that individual differences may overcome age effects that inhibit acquisition, especially of properties unique to the L2, the relationship of aptitude to sensitivity to violations will also be explored separately for number and gender. The analysis of correlations between verbal aptitude and characteristics of ERP components (such as amplitude) is an important contribution of this research, both in terms of individual differences and in ERP research into L2 acquisition.

6.2.2 Nonverbal Aptitude

While several cognitive constructs arguably related to domain-general processing have been tapped in second-language research, the majority of these constructs are operationalized in the literature using at least partially domain-specific (verbal) measures. The primary question raised here has been whether or not a general measure of nonverbal intelligence or aptitude outside of the verbal domain can provide evidence for the role of domain-general cognitive factors in L2 processing.

Research Question 4:

Does nonverbal aptitude modulate sensitivity to agreement violations and/or to number and gender agreement differentially?

The original formulation of the *Fundamental Difference Hypothesis* (Bley-Vroman, 1989, 1990) suggests that this should be the case, leading to the prediction that nonverbal aptitude scores will be correlated with measures of sensitivity to agreement violations. However, while full-access theories like FTFA (Schwartz and Sprouse, 1994, 1996) do not make predictions in this regard, it could be argued that domain-specific as opposed to domain-general capacities would be more likely to be correlated with L2 performance if L2 acquisition is guided by UG; in this case, verbal aptitude may correlate more strongly than nonverbal aptitude with measures of sensitivity to agreement features. Just as with verbal aptitude, the relationship of nonverbal aptitude to sensitivity to violations will be explored separately for number and gender.

CHAPTER 7:

METHODS AND MATERIALS

This chapter presents the methods of the present study, including participants, materials, and procedures. Methods of data analysis are also described.

7.1 Participants

A total of 50 people participated in the study: 32 English-speaking learners of Spanish and a control group of 18 native Spanish speakers. Due to the simultaneous collection of EEG data, all of the participants who were recruited for the study were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971), had normal or corrected-to-normal vision, and reported no neurological impairment. All gave informed consent and were paid $5 per half-hour for their participation in the study. Data from six native speakers and four L2 participants were excluded due to excessive artifacts in the EEG data. Three L2 participants were excluded due to substantial previous exposure to Spanish or to a Romance language other than Spanish which also instantiates gender and number agreement. An additional L2 participant indicated that she had mistakenly rejected half of the stimuli, thinking that the structure used in the wrap-up material at the end of the sentence was not a grammatical structure in Spanish. Since that material was not the focus of investigation here, data from this participant were not included in the analysis. Therefore, the final number of participants included 24 learners and 12 native speakers.

7.1.1 English-Speaking Learners of Spanish

Participants in the learner group were students at the University of Kansas who were recruited from fourth-semester Spanish classes by means of a posting on the class website.

Fourth-semester Spanish classes at the University of Kansas are taught using an online curriculum (the *Acceso* Project, http://www2.ku.edu/~spanish/acceso/) designed by the department. This curriculum uses authentic language materials, with explicit grammar instruction and a focus on communication for functional tasks. A language background survey confirmed that all learners spoke English as their native language. The survey also collected information on the first age of exposure to Spanish, amount of time spent in a Spanish-speaking country, and any previous exposure they may have had to Romance languages other than Spanish. It was concluded that all learners included in the final analysis (n=24) could be considered late learners of Spanish[7]. A summary of the findings of the survey is given in Table 2. Additionally, all participants that were included in the analysis tested in the low proficiency range on a test composed of the vocabulary section of the MLA Cooperative Foreign Language test (Educational Testing Service, Princeton, N.J.) and a cloze section from the Diploma de Español como Lengua Extranjera (DELE) test (Spanish Embassy, Washington D.C.), to be described below.

Table 2. Profile of L2 group characteristics, in years.

	Age at testing (n=24)	**Years of Spanish study[a] (n=24)**	**Age of exposure to Spanish (n=22[b])**
Mean	20.04	4.69	14.32
Standard Deviation	2.22	2.01	2.73
Range	18-29	1-8	9-20

[a] This response includes classes taken in junior high and high school.
[b] Some participants did not give a numeric response.

[7] Statistical tests were performed in order to ensure that there were no significant effects due to the point in the semester when individual learners were tested. An analysis of correlations between scaled test dates and proficiency scores, as well as measures of sensitivity in the grammaticality judgment task, showed a significant correlation for test dates only with proficiency scores, and it was negative, r=-.544, *p*=.006. Even when excluding one outlier who scored low on the proficiency test very late in the semester, the correlation was still significant, r=-.461, *p*=.027. The later in the semester that these students were tested, the worse they performed on the proficiency test, which could reflect a propensity of better students to volunteer for a study right away; however, no significant correlation of test date was found for any of the grammaticality judgment measures.

7.1.2 Native Speakers of Spanish

A control group of twelve native Spanish speakers were also included in the analysis. These participants were recruited in one of three ways: by word of mouth, by visiting the Spanish Round Table discussions at the University of Kansas, and by emailing native Spanish speakers who had previously registered themselves in a database of possible participants for research studies. A background questionnaire confirmed that all of these participants were native speakers of Spanish from either Costa Rica or a South American country.For all native speakers, the average age at testing, length of residence in the United States, and age of exposure to English are provided in Table 3.

Table 3. Characteristics of native speakers, in years.

	Age at testing (n=12)	Length of residence in the U.S. (n=12)	Age of exposure to English (n=11[a])
Mean	26.83	3.35	12.27
Standard Deviation	7.85	3.98	6.34
Range	18-41	0.5-14.5	3-21

[a] One participant did not give a numeric response.

7.2 Materials

7.2.1 Stimuli

A grammaticality judgment task was administered to all participants during EEG recording as described under Procedures below. A total of 240 test sentences were constructed, including 120 for each of two triplet sets of stimuli. The first set to be described here were adapted from the stimuli in Experiment 2 of the Alemán Bañón et al. (2012) study of morphosyntactic agreement in native speakers of Spanish, described in Chapter 4. The grammatical sentence in each triplet contained an adjectival predicate agreeing in number and gender with the subject of the sentence (32a), which was always singular. The ungrammatical versions of these sentences varied the morphological suffixes on the adjectives to provide a mismatch in either gender (32b) or number (32c) between the subject noun (underlined here) and the adjective in the predicate (in bold).

(32) a. ***Grammatical Number and Gender (N-Adj)***

La	isla	es	*rocosa*	y	la	península	también.
$The_{fem.sg}$	$island_{fem.sg}$	is_{3sg}	$rocky_{fem.sg}$	and	the	peninsula	too.

'The island is rocky and the peninsula too.'

b. ***Ungrammatical Number (*N-Adj NUM)***

La	isla	es	**rocosas*	y	la	península	también.
$The_{fem.sg}$	$island_{fem.sg}$	is_{3sg}	$rocky_{fem.pl}$	and	the	peninsula	too.

c. ***Ungrammatical Gender (*N-Adj GEN)***

La	isla	es	**rocoso*	y	la	península	también.
$The_{fem.sg}$	$island_{fem.sg}$	is_{3sg}	$rocky_{masc.sg}$	and	the	peninsula	too.

All sentences were presented in the present tense. With regard to gender, approximately half of the sentence subjects were feminine and the rest were masculine. A total of 60 adjectives were used, each appearing in two different triplets. It should also be noted that none of the sentences contained violations of both number and gender agreement. The triplet sets for these conditions involving Noun-Adjective agreement can be found in Appendix 1.

The second set of stimuli consisted of both grammatical and ungrammatical sentences involving Subject-Verb agreement. Sample stimuli are given in (33) below, with target words in bold. In the grammatical sentences (33a), the verb (in bold) agrees with the subject NP in number. The ungrammatical sentences (as in (33b)) contained a violation of number agreement on the verb that was created by providing a plural suffix. A filler (33c) was formed by replacing the lexical verb in these sentences with a copula; all fillers were grammatical in order to balance the number of grammatical and ungrammatical stimuli over the whole experiment. The full set of stimuli for the Subject-Verb conditions can be found in Appendix 2.

(33) a. ***Grammatical Subject-Verb agreement (S-V)***

La	viajera	agotada	*descansa*	en	el	hotel.
The_{3sg}	$traveler_{3sg}$	$exhausted_{3sg}$	$rest_{3sg}$	in	the	hotel.

'The exhausted traveler rests in the hotel.'

b. ***Ungrammatical Subject-Verb (*S-V)***

La	viajera	agotada	**descansan*	en	el	hotel.
The_{3sg}	$traveler_{3sg}$	$exhausted_{3sg}$	$rest_{3pl}$	in	the	hotel.

c. ***Grammatical Filler***

La	viajera	agotada	está	en	el	hotel.
The_{3sg}	$traveler_{3sg}$	$exhausted_{3sg}$	is_{3sg}	in	the	hotel.

'The exhausted traveler is in the hotel.'

Critically, all three types of agreement tested here (N-Adj Number, N-Adj Gender, and S-V) must be calculated across a phrase boundary, making comparisons possible across conditions. By including an adjective in the subject NP for the Subject-Verb stimuli (as in (33) above), the stimuli were also matched across experiments both in terms of the linear distance between the subject and target word (one intervening word) and in the quantity of number cues preceding the target word (three cues). It was not possible, however, to equalize the number of gender cues preceding the target word, since verbs in Spanish are not marked for gender. Across both experiments, critical words (subjects, adjectives, and verbs) were followed by additional material in order to avoid interference from end-of-sentence ERP effects (Hagoort, 2003). The use of a short functional word immediately following the target word (either a preposition or a conjunction) was meant to inhibit as much interference as possible in the late ERP responses to the target word. Finally, only nouns displaying the canonical gender markings *–o* and *-a* (e.g., *espejo*$_{masc}$/*espada*$_{fem}$) and the simple number marking [-s] (e.g., *vestidos*) were used in the subject NPs, and all nouns in the gender agreement conditions were inanimate so as to avoid bias toward natural gender.

In an attempt to eliminate any effects of unfamiliar vocabulary, efforts were made to ensure that the nouns, adjectives, and verbs used across all conditions were understandable by learners, either because they appeared in textbooks or because they are cognates with English. Because the verbs and adjectives appeared in the critical region where violations occurred, they were included in a computerized vocabulary recognition task completed by learners after the experimental task was done.

The 240 sentence sets (120 Noun-Adjective, 120 Subject-Verb and fillers) were counter-balanced across three lists in a Latin Square design such that participants read 40 tokens of each condition and never saw more than one version of each target item. Target sentences were mixed with fillers and presented in random order. Presentation of grammatical and ungrammatical sentences was therefore also randomized across each list. Each list was viewed by exactly four native speakers and eight learners whose data have been included for analysis. To ensure that there had been no advantage for participants receiving any one list over the others, a one-way ANOVA was conducted following the experiment to compare accuracy rates for each list. No significant differences were found, $F(2,33)=.367$, $p=.696$[8].

7.2.2 Gender Assignment Task

All participants also completed an offline task in order to ensure that they were able to correctly assign the gender of each of the 60 nouns that appeared in the subject NP in the stimuli targeting N-Adj Gender agreement. In this task, each noun was presented briefly in random order, followed by a prompt to which participants had to respond by mouse click, identifying the appropriate definite article to use before each noun - “El” for masculine nouns, and “La” for feminine nouns.

[8] Test results for list differences by condition were also non-significant.

7.2.3 Vocabulary Recognition Task

In addition to the two tests described above, which were administered to both the native speaker controls and the learners, the learners were also required to take the remainder of the tests to be described in this section. A vocabulary recognition task was conducted in order to ensure that learners were familiar with the critical words used in the grammaticality judgment task. The task included all of the adjectives from the Noun-Adjective conditions (as in (32) above), as well as all of the verbs from the Subject-Verb conditions (as in (33)), for a total of 120 items. All items were presented in random order. Learners responded by mouse click to choose the correct alternative between two possible English meanings for each Spanish word.

7.2.4 Verbal Aptitude

7.2.4.1 Modern Language Aptitude Test

The measure of verbal aptitude chosen for this study was the short form of the Modern Language Aptitude Test (MLAT) (Carroll and Sapon, 1959), which consists of Parts 3, 4, and 5 of the longer version of the test. The MLAT3 (Spelling Clues) is a multiple-choice test that asks participants to select appropriate synonyms of words that are spelled as they are pronounced rather than by conventional orthography (e.g., an appropriate synonym for *ritn* might be 'printed', since *ritn* can represent the pronunciation of *written*). This test is designed to assess phonetic coding ability and memory for vocabulary. There are 50 items in this section, but only 5 minutes are allowed. The MLAT4 (Words in Sentences) is a test of grammatical sensitivity or a broader language analytic ability (Skehan, 1998), including sensitivity to grammatical roles and the ability to make analogies at a grammatical level. Participants must select from among several underlined words in each test sentence the one choice that functions in the same way as the underlined word in an example sentence. For example, if the subject is underlined in the example sentence, the participant's correct response would also indicate the subject of the test sentence out of the multiple choices of underlined words. Participants are allowed 20 minutes to complete 45 items. The MLAT5 (Paired Associates) requires the rapid learning of a list of 24 vocabulary words in an adapted language, along with their associated English meanings. Following a 2-minute period of vocabulary memorization and a subsequent 2-minute practice period, participants have 4 minutes to choose the correct meaning from multiple choices given for each of the 24 items. The MLAT5 indexes associative memory and requires the storage and retrieval of a large amount of material in a short period of time.

7.2.4.2 LLAMA Aptitude Tests

As a follow-up to the MLAT testing described above, a few participants returned to take a subset of the LLAMA aptitude tests (Meara, 2005). Roughly based on the MLAT4 (Words in Sentences) and MLAT5 (Paired Associates), the LLAMA_F and LLAMA_B are based on pictures rather than English. The LLAMA_F is designed to test grammatical inferencing using pictures with matching sentences. The LLAMA_B requires rapid vocabulary learning using nonsensical pictures, after which participants are tested on their ability to match each new word to its picture. The LLAMA Language Aptitude Tests were developed at the University of Wales Swansea in an effort to provide an aptitude battery that did not require L1 input and was therefore suitable for learners from any L1.

7.2.5 Nonverbal Aptitude: Raven's Advanced Progressive Matrices

In order to contrast the contributions of skills that are domain-specific to language with those that are more domain-general, a test of nonverbal aptitude was also conducted. The Raven Advanced Progressive Matrices (RAVEN) (Raven, 1965) is a multiple-choice test of nonverbal intelligence and reasoning skills, including the ability to decompose complex problems, search for rules and manage those rules in working memory (Carpenter, Just, and Shell, 1990). In a multidimensional scaling analysis by Snow, Kyllonen, and Marshalek (1984), where a wide variety of domain-general and domain-specific tests were placed in concentric circles representing the closeness of correlations among the tests, the RAVEN occupied a central position, demonstrating that it is an optimal test for measuring domain-general reasoning. Each of the 12 practice items in Set I of the RAVEN and the 36 test items in Set II consists of a visual pattern with a piece missing, followed by an array of choices of patterns to fill the missing area in the test pattern. The test set can be administered as timed (usually for 40 minutes) or untimed; here it was administered in a fairly short amount of time due to time constraints and the possibility of participant fatigue during the testing session. The time allowed for the practice set was five minutes, and twenty minutes were allowed for the test set[9].

7.2.6 Spanish Proficiency Test

A short, written Spanish proficiency test was administered to learners that included the vocabulary section of the MLA Cooperative Foreign Language test (Educational Testing Service, Princeton, N.J.) and a cloze section from the Diploma de Español como Lengua Extranjera (DELE) test (Spanish Embassy, Washington D.C.). This test was chosen because it had previously been used in several other studies of L2 learners and heritage speakers (Alemán Bañón et al., 2012; White et al., 2004; McCarthy, 2008; Montrul and Slabakova, 2003), allowing the comparison of learner groups across studies. The test includes a total of 50 multiple choice items. A score below 30 is considered to indicate low proficiency, intermediate proficiency ranges from 30-39 points, and advanced proficiency is reflected by scores at 40 points or higher.

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

In both sections of the test, participants had to choose the word that correctly fills the blank in a sentence or passage from among the choices given. The items in this test involve knowledge of specific vocabulary, use of functional words, and/or specific grammatical knowledge. The test was untimed, but all participants finished it in about 30 minutes.

7.3 Procedures

7.3.1 Learners

All of the learners involved in the study attended at least two experimental sessions lasting a total of approximately five hours. The first session was administered in the Second Language Acquisition Laboratory at the University of Kansas. During this session, learners provided informed consent for the entire study and completed a language background survey. The MLAT, RAVEN, and proficiency tests were then administered as described below, with breaks between each test. Before leaving the test session, learners scheduled their second session with the researcher. Participants were paid for their participation at the end of each session.

Given the intense nature of this testing session, the order of the three tests was varied between participants in order to eliminate the effects of participant fatigue on any single set of test scores. Half of the learners started their testing session with the Spanish proficiency test, the other half ended with it. In order to avoid any confound introduced by the order of testing languages, the Spanish test was not given between the other two tests. The order of the MLAT and RAVEN was also randomized, resulting in four different test orders, as presented in Table 4.

A one-way analysis of variance (ANOVA) was conducted after the testing to explore possible differences in test outcomes based on test order. No significant effects were found for the MLAT Total score, $F(3,20)=.279$, $p=.840$, for MLAT3, $F(3,20)=2.147$, $p=.126$, for MLAT4, $F(3,20)=2.350$, $p=.103$, for MLAT5, $F(3,20)=.210$, $p=.888$. Neither were there any effects for the

Table 4. Test order between groups of L2 participants.

		Group 1 (n=6)	Group 2 (n=6)	Group 3 (n=6)	Group 4 (n=6)
Test Order	**1**	RAVEN	MLAT	Proficiency	Proficiency
	2	MLAT	RAVEN	RAVEN	MLAT
	3	Proficiency	Proficiency	MLAT	RAVEN

RAVEN, $F(3,20)=1.375$, $p=.279$, or the proficiency test, $F(3,20)=2.146$, $p=.126$. Accordingly, no statistical adjustments were made for test order in subsequent analyses.

The second experimental session was conducted in the Neurolinguistics and Language Processing Laboratory at the University of Kansas. During this session, each learner was tested using the experimental stimuli in a grammaticality judgment task during simultaneous EEG recording, described below. The Edinburgh Handedness Inventory was completed first, in order to verify right-handedness (Oldfield, 1971). Participants were then fitted with an electrode cap and additional electrodes above, below, and to the outside of each eye and behind each ear. Participants were seated comfortably in front of a CRT monitor in a dimly-lit, sound-attenuated experiment room. The Paradigm experimental control system designed by Perception Research Systems, Inc. (Tagliaferri, 2005) was employed for randomized stimulus presentation and EEG interfacing. In order to familiarize the participant with the task, each recording session began with a practice session of nine trials, which included items targeting areas of grammar not investigated in the experiment. No words appeared in both the practice and experimental items. Participants received feedback after each of the first three practice items. For all practice and experimental items, each trial was preceded by a blank screen for 500ms, allowing the participant time to blink, followed by a fixation cross for 500ms and then a 300-ms pause, after which the stimulus sentence was presented word-by-word in an RSVP (Rapid Serial Visual Presentation) paradigm. Each word appeared in black text on a dark gray screen for 450ms, with a pause between each word lasting 300ms. A pause of 1000ms followed the final word of the sentence before a response prompt appeared on the screen, asking participants to indicate whether the sentence was “Bien” (good) or “Mal” (bad). Participants were instructed to respond by button push for “Bien” if they felt the Spanish sentence was grammatical and “Mal” if they felt it was ungrammatical. The accuracy of the participant’s response was recorded by the software. After every 40 trials, participants were prompted to take a break. The task averaged around 55 minutes for learners to complete, including breaks.

Following the experimental task, the electrode cap was removed. Participants were given a break, followed by a short computerized test that included both the gender assignment and the vocabulary recognition tasks, in that order. Instructions and six practice items immediately preceded each task. No feedback was provided for the practice items. Again, the test items were presented in random order and responses were recorded by the Paradigm experimental control software (Tagliaferri, 2005). For the gender assignment task, each trial was preceded by a 300-ms pause and 500-ms fixation cross. In order to mimic the conditions of the grammaticality judgment task, each Spanish noun was presented for 450ms in black text on a dark gray screen. Presentation of the noun was followed by a prompt that presented the masculine determiner “El” on the left of the screen and the feminine determiner “La” on the right. Participants responded by mouse click in the area of the determiner of choice. For each item in the vocabulary recognition task, a Spanish word was presented in lower-case letters, with two possible English translations for that word in capital letters below it. Participants were asked to use the mouse to select the appropriate translation for each item. A 1000-ms pause occurred between each trial. Participants finished these computerized tasks in approximately 10 minutes. The total time for this session was around 2½ hours, including 10 minutes for paperwork, 45-60 minutes to set the electrode cap, approximately one hour for the EEG task, a break, and then another 10 minutes for the gender assignment and vocabulary recognition tasks.

Additionally, 14 learners participated in a follow-up session after the EEG session had been completed. Data from 3 of these participants was lost due to technical failures, resulting in data from 11 participants being included for analysis. The follow-up session was conducted in the Second Language Acquisition Laboratory at the University of Kansas. In this session, which lasted less than 30 minutes, participants completed the LLAMA_F and LLAMA_B picture-based aptitude tests (Meara, 2005). Participants first provided their informed consent and then received instructions on the use of the computer interface for the LLAMA tests. In the LLAMA_F, participants had 5 minutes to infer the meaning of morphemes as they viewed a series of pictures with sentences to describe them. Following the 5-minute viewing period, participants completed a series of 20 test items. For each item, a novel picture appeared, along with two sentences. Participants were asked to choose which sentence best matched the picture. The LLAMA_B presented 20 nonsensical pictures. During a 2-minute viewing period, a word from a Central American language which had been assigned to each picture could be viewed by clicking on the picture. Afterwards, participants were tested on their ability to match each word to its picture. The tests for both the LLAMA_F and the LLAMA_B were self-paced.

7.3.2 Native Speakers

The native speakers in this study were not required to take the aptitude and proficiency tests, so they only needed to attend one session. At the beginning of this session, they provided informed consent and then completed a background questionnaire and the Edinburgh Handedness Inventory (Oldfield, 1971). After being fitted with the electrode cap, they completed the grammaticality judgment test, which was administered in the same way as described above for the learner group. Including breaks, native speakers averaged 45 minutes to complete the grammaticality judgment task. After a short break, they were asked to complete the gender assignment task in an effort to ensure the validity of all the items included in that task. The total amount of time for this session was approximately 2½ hours. Each participant was compensated financially for their time.

7.4 EEG Recording

During presentation of the grammaticality judgment task, brain potentials were recorded continuously using an electrode cap (Electro-Cap International, Inc.) containing 32 Ag/AgCl scalp electrodes arrayed in a modified 10-20 layout (midline: FPZ, FZ, FCZ, CZ, CPZ, PZ, OZ; lateral: F7/8, F3/4, FT7/8, FC3/4, T3/4, C3/4, TP7/8, CP3/4, T5/6, P3/4, O1/2) as shown in Figure 3. Additional electrodes were placed on the left and right outer canthus to monitor eye-movements and above and below each eye to monitor eye-blinks. Electrode AFZ served as the

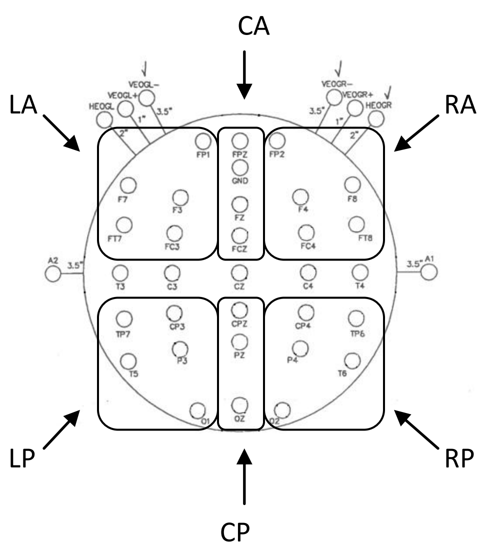

Figure 3. Map of electrodes in the 32-channel cap, showing electrodes in each region of interest, including Left Anterior (LA), Central Anterior (CA), Right Anterior (RA), Left Posterior (LP), Central Posterior (CP), and Right Posterior (RP).

ground, and an electrode at the left mastoid provided an online reference. Impedances for each scalp electrode were kept below 5 kOhms. The recordings were amplified by a Neuroscan Synamps2 amplifier (Compumedics Neuroscan, Inc.) with a bandpass of 0.1 to 200 Hz and digitized at a sampling rate of 1000 Hz. Data were re-referenced offline to the linked mastoid electrodes. Following artifact rejection, data was filtered using a 30 Hz low-pass filter.

7.5 Data Analysis

Similar analyses, described below, were conducted on behavioral and ERP data in order to test both for effects of Grammaticality (grammatical versus ungrammatical stimuli in each type of agreement) and for effects of AgreementType (number agreement on verbs versus adjectives, number versus gender agreement on adjectives). This section also describes the analysis of correlations between behavioral/ERP responses and the various measures of individual differences employed in the study, including those for verbal aptitude, nonverbal aptitude, and proficiency.

7.5.1 Behavioral Data Analysis

During simultaneous EEG recording, participants gave grammaticality judgments on sentences containing either grammatical or ungrammatical instances of three types of agreement violations across phrases: number agreement between subjects and verbs, number agreement on adjectives, and gender agreement on adjectives. Acceptance rates reflecting the percentage of items that were accepted as grammatical were calculated for each participant for each experimental condition [grammatical Subject-Verb (S-V), ungrammatical Subject-Verb (*S-V), grammatical Noun-Adj (N-Adj), ungrammatical Noun-Adj with respect to number (*N-Adj NUM), ungrammatical Noun-Adj with respect to gender (*N-Adj GEN)]. Thus, it was expected that if the participant performed well, the acceptance rate for grammatical conditions would approach 100% and for ungrammatical conditions, 0%. To test for effects of Grammaticality, a repeated-measures analysis of variance (ANOVA) with Grammaticality as the within-subjects factor was conducted on acceptance rates separately for each ungrammatical condition against its grammatical counterpart.

In the analysis of Grammaticality effects presented above, it was not appropriate to include a test for AgreementType, since the same stimuli were used as grammatical counterparts for both the number and the gender conditions. In order to directly compare sensitivity to number violations in the Subject-Verb versus the Noun-Adj Number conditions, as well as differences in the Noun-Adj Number and Gender conditions, a separate analysis was needed. Therefore, a series of repeated-measures ANOVAs with AgreementType as the within-subjects factor were carried out on the effect sizes across types of agreement, as measured by d' scores. D' scores are a measure of sensitivity to signals that reflect standardized differences in average acceptance rates for a signal (in this case ungrammatical stimuli) versus a control condition (grammatical stimuli). D' scores seek to remove acceptance bias while preserving the differences between conditions, as well as differences between participants. Here, d' scores were calculated for each participant for each type of agreement violation based on their acceptance rates in both the grammatical and ungrammatical conditions. The formula for d' used here is specific to forced-choice experimental paradigms, d' = NORMSINV(H)-NORMSINV(FA))/SQRT(2), where H = the Hit rate for choosing Grammatical when the stimulus is grammatical, and FA = the False Alarm rate for choosing Grammatical when the stimulus is ungrammatical. Following standard procedures, Hit rates of 1 (corresponding to 100% acceptance rates) were corrected to 1 – 1/480, where 480 represents twice the number of items in the test. False Alarm rates of zero were corrected to 1/480. A d' score near zero represents performance at chance, while perfect performance in this analysis results in a d’ score of approximately 4.0. Analyses for AgreementType were conducted separately for number agreement on verbs versus adjectives (S-V versus N-Adj Number) and for number and gender agreement on adjectives (N-Adj Number versus N-Adj Gender).

7.5.2 EEG Data Analysis

The EEG data were analyzed using the Neuroscan Edit software (Compumedics Neuroscan, Inc.). Pre-processing of the EEG data included the manual rejection of trials characterized by excessive muscle artifacts or eye movements and blinks, resulting in the loss of 17.4% of trials completed by learners overall, as well as 16.9% of trials completed by native speakers. Event-related potentials, time-locked to the onset of the target word in each sentence, were averaged off-line for each subject at each electrode site and were also baseline-corrected relative to a 200ms pre-stimulus interval. Based on visual inspection of grand-averaged waveforms for each type of agreement in preliminary native speaker and learner data, each epoch extended to 1000ms post-stimulus.

In order to determine the time-windows of interest, grand-averaged waveforms were generated for each experimental condition. After visual inspection of the waveforms, the following time windows were selected for analysis: 150-250ms, 250-450ms - relevant to the LAN component (Friederici, 2002), and 450-950ms, which corresponds to similar time windows where previous studies have observed a P600 response (Osterhout & Holcomb, 1992; Hagoort et al., 1993; Friederici, 2002). Since some studies have reported both an early and a late component of the P600 time window (Hagoort & Brown, 2000; Barber & Carreiras, 2005), the latter time window was also analyzed from 450-700ms and from 700-950ms.

For each time window, mean amplitudes were calculated for each condition across six different regions of electrodes: Left Anterior (FP1, F7, F3, FT7, FC3), Right Anterior (FP2, F8, F4, FT8, FC4), Left Posterior (TP7, CP3, T5, P3, O1), Right Posterior (TP6, CP4, T6, P4, O2), Central Anterior (FPZ, FZ, FCZ), and Central Posterior (CPZ, PZ, OZ). Native speakers and learners were analyzed separately. A series of ANOVAs using repeated measures was conducted on mean amplitudes for each group and for each type of agreement, including the within-subjects factors of Grammaticality (Grammatical, Ungrammatical), Laterality (Left, Central, Right), and Anteriority (Anterior, Posterior). Where violations of sphericity were present, Greenhouse-Geisser values are reported. Follow-up tests were conducted for all possible effects and interactions, whether significant or marginal. Each ungrammatical condition was compared to its grammatical counterpart, but it was not statistically expedient to include all three agreement types in the ANOVA with Agreement Type as a factor, since two types of agreement (N-Adj Number, N-Adj Gender) relied on comparison with the same grammatical stimuli. Similarly to the analysis of behavioral data presented above, a separate series of repeated-measures ANOVAs was performed for AgreementType using the effect sizes (mean amplitudes for ungrammatical minus grammatical conditions) of the components in each time window. Analyses for effects of AgreementType in the first series of ANOVAs tested number agreement on verbs versus adjectives (S-V, N-Adj Number), and in the second series the two types of agreement on adjectives were compared (N-Adj Number, N-Adj Gender).

7.5.3 Analysis of Individual Differences

Aptitude scores were recorded for the Raven Advanced Progressive Matrices (RAVEN) and the total score on the Modern Languages Aptitude Test (MLAT Total), as well as each of its subtests (MLAT 3: Spelling Clues; MLAT4: Words in Sentences; MLAT 5: Paired Associates).

Pearson correlation coefficients and their *p*-values were calculated in order to assess first the intercorrelation of these independent variables and then their relationship to d' scores for each type of agreement (S-V, N-Adj Number, N-Adj Gender). Correlations with the size of the P600 effect were also tested, using the mean amplitude differences between ungrammatical and grammatical conditions over the time windows of interest. Due to its correlation with aptitude in previous studies as well as its purported role in L2 processing, proficiency was also included in this analysis. Additionally, eleven learners participated in a follow-up study of L1-independent measures of verbal aptitude, including the LLAMA_F and LLAMA_B, which are picture-based measures similar to the MLAT4 and MLAT5, respectively. Correlations of these variables with other aptitude/proficiency measures, d' scores in the behavioral results, and mean amplitude differences for the P600 time windows are also reported.

CHAPTER 8:

NATIVE SPEAKERS' SENSITIVITY TO AGREEMENT VIOLATIONS

Results for the native speaker group are presented in this chapter, where behavioral responses in the grammaticality judgment task are first analyzed for Grammaticality effects, and differences between types of agreement are then investigated. Following that, the ERP responses to the three types of agreement are reported, along with tests for effects of Grammaticality and AgreementType. A summary of all behavioral and ERP results for the native speakers is also presented.

8.1 Behavioral Results

8.1.1 Grammaticality Judgments

The native speakers who participated in the study demonstrated a high level of accuracy with regard to all conditions, with high acceptance rates for all three grammatical conditions and low acceptance rates for all ungrammatical conditions. The mean acceptance rates for native speakers for each condition are presented in Figure 4[10]. For Subject-Verb agreement, the average acceptance rates were 96% (SD=3.62; Range=90-100) for grammatical items and 3% (SD=3.34; Range=0-10) for ungrammatical items. Native speakers accepted 97% (SD=3.43; Range=90-100) of sentences in the condition involving grammatical number and gender agreement on Adjectives (N-Adj), but only 3% (SD=3.02; Range=0-10) of the sentences containing violations of number agreement (N-Adj NUM) and 4% (SD=3.76; Range=0-10) of violations of gender agreement (N-Adj GEN).

[10] See Appendix 3 for a table of means for individual native speakers.

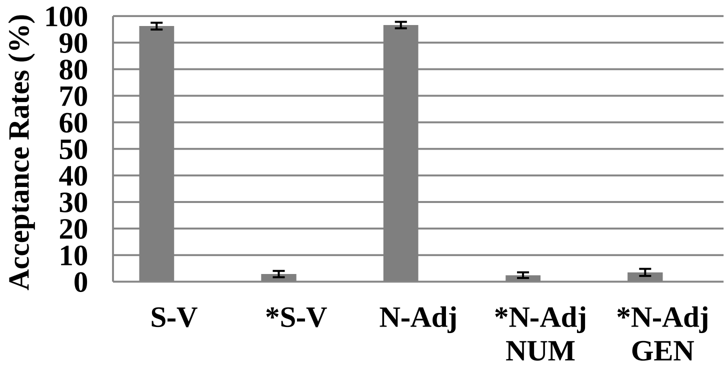

Figure 4. Mean acceptance rates for native speakers (NS) across conditions.

8.1.2 Analysis of Variance

8.1.2.1 Grammaticality Effects

To confirm native speakers' sensitivity to agreement violations, a repeated-measures analysis of variance (ANOVA) with Grammaticality as the within-subjects factor was conducted separately for each ungrammatical condition against its grammatical counterpart using the acceptance rates calculated above. When the grammatical Subject-Verb agreement condition (S-V) was compared to the ungrammatical condition (*S-V), there was a significant effect for Grammaticality, $F(1,11)=4118.925$, $p<.001$. Similarly, when the grammatical (N-Adj) and ungrammatical number (*N-Adj NUM) conditions were compared, a significant effect for Grammaticality was present, $F(1,11)=3848.192$, $p<.001$. A Grammaticality effect was also found when the grammatical (N-Adj) and ungrammatical gender (*N-Adj GEN) conditions were compared, $F(1,11)=6054.818$, $p<.001$. Results of these analyses are presented in Table 5.

Table 5. Results of Repeated Measures ANOVAs on native speakers' acceptance rates for each type of agreement.

		S-V Number	**N-Adj Number**	**N-Adj Gender**
Grammaticality	F (*p*)	4118.925* (.000)	3848.192* (.000)	6054.818* (.000)

* Effect/interaction is significant at the *p*<.05 level.

8.1.2.2 Effects of Agreement Type

The d' scores calculated as described above showed that native speakers were equally sensitive to all three types of violations. Average d' scores were 2.994 (SD=.691) for Subject-Verb agreement, 3.023 (SD=.519) for N-Adj Number agreement and 2.918 (SD=.481) for N-Adj Gender agreement. Individual d' scores for native speakers can be found in Appendix 4. In order to compare sensitivity to different types of agreement violations, a series of repeated-measures ANOVAs were conducted on d' scores, with AgreementType as the within-subjects factor. No effects of AgreementType were uncovered, either when Subject-Verb and N-Adj Number agreement were compared, $F(1,11)=.044$, $p=.837$, or when comparing N-Adj Number and Gender agreement, $F(1,11)=.350$, $p=.566$.

8.1.3 Gender Assignment

The gender assignment task was administered to both native speakers and learners following the experimental task. By making a choice between masculine and feminine determiners, participants identified the gender of each noun used as a subject in the target stimuli. Additionally, for this task, participants had to be able to assign the appropriate gender under conditions mimicking those of the experimental task. The performance of the native speakers served to validate the test since they gave the expected responses at a mean rate of 99% (SD=1.09), with a range of 97-100%.

8.1.4 Summary

As could be expected, native speakers performed very well on the grammaticality judgment task, accepting grammatical sentences at a high rate and ungrammatical sentences at a very low rate for all three types of agreement. No differences were observed in native speakers' sensitivity to violations of number agreement on verbs versus adjectives, nor to violations of number versus gender on adjectives, as measured by d’ scores. Additionally, no problems were noted in terms of gender assignment of any of the nouns used in the experiment when tested in the gender assignment task.

8.2 ERP Responses to Agreement Violations

Before investigating learners’ electrophysiological responses to the agreement violations present in the stimuli, it was necessary to determine the nature of ERP components in native speakers in response to the same violations. Native speaker ERP data for the time windows selected (150-250ms, 250-450ms, 450-950ms, 450-700ms, and 700-950ms) were submitted to a 2 x 3 x 2 repeated-measures Analysis of Variance (ANOVA), including the within-subjects factors of Grammaticality (Grammatical, Ungrammatical), Laterality (Left, Central, Right), and Anteriority (Anterior, Posterior). Significant (p<.05) and marginal (.05 <p<.1) effects and interactions are reported only if they involve Grammaticality. In the presence of a significant higher-level interaction, lower-level interactions and main effects are not interpreted. Post-hoc tests for interpreted effects are reported. In all cases, where necessary due to violations of sphericity, Greenhouse-Geisser corrected p-values are reported.

8.2.1 Number Agreement on Verbs

Number agreement on verbs was tested in the Subject-Verb agreement conditions, for which grand-averaged ERP responses for native speakers are plotted in Figure 5 at representative electrodes in each region of interest. Visual inspection of these waveforms revealed that number violations on verbs yielded more positive waveforms than their grammatical counterparts in the 450-950ms time window. This broadly-distributed positivity peaked at roughly 650ms, with an onset at approximately 450ms. Also visible was a possible positivity in early time windows in response to ungrammatical versus grammatical stimuli.

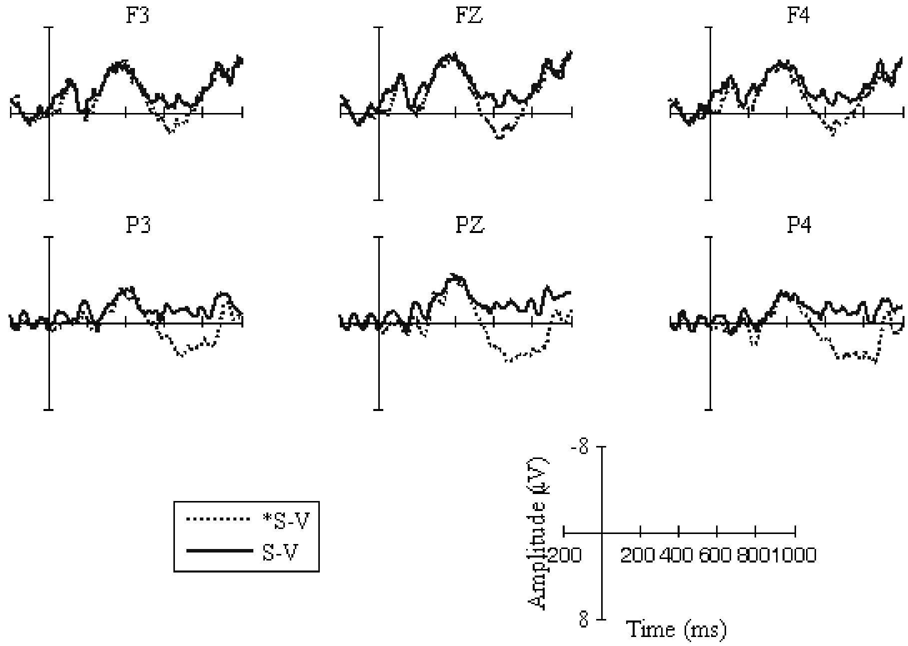

Figure 5. Grand average ERPs for native speakers in response to grammatical (solid line) and ungrammatical (dashed line) Subject-Verb agreement at representative electrodes for each region of interest.

Results of the omnibus ANOVA for each time window of interest are presented in Table 6 on the next page, and Table 7 breaks the P600 time window down into two stages, one from 450-700ms and the other from 700-950ms. Of primary interest are the Grammaticality x Laterality x Anteriority interactions that were present in all time windows, which will be explored in the analyses that follow. Grammaticality effects were also analyzed region-by-region where necessary.

8.2.1.1 150-250ms

Table 8 displays mean amplitudes in the 150-250ms time window in each electrode region for native speakers in the grammatical and ungrammatical Subject-Verb conditions (S-V and *S-V, respectively). An analysis of mean amplitudes using a three-way Repeated Measures ANOVA revealed a significant Grammaticality x Laterality x Anteriority interaction,

Table 6. Results of repeated-measures ANOVAs on native speakers' mean amplitudes for Subject-Verb agreement in each of the three primary time windows of interest.

	150-250ms	**250-450ms**	**450-950ms**
Grammaticality x Laterality x Anteriority	***F*[1.024,11.264]=9.675 *p*=.009***	***F*[1.034,11.377]=5.076 *p*=.044***	***F*[1.169,12.860]=5.280 *p*=.035***
Grammaticality x Laterality	***F*[2,22]=3.029 *p*=.069†**	***F*[1.271,13.982]=3.388 *p*=.080†**	*F*[1.303,14.334]=2.103 *p*=.167
Grammaticality x Anteriority	*F*[1,11]=2.232 *p*=.163	*F*[1,11]=1.268 *p*=.284	***F*[1,11]=10.626 *p*=.008***
Laterality x Anteriority	*F*[1.109,12.202]=1.939 *p*=.189	***F*[1.135,12.490]=6.839 *p*=.019***	***F*[1.161,12.770]=4.038 *p*=.062†**
Grammaticality	***F*[1,11]=3.623 *p*=.083†**	*F*[1,11]=1.716 *p*=.217	***F*[1,11]=11.346 *p*=.006***
Laterality	***F*[2,22]=2.825 *p*=.081†**	***F*[1.254,13.793]=17.169 *p*=.001***	***F*[2,22]=5.322 *p*=.013***
Anteriority	***F*[1,11]=5.652 *p*=.037***	***F*[1,11]=34.008 *p*<.001***	***F*[1,11]=40.749 *p*<.001***

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

Table 7. Results of repeated-measures ANOVAs on native speakers' mean amplitudes for Subject-Verb agreement in the early and late P600 time windows.

	450-700ms	**700-950ms**
Grammaticality x Laterality x Anteriority	***F*[1.138,12.522]=4.849 *p*=.043***	***F*[1.241,13.653]=6.316 *p*=.020***
Grammaticality x Laterality	*F*[1.169,12.858]=2.635 *p*=.126	*F*[2,22]=1.569 *p*=.231
Grammaticality x Anteriority	***F*[1,11]=5.555 *p*=.038***	***F*[1,11]=13.587 *p*=.004***
Laterality x Anteriority	***F*[1.127,12.400]=6.416 *p*=.023***	***F*[1.242,13.663]=3.459 *p*=.078†**
Grammaticality	***F*[1,11]=14.294 *p*=.003***	***F*[1,11]=6.776 *p*=.025***
Laterality	***F*[2,22]=5.794 *p*=.010***	***F*[2,22]=4.426 *p*=.024***
Anteriority	***F*[1,11]=20.784 *p*=.001***	***F*[1,11]=51.814 *p*<.001***

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

F(1.024,11.264)=9.675, p=.009. The nature of this interaction was further explored by conducting separate follow-up analyses within each level of anteriority (anterior, posterior). For the anterior electrodes, analyses revealed a Grammaticality x Laterality interaction, F(1.795,19.744)=4.004, p=.038, but no main effect of Grammaticality, and post-hoc tests for the Grammaticality x Laterality interaction in each anterior region (left, central, right) revealed no main effects of Grammaticality. Follow-up analysis of the posterior electrodes also revealed a Grammaticality x Laterality interaction, F(1.215,13.365)=6.937, p=.016, but in this case there was also a main effect of Grammaticality, F(1,11)=6.118, p=.031. Post-hoc tests revealed a main effect of Grammaticality only in the Left Posterior region, F(1,11)=8.369, p=.015, indicating that violations of Subject-Verb agreement yielded significantly more positive waveforms than their grammatical counterparts.

Table 8. Mean amplitudes from 150-250ms in each region of interest for native speakers in the grammatical and ungrammatical Subject-Verb conditions.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
S-V	Avg	-.460	-1.154	-1.250	-.875	-.511	-.227
	(SD)	(2.467)	(2.676)	(2.134)	(1.650)	(2.059)	(1.587)
***S-V**	Avg	-.513	-.414	-.254	2.920	.500	.262
	(SD)	(1.516)	(2.157)	(1.208)	(4.163)	(2.442)	(1.736)

8.2.1.2 250-450ms

Mean amplitudes are displayed in Table 9 below for native speakers in the Subject-Verb conditions during the 250-450ms time window in each electrode region. For this time window, the three-way Repeated Measures ANOVA also revealed a significant Grammaticality x Laterality x Anteriority interaction, F(1.034,11.377)=5.076, p=.044. Follow-up analysis of the anterior electrodes revealed a Grammaticality x Laterality interaction, F(1.693,18.623)=4.364, p=.033, but no main effect of Grammaticality, and separate post-hoc tests for each anterior region (left, central, right) revealed no effects of Grammaticality. Follow-up analysis of the posterior electrodes revealed a marginal Grammaticality x Laterality interaction, $F(1.089,11.976)=4.346$, $p=.057$. Post-hoc tests for the Grammaticality x Laterality interaction revealed a main effect of Grammaticality in the Left Posterior region, $F(1,11)=4.849$, $p=.050$, indicating that the positivity observed in the Left Posterior region in the earlier time window in response to ungrammatical Subject-Verb stimuli was still present in the 250-450ms time window.

Table 9. Mean amplitudes from 250-450ms in each region of interest for native speakers in the grammatical and ungrammatical Subject-Verb conditions.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
S-V	Avg	-2.784	-3.751	-3.049	-1.043	-2.313	-.856
	(SD)	(1.707)	(1.658)	(1.231)	(1.424)	(1.837)	(1.366)
***S-V**	Avg	-2.879	-3.788	-2.409	1.421	-2.239	-1.034
	(SD)	(1.517)	(1.799)	(1.335)	(3.596)	(1.880)	(1.384)

8.2.1.3 450-950ms

Table 10 displays mean amplitudes in the 450-950ms time window in each electrode region for native speakers in the Subject-Verb conditions. The three-way Repeated Measures ANOVA revealed a significant Grammaticality x Laterality x Anteriority interaction, $F(1.169,12.860)=5.280$, $p=.035$. Follow-up analysis of the anterior electrodes revealed a Grammaticality x Laterality interaction, $F(1.739,19.127)=4.710$, $p=.026$. Separate post-hoc tests for the Grammaticality x Laterality interaction in each anterior region (left, central, right) revealed a marginal effect of Grammaticality in the Right Anterior region, $F(1,11)=3.797$, $p=.077$, where ungrammatical stimuli yielded slightly more positive mean amplitudes than grammatical stimuli. This positivity was also present across the posterior electrodes, where follow-up analysis revealed a marginal Grammaticality x Laterality interaction,

$F(1.144,12.582)=4.061$, $p=.062$, as well as a main effect of Grammaticality, $F(1,11)=20.486$, $p=.001$. Post-hoc tests revealed a main effect of Grammaticality in all three posterior regions (Left Posterior: $F(1,11)=13.897$, $p=.003$; Central Posterior: $F(1,11)=15.153$, $p=.003$; Right Posterior: $F(1,11)=14.510$, $p=.003$), reflecting a significantly greater positivity in response to ungrammatical versus grammatical stimuli in the posterior regions.

Table 10. Mean amplitudes from 450-950ms in each region of interest for native speakers in the grammatical and ungrammatical Subject-Verb conditions.

		Left Anterior	**Central Anterior**	**Right Anterior**	**Left Posterior**	**Central Posterior**	**Right Posterior**
S-V	Avg	-1.381	-2.019	-1.916	-.610	-1.124	-.676
	(SD)	(1.642)	(1.934)	(1.317)	(1.038)	(1.672)	(1.159)
***S-V**	Avg	-.958	-.953	-.684	3.449	1.439	1.046
	(SD)	(1.352)	(1.548)	(1.357)	(3.840)	(1.128)	(.794)

8.2.1.3.1 450-700ms

The time window of interest with regard to possible P600 effects (450-950ms) was further divided into an early and a late window for analysis. Native speakers' mean amplitudes for the Subject-Verb conditions in the early P600 time window (450-700ms) are displayed in Table 11. Analysis of means using a three-way Repeated Measures ANOVA revealed a significant Grammaticality x Laterality x Anteriority interaction, $F(1.138,12.522)=4.849$, $p=.043$. Follow-up analysis of the anterior electrodes revealed a Grammaticality x Laterality interaction, $F(1.635,17.985)=7.605$, $p=.006$, and a main effect of Grammaticality, $F(1,11)= 5.190$, $p=.044$. Separate post-hoc tests for each anterior region (left, central, right) revealed a marginal effect of Grammaticality in both the Central Anterior region, $F(1,11)=9.257$, $p=.011$, and the Right Anterior region, $F(1,11)=4.949$, $p=.048$, indicating that the positivity observed only marginally in the Right Anterior region over the entire P600 time window was in fact stronger and more broadly distributed in the earlier portion of that window from 450-700ms.

Follow-up analysis in the posterior regions again revealed only a marginal Grammaticality x Laterality interaction, $F(1.107,12.181)=3.683$, $p=.076$, but a significant Grammaticality effect, $F(1,11)=19.564$, $p=.001$. Post-hoc tests revealed a main effect of Grammaticality in all three posterior regions (Left Posterior: $F(1,11)=12.849$, $p=.004$; Central Posterior: $F(1,11)=12.271$, $p=.005$; Right Posterior: $F(1,11)=11.836$, $p=.006$), where ERPs to ungrammatical versus grammatical stimuli were also significantly more positive.

Table 11. Mean amplitudes from 450-700ms in each region of interest for native speakers in the grammatical and ungrammatical Subject-Verb conditions.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
S-V	Avg	-1.087	-1.637	-1.592	-.435	-1.188	-.729
	(SD)	(1.481)	(1.862)	(1.214)	(1.039)	(1.722)	(1.128)
***S-V**	Avg	-.343	.141	-.345	3.311	.932	.784
	(SD)	(1.335)	(1.674)	(1.368)	(3.790)	(1.671)	(1.238)

8.2.1.3.2 700-950ms

Table 12 displays mean amplitudes in the late P600 time window (700-950ms) in each electrode region for native speakers in the grammatical and ungrammatical Subject-Verb conditions. The three-way Repeated Measures ANOVA revealed that the significant Grammaticality x Laterality x Anteriority interaction was still present, $F(1.241,13.653)=6.316$, $p=.020$. Follow-up analysis of the anterior electrodes revealed a Grammaticality x Laterality interaction, $F(1.644,18.081)=5.634$, $p=.016$, but no effect of Grammaticality, and separate post-hoc tests for each Anteriority (left, central, right) revealed no effects of Grammaticality. Follow-up analysis in the posterior regions again revealed a marginal Grammaticality x Laterality interaction, $F(1.261,13.868)=4.260$, $p=.051$, and a significant main effect of Grammaticality, $F(1,11)=14.872$, $p=.003$. Post-hoc tests revealed a main effect of Grammaticality in all three posterior regions (Left Posterior: $F(1,11)=13.096$, $p=.004$; Central

Posterior: $F(1,11)=11.134$, $p=.007$; Right Posterior: $F(1,11)=10.018$, $p=.009$), where ERPs to ungrammatical stimuli were significantly more positive than to grammatical stimuli.

Table 12. Mean amplitudes from 700-950ms in each region of interest for native speakers in the grammatical and ungrammatical Subject-Verb conditions.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
S-V	Avg	-1.671	-2.397	-2.238	-.784	-1.061	-.624
	(SD)	(1.907)	(2.131)	(1.523)	(1.209)	(1.875)	(1.390)
***S-V**	Avg	-1.567	-2.039	-1.016	3.591	1.951	1.310
	(SD)	(1.727)	(2.022)	(1.752)	(4.123)	(1.965)	(1.366)

8.2.1.4 Summary

Native speakers demonstrated electrophysiological sensitivity to ungrammatical stimuli in comparison with grammatical Subject-Verb agreement. Despite earlier studies that showed a left anterior negativity (LAN) in response to L1 morphosyntactic anomalies, the native speakers here demonstrated an early positivity in the Left Posterior region that was present in both the 150-250ms and 250-450ms time windows. However, the late positivity often seen in native speakers in response to morphosyntactic anomalies was also present here. The response was broadly distributed in both hemispheres in the 450-700ms time window and more posteriorly distributed from 700-950ms.

8.2.2 Number Agreement on Adjectives

Native speakers' grand-averaged ERP responses for conditions involving Noun-Adjective Number agreement are plotted in Figure 6 at representative electrodes in each region of interest. Visual inspection revealed that number violations on Adjectives also yielded more positive waveforms than their grammatical counterparts in the 450-950ms time window, particularly in the posterior electrode regions. This broadly-distributed positivity peaked at roughly 700ms, with an onset at approximately 500ms.

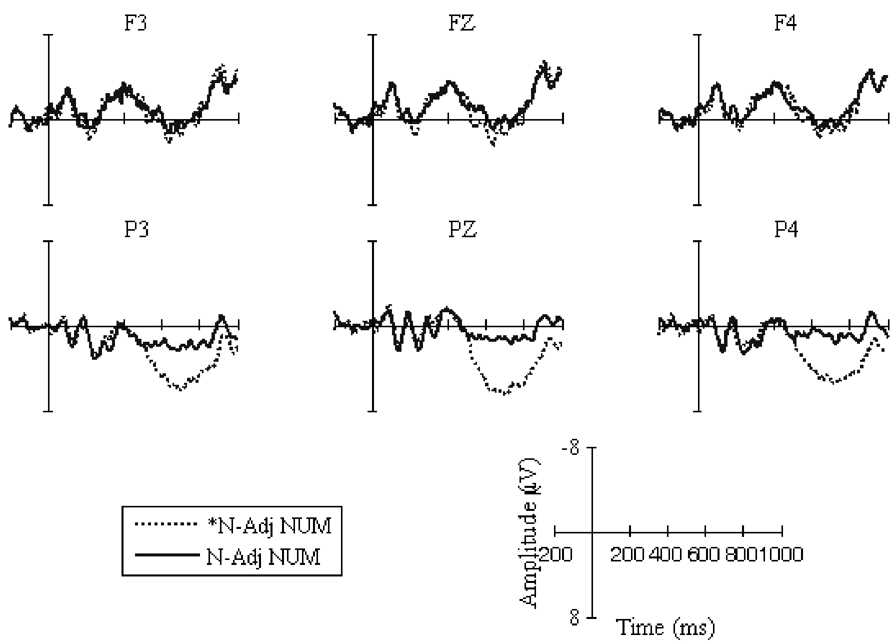

Figure 6. Grand average ERPs for native speakers in response to grammatical (solid line) and ungrammatical (dashed line) Noun-Adjective Number agreement at representative electrodes for each region of interest.

Mean amplitudes for the grammatical and ungrammatical Noun-Adjective Number conditions were also submitted to a 2 x 3 x 2 repeated-measures ANOVA for each time window of interest, with within-subject factors of Grammaticality (Grammatical, Ungrammatical), Laterality (Left, Central, Right), and Anteriority (Anterior, Posterior). Results of the omnibus ANOVA are presented in Table 13 for the three overall time windows, and in Table 14 for the two P600 time windos. No effects of Grammaticality were present in early time windows, but Grammaticality x Laterality x Anteriority interactions in the early, late, and overall P600 time windows will be investigated. The analyses for each time window are followed by a region-by-region analysis of Grammaticality effects, in order to determine whether limited effects may have been present.

Table 13. Results of repeated-measures ANOVAs on native speakers' mean amplitudes for Noun-Adjective Number agreement in each of three primary time windows of interest.

	150-250ms	**250-450ms**	**450-950ms**
Grammaticality x Laterality x Anteriority	$F[1.189,13.075]=1.772$ $p=.208$	$F[1.280,14.083]=2.925$ $p=.102$	**$F[1.373,15.101]=7.631$ $p=.009$***
Grammaticality x Laterality	$F[1.243,13.671]=2.707$ $p=.118$	$F[1.296,14.261]=2.022$ $p=.176$	**$F[1.410,15.508]=6.726$ $p=.013$***
Grammaticality x Anteriority	$F[1,11]=2.145$ $p=.171$	$F[1,11]=.779$ $p=.396$	**$F[1,11]=15.930$ $p=.002$***
Laterality x Anteriority	**$F[1.073,11.808]=5.506$ $p=.035$***	**$F[1.159,12.745]=4.984$ $p=.040$***	**$F[1.079,11.864]=6.091$ $p=.028$***
Grammaticality	$F[1,11]=.407$ $p=.537$	$F[1,11]=.390$ $p=.545$	**$F[1,11]=10.648$ $p=.008$***
Laterality	**$F[2,22]=3.916$ $p=.035$***	**$F[1.284,14.128]=8.250$ $p=.009$***	**$F[1.136,12.498]=5.358$ $p=.035$***
Anteriority	**$F[1,11]=6.387$ $p=.028$***	**$F[1,11]=18.494$ $p=.001$***	**$F[1,11]=24.206$ $p<.001$***

* Effect/interaction is significant at the $p<.05$ level.
† Effect is marginally significant ($.10>p>.05$).

Table 14. Results of the repeated-measures ANOVAs on native speakers' mean amplitudes for Noun-Adjective Number agreement in the early and late P600 time windows.

	450-700ms	**700-950ms**
Grammaticality x Laterality x Anteriority	**$F[1.339,14.728]=5.669$ $p=.024$***	**$F[2,22]=8.998$ $p=.001$***
Grammaticality x Laterality	**$F[2,22]=8.889$ $p=.001$***	**$F[1.384,15.221]=4.082$ $p=.051$†**
Grammaticality x Anteriority	**$F[1,11]=10.084$ $p=.009$***	**$F[1,11]=15.751$ $p=.002$***
Laterality x Anteriority	**$F[1.127,12.399]=4.982$ $p=.041$***	**$F[1.081,11.888]=7.384$ $p=.017$***
Grammaticality	**$F[1,11]=7.683$ $p=.018$***	**$F[1,11]=8.352$ $p=.015$***
Laterality	**$F[1.119,12.304]=3.789$ $p=.071$†**	**$F[1.227,13.494]=6.959$ $p=.016$***
Anteriority	**$F[1,11]=16.330$ $p=.002$***	**$F[1,11]=31.273$ $p<.001$***

* Effect/interaction is significant at the $p<.05$ level.
† Effect is marginally significant ($.10>p>.05$).

8.2.2.1 150-250ms

Following the tables reporting the omnibus ANOVA results, Table 15 displays mean amplitudes in the 150-250ms time window in each electrode region for native speakers in the grammatical and ungrammatical conditions involving Noun-Adjective number agreement. An analysis of mean amplitudes using a three-way Repeated Measures ANOVA revealed no effect of Grammaticality, nor any interactions with Grammaticality.

Table 15. Mean amplitudes from 150-250ms in each region of interest for native speakers in the conditions involving grammatical and ungrammatical Noun-Adjective Number agreement.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
N-Adj	Avg	-.173	-.127	.180	3.070	.210	.838
	(SD)	(1.542)	(1.756)	(.980)	(3.961)	(1.147)	(1.175)
***N-Adj NUM**	Avg	.453	.702	.532	2.487	.757	.851
	(SD)	(1.523)	(2.174)	(1.549)	(2.893)	(2.780)	(2.011)

8.2.2.2 250-450ms

Mean amplitudes are displayed in Table 16 for native speakers in the grammatical and ungrammatical conditions involving Noun-Adjective number agreement during the 250-450ms time window. Here, the three-way repeated-measures ANOVA again revealed no effect and no interactions with Grammaticality, nor any interactions with Grammaticality. The region-by-region analyses revealed only a marginal effect of Grammaticality in the Left Posterior region, $F(1,11)=3.417$, $p=.092$.

Table 16. Mean amplitudes from 250-450ms in each region of interest for native speakers in the conditionsb involving grammatical and ungrammatical Noun-Adjective Number agreement.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
N-Adj	Avg	-1.414	-2.187	-.549	2.836	-.560	.935
	(SD)	(1.833)	(2.585)	(1.522)	(3.296)	(1.690)	(1.862)
***N-Adj NUM**	Avg	-1.436	-2.059	-.916	1.793	-.478	.529
	(SD)	(1.592)	(2.432)	(1.638)	(2.351)	(2.233)	(1.967)

8.2.2.3 450-950ms

Table 17 displays mean amplitudes in the 450-950ms time window in each electrode region for native speakers in the Noun-Adjective conditions involving number agreement. The three-way Repeated Measures ANOVA revealed a significant Grammaticality x Laterality x Anteriority interaction, $F(1.373,15.101)=7.631$, $p=.009$. Follow-up analysis of the anterior electrodes showed no interactions and no effect of Grammaticality. This was not the case in the posterior regions, where follow-up analyses revealed a significant Grammaticality x Laterality interaction, $F(1.233,13.559)=8.740$, $p=.008$, as well as a main effect of Grammaticality, $F(1,11)=20.824$, $p=.001$. Post-hoc tests revealed a main effect of Grammaticality that was marginal in the Left Posterior region, $F(1,11)=3.661$, $p=.082$, and significant in the Central Posterior region, $F(1,11)=28.355$, $p<.001$, as well as the Right Posterior region, $F(1,11)=14.662$, $p=.003$. In all three regions, responses to ungrammatical stimuli were more positive than those to grammatical stimuli.

Table 17. Mean amplitudes from 450-950ms in each region of interest for native speakers in the conditions involving grammatical and ungrammatical Noun-Adjective Number agreement.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
N-Adj	Avg	-.749	-1.498	-.083	2.986	.0556	.960
	(SD)	(1.482)	(2.006)	(1.027)	(3.920)	(1.205)	(1.115)
***N-Adj NUM**	Avg	-.161	-.877	.625	4.106	3.504	2.707
	(SD)	(1.078)	(1.882)	(1.129)	(2.855)	(2.118)	(1.749)

8.2.2.3.1 450-700ms

The late positivity observed for native speakers in the overall P600 time window for Noun-Adjective Number violations was further investigated for early and late time windows. Mean amplitudes for the grammatical and ungrammatical conditions from 450-700ms are displayed in Table 18. The three-way Repeated Measures ANOVA revealed a significant

Grammaticality x Laterality x Anteriority interaction, $F(1.339,14.728)=5.669$, $p=.024$. Just as in the overall P600 time window, follow-up analysis of the anterior electrodes showed no interactions and no effect of Grammaticality. Follow-up analysis in the posterior regions revealed a significant Grammaticality x Laterality interaction, $F(1.247,13.714)=9.407$, $p=.006$, and a main effect of Grammaticality, $F(1,11)=16.138$, $p=.002$. Post-hoc tests in the Left Posterior region showed no effect of Grammaticality, but there was a main effect of Grammaticality in the Central Posterior region, $F(1,11)=33.483$, $p<.001$, and in the Right Posterior region, $F(1,11)=15.318$, $p=.002$), where ERPs to ungrammatical versus grammatical stimuli were significantly more positive.

Table 18. Mean amplitudes from 450-700ms in each region of interest for native speakers in the conditions involving grammatical and ungrammatical Noun-Adjective Number agreement.

		Left Anterior	**Central Anterior**	**Right Anterior**	**Left Posterior**	**Central Posterior**	**Right Posterior**
N-Adj	Avg	-.251	-.900	.340	3.144	.260	1.049
	(SD)	(1.536)	(2.079)	(1.105)	(3.783)	(1.353)	(1.219)
***N-Adj NUM**	Avg	.353	.045	1.011	3.823	3.194	2.541
	(SD)	(1.806)	(2.901)	(1.724)	(2.590)	(1.789)	(1.900)

8.2.2.3.2 700-950ms

Table 19 displays mean amplitudes in the late P600 time window (700-950ms) in each electrode region for native speakers in the grammatical and ungrammatical Noun-Adjective Number conditions. The three-way Repeated Measures ANOVA revealed that the significant Grammaticality x Laterality x Anteriority interaction was still present, $F(2,22)=8.998$, $p=.001$. Follow-up analysis of the anterior electrodes showed no interactions and no effect of Grammaticality. However, follow-up analysis in the posterior regions revealed that the Grammaticality x Laterality interaction present in the early P600 window was still present in the late P600 window, $F(1.209,13.297)=7.558$, $p=.013$, as well as a main effect of Grammaticality,

$F(1,11)=15.112$, $p=.003$. Post-hoc tests revealed a main effect of Grammaticality in all three posterior regions (Left Posterior: $F(1,11)=6.456$, $p=.027$; Central Posterior: $F(1,11)=17.775$, $p=.001$; Right Posterior: $F(1,11)=9.525$, $p=.010$), where ERPs to ungrammatical stimuli were significantly more positive than to grammatical stimuli.

Table 19. Mean amplitudes from 700-950ms in each region of interest for native speakers in the conditions involving grammatical and ungrammatical Noun-Adjective Number agreement.

		Left Anterior	**Central Anterior**	**Right Anterior**	**Left Posterior**	**Central Posterior**	**Right Posterior**
N-Adj	Avg	-1.244	-2.092	-.503	2.831	-.148	.871
	(SD)	(1.549)	(2.090)	(1.099)	(4.095)	(1.175)	(1.226)
***N-Adj NUM**	Avg	-.667	-1.790	.245	4.395	3.821	2.877
	(SD)	(.908)	(1.459)	(1.241)	(3.363)	(3.182)	(2.160)

8.2.2.4 Summary

Native speakers' ERP responses to violations of Noun-Adjective Number agreement did not include the early positivity seen in the Subject-Verb conditions, nor was the canonical LAN present. However, a P600 effect was present which was strongest in the Central and Right Posterior regions in the 450-700ms window and across all three posterior regions from 700-950ms, reflecting significantly more positive responses to violations of Noun-Adjective Number agreement than to their grammatical counterparts.

8.2.3 Gender Agreement on Adjectives

Finally, native speakers' grand-averaged ERP responses for conditions involving Noun-Adjective Gender agreement are plotted in Figure 7 at representative electrodes in each region of interest. Visual inspection reveals that gender violations on Adjectives yielded more positive waveforms than their grammatical counterparts in the 450-950ms time window, just as with number violations on both verbs and adjectives, presented earlier. The positivity for gender violations peaks 650ms, with an onset at approximately 350ms, and somewhat earlier in anterior electrodes.

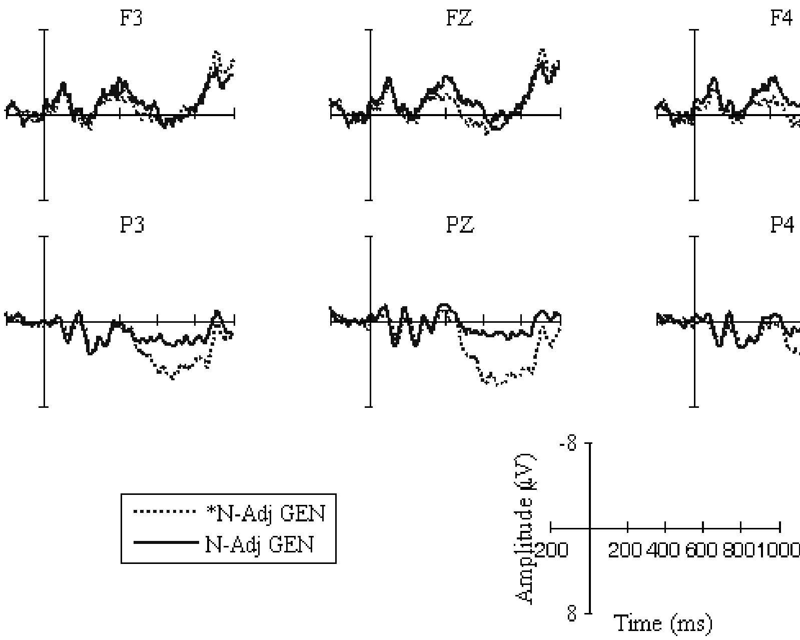

Figure 7. Grand average ERPs for native speakers in response to grammatical (solid line) and ungrammatical (dashed line) Noun-Adjective Gender agreement at representative electrodes for each region of interest.

Results are presented in Table 20 for the series of 2 x 3 x 2 repeated-measures ANOVAs that were conducted in each main time window of interest on the mean amplitudes for the grammatical and ungrammatical Noun-Adjective Gender conditions, and in Table 21 for the P600 time windows. Again, within-subject factors included Grammaticality (Grammatical, Ungrammatical), Laterality (Left, Central, Right), and Anteriority (Anterior, Posterior). Interactions involving Grammaticality will be explored in the following sections. The analyses for each time window are also followed by a region-by-region analysis of Grammaticality effects, in order to determine whether limited effects may have been present.

Table 20. Results of repeated-measures ANOVAs on native speakers' mean amplitudes for Noun-Adjective Gender agreement in each of three primary time windows of interest.

	150-250ms	**250-450ms**	**450-950ms**
Grammaticality x Laterality x Anteriority	*F*[1.216,13.373]=2.515 *p*=.133	*F*[1.450,15.950]=2.412 *p*=.132	*F*[2,22]=1.719 *p*=.203
Grammaticality x Laterality	*F*[1.218,13.403]=2.385 *p*=.143	***F*[1.438,15.823]=3.139 *p*=.084†**	***F*[2,22]=4.227 *p*=.028***
Grammaticality x Anteriority	*F*[1,11]=2.128 *p*=.173	*F*[1,11]=1.919 *p*=.193	*F*[1,11]=3.153 *p*=.103
Laterality x Anteriority	***F*[1.034,11.369]=3.828 *p*=.074†**	***F*[1.075,11.827]=3.927 *p*=.069†**	***F*[1.054,11.597]=4.780 *p*=.049***
Grammaticality	*F*[1,11]=.000 *p*=.988	*F*[1,11]=.394 *p*=.543	***F*[1,11]=5.938 *p*=.033***
Laterality	***F*[2,22]=2.704 *p*=.089†**	***F*[1.303,14.335]=5.172 *p*=.031***	***F*[1.142,12.563]=4.774 *p*=.045***
Anteriority	***F*[1,11]=5.704 *p*=.036***	***F*[1,11]=13.594 *p*=.004***	***F*[1,11]=17.065 *p*=.002***

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

Table 21. Results of repeated-measures ANOVAs on native speakers' mean amplitudes for Noun-Adjective Gender agreement in the early and late P600 time windows.

	450-700ms	**700-950ms**
Grammaticality x Laterality x Anteriority	*F*[2,22]=1.673 *p*=.211	*F*[2,22]=2.135 *p*=.142
Grammaticality x Laterality	***F*[1.447,15.913]=7.375 *p*=.009***	*F*[2,22]=1.904 *p*=.173
Grammaticality x Anteriority	*F*[1,11]=.593 *p*=.457	***F*[1,11]=5.220 *p*=.043***
Laterality x Anteriority	***F*[1.143,12.573]=4.094 *p*=.061†**	***F*[1.058,11.633]=5.448 *p*=.037***
Grammaticality	***F*[1,11]=4.433 *p*=.059†**	***F*[1,11]=7.372 *p*=.020***
Laterality	***F*[1.131,12.439]=3.151 *p*=.097†**	***F*[1.190,13.091]=6.800 *p*=.018***
Anteriority	***F*[1,11]=11.401 *p*=.006***	***F*[1,11]=22.118 *p*=.001***

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

8.2.3.1 150-250ms

Native speakers' mean amplitudes by electrode region in the 150-250ms time window for grammatical and ungrammatical conditions involving Noun-Adjective gender agreement (N-Adj and *N-Adj GEN, respectively) are displayed in Table 22 (following the omnibus ANOVA tables below). An analysis of mean amplitudes using a three-way Repeated Measures ANOVA revealed no effect of Grammaticality, nor were there any interactions with Grammaticality.

Table 22. Mean amplitudes from 150-250ms in each region of interest for native speakers in the conditions involving grammatical and ungrammatical Noun-Adjective Gender agreement.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
N-Adj	Avg	-.173	-.127	.180	3.070	.210	.838
	(SD)	(1.542)	(1.756)	(.980)	(3.961)	(1.147)	(1.175)
***N-Adj GEN**	Avg	.277	.239	.388	1.889	.460	.780
	(SD)	(1.414)	(1.870)	(1.091)	(3.678)	(1.256)	(1.557)

8.2.3.2 250-450ms

Mean amplitudes are displayed in Table 23 for native speakers in the grammatical and ungrammatical conditions involving Noun-Adjective gender agreement during the 250-450ms time window in each electrode region. For this time window, the three-way Repeated Measures ANOVA revealed a marginal Grammaticality x Laterality interaction, $F(1.438,15.823)= 3.139$, $p=.084$. Post-hoc analyses yielded a marginal effect of Grammaticality only in the central electrodes, $F(1,11)=4.514$, $p=.057$, where amplitudes of ERP responses to ungrammatical versus grammatical stimuli were marginally more positive. The region-by-region analyses revealed that this effect was only marginally present in the Central Anterior region, $F(1,11)=4.031$, $p=.070$.

Table 23. Mean amplitudes from 250-450ms in each region of interest for native speakers in the conditions involving grammatical and ungrammatical Noun-Adjective Gender agreement.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
N-Adj	Avg	-1.414	-2.187	-.549	2.836	-.560	.935
	(SD)	(1.833)	(2.585)	(1.522)	(3.296)	(1.690)	(1.862)
***N-Adj GEN**	Avg	-.888	-1.240	-.372	1.833	-.140	.970
	(SD)	(1.700)	(2.894)	(1.572)	(3.100)	(1.763)	(1.926)

8.2.3.3 450-950ms

Table 24 displays mean amplitudes in the 450-950ms time window in each electrode region for native speakers in the conditions involving gender agreement on Adjectives. The three-way Repeated Measures ANOVA revealed a Grammaticality x Laterality interaction, $F(2,22)=4.227$, $p=.028$, as well as a main effect of Grammaticality, $F(1,11)=5.938$, $p=.033$. Post-hoc analyses yielded a main effect of Grammaticality in the central electrodes, $F(1,11)=12.246$, $p=.005$, and in the electrodes in the right hemisphere, $F(1,11)=7.715$, $p=.018$, reflecting a significantly greater positivity in response to ungrammatical versus grammatical stimuli with a distribution in the central and right electrodes. The region-by-region analyses revealed that the positivity was only present in the posterior regions (Central Posterior: $F(1,11)=15.504$, $p=.002$; Right Posterior: $F(1,11)=11.709$, $p=.006$).

Table 24. Mean amplitudes from 450-950ms in each region of interest for native speakers in the conditions involving grammatical and ungrammatical Noun-Adjective Gender agreement.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
N-Adj	Avg	-.749	-1.498	-.083	2.986	.0556	.960
	(SD)	(1.482)	(2.006)	(1.027)	(3.920)	(1.205)	(1.115)
***N-Adj GEN**	Avg	-.219	-.662	.532	3.682	2.227	2.555
	(SD)	(1.662)	(2.705)	(1.784)	(3.585)	(1.853)	(1.788)

8.2.3.3.1 450-700ms

Native speakers' mean amplitudes are displayed in Table 25 for the 450-700ms time window in each electrode region for conditions involving gender agreement on Adjectives. The three-way Repeated Measures ANOVA revealed a Grammaticality x Laterality interaction, $F(1.447,15.913)=7.375$, $p=.009$, and a marginal effect of Grammaticality, $F(1,11)=4.433$, $p=.059$. Post-hoc analyses yielded a main effect of Grammaticality in both the central electrodes, $F(1,11)=9.619$, $p=.010$, and electrodes in the right hemisphere, $F(1,11)=5.473$, $p=.039$, driven by significantly more positive waveforms in response to ungrammatical versus grammatical stimuli. Region-by-region analyses indicated that this positivity was significant in the Right Posterior region, $F(1,11)=6.634$, $p=.026$ and both central regions (Central Anterior: $F(1,11)=5.978$, $p=.033$; Central Posterior: $F(1,11)=10.150$, $p=.009$), and it was marginally present in the Left Anterior region, $F(1,11)=3.577$, $p=.085$, and in the Right Anterior regions as well, $F(1,11)=3.409$, $p=.092$.

Table 25. Mean amplitudes from 450-700ms in each region of interest for native speakers in the conditions involving grammatical and ungrammatical Noun-Adjective Gender agreement.

		Left Anterior	**Central Anterior**	**Right Anterior**	**Left Posterior**	**Central Posterior**	**Right Posterior**
N-Adj	Avg	-.251	-.900	.340	3.144	.260	1.049
	(SD)	(1.536)	(2.079)	(1.105)	(3.783)	(1.353)	(1.219)
***N-Adj GEN**	Avg	.530	.538	1.166	3.578	2.288	2.593
	(SD)	(1.837)	(3.216)	(2.033)	(3.546)	(2.144)	(2.394)

8.2.3.3.2 700-950ms

Table 26 displays native speakers' mean amplitudes in the 700-950ms time window in each electrode region for Noun-Adjective Gender agreement. The three-way repeated-measures ANOVA revealed a Grammaticality x Anteriority interaction, $F(1,11)=5.220$, $p=.043$, and a main effect of Grammaticality, $F(1,11)=7.372$, $p=.020$. Post-hoc analysis of the anterior regions revealed no effect of Grammaticality, but this was not the case in the posterior electrodes, which yielded a main effect of Grammaticality, $F(1,11)=9.628$, $p=.010$, reflecting a significantly greater positivity in response to ungrammatical versus grammatical stimuli in the posterior regions during this time window. Region-by-region analyses revealed that this positivity was present in both the Central Posterior region, $F(1,11)=17.881$, $p=.001$, and the Right Posterior region, $F(1,11)=12.128$, $p=.005$.

Table 26. Mean amplitudes from 700-950ms in each region of interest for native speakers in the conditions involving grammatical and ungrammatical Noun-Adjective Gender agreement.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
N-Adj	Avg	-1.244	-2.092	-.503	2.831	-.148	.871
	(SD)	(1.549)	(2.090)	(1.099)	(4.095)	(1.175)	(1.226)
***N-Adj GEN**	Avg	-.964	-1.854	-.098	3.790	2.171	2.521
	(SD)	(1.657)	(2.443)	(1.740)	(3.830)	(2.019)	(1.698)

8.2.3.4 Summary

Native speakers' ERPs demonstrated sensitivity to violations of gender on Adjectives. While there was a marginal and narrowly-distributed positivity to gender violations in the 250-450ms time window, the only significant effects were found in late time windows, where responses to violations indicated a canonical P600 that was distributed across central and right electrodes in both the 450-700ms and the 700-950ms windows.

8.3 Comparison of ERP Responses across Agreement Types

A separate statistical analysis was performed for Agreement Type using the effect sizes (mean amplitudes for ungrammatical minus grammatical conditions) of the components in each time window. This analysis was performed separately for two comparisons of interest. First, the two conditions involving number agreement (Subject-Verb Number, N-Adj Number) were compared in each time window using a 2 x 3 x 2 repeated-measures ANOVA with within-subject factors of AgreementType (S-V, N-Adj Number), Laterality (Left, Central, Right), and Anteriority (Anterior, Posterior). Secondly, the same analysis was performed in order to compare number and gender agreement on Adjectives (AgreementType: N-Adj Number and N-Adj Gender). Both analyses are presented below.

8.3.1 Number Agreement on Verbs versus Adjectives

When Subject-Verb and Noun-Adjective number agreement were compared, significant AgreementType x Laterality x Anteriority interactions were present in all time windows. Results are presented in Table 27 for the 150-250ms, 250-450ms, and 450-950ms time windows. Table 28 presents the results of the same analysis in the early and late P600 time windows, 450-700ms and 700-950ms, respectively.

Table 27. Results of repeated-measures ANOVA comparing native speakers' mean difference amplitudes for number versus gender agreement on adjectives in each primary time window.

	150-250ms	250-450ms	450-950ms
AgreementType x Laterality x Anteriority	$F[2,22]=1.007$ $p=.381$	$F[2,22]=.532$ $p=.595$	$\mathbf{F[2,22]=11.229}$ $\mathbf{p<.001}$*****
AgreementType x Laterality	$F[2,22]=.735$ $p=.491$	$F[2,22]=.441$ $p=.649$	$F[1.289,14.183]=1.295$ $p=.287$
AgreementType x Anteriority	$F[1,11]=.043$ $p=.840$	$F[1,11]=2.047$ $p=.180$	$\mathbf{F[1,11]=4.267}$ $\mathbf{p=.063}$**†**
Laterality x Anteriority	$F[1.129,12.414]=2.414$ $p=.144$	$F[1.267,13.938]=2.948$ $p=.102$	$\mathbf{F[2,22]=4.086}$ $\mathbf{p=.031}$*****
AgreementType	$F[1,11]=.450$ $p=.516$	$F[1,11]=.862$ $p=.373$	$F[1,11]=.284$ $p=.605$
Laterality	$F[1.192,13.117]=2.945$ $p=.105$	$\mathbf{F[1.298,14.276]=3.172}$ $\mathbf{p=.088}$**†**	$\mathbf{F[2,22]=6.429}$ $\mathbf{p=.006}$*****
Anteriority	$F[1,11]=2.398$ $p=.150$	$F[1,11]=1.437$ $p=.256$	$\mathbf{F[1,11]=8.646}$ $\mathbf{p=.013}$*****

* Effect/interaction is significant at the $p<.05$ level.
† Effect is marginally significant ($.10>p>.05$).

Table 28. Results of repeated-measures ANOVA comparing native speakers' mean difference amplitudes for number versus gender agreement on adjectives in the early and late P600 time windows.

	450-700ms	**700-950ms**
AgreementType x Laterality x Anteriority	***F*[2,22]=7.722 *p*=.003***	***F*[2,22]=8.650 *p*=.002***
AgreementType x Laterality	*F*[1.392,15.307]=.757 *p*=.440	*F*[1.367,15.041]=1.630 *p*=.227
AgreementType x Anteriority	***F*[1,11]=6.320 *p*=.029***	*F*[1,11]=2.624 *p*=.134
Laterality x Anteriority	***F*[2,22]=3.090 *p*=.066†**	***F*[2,22]=5.351 *p*=.013***
AgreementType	*F*[1,11]=.006 *p*=.939	*F*[1,11]=.878 *p*=.369
Laterality	***F*[2,22]=9.737 *p*=.001***	***F*[2,22]=3.422 *p*=.061†**
Anteriority	***F*[1,11]=3.535 *p*=.087†**	***F*[1,11]=10.693 *p*=.007***

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

8.3.1.1 150-250ms

In the 150-250ms time window, the AgreementType x Laterality x Anteriority interaction was significant, F(1.048,11.532)=7.504, p=.018. Follow-up analysis in the anterior regions revealed an AgreementType x Laterality interaction, F(2,22)=4.091, p=.031, but no main effect of AgreementType. Similarly, no effect of AgreementType was found in the three anterior regions in post-hoc tests for the AgreementType x Laterality interaction. An AgreementType x Laterality interaction was also present in the posterior regions, F(1.232,13.554)=5.701, p=.027. Post-hoc analyses revealed a significant effect of AgreementType in the Left Posterior region, F(1,11)=6.246, p=.030, where a greater difference between grammatical and ungrammatical stimuli was noted for Subject-Verb agreement as opposed to Noun-Adjective Number agreement.

8.3.1.2 250-450ms

The AgreementType x Laterality x Anteriority interaction was also significant in the 250-450ms time window, $F(1.069,11.755)=5.052$, $p=.043$. Follow-up analysis in the 250-450ms time window revealed a significant AgreementType x Laterality interaction in anterior regions, $F(2,22)=6.974$, $p=.005$, but no main effect of AgreementType. Post-hoc tests for the AgreementType x Laterality interaction revealed only a marginal effect of AgreementType in the Right Anterior region, $F(1,11)=3.419$, $p=.092$. In the posterior regions, an AgreementType x Laterality interaction was also present, $F(2,22)=4.093$, $p=.031$. In post-hoc tests, an effect of AgreementType was found only in the Left Posterior region, $F(1,11)=5.569$, $p=.038$, reflecting a greater difference between grammatical and ungrammatical Subject-Verb agreement violations in comparison to Noun-Adjective number agreement.

8.3.1.3 450-950ms

In the 450-950ms time window, there was also a significant AgreementType x Laterality x Anteriority interaction, $F(1.130,12.432)=5.792$, $p=.029$. Follow-up analysis in the anterior electrode regions revealed no interaction and no main effect of AgreementType, $F(1,11)=.090$, $p=.769$. In posterior regions, a significant AgreementType x Laterality interaction was present, $F(1.083,11.915)=5.169$, $p=.040$. Post-hoc tests revealed only a marginal effect of AgreementType in the Left Posterior region, $F(1,11)=3.808$, $p=.077$, where the difference in mean amplitudes for grammatical and ungrammatical stimuli was greater for Subject-Verb agreement.

8.3.1.3.1 450-700ms

In the early P600 time window (450-700ms), the AgreementType x Laterality x Anteriority interaction was significant, $F(1.137,12.509)=6.022$, $p=.027$. Follow-up analysis in the anterior electrode regions revealed no interaction and no main effect of AgreementType, $F(1,11)=.327$, $p=.579$. In posterior regions, a significant AgreementType x Laterality interaction was present, $F(1.097,12.070)=5.311$, $p=.037$. Post-hoc tests revealed only a marginal effect of AgreementType in the Left Posterior region, $F(1,11)=4.078$, $p=.068$, where the difference in mean amplitudes for grammatical and ungrammatical stimuli was greater for Subject-Verb agreement, consistent with results for the overall P600 time window.

8.3.1.3.2 700-950ms

In the late P600 time window (700-950ms), the AgreementType x Laterality x Anteriority interaction was significant, $F(1.166,12.831)=5.375$, $p=.033$. Follow-up analysis in the anterior electrode regions revealed no interaction and no main effect of AgreementType, $F(1,11)=.000$, $p=.984$. In posterior regions, a significant AgreementType x Laterality interaction was present, $F(1.108,12.192)=4.795$, $p=.046$. Post-hoc tests revealed only a marginal effect of AgreementType in the Left Posterior region, $F(1,11)=3.383$, $p=.093$, indicating that the difference in mean amplitudes for grammatical and ungrammatical stimuli was also greater for Subject-Verb agreement in the late P600 time window.

8.3.1.4 Summary

When responses to number agreement on verbs were compared to those for number agreement on adjectives, a significant interaction between AgreementType, Laterality, and Anteriority was present in all time windows. This interaction was driven largely by differences in the Left Posterior region that were often marginal, but that indicate slightly greater differences between grammatical and ungrammatical Subject-Verb agreement as opposed to Noun-Adjective number agreement. This difference between AgreementTypes reflects the fact that Grammaticality effects, as previously tested, were not as broadly distributed for N-Adj Number agreement in most time windows, but an effect of AgreementType was also found in the 700-950ms time window where significant Grammaticality effects had been found to be equally distributed for both types of number agreement.

8.3.2 Number versus Gender Agreement on Adjectives

In the same way, the two violations on Adjectives were also compared in order to determine whether there were differences between number and gender agreement on Adjectives. The omnibus ANOVA results are presented in Table 29 for the 150-250ms, 250-450ms, and 450-950ms time windows. Table 30 presents the results of the same analysis in the early and late P600 time windows, 450-700ms and 700-950ms, respectively. Significant AgreementType x Laterality x Anteriority interactions were only present in the P600 time windows (overall, early, and late) and will be further explored below.

Table 29. Results of repeated-measures ANOVAs comparing native speakers' mean difference amplitudes for number and gender agreement on adjectives in each primary time window.

	150-250ms	**250-450ms**	**450-950ms**
AgreementType x Laterality x Anteriority	$F[2,22]=1.007$ $p=.381$	$F[2,22]=.532$ $p=.595$	$\boldsymbol{F[2,22]=11.229}$ $\boldsymbol{p<.001}$*
AgreementType x Laterality	$F[2,22]=.735$ $p=.491$	$F[2,22]=.441$ $p=.649$	$F[1.289,14.183]=1.295$ $p=.287$
AgreementType x Anteriority	$F[1,11]=.043$ $p=.840$	$F[1,11]=2.047$ $p=.180$	$\boldsymbol{F[1,11]=4.267}$ $\boldsymbol{p=.063}$†
Laterality x Anteriority	$F[1.129,12.414]=2.414$ $p=.144$	$F[1.267,13.938]=2.948$ $p=.102$	$\boldsymbol{F[2,22]=4.086}$ $\boldsymbol{p=.031}$*
AgreementType	$F[1,11]=.450$ $p=.516$	$F[1,11]=.862$ $p=.373$	$F[1,11]=.284$ $p=.605$
Laterality	$F[1.192,13.117]=2.945$ $p=.105$	$\boldsymbol{F[1.298,14.276]=3.172}$ $\boldsymbol{p=.088}$†	$\boldsymbol{F[2,22]=6.429}$ $\boldsymbol{p=.006}$*
Anteriority	$F[1,11]=2.398$ $p=.150$	$F[1,11]=1.437$ $p=.256$	$\boldsymbol{F[1,11]=8.646}$ $\boldsymbol{p=.013}$*

* Effect/interaction is significant at the $p<.05$ level.
† Effect is marginally significant ($.10>p>.05$).

Table 30. Results of repeated-measures ANOVA comparing native speakers' mean difference amplitudes for number and gender agreement on adjectives in the early and late P600 time windows.

	450-700ms	**700-950ms**
AgreementType x Laterality x Anteriority	**F[2,22]=7.722 p=.003***	**F[2,22]=8.650 p=.002***
AgreementType x Laterality	F[1.392,15.307]=.757 p=.440	F[1.367,15.041]=1.630 p=.227
AgreementType x Anteriority	**F[1,11]=6.320 p=.029***	F[1,11]=2.624 p=.134
Laterality x Anteriority	**F[2,22]=3.090 p=.066†**	**F[2,22]=5.351 p=.013***
AgreementType	F[1,11]=.006 p=.939	F[1,11]=.878 p=.369
Laterality	**F[2,22]=9.737 p=.001***	**F[2,22]=3.422 p=.061†**
Anteriority	**F[1,11]=3.535 p=.087†**	**F[1,11]=10.693 p=.007***

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

8.3.2.1 150-250ms

In the 150-250ms time window, the 2 x 3 x 2 repeated-measures ANOVA found no interactions involving the factors of AgreementType (Number, Gender), Laterality (Left, Central, Right), or Anteriority (Anterior, Posterior), nor was there a main effect of Agreement Type.

8.3.2.2 250-450ms

Results were similar in the 250-450ms time window, where no main effect of AgreementType and no interactions involving AgreementType were found.

8.3.2.3 450-950ms

In the 450-950ms time window, there was a significant AgreementType x Laterality x Anteriority interaction, $F(2,22)=11.229$, $p<.001$. Follow-up analysis in the anterior electrode regions revealed no interaction and no main effect of AgreementType, $F(1,11)=.002$, $p=.968$. In posterior regions, a significant AgreementType x Laterality interaction was present,

$F(1.425,15.670)=5.946$, $p=.019$. In post-hoc tests conducted separately on each posterior region, no effects of AgreementType were present.

8.3.2.3.1 450-700ms

In the early P600 time window (450-700ms), the AgreementType x Laterality x Anteriority interaction was significant, $F(2,22)=7.722$, $p=.003$. Follow-up analysis in the anterior electrode regions revealed no interaction and no main effect of AgreementType, $F(1,11)=.270$, $p=.614$. In posterior regions, a significant AgreementType x Laterality interaction was also present, $F(1.388,15.268)=4.991$, $p=.031$, but no main effect of AgreementType. No effects of AgreementType were present in post-hoc tests for the AgreementType x Laterality interaction.

8.3.2.3.2 700-950ms

In the late P600 time window (700-950ms), the AgreementType x Laterality x Anteriority interaction was significant, $F(2,22)=8.650$, $p=.002$. Follow-up analysis in the anterior electrode regions revealed no interaction and no main effect of AgreementType, $F(1,11)=.168$, $p=.690$. In posterior regions, a significant AgreementType x Laterality interaction was present, $F(1.342,14.763)=5.277$, $p=.028$. Post-hoc tests revealed only a marginal effect of AgreementType in the Central Posterior region, $F(1,11)=3.448$, $p=.090$, where the difference in mean amplitudes for grammatical and ungrammatical stimuli was marginally greater for number agreement on Adjectives than for gender agreement.

8.3.2.4 Summary

Responses to number and gender agreement on Adjectives were statistically similar, in that very few effects or interactions involving AgreementType were significant in any time window. While there was a significant AgreementType x Laterality x Anteriority interaction present after 450ms, the only effect of AgreementType present in post-hoc tests appeared as a marginal difference in the central electrodes in the late P600 time window (700-950ms).

8.4 Summary of All Results for Native Speakers

Unsurprisingly, native speakers were sensitive to violations of all three types of agreement presented in the study. Effects of Grammaticality were statistically significant in both the grammaticality judgments made by native speakers and in neurophysiological responses to stimuli, where a P600 was evident for all three types of agreement. The P600 for violations of number agreement on verbs was broadly distributed across anterior and posterior electrodes in the early P600 time window (450-700ms) but more posterior in the late P600 time window (700-950ms), consistent with a more frontal early-stage P600 (Friederici et al., 2002; Kaan and Swaab, 2003; Molinaro et al., 2011). It was preceded by a narrowly-distributed positivity in the Left Posterior region in early time windows (150-250ms, 250-450ms) that does not seem to correspond to any syntax-related ERP components. The P600 response to number violations on Adjectives was not as broadly distributed, but a more frontal P600 in the early time window is also demonstrated in response to gender violations, although the effect in the anterior electrodes was marginal. Even though the P600 response to gender violations was marginally more broadly distributed than the response to number violations in the early time window, it was less broadly distributed in the late time window, where there was an effect for gender agreement only in the Central and Right Posterior regions versus in all posterior regions for both types of number agreement. Finally, tests for effect of AgreementType resulted in marginal but persistent differences for Subject-Verb agreement in comparison with Noun-Adjective Number agreement across all time windows, confirming that response for number violations on verbs was more broadly distributed. A narrow and marginal difference was also found for number and gender agreement on adjectives. These results are summarized in Table 31 below, followed by a summary of the region-by-region analysis in Table 32.

Table 31. Summary of ERP results for native speakers across agreement types.

<table>
<tr><th></th><th colspan="2">S-V</th><th colspan="2">N-Adj Number</th><th colspan="2">N-Adj Gender</th></tr>
<tr><td rowspan="2">150-250ms</td><td colspan="2">Positivity
(LP)</td><td colspan="2">-----</td><td colspan="2">-----</td></tr>
<tr><td colspan="3">Greater for S-V (LP)</td><td colspan="3">No difference</td></tr>
<tr><td rowspan="2">250-450ms</td><td colspan="2">Positivity
(LP)</td><td colspan="2">-----</td><td colspan="2">Marginal positivity
(Central)</td></tr>
<tr><td colspan="3">Greater for S-V (RA†, LP)</td><td colspan="3">No difference</td></tr>
<tr><td rowspan="2">450-950ms</td><td colspan="2">P600
(RA†, LP, CP, RP)</td><td colspan="2">P600
(LP†, CP, RP)</td><td colspan="2">P600
(Central, Right)</td></tr>
<tr><td colspan="3">Marginally greater for S-V (LP†)</td><td colspan="3">No difference</td></tr>
<tr><td rowspan="2">a. 450-700ms</td><td colspan="2">P600
(RA†, CA†, LP, CP, RP)</td><td colspan="2">P600
(CP, RP)</td><td colspan="2">P600
(Central, Right)</td></tr>
<tr><td colspan="3">Marginally greater for S-V (LP†)</td><td colspan="3">No difference</td></tr>
<tr><td rowspan="2">b. 700-950ms</td><td colspan="2">P600
(LP, CP, RP)</td><td colspan="2">P600
(LP, CP, RP)</td><td colspan="2">P600
(Posterior)</td></tr>
<tr><td colspan="3">Marginally greater for S-V (LP†)</td><td colspan="3">Marginally greater for NUM (CP†)</td></tr>
</table>

† Correlation is marginally significant (.10>*p*>.05). All other correlations are significant (*p*<.05).

Table 32. Summary of region-by-region analysis for native speakers across agreement types.

	S-V	**N-Adj Number**	**N-Adj Gender**
150-250ms	LP	-----	-----
250-450ms	LP	LP†	CA†
450-950ms	RA†, LP, CP, RP	LP†, CP, RP	CP, RP
a. 450-700ms	RA†, CA†, LP, CP, RP	CP, RP	LA†, RA†, CA, CP, RP
b. 700-950ms	LP, CP, RP	LP, CP, RP	CP, RP

† Correlation is marginally significant (.10>*p*>.05). All other correlations are significant (*p*<.05).

CHAPTER 9:

LEARNER SENSITIVITY TO AGREEMENT VIOLATIONS

In this chapter, results for the learner group are presented, first in terms of the behavioral responses in the grammaticality judgment task, where both Grammaticality effects and differences between types of agreement are analyzed. Following that, the ERP responses to the three types of agreement are reported, along with tests for effects of Grammaticality and AgreementType. A summary of all results for the learners is also presented.

9.1 Behavioral Results

9.1.1 Grammaticality Judgments

When acceptance rates for each condition in the grammaticality judgment task were calculated for learners, generally high acceptance rates were present for both grammatical conditions, with lower acceptance rates for ungrammatical conditions. For Subject-Verb agreement, the average acceptance rates were 88% (SD=10.68; Range=55-100) for grammatical items and 17% (SD=19.52; Range=0-77.5) for ungrammatical items. Learners accepted 91% (SD=10.75; Range=60-100) of sentences in the condition involving grammatical number and gender agreement on Adjectives (N-Adj) and averaged 18% (SD=24.76; Range=0-92.5) and 34% (SD=25.35; Range=2.5-87.5) acceptance rates for number (*N-Adj NUM) and gender (*N-Adj GEN) violations, respectively. The broad ranges present for each agreement type, particularly with regard to acceptance of sentences containing violations, indicate individual variability, with some learners demonstrating low sensitivity to violations while others perform quite well. Indeed, for the ungrammatical Subject-Verb stimuli, 13 out of 24 learners demonstrated acceptance rates within the range of the native speakers. For the ungrammatical Noun-Adjective Number condition, 15 learners demonstrated acceptance rates within the range of the native speakers, and there were seven learners with acceptance rates within the range of native speakers for the ungrammatical Noun-Adjective Gender condition. The mean acceptance rates for learners in each condition are presented in Figure 8[11].

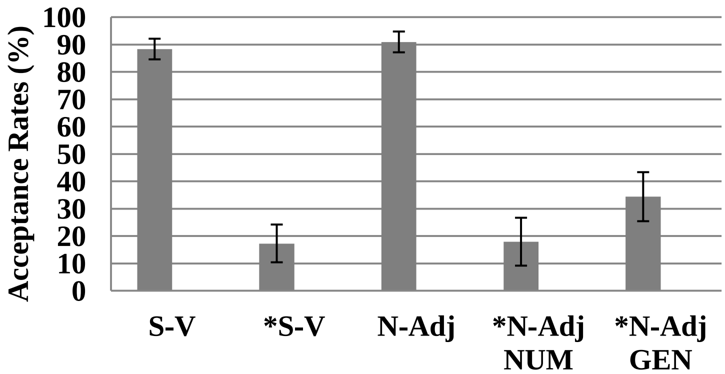

Figure 8. Mean acceptance rates for learners (L2) across conditions.

9.1.2 Analysis of Variance

9.1.2.1 Grammaticality Effects

Similarly to the analysis for native speakers described above, learner sensitivity to violations was tested using a repeated-measures analysis of variance (ANOVA) with Grammaticality as the within-subjects factor, with separate analyses of acceptance rates for each ungrammatical condition against its grammatical counterpart. Grammaticality effects were found for all three types of violations. First, when the grammatical Subject-Verb agreement condition (S-V) was compared to the ungrammatical condition (*S-V), there was a significant effect, $F(1,23)=162.690$, $p<.001$. Similarly, when the grammatical N-Adj and ungrammatical N-Adj Number conditions were compared, a significant effect for Grammaticality was also present, $F(1,23)=118.503$, $p<.001$. Finally, a Grammaticality effect was also found when the grammatical N-Adj and ungrammatical N-Adj Gender conditions were compared, $F(1,23)=77.169$, $p<.001$.

[11] See Appendix 5 for a table of means for individual learners.

The results of these analyses are presented in Table 33. In all cases, acceptance rates were higher for grammatical versus ungrammatical stimuli.

Table 33. Results of Repeated Measures ANOVAs on learners' acceptance rates for each type of agreement.

		S-V	N-Adj Number	N-Adj Gender
Grammaticality	F	162.690*	118.503*	77.169*
	(*p*)	(.000)	(.000)	(.000)

* Effect/interaction is significant at the p<.05 level.

9.1.2.2 Effects of Agreement Type

In order to directly compare learner sensitivity to number violations in the Subject-Verb versus the Noun-Adjective conditions (addressing research question 1), as well as differences in the N-Adj Number and Gender conditions (research question 2), d' scores for each participant were calculated to reflect the standardized differences in the average acceptance rates between each ungrammatical condition and its grammatical counterpart. Again, a d' score near zero represents performance at chance, while perfect performance in this analysis results in a d' score of approximately 4.0. The average d' scores for the learner group were 1.885 (SD=.931) for Subject-Verb agreement, 2.146 (SD=1.201) for N-Adj Number agreement and 1.515 (SD=.892) for N-Adj Gender agreement. Individual d' scores for learners are presented in Appendix 6. A series of repeated-measures ANOVAs was conducted on d' scores, with AgreementType as the within-subjects factor. When Subject-Verb and N-Adj Number agreement were compared, a marginal effect of AgreementType was present, $F(1,23)=3.917$, $p=.060$. Learners were slightly less sensitive to violations of the shared number feature on verbs (similar to the L1) than on adjectives (different than the L1), but this difference was not significant. When N-Adj Number and Gender agreement were compared, a significant effect of AgreementType was present, $F(1,23)=24.203$, $p<.001$, suggesting that the learners were significantly worse at detecting violations of gender agreement (unique to the L2) than number agreement on adjectives (shared feature).

9.1.3 Gender Assignment

In order to confirm that learners were aware of the appropriate gender category of each critical noun used in the experiment, a gender assignment task was administered following the experimental task. Learners correctly identified the masculine or feminine determiner for critical nouns used in the experiment, responding with a mean accuracy rate of 98% (SD=3.38) and a range of 88-100%. Thus, it was assumed that overall errors observed in the gender condition in the experimental task were not due to inability on the part of the learners to correctly assign gender to the particular nouns used in the stimuli. Since participants also had to be able to assign the appropriate gender under the same conditions of timing and presentation of stimuli in the grammaticality judgment task, the conditions of the experimental task were also ruled out as a factor in learner performance.

9.1.4 Vocabulary Recognition

Following the gender assignment task, learners were tested on their knowledge of the adjectives and verbs used in the predicates in the experimental sentences. During this vocabulary recognition task, participants had to choose between two English meanings for the word presented in Spanish. The mean accuracy rate for all items, including both adjectives and verbs, was 92% (SD=5.09), with a range of 80-98%. The mean was only slightly higher for verbs (Mean=93%, SD=4.91, range=80-100) than for adjectives (Mean=91%, SD=7.34, range=77-100). It was concluded that learners were sufficiently familiar with the vocabulary used in the experimental task.

9.1.5 Summary

In summary, learners generally showed a pattern of high acceptance rates for grammatical sentences and significantly lower acceptance rates for ungrammatical sentences, demonstrating the ability of many learners to discriminate between grammatical and ungrammatical stimuli across all three types of agreement. While there was only a marginal difference in learner sensitivity to the two types of number agreement, learners were slightly more sensitive to number violations on adjectives versus verbs. However, over half of the learners were within the range of the native speakers with regard to rejecting ungrammatical number in both conditions. Finally, while learners were significantly less sensitive to gender than number agreement on adjectives, they still showed a grammaticality effect for the gender conditions, and almost 1/3 of the learners (7 out of 24) performed within the range of the native speakers in rejecting ungrammatical gender.

9.2 ERP Responses to Agreement Violations

In order to investigate learners' sensitivity to agreement violations in terms of brain responses, ERP data for the time windows selected (150-250ms, 250-450ms, 450-950ms, 450-700ms, and 700-950ms) were submitted to the same analysis presented above for native speakers. A 2 x 3 x 2 repeated-measures Analysis of Variance (ANOVA) was conducted, with the within-subjects factors of Grammaticality (Grammatical, Ungrammatical), Laterality (Left, Central, Right), and Anteriority (Anterior, Posterior). Again, significant ($p<.05$) and marginal ($.050 < p < .100$) effects and interactions are reported only if they involve Grammaticality, and lower-level interactions and main effects are generally not interpreted in the presence of a significant higher-level interaction. Post-hoc tests for interpreted effects are reported. Additionally, where necessary to further investigate effects that may be just emerging in this learner group, region-by-region analyses are also presented. In all analyses, Greenhouse-Geisser corrected *p*-values are reported where violations of sphericity were present.

9.2.1 Number Agreement on Verbs

Grand-averaged ERP responses for learners in the conditions involving Subject-Verb agreement are plotted in Figure 9 at representative electrodes in each region of interest. Visual inspection revealed that the waveforms for number violations on verbs diverged slightly from their grammatical counterparts in at least the 450-950ms time window. This potential positivity was broadly-distributed and peaked at roughly 650ms, with an onset at approximately 350ms and an offset at approximately 950ms.

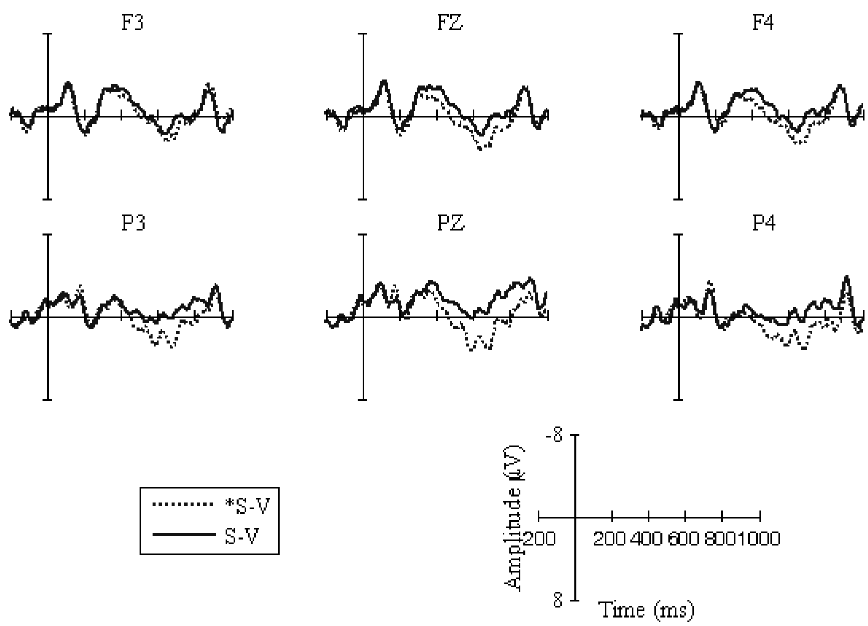

Figure 9. Grand average ERPs for learners in response to grammatical (solid line) and ungrammatical (dashed line) Subject-Verb agreement at representative electrodes for each region of interest.

The omnibus ANOVA results for AgreementType are presented in Table 34 for the 150-250ms, 250-450ms, and 450-950ms time windows. Table 35 presents the results of the same analysis in the early and late P600 time windows, 450-700ms and 700-950ms, respectively. Of

Table 34. Results of repeated-measures ANOVAs on learners' mean amplitudes for Subject-Verb agreement in each of the three primary time windows of analysis.

	150-250ms	**250-450ms**	**450-950ms**
Grammaticality x Laterality x Anteriority	$F[1.664,38.276]=.319$ $p=.689$	$F[1.631,37.510]=.225$ $p=.755$	**$F[2,46]=3.181$ $p=.051$†**
Grammaticality x Laterality	$F[1.359,31.262]=.729$ $p=.440$	**$F[2,46]=3.522$ $p=.038$***	**$F[1.516,34.865]=5.543$ $p=.014$***
Grammaticality x Anteriority	$F[1,23]=.974$ $p=.334$	$F[1,23]=.256$ $p=.618$	$F[1,23]=1.674$ $p=.209$
Laterality x Anteriority	$F[2,46]=2.370$ $p=.105$	**$F[2,46]=15.402$ $p<.001$***	**$F[2,46]=9.376$ $p<.001$***
Grammaticality	$F[1,23]=.010$ $p=.920$	$F[1,23]=.699$ $p=.412$	**$F[1,23]=9.050$ $p=.006$***
Laterality	**$F[2,46]=2.517$ $p=.092$†**	**$F[1.566,36.021]=5.875$ $p=.010$***	$F[1.427,32.810]=.443$ $p=.579$
Anteriority	**$F[1,23]=20.497$ $p<.001$***	$F[1,23]=2.247$ $p=.147$	$F[1,23]=.005$ $p=.943$

* Effect/interaction is significant at the $p<.05$ level.
† Effect is marginally significant ($.10>p>.05$).

Table 35. Results of the 2 x 3 x 2 Repeated Measures ANOVAs on learners' mean amplitudes for Subject-Verb agreement in the early and late P600 time windows.

	450-700ms	**700-950ms**
Grammaticality x Laterality x Anteriority	**$F[2,46]=3.006$ $p=.059$†**	**$F[2,46]=2.685$ $p=.081$†**
Grammaticality x Laterality	**$F[1.641,37.742]=6.262$ $p=.007$***	**$F[1.444,33.221]=3.976$ $p=.040$***
Grammaticality x Anteriority	$F[1,23]=1.015$ $p=.324$	$F[1,23]=2.048$ $p=.166$
Laterality x Anteriority	**$F[2,46]=8.428$ $p=.001$***	**$F[2,46]=7.943$ $p=.001$***
Grammaticality	**$F[1,23]=7.801$ $p=.010$***	**$F[1,23]=6.957$ $p=.015$***
Laterality	$F[1.605,36.917]=1.362$ $p=.265$	$F[1.420,32.667]=1.785$ $p=.190$
Anteriority	$F[1,23]=.065$ $p=.801$	$F[1,23]=.084$ $p=.774$

* Effect/interaction is significant at the $p<.05$ level.
† Effect is marginally significant ($.10>p>.05$).

primary interest are the marginal Grammaticality x Laterality x Anteriority interactions that were present in the P600 time windows, as well as significant interactions of Grammaticality and Laterality from 250-450ms and in the P600 time windows. Analyses of each time window will be explored below. In addition, where more limited effects may have been emerging for this group of learners, a region-by-region analysis of Grammaticality effects is also included.

9.2.1.1 150-250ms

Table 36 displays mean amplitudes in the 150-250ms time window in each electrode region for native speakers in the grammatical and ungrammatical Subject-Verb conditions (S-V and *S-V, respectively). An analysis of mean amplitudes using a three-way Repeated Measures ANOVA revealed no interactions with Grammaticality and no main effect of Grammaticality. Similarly, a region-by-region analysis showed no effects of Grammaticality.

Table 36. L2 mean amplitudes in the grammatical and ungrammatical Subject-Verb conditions for each region of interest from 150-250ms.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
S-V	Avg	1.164	.950	.432	-.614	-.660	-.889
	(SD)	(2.340)	(2.621)	(2.264)	(1.793)	(2.083)	(1.637)
***S-V**	Avg	1.168	1.225	.598	-1.017	-.819	-.990
	(SD)	(2.687)	(2.772)	(2.250)	(2.454)	(2.765)	(1.906)

9.2.1.2 250-450ms

Mean amplitudes are displayed in Table 37 for learners in the Subject-Verb conditions during the 250-450ms time window in each electrode region. For this time window, the three-way Repeated Measures ANOVA revealed a significant Grammaticality x Laterality interaction, $F(2,46)=3.522$, $p=.038$. Follow-up analysis was performed for each level of Laterality (left, central, right), but no effects of Grammaticality were found. The same was true in an analysis of independent regions that revealed no effects of Grammaticality.

Table 37. L2 mean amplitudes in the grammatical and ungrammatical Subject-Verb conditions for each region of interest from 250-450ms.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
S-V	Avg	-1.492	-1.662	-1.656	-.700	-1.741	-.025
	(SD)	(2.921)	(3.049)	(2.129)	(1.912)	(2.262)	(1.608)
***S-V**	Avg	-1.419	-1.014	-1.086	-.820	-1.291	.189
	(SD)	(2.508)	(2.664)	(2.219)	(2.586)	(3.261)	(2.112)

9.2.1.3 450-950ms

Table 38 displays mean amplitudes for the L2 group in the 450-950ms time window for each electrode region in the Subject-Verb conditions. The three-way Repeated Measures ANOVA revealed a marginal Grammaticality x Laterality x Anteriority interaction, $F(2,46)=3.181$, $p=.051$, as well as a Grammaticality x Laterality interaction, $F(1.516,34.865)=5.543$, $p=.014$. Follow-up analysis for the marginal 3-way interaction was conducted separately for the anterior and posterior electrodes. For the anterior electrodes, a significant Grammaticality x Laterality interaction was present, $F(1.483,34.102)=4.006$, $p=.038$. Separate post-hoc tests for the Grammaticality x Laterality interaction in each anterior region (left, central, right) revealed a marginal effect of Grammaticality in the Right Anterior region, $F(1,23)=3.790$, $p=.064$, where ungrammatical stimuli yielded slightly more positive mean amplitudes than grammatical stimuli. This positivity was also present across the posterior electrodes, where follow-up analysis revealed a significant Grammaticality x Laterality interaction, $F(2,46)=6.094$, $p=.004$, as well as a main effect of Grammaticality, $F(1,11)=10.061$, $p=.004$. Post-hoc tests revealed a main effect of Grammaticality in all three posterior regions (Left Posterior: $F(1,23)=5.831$, $p=.024$; Central Posterior: $F(1,23)=12.584$, $p=.002$; Right Posterior: $F(1,23)=8.186$, $p=.009$), reflecting a significantly greater positivity in response to ungrammatical versus grammatical stimuli in the posterior regions.

Table 38. L2 mean amplitudes in the grammatical and ungrammatical Subject-Verb conditions for each region of interest from 450-950ms.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
S-V	Avg	.071	.060	-.628	-.334	-1.031	-.395
	(SD)	(2.201)	(2.399)	(1.907)	(1.529)	(1.799)	(1.347)
***S-V**	Avg	.104	.748	.130	.558	.740	.778
	(SD)	(2.316)	(2.459)	(2.301)	(2.105)	(2.739)	(1.907)

9.2.1.3.1 450-700ms

As in the previous analyses for the native speakers, the P600 time window for learners was further divided into an early and a late window for analysis, at 450-700ms and 700-950ms, respectively. Learners' mean amplitudes for the Subject-Verb conditions in the early P600 time window (450-700ms) are displayed in Table 39. Analysis of means using a three-way repeated-measures ANOVA revealed a marginal Grammaticality x Laterality x Anteriority interaction, $F(2,46)=3.006$, $p=.059$. Follow-up analysis of the anterior electrodes revealed a Grammaticality x Laterality interaction, $F(1.400,32.203)=4.568$, $p=.029$. Separate post-hoc tests for each anterior region (left, central, right) revealed a significant effect of Grammaticality only in the Right Anterior region, $F(1,23)=4.848$, $p=.038$, indicating that the positivity observed only marginally in the Right Anterior region over the entire P600 time window was stronger and more broadly distributed in the earlier portion of that window from 450-700ms, mirroring the performance of native speakers. Follow-up analysis in the posterior regions revealed a significant Grammaticality x Laterality interaction, $F(2,46)=6.087$, $p=.005$, along with a significant Grammaticality effect, $F(1,23)=9.773$, $p=.005$. Post-hoc tests revealed a main effect of Grammaticality in all three posterior regions that was also similar to that of native speakers (Left Posterior: $F(1,23)=6.589$, $p=.017$; Central Posterior: $F(1,23)=12.073$, $p=.002$; Right Posterior:

F(1,23)=7.179, *p*=.013), where ERPs to ungrammatical versus grammatical stimuli were also significantly more positive.

Table 39. L2 mean amplitudes in the grammatical and ungrammatical Subject-Verb conditions for each region of interest from 450-700ms.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
S-V	Avg	.313	.513	-.133	.050	-.380	.263
	(SD)	(2.853)	(3.040)	(2.084)	(1.903)	(2.068)	(1.422)
***S-V**	Avg	.499	1.497	.797	1.082	1.561	1.433
	(SD)	(2.312)	(2.415)	(2.002)	(2.530)	(3.416)	(2.198)

9.2.1.3.2 700-950ms

Table 40 displays mean amplitudes in the late P600 time window (700-950ms) in each electrode region for native speakers in the grammatical and ungrammatical Subject-Verb conditions. The three-way Repeated Measures ANOVA revealed that the significant Grammaticality x Laterality x Anteriority interaction was still marginally present, *F*(2,46)=2.685, *p*=.081. Follow-up analysis of the anterior electrodes revealed a marginal Grammaticality x Laterality interaction, *F*(1.578,36.303)=2.695, *p*=.092, but no effect of Grammaticality, and separate post-hoc tests for each Anteriority (left, central, right) revealed no effects of Grammaticality. Follow-up analysis in the posterior regions revealed a significant Grammaticality x Laterality interaction, *F*(2,46)=5.086, *p*=.010, and a significant main effect of Grammaticality, *F*(1,11)=14.872, *p*=.003. Post-hoc tests revealed a marginal effect of Grammaticality in the Left Posterior region, *F*(1,23)=3.636, *p*=.069, while significant effects of Grammaticality were present in the Central Posterior region, *F*(1,23)=9.621, *p*=.005, and in the Right Posterior region, *F*(1,23)=7.456, *p*=.012, where ERPs to ungrammatical stimuli were significantly more positive than to grammatical stimuli.

Table 40. L2 mean amplitudes in the grammatical and ungrammatical Subject-Verb conditions for each region of interest from 700-950ms.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
S-V	Avg	-.171	-.391	-1.122	-.718	-1.681	-1.053
	(SD)	(2.385)	(2.723)	(2.504)	(1.703)	(2.210)	(1.722)
***S-V**	Avg	-.286	.005	-.533	.038	-.074	.126
	(SD)	(2.712)	(3.000)	(2.976)	(2.229)	(2.778)	(2.119)

9.2.1.4 Summary

With regard to Subject-Verb agreement, learners demonstrated electrophysiological sensitivity to agreement violations. While no effects of Grammaticality were found in the early time windows, a late positivity similar to that of native speakers was present for learners as well. This positivity was broadly distributed in both hemispheres in the 450-700ms time window and more posteriorly distributed from 700-950ms.

9.2.2 Number Agreement on Adjectives

Learners' grand-averaged ERP responses for conditions involving Noun-Adjective Number agreement are plotted in Figure 10 at representative electrodes in each region of interest. Visual inspection reveals that number violations on Adjectives also yielded more positive waveforms than their grammatical counterparts in the 450-950ms time window. This broadly-distributed positivity lasts from approximately 450-800ms in the anterior regions, beginning earlier and ending later in the posterior regions, roughly from 350-900ms, with a peak at roughly 600ms.

Mean amplitudes for the grammatical and ungrammatical Noun-Adjective Number conditions were submitted to a 2 x 3 x 2 repeated-measures ANOVA for each time window of interest, with within-subject factors of Grammaticality (Grammatical, Ungrammatical), Laterality (Left, Central, Right), and Anteriority (Anterior, Posterior). The omnibus ANOVA

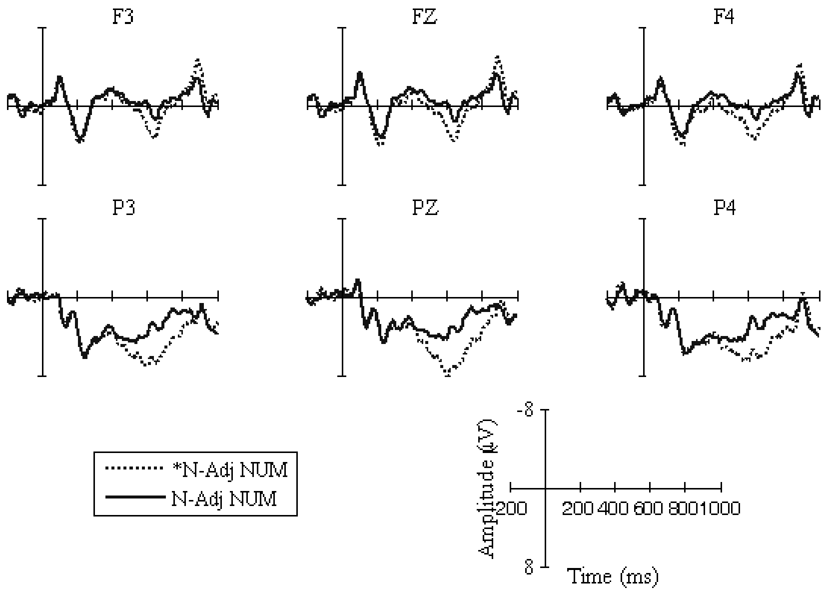

Figure 10. Grand average ERPs for native speakers in response to grammatical (solid line) and ungrammatical (dashed line) Noun-Adjective Number agreement at representative electrodes for each region of interest.

results are presented in Table 41 for the 150-250ms, 250-450ms, and 450-950ms time windows. Table 42 presents the results of the same analysis in the early and late P600 time windows, 450-700ms and 700-950ms, respectively. Fewer interactions were revealed in the Noun-Adjective Number agreement conditions, probably due to the fact that the P600 response was so broadly distributed. In addition, for each time window, the possibility of limited effects is investigated through a region-by-region analysis of Grammaticality effects.

9.2.2.1 150-250ms

Table 43 displays mean amplitudes in the 150-250ms time window in each electrode region for learners in the grammatical and ungrammatical conditions involving Noun-Adjective number agreement (N-Adj and *N-Adj NUM, respectively). An analysis of mean amplitudes using a three-way Repeated Measures ANOVA revealed a marginal interaction of

Table 41. Results of repeated-measures ANOVAs on learners' mean amplitudes for Noun-Adjective Number agreement in each primary time window of analysis.

	150-250ms	**250-450ms**	**450-950ms**
Grammaticality x Laterality x Anteriority	*F*[2,46]=.542, *p*=.585	*F*[1.608,36.975]=2.076, *p*=.148	*F*[2,46]=1.386, *p*=.260
Grammaticality x Laterality	*F*[2,46]=1.018, *p*=.369	*F*[2,46]=1.563, *p*=.220	*F*[2,46]=2.062, *p*=.139
Grammaticality x Anteriority	***F*[1,23]=3.016, *p*=.096†**	*F*[1,23]=.421, *p*=.523	*F*[1,23]=2.355, *p*=.139
Laterality x Anteriority	*F*[2,46]=.279, *p*=.758	***F*[2,46]=3.030, *p*=.058***	***F*[2,46]=3.279, *p*=.047***
Grammaticality	*F*[1,23]=1.808, *p*=.192	*F*[1,23]=1.145, *p*=.296	***F*[1,23]=11.810, *p*=.002***
Laterality	***F*[1.672,38.456]=3.697, *p*=.041***	***F*[2,46]=5.495, *p*=.007***	*F*[2,46]=.820, *p*=.447
Anteriority	*F*[1,23]=.126, *p*=.726	***F*[1,23]=23.177, *p*<.001***	***F*[1,23]=61.145, *p*<.001***

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

Table 42. Results of repeated-measures ANOVAs on learners' mean amplitudes for Noun-Adjective Number agreement in the early and late P600 time windows.

	450-700ms	**700-950ms**
Grammaticality x Laterality x Anteriority	*F*[2,46]=1.126, *p*=.333	*F*[2,46]=1.382, *p*=.261
Grammaticality x Laterality	***F*[2,46]=5.859, *p*=.005***	*F*[2,46]=.654, *p*=.525
Grammaticality x Anteriority	*F*[1,23]=.915, *p*=.349	***F*[1,23]=3.423, *p*=.077†**
Laterality x Anteriority	***F*[2,46]=2.499, *p*=.093†**	***F*[2,46]=3.071, *p*=.056†**
Grammaticality	***F*[1,23]=20.381, *p*<.001***	*F*[1,23]=2.642, *p*=.118
Laterality	***F*[2,46]=2.617, *p*=.084†**	*F*[1.667,38.330]=.254, *p*=.737
Anteriority	***F*[1,23]=41.173, *p*<.001**	***F*[1,23]=62.288, *p*<.001***

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

Grammaticality and Anteriority, $F(1,23)=3.016$, $p=.096$, but no main effect of Grammaticality. Follow-up tests at each level of Anteriority (anterior, posterior) revealed no effects of Grammaticality. The region-by-region analysis showed only a marginal effect of Grammaticality in the Central Anterior region, $F(1,23)=3.287$, $p=.083$.

Table 43. L2 mean amplitudes from 150-250ms for each region of interest in the conditions involving grammatical and ungrammatical Noun-Adjective agreement with regard to Number.

		Left Anterior	**Central Anterior**	**Right Anterior**	**Left Posterior**	**Central Posterior**	**Right Posterior**
N-Adj	Avg	1.909	2.167	1.557	2.421	2.557	2.079
	(SD)	(2.361)	(2.965)	(2.846)	(2.185)	(2.735)	(2.277)
***N-Adj NUM**	Avg	2.529	3.079	2.252	2.389	2.820	2.337
	(SD)	(2.775)	(3.412)	(2.843)	(1.944)	(1.938)	(1.749)

9.2.2.2 250-450ms

Mean amplitudes are displayed in Table 44 for native speakers in the grammatical and ungrammatical conditions involving Noun-Adjective number agreement during the 250-450ms time window in each electrode region. For this time window, the three-way Repeated Measures ANOVA again revealed no effect of Grammaticality, nor any interactions with Grammaticality. Similarly, no Grammaticality effects were present in the analysis of independent regions.

Table 44. L2 mean amplitudes from 250-450ms for each region of interest in the conditions involving grammatical and ungrammatical Noun-Adjective agreement with regard to Number.

		Left Anterior	**Central Anterior**	**Right Anterior**	**Left Posterior**	**Central Posterior**	**Right Posterior**
N-Adj	Avg	-.925	-.717	-.411	2.846	2.357	3.502
	(SD)	(2.692)	(3.463)	(2.982)	(2.781)	(3.565)	(2.807)
***N-Adj NUM**	Avg	-.911	-.226	.226	2.984	2.502	3.688
	(SD)	(2.613)	(3.467)	(2.923)	(2.723)	(3.668)	(2.780)

9.2.2.3 450-950ms

Table 45 displays mean amplitudes in the 450-950ms time window in each electrode region for learners in the Noun-Adjective conditions involving number agreement. The three-way Repeated Measures ANOVA revealed a significant main effect of Grammaticality, $F(1,23)=11.810$, $p=.002$, where responses to ungrammatical stimuli were more positive than those to grammatical stimuli overall. The region-by-region analysis revealed that this positivity was present only in the posterior electrodes (Left Posterior: $F(1,23)=20.499$, $p<.001$; Central Posterior: $F(1,23)=19.220$, $p<.001$; Right Posterior: $F(1,23)=20.138$, $p<.001$).

Table 45. L2 mean amplitudes from 450-950ms for each region of interest in the conditions involving grammatical and ungrammatical Noun-Adjective agreement with regard to Number.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
N-Adj	Avg	-.768	-.658	-.482	2.158	2.108	2.091
	(SD)	(2.174)	(2.618)	(1.884)	(1.663)	(2.284)	(1.726)
***N-Adj NUM**	Avg	-.381	.020	.279	3.426	3.746	3.390
	(SD)	(2.102)	(2.593)	(1.931)	(2.041)	(2.522)	(1.727)

9.2.2.3.1 450-700ms

The late positivity observed for learners in the overall P600 time window for Noun-Adjective Number violations was further investigated for early and late time windows. Mean amplitudes for the grammatical and ungrammatical conditions from 450-700ms are displayed in Table 46. The three-way Repeated Measures ANOVA revealed a significant Grammaticality x Laterality interaction, $F(2,46)=5.859$, $p=.005$, along with a significant main effect of Grammaticality, $F(1,23)=20.381$, $p<.001$. Follow-up analysis revealed an effect of Grammaticality in all three levels of Laterality (Left: $F(1,23)=14.178$, $p=.001$; Central: $F(1,23)=22.502$, $p<.001$; Right: $F(1,23)=17.932$, $p<.001$), where ERPs to ungrammatical versus grammatical stimuli were significantly more positive. The region-by-region analysis revealed that this positivity was marginal in the Left Anterior electrodes, $F(1,23)=3.971$, $p=.058$, and significant in all other regions (Central Anterior: $F(1,23)=7.383$, $p=.012$; Right Anterior:

$F(1,23)=7.770$, $p=.010$; Left Posterior: $F(1,23)=18.861$, $p<.001$; Central Posterior: $F(1,23)=26.429$, $p<.001$; Right Posterior: $F(1,23)=20.052$, $p<.001$).

Table 46. L2 mean amplitudes from 450-700ms for each region of interest in the conditions involving grammatical and ungrammatical Noun-Adjective agreement with regard to Number.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
N-Adj	Avg	-.410	-.163	.023	2.847	2.815	2.828
	(SD)	(2.512)	(3.233)	(2.555)	(2.119)	(2.944)	(2.249)
***N-Adj NUM**	Avg	.487	1.366	1.304	4.317	4.973	4.348
	(SD)	(2.113)	(2.714)	(2.007)	(2.811)	(3.399)	(2.302)

9.2.2.3.2 700-950ms

Table 47 displays mean amplitudes in the late P600 time window (700-950ms) in each electrode region for learners in response to grammatical and ungrammatical Noun-Adjective Number conditions. The analysis of mean amplitudes using a three-way Repeated Measures ANOVA revealed a marginal interaction of Grammaticality and Anteriority, $F(1,23)=3.423$, $p=.077$, but no main effect of Grammaticality. Follow-up tests for the anterior electrodes revealed no effect of Grammaticality, but this was not the case in the posterior electrodes, where there was a significant main effect of Grammaticality, $F(1,23)=14.720$, $p=.001$. The region-by-region analysis revealed that the Grammaticality effects were present in all three posterior regions (Left Posterior: $F(1,23)=14.701$, $p=.001$; Central Posterior: $F(1,23)=8.746$, $p=.007$; Right Posterior: $F(1,23)=13.887$, $p=.001$).

Table 47. L2 mean amplitudes from 700-950ms for each region of interest in the conditions involving grammatical and ungrammatical Noun-Adjective agreement with regard to Number.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
N-Adj	Avg	-1.124	-1.152	-.985	1.469	1.400	1.354
	(SD)	(2.097)	(2.405)	(1.711)	(1.613)	(2.113)	(1.503)
***N-Adj NUM**	Avg	-1.244	-1.322	-.743	2.537	2.523	2.436
	(SD)	(2.520)	(3.022)	(2.450)	(1.848)	(2.240)	(1.626)

9.2.2.4 Summary

Learners' ERP responses to violations of Noun-Adjective Number agreement demonstrated a significant positivity that was broadly distributed in the early P600 time window (450-700ms) and more posterior in the late P600 time window (700-950ms). No effects were present in the early time windows.

9.2.3 Gender Agreement on Adjectives

Finally, learners' grand-averaged ERP responses for conditions involving Noun-Adjective Gender agreement are plotted in Figure 11 at representative electrodes in each region of interest. Visual inspection revealed a possible divergence in the posterior electrodes, where waveforms for gender violations were more positive than their grammatical counterparts for a brief interval during the 450-950ms time window. This potential positivity for gender violations peaked at about 600ms, similar to the P600 responses to number violations on Adjectives presented above, but here the overall divergence lasted from only 550-900ms.

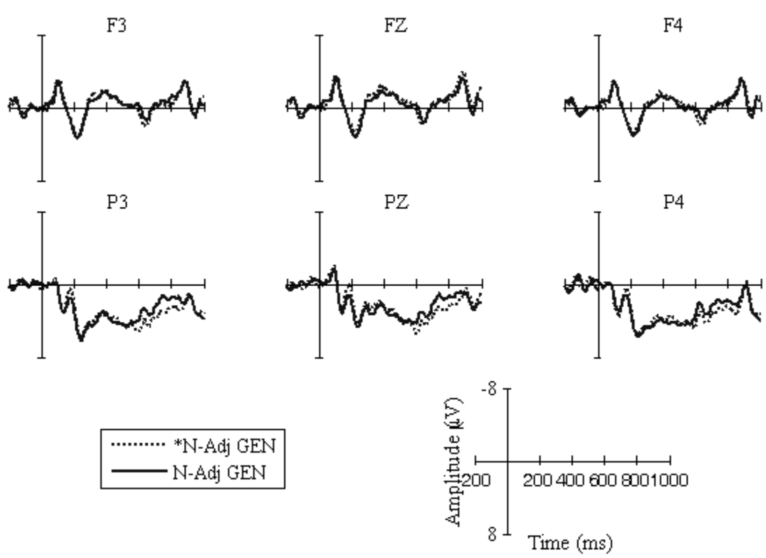

Figure 11. Grand average ERPs for native speakers in response to grammatical (solid line) and ungrammatical (dashed line) Noun-Adjective Gender agreement at representative electrodes for each region of interest.

The same series of 2 x 3 x 2 repeated-measures ANOVAs were conducted in each time window of interest on the mean amplitudes for the grammatical and ungrammatical Noun-Adjective Gender conditions. Again, within-subject factors included Grammaticality (Grammatical, Ungrammatical), Laterality (Left, Central, Right), and Anteriority (Anterior, Posterior). Results are presented in Table 48 for the 150-250ms, 250-450ms, and 450-950ms time windows. Table 49 presents the results for the early and late P600 time windows, 450-700ms and 700-950ms, respectively. As will be evident in the following sections, no interactions or effects of Grammaticality were found for gender violations on Adjectives. The analyses for each time window are followed by a region-by-region analysis of Grammaticality effects, in order to determine whether limited effects may have been present.

Table 48. Results of repeated-measures ANOVAs on learners' mean amplitudes for Noun-Adjective Gender agreement in each primary time window of analysis.

	150-250ms	**250-450ms**	**450-950ms**
Grammaticality x Laterality x Anteriority	$F[2,46]=.729$ $p=.488$	$F[2,46]=.180$ $p=.836$	$F[2,46]=.320$ $p=.727$
Grammaticality x Laterality	$F[2,46]=2.310$ $p=.111$	$F[2,46]=.624$ $p=.540$	$F[2,46]=.566$ $p=.545$
Grammaticality x Anteriority	$F[1,23]=.571$ $p=.458$	$F[1,23]=.030$ $p=.863$	$F[1,23]=.573$ $p=.457$
Laterality x Anteriority	$F[2,46]=.307$ $p=.737$	**$F[2,46]=3.588$ $p=.036$***	$F[2,46]=2.166$ $p=.126$
Grammaticality	$F[1,23]=1.301$ $p=.266$	$F[1,23]=.624$ $p=.438$	$F[1,23]=1.352$ $p=.257$
Laterality	$F[2,46]=1.323$ $p=.276$	**$F[2,46]=5.140$ $p=.010$***	$F[2,46]=.275$ $p=.761$
Anteriority	$F[1,23]=.600$ $p=.447$	**$F[1,23]=26.357$ $p<.001$***	**$F[1,23]=58.422$ $p<.001$***

* Effect/interaction is significant at the $p<.05$ level.

Table 49. Results of repeated-measures ANOVAs on learners' mean amplitudes for Noun-Adjective Gender agreement in the early and late P600 time windows.

	450-700ms	700-950ms
Grammaticality x Laterality x Anteriority	*F*[2,46]=.310 *p*=.735	*F*[2,46]=.379 *p*=.687
Grammaticality x Laterality	*F*[2,46]=.254 *p*=.777	*F*[2,46]=.720 *p*=.492
Grammaticality x Anteriority	*F*[1,23]=.134 *p*=.718	*F*[1,23]=1.070 *p*=.312
Laterality x Anteriority	*F*[2,46]=1.573 *p*=.218	*F*[2,46]=1.955 *p*=.153
Grammaticality	*F*[1,23]=.659 *p*=.425	*F*[1,23]=1.834 *p*=.189
Laterality	*F*[1.648,37.893]=.298 *p*=.702	*F*[2,46]=.544 *p*=.584
Anteriority	***F*[1,23]=36.330 *p*<.001***	***F*[1,23]=69.257 *p*<.001***

* Effect/interaction is significant at the *p*<.05 level.

9.2.3.1 150-250ms

Learners' mean amplitudes by electrode region in the 150-250ms time window for grammatical and ungrammatical conditions involving Noun-Adjective gender agreement (N-Adj and *N-Adj GEN, respectively) are displayed in Table 50. An analysis of mean amplitudes using a three-way Repeated Measures ANOVA revealed no effect of Grammaticality, nor any interactions with Grammaticality. However, the region-by-region analysis did show a significant effect of Grammaticality in the Central Posterior region, $F(1,23)=5.392$, $p=.029$, as well as a marginal effect in the Left Posterior electrodes, $F(1,23)=3.519$, $p=.073$. In both cases, the waveforms for the ungrammatical conditions are more negative than those for the grammatical stimuli.

Table 50. L2 mean amplitudes from 150-250ms for each region of interest in the conditions involving grammatical and ungrammatical Noun-Adjective agreement with regard to Gender.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
N-Adj	Avg	1.909	2.167	1.557	2.421	2.557	2.079
	(SD)	(2.361)	(2.965)	(2.846)	(2.185)	(2.735)	(2.277)
***N-Adj GEN**	Avg	1.756	1.822	1.510	1.980	1.892	1.921
	(SD)	(2.472)	(3.212)	(2.801)	(1.760)	(2.335)	(1.754)

9.2.3.2 250-450ms

Mean amplitudes are displayed in Table 51 for learners in the grammatical and ungrammatical conditions involving Noun-Adjective gender agreement during the 250-450ms time window in each electrode region. For this time window, the three-way Repeated Measures ANOVA again revealed no effect of Grammaticality, nor any interactions with Grammaticality. Additionally, no effects of Grammaticality were present in an analysis of independent regions.

Table 51. L2 mean amplitudes from 250-450ms for each region of interest in the conditions involving grammatical and ungrammatical Noun-Adjective agreement with regard to Gender.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
N-Adj	Avg	-.925	-.717	-.411	2.846	2.357	3.502
	(SD)	(2.692)	(3.463)	(2.982)	(2.781)	(3.565)	(2.807)
***N-Adj GEN**	Avg	-1.123	-1.094	-.624	2.787	2.016	3.289
	(SD)	(2.636)	(3.654)	(3.096)	(2.585)	(3.414)	(2.367)

9.2.3.3 450-950ms

Table 52 displays mean amplitudes in the 450-950ms time window in each electrode region for native speakers in the conditions involving gender agreement on Adjectives. The three-way Repeated Measures ANOVA revealed no interactions with Grammaticality, and no main effect of Grammaticality. However, the region-by-region analysis did reveal a significant effect of Grammaticality in the Left Posterior region, $F(1,23)=8.557$, $p=.008$.

Table 52. L2 mean amplitudes from 450-950ms for each region of interest in the conditions involving grammatical and ungrammatical Noun-Adjective agreement with regard to Gender.

		Left Anterior	**Central Anterior**	**Right Anterior**	**Left Posterior**	**Central Posterior**	**Right Posterior**
N-Adj	Avg	-.768	-.658	-.482	2.158	2.108	2.091
	(SD)	(2.174)	(2.618)	(1.884)	(1.663)	(2.284)	(1.726)
***N-Adj GEN**	Avg	-.426	-.624	-.279	2.697	2.528	2.519
	(SD)	(2.029)	(2.704)	(2.014)	(1.633)	(2.371)	(1.810)

9.2.3.3.1 450-700ms

Learners' mean amplitudes are displayed in Table 53 for the 450-700ms time window in each electrode region for conditions involving gender agreement on Adjectives. Just as in the overall P600 time window, the three-way Repeated Measures ANOVA revealed no effect of Grammaticality, nor any interactions with Grammaticality. No Grammaticality effects were present in an analysis of each independent region.

Table 53. L2 mean amplitudes from 450-700ms for each region of interest in the conditions involving grammatical and ungrammatical Noun-Adjective agreement with regard to Gender.

		Left Anterior	**Central Anterior**	**Right Anterior**	**Left Posterior**	**Central Posterior**	**Right Posterior**
N-Adj	Avg	-.410	-.163	.023	2.847	2.815	2.828
	(SD)	(2.512)	(3.233)	(2.555)	(2.119)	(2.944)	(2.249)
***N-Adj GEN**	Avg	-.069	-.060	.185	3.229	3.162	3.139
	(SD)	(2.541)	(3.337)	(2.741)	(2.057)	(2.981)	(2.126)

9.2.3.3.2 700-950ms

Learners' mean amplitudes are displayed in Table 54 for the 700-950ms time window in each electrode region for conditions involving gender agreement on Adjectives. The three-way Repeated Measures ANOVA revealed no effect of Grammaticality, nor any interactions with Grammaticality. The region-by-region analysis yielded a Grammaticality effect that was significant in the Left Posterior region, $F(1,23)=8.651$, $p=.007$, and marginal in the Right Posterior region, $F(1,23)=3.196$, $p=.087$. In both cases, waveforms for ungrammatical versus grammatical Noun-Adjective Gender agreement were more positive in the late P600 time window. This positivity was not present in the 450-700ms time window, indicating a late response to gender violations in comparison with number violations. (Although there was a numerical positivity in the Central Posterior electrodes, as can be seen in the table below, the difference in mean amplitudes did not reach significance for that region.)

Table 54. L2 mean amplitudes from 700-950ms for each region of interest in the conditions involving grammatical and ungrammatical Noun-Adjective agreement with regard to Gender.

		Left Anterior	**Central Anterior**	**Right Anterior**	**Left Posterior**	**Central Posterior**	**Right Posterior**
N-Adj	Avg	-1.124	-1.152	-.985	1.469	1.400	1.354
	(SD)	(2.097)	(2.405)	(1.711)	(1.613)	(2.113)	(1.503)
***N-Adj GEN**	Avg	-.781	-1.186	-.742	2.167	1.895	1.901
	(SD)	(2.016)	(2.635)	(1.648)	(1.853)	(2.412)	(1.793)

9.2.3.4 Summary

Sensitivity to gender agreement violations in this group of learners is emerging. Even though the P600 effect did not attain statistical significance in the omnibus ANOVA, the means in the P600 time windows for ungrammatical versus grammatical stimuli are numerically more positive, and the region-by-region analysis revealed Grammaticality effects that were significant in the Left Posterior region and marginal in the Right Posterior region. Additionally, the region-by-region analysis revealed a limited negativity in the 150-250ms time window.

9.3 Comparison of ERP Responses across Agreement Types

In order to statistically compare the two types of number agreement, as well as the two types of agreement occurring on Adjectives, a separate statistical analysis was performed for Agreement Type using the effect sizes (mean amplitudes for ungrammatical minus grammatical conditions) of the components in each time window. First, the two conditions involving number agreement (Subject-Verb, N-Adj Number) were compared in each time window using a 2 x 3 x 2 repeated-measures ANOVA with within-subject factors of AgreementType (S-V, N-Adj Number), Laterality (Left, Central, Right), and Anteriority (Anterior, Posterior). Secondly, the same analysis was performed in order to compare number and gender agreement on Adjectives (AgreementType: N-Adj Number and N-Adj Gender). Both analyses are presented below.

9.3.1 Number Agreement on Verbs versus Adjectives

When Subject-Verb and Noun-Adjective number agreement were compared, no effects of AgreementType and no interactions involving AgreementType were present in any time window, indicating that learners were equally sensitive to violations of number on verbs and adjectives. The omnibus ANOVA results for AgreementType are presented in Table 55 for the 150-250ms, 250-450ms, and 450-950ms time windows. Table 56 presents the results of the same analysis in the early and late P600 time windows, 450-700ms and 700-950ms, respectively.

Table 55. Results of repeated-measures ANOVAs comparing learners' mean difference amplitudes for number agreement on Verbs versus Adjectives in each primary time window.

	150-250ms	**250-450ms**	**450-950ms**
AgreementType x Laterality x Anteriority	$F[2,46]=.027$ $p=.973$	$F[1.669,38.393]=.938$ $p=.385$	$F[2,46]=.341$ $p=.713$
AgreementType x Laterality	$F[1.496,34.397]=.026$ $p=.944$	$F[1.662,38.226]=.576$ $p=.536$	$F[2,46]=1.221$ $p=.304$
AgreementType x Anteriority	$F[1,23]=.180$ $p=.675$	$F[1,23]=.001$ $p=.974$	$F[1,23]=.000$ $p=.989$
Laterality x Anteriority	$F[1.628,37.451]=.915$ $p=.391$	$F[1.570,36.118]=1.425$ $p=.252$	**$F[2,22]=3.169$ $p=.051$†**
AgreementType	$F[1,23]=.953$ $p=.339$	$F[1,23]=.009$ $p=.925$	$F[1,23]=.090$ $p=.767$
Laterality	$F[1.523,35.036]=2.077$ $p=.150$	**$F[2,46]=5.674$ $p=.006$***	**$F[1.641,37.738]=7.568$ $p=.003$***
Anteriority	**$F[1,23]=3.689$ $p=.067$†**	$F[1,23]=1.033$ $p=.320$	$F[1,23]=2.438$ $p=.132$

* Effect/interaction is significant at the $p<.05$ level.
† Effect is marginally significant ($.10>p>.05$).

Table 56. Results of repeated-measures ANOVAs comparing learners' mean difference amplitudes for number agreement on verbs versus adjectives in the early and late P600 time windows.

	450-700ms	**700-950ms**
AgreementType x Laterality x Anteriority	$F[2,46]=.505$ $p=.607$	$F[2,46]=.316$ $p=.731$
AgreementType x Laterality	$F[1.662,38.225]=.288$ $p=.711$	$F[1.672,38.452]=2.320$ $p=.120$
AgreementType x Anteriority	$F[1,23]=.120$ $p=.732$	$F[1,23]=.138$ $p=.714$
Laterality x Anteriority	**$F[2,22]=3.032$ $p=.058$†**	**$F[2,22]=2.722$ $p=.076$†**
AgreementType	$F[1,23]=.894$ $p=.354$	$F[1,23]=.192$ $p=.665$
Laterality	**$F[2,22]=12.719$ $p<.001$***	**$F[1.506,34.644]=3.296$ $p=.062$†**
Anteriority	$F[1,23]=1.244$ $p=.276$	**$F[1,23]=3.465$ $p=.075$†**

* Effect/interaction is significant at the $p<.05$ level.
† Effect is marginally significant ($.10>p>.05$).

9.3.2 Number versus Gender Agreement on Adjectives

In the same way, the number and gender violations on Adjectives were also compared. The omnibus ANOVA results for AgreementType are presented in Table 57 for the 150-250ms, 250-450ms, and 450-950ms time windows. Table 58 presents the results of the same analysis in the early and late P600 time windows, 450-700ms and 700-950ms, respectively. Interactions and effects will be further explored below.

9.3.2.1 150-250ms

In the 150-250ms time window, the 2 x 3 x 2 repeated-measures ANOVA yielded a marginal interaction of AgreementType and Laterality, $F[1.456,33.483]=3.307$, $p=.063$, as well as a significant effect of AgreementType, $F[1,23]=12.824$, $p=.002$. Follow-up analyses indicated

Table 57. Results of repeated-measures ANOVAs comparing learners' mean difference amplitudes for number versus gender agreement on Adjectives in each primary time window.

	150-250ms	**250-450ms**	**450-950ms**
AgreementType x Laterality x Anteriority	$F[2,46]=.012$ $p=.988$	$F[2,46]=1.603$ $p=.212$	$F[2,46]=1.744$ $p=.186$
AgreementType x Laterality	$\mathbf{F[1.456,33.483]=3.307}$ $\mathbf{p=.063}$†	$F[1.316,30.270]=2.570$ $p=.111$	$\mathbf{F[1.484,34.142]=3.525}$ $\mathbf{p=.053}$†
AgreementType x Anteriority	$F[1,23]=.822$ $p=.374$	$F[1,23]=.507$ $p=.483$	$F[1,23]=2.707$ $p=.114$
Laterality x Anteriority	$F[2,46]=1.004$ $p=.374$	$F[2,46]=1.252$ $p=.295$	$F[2,46]=.728$ $p=.489$
AgreementType	$\mathbf{F[1,23]=12.824}$ $\mathbf{p=.002}$*	$\mathbf{F[1,23]=6.666}$ $\mathbf{p=.017}$*	$\mathbf{F[1,23]=7.642}$ $\mathbf{p=.011}$*
Laterality	$F[2,46]=788$ $p=.461$	$F[2,46]=.360$ $p=.700$	$F[2,46]=.077$ $p=.926$
Anteriority	$F[1,23]=2.309$ $p=.142$	$F[1,23]=.091$ $p=.766$	$F[1,23]=1.645$ $p=.212$

* Effect/interaction is significant at the $p<.05$ level.
† Effect is marginally significant ($.10>p>.05$).

Table 58. Results of 3-way ANOVA comparing learners' mean difference amplitudes for number versus gender agreement on Adjectives in the early and late P600 time windows.

	450-700ms	**700-950ms**
AgreementType x Laterality x Anteriority	$F[2,46]=2.105$ $p=.133$	$F[1.677, 38.567]=.941$ $p=.384$
AgreementType x Laterality	$\mathbf{F[2,46]=10.468}$ $\mathbf{p<.001}$*	$F[1.403,32.269]=.836$ $p=.405$
AgreementType x Anteriority	$F[1,23]=.932$ $p=.344$	$\mathbf{F[1,23]=3.872}$ $\mathbf{p=.061}$†
Laterality x Anteriority	$F[2,46]=.483$ $p=.620$	$F[2,46]=.896$ $p=.415$
AgreementType	$\mathbf{F[1,23]=11.868}$ $\mathbf{p=.002}$*	$F[1,23]=.378$ $p=.545$
Laterality	$F[2,46]=1.133$ $p=.331$	$F[2,46]=.605$ $p=.550$
Anteriority	$F[1,23]=.562$ $p=.461$	$F[1,23]=2.572$ $p=.122$

* Effect/interaction is significant at the $p<.05$ level.
† Effect is marginally significant ($.10>p>.05$).

a significant effect of AgreementType at all three levels of Laterality (Left: $F[1,23]=8.705$, $p=.007$; Central: $F[1,23]=17.357$, $p<.001$; Right: $F[1,23]=4.402$, $p=.047$). In all cases, mean differences for Noun-Adjective Number agreement were greater than for Noun-Adjective Gender agreement.

9.3.2.2 250-450ms

The 2 x 3 x 2 repeated-measures ANOVA also yielded a significant effect of AgreementType in the 250-450ms time window, $F[1,23]=6.666$, $p=.017$, where mean differences for number agreement were still greater than for gender agreement.

9.3.2.3 450-950ms

In the overall P600 time window, the 2 x 3 x 2 repeated-measures ANOVA yielded a marginal interaction of AgreementType and Laterality, $F[1.484,34.142]=3.525$, $p=.053$, as well as a significant effect of AgreementType, $F[1,23]=7.642$, $p=.011$. Follow-up analyses indicated a significant effect of AgreementType in the Central electrode regions, $F[1,23]=9.803$, $p=.005$, and in the Right electrode regions, $F[1,23]=7.265$, $p=.013$, where the differences in mean amplitudes for grammatical and ungrammatical stimuli ware greater for number agreement than for gender agreement.

9.3.2.3.1 450-700ms

In the early P600 time window from 450-700ms, the 2 x 3 x 2 repeated-measures ANOVA revealed a significant interaction of AgreementType and Laterality, $F[2,46]=10.468$, $p<.001$, as well as a significant effect of AgreementType, $F[1,23]=11.868$, $p=.002$. Follow-up analyses indicated a significant effect of AgreementType at all three levels of Laterality (Left: $F[1,23]=6.343$, $p=.019$; Central: $F[1,23]=15.420$, $p=.001$; Right: $F[1,23]=11.382$, $p=.003$). In all cases, mean differences for Noun-Adjective Number agreement were greater than for Noun-Adjective Gender agreement.

9.3.2.3.2 700-950ms

In the late P600 time window (700-950ms), the 2 x 3 x 2 repeated-measures ANOVA revealed only a marginal AgreementType x Anteriority interaction, $F(1,23)=3.872$, $p=.061$, but no main effect of AgreementType was present. Follow-up analysis revealed that a marginal effect of AgreementType was present only in the posterior electrode regions, $F[1,23]=3.577$, $p=.071$.

9.3.2.4 Summary

No significant differences were observed between number agreement on verbs and on adjectives. However, throughout all time windows of interest, number agreement on adjectives elicited a greater response than did gender agreement on adjectives. This difference was present in the earliest time window and continued to be significant through 700ms. In the late P600 time window (700-950ms), the greater response for number agreement was only marginal, which may be due to the fact that the P600 for gender, although limited in distribution, began to emerge in this time window.

9.4 Summary of All Results for Learners

Overall, learners demonstrated sensitivity to the agreement violations presented in the study. Grammaticality effects were present in the behavioral results for all three types of agreement. Learners were marginally more sensitive to number violations on Adjectives than on verbs, and there were significant differences when d’ scores for number and gender violations on adjectives were compared, but for each type of agreement, there were several learners who performed within the range of the native speakers. With regard to ERP responses, effects for Grammaticality were also statistically significant in the P600 time window for all agreement types, although the response for gender violations was later and more limited in distribution. Similar to the results for native speakers, the P600 for violations of number agreement on verbs was broadly distributed from 450-700ms, but was only present on posterior electrodes from 700-950ms. The P600 response to number violations on adjectives demonstrated the same distribution. In both cases, results were consistent with claims based on native speaker data that the early P600 phase is more frontally distributed than the late phase (Molinaro et al., 2011). For gender violations on adjectives, no overall effect of Grammaticality was found in the omnibus ANOVAs; however, the region-by-region analysis showed a significant effect of Grammaticality from 700-950ms in the Left Posterior region, as well as a marginal effect in the Right Posterior region. As for the early time windows, the significant positivities observed in native speakers for Subject-Verb and Noun-Adjective Gender agreement were not present in the L2 group. Finally, tests for effect of AgreementType showed that there were no significant differences between Subject-Verb and Noun-Adjective Number agreement, while differences were present for Noun-Adjective Number versus Gender agreement, where differences between grammatical and ungrammatical stimuli were greater for number agreement. These results are summarized in Table 59 below, followed by the results of the region-by-region analyses in Table 60. Unless otherwise indicated, all effects represent more positive-going waveforms in response to ungrammatical versus grammatical stimuli.

Table 59. Summary of ERP results for learners across agreement types.

<table>
<tr><th></th><th colspan="2">S-V</th><th colspan="2">N-Adj Number</th><th colspan="2">N-ADJ Gender</th></tr>
<tr><td rowspan="2">150-250ms</td><td colspan="2">-----</td><td colspan="2">-----</td><td colspan="2">-----</td></tr>
<tr><td colspan="3">No difference</td><td colspan="3">Greater for NUM</td></tr>
<tr><td rowspan="2">250-450ms</td><td colspan="2">-----</td><td colspan="2">-----</td><td colspan="2">-----</td></tr>
<tr><td colspan="3">No difference</td><td colspan="3">Greater for NUM</td></tr>
<tr><td rowspan="2">450-950ms</td><td colspan="2">P600
(RA†, LP, CP, RP)</td><td colspan="2">P600</td><td colspan="2">-----</td></tr>
<tr><td colspan="3">No difference</td><td colspan="3">Greater for NUM
(Central, Right regions only)</td></tr>
<tr><td rowspan="2">a. 450-700ms</td><td colspan="2">P600
(RA, LP, CP, RP)</td><td colspan="2">P600
(Left, Central, Right)</td><td colspan="2">-----</td></tr>
<tr><td colspan="3">No difference</td><td colspan="3">Greater for NUM
(Left, Central, Right regions)</td></tr>
<tr><td rowspan="2">b. 700-950ms</td><td colspan="2">P600
(LP†, CP, RP)</td><td colspan="2">P600
(Posterior)</td><td colspan="2">-----</td></tr>
<tr><td colspan="3">No difference</td><td colspan="3">Marginally greater for NUM
(Posterior regions only)</td></tr>
</table>

† Correlation is marginally significant (.10>*p*>.05). All other correlations are significant (*p*<.05).

Table 60. Summary of region-by-region analysis for learners across agreement types.

	S-V	N-Adj Number	N-Adj Gender
150-250ms	-----	CA†	LP†, CP (-)
250-450ms	-----	-----	-----
450-950ms	RA†, LP, CP, RP	LP, CP, RP	LP
a. 450-700ms	RA, LP, CP, RP	LA†, CA, RA, LP, CP, RP	-----
b. 700-950ms	LP†, CP, RP	LP, CP, RP	LP, RP†

† Correlation is marginally significant (.10>*p*>.05). All other correlations are significant (*p*<.05).

CHAPTER 10:

ANALYSIS OF INDIVIDUAL DIFFERENCES

In the previous chapters, analyses of both behavioral and electrophysiological evidence have shown that learners are sensitive to ungrammaticality in sentences containing various types of agreement-related errors in the L2, but that unique L2 features pose more difficulty for learners at low levels of proficiency. In addition to testing L1/L2 similarity, this experiment was also designed to investigate another factor in L2 acquisition: individual differences. L2 participants completed tests for verbal aptitude (MLAT), nonverbal aptitude (RAVEN), and Spanish proficiency (MLA/DELE). This chapter first presents descriptive statistics for these independent variables along with tests for correlations between them. Following that, analyses are presented which test for correlations between these measures of individual differences and the dependent measures obtained in the study, including d’ scores on the grammaticality judgment task and ERP responses. While the primary focus is on measures of verbal and nonverbal aptitude, proficiency test scores were also included for analysis, given the fact that previous studies had reported a moderate correlation between verbal aptitude scores and L2 proficiency measures (Carroll, 1981).

10.1 Independent Variables

10.1.1 Verbal Aptitude

10.1.1.1 Modern Languages Aptitude Test

Verbal aptitude was tested using the Short Form of the Modern Languages Aptitude Test (MLAT) described in *Methods* above, which consists of three sections. The participants’ total MLAT scores ranged from 40 to 83 points out of a possible total of 119 points. The mean score was 60.21 (SD=10.84). As could be expected for a standardized test, the scores were clustered around the 50^{th} percentile, with a mean of 44^{th} percentile. Data were normally distributed, as determined by the Shapiro-Wilk test for normality in small sample sizes (p=.407). The mean score for the MLAT 3 (Spelling Clues) was 18.50 (SD=5.84) out of 50 points, with a range of 8 to 28 points. Scores on the MLAT 3 were also normally distributed (p=.119). On the MLAT 4 (Words in Sentences), the mean score was 22.17 (SD=5.55) out of 45 points, with a range of 14 to 35 points and normal distribution (p=.123). The mean score for the MLAT 5 (Paired Associates) was 19.54 (SD=5.61) out of 24 points, with a range of 4 to 24 points. Here the data were not normally distributed (p<.001), given that many learners were able to answer all items accurately within the given time limit.

10.1.1.2 LLAMA Aptitude Tests

In a follow-up study for which results will be included here, eleven participants provided scores on the LLAMA_B and LLAMA_F, which were used in order to provide an L1-independent measure of verbal aptitude since these tests rely on associations between new words and pictures rather than between new words and their meanings in an existing language. Scores are reported here as percentages. The mean score for the LLAMA_F (which tests grammatical inferencing) was 56.36 (SD=27.30), with a range of 10 to 90 percent. The participants' LLAMA_B scores, which indicate vocabulary learning ability, ranged from 30 to 85 percent. The mean score was 63.18 (SD=17.65).

10.1.2 Nonverbal Aptitude

Raven's Advanced Progressive Matrices (RAVEN) were used to test for nonverbal aptitude for reasons defined in the *Methods* section. The mean score on the RAVEN was 22.13 (SD=3.19) out of a possible 36 points. Scores on this measure ranged from 16 to 28 points. No standardized results can be reported due to the reduced time allowed for the test. The Shapiro-Wilk test for normality showed that scores on the RAVEN were normally distributed (p=.751).

10.1.3 Proficiency

Spanish proficiency was tested using a combination of two tests which has been widely used in studies of L2 agreement. The combination MLA/DELE test, described previously, involved a total of 50 multiple choice items. The scores for the L2 participants in this study ranged from 11 to 27 points, so all scores were in the range designated as low proficiency (0-29 points). The mean score was 18.79 (SD=4.50). Scores were normally distributed (p=.438).

10.1.4 Summary

All learners who participated in the study demonstrated low proficiency in Spanish, but were normally distributed on almost all measures of verbal and nonverbal aptitude. The descriptive statistics for these measures of aptitude and proficiency are summarized in Table 61.

Table 61. Observed measures of aptitude and proficiency, along with the maximum possible scores for each measure.

	Mean	SD	Range	Max
MLAT3	18.50	5.84	8-28	out of 50
MLAT4	22.17	5.55	14-35	out of 45
MLAT5	19.54	5.61	4-24	out of 24
MLAT Total	60.21	10.83	40-83	out of 119
LLAMA_F	56.36	27.30	10-90	out of 100
LLAMA_B	63.18	17.65	30-85	out of 100
RAVEN	22.13	3.19	16-28	out of 36
Proficiency	18.79	4.50	11-27	out of 50

10.2 Relationships between Aptitude and Proficiency

In order to examine the interrelationship of verbal aptitude, nonverbal aptitude, and L2 proficiency, correlations between all independent variables were investigated. The results (see

Table 62 below) indicated that verbal and nonverbal aptitude could largely be tested as separate constructs by the MLAT and RAVEN, since there were no significant correlations between the RAVEN and any verbal aptitude scores on the MLAT (MLAT3: r=-.274, p=.195; MLAT4: r=.168, p=.432; MLAT5: r=.154, p=.473; MLAT Total: r=.018, p=.933). A marginally positive correlation was found between the LLAMA_F and the RAVEN (r=.535, p=.090), but not between the RAVEN and the LLAMA_B (r=.414, p=.206). With regard to proficiency, three verbal aptitude measures were either marginally or significantly correlated with proficiency scores on the MLA/DELE test (MLAT4: r=.359, p=.085; MLAT5: r=.597, p=.002; MLAT Total: r=.565, p=.004). The MLAT 3 was not correlated with proficiency (MLAT3: r=.135, p=.530), nor were the LLAMA_F (r=.078, p=.821) and LLAMA_B tests (r=-.301, p=.369). Interestingly, nonverbal aptitude demonstrated no significant correlation with proficiency (r=.289, p=.170).

Table 62. Intercorrelations of aptitude and proficiency measures.

	MLAT3	**MLAT4**	**MLAT5**	**MLAT Total**	**LLAMA _F**	**LLAMA _B**	**RAVEN**
MLAT4	-.066 (.760)						
MLAT5	.165 (.440)	.233 (.273)					
MLAT Total	**.591* (.002)**	**.597* (.002)**	**.726* (.000)**				
LLAMA_F	-.125 (.714)	.058 (.865)	-.003 (.992)	-.035 (.919)			
LLAMA_B	-.434 (.182)	**.541† (.086)**	.077 (.822)	.075 (.827)	.462 (.152)		
RAVEN	-.274 (.195)	.168 (.432)	.154 (.473)	.018 (.933)	**.535† (.090)**	.414 (.206)	
Proficiency	.135 (.530)	**.359† (.085)**	**.597* (.002)**	**.565* (.004)**	.078 (.821)	-.301 (.369)	.289 (.170)

* Correlation is significant at the p<.05 level.
† Effect/interaction is marginally significant.

Given the apparent relationship of proficiency and verbal aptitude scores on the MLAT 4, MLAT 5, and overall MLAT, partial correlations are reported where necessary in the analyses of relationships between these variables and learner sensitivity to agreement violations.

10.3 Individual Differences and Behavioral Sensitivity to Agreement Violations

In order to address research questions regarding the role of individual differences in both behavioral and neural detection of agreement violations, this section investigates correlations between behavioral measures and aptitude/proficiency scores for learners, while the following section reports analyses of correlations between aptitude/proficiency and ERP responses. In each of these two chapters, analyses are presented separately for verbal and nonverbal aptitude, as well as proficiency.

10.3.1 Verbal Aptitude

10.3.1.1 Modern Language Aptitude Test

Scores for all measures of verbal aptitude on the MLAT were tested for correlations with d' scores for each type of agreement (S-V, N-Adj Number, N-Adj Gender). The resulting Pearson correlation coefficients are presented in Table 63 and described below. In general, the strongest relationship was found between subtests of the MLAT verbal aptitude battery and d' scores for Subject-Verb agreement.

A significant positive correlation was present between Subject-Verb d' scores and the MLAT Total scores (r=.420, p=.041), with the MLAT accounting for roughly 18% of variance in the d' scores, a correlation that was largely driven by significant positive correlations with the MLAT4 (r=.473, p=.020) and MLAT5 (r=.407, p=.048). These correlations were present with roughly the same strength of correlation as has been reported previously for the MLAT and

Table 63. Correlations between d' scores and measures of verbal aptitude on the MLAT test.

		S-V	N-Adj Number	N-Adj Gender
MLAT3	*r*	-.060	-.224	-.022
	(*p*)	(.780)	(.292)	(.917)
MLAT4	*r*	**.473***	**.380†**	.286
	(*p*)	**(.020)**	**(.067)**	(.175)
MLAT5	*r*	**.407***	.261	.262
	(*p*)	**(.048)**	(.217)	(.216)
MLAT Total	*r*	**.420***	.209	.270
	(*p*)	**(.041)**	(.327)	(.202)

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

various proficiency measures in classroom learners (Carroll, 1981). The MLAT3 was not significantly correlated with Subject-Verb d' scores (MLAT3: r=-.060, p=.780).

The MLAT4 was also marginally and positively correlated with d' scores for Noun-Adjective Number agreement (r=.380, p=.067). Here, no significant correlations were found for the MLAT3 (r=-.224, p=.292), MLAT5 (r=.261, p=.217), or MLAT Total scores (r=.209, p=.327).

As can also be seen in Table 63, no correlations were found between any measure of verbal aptitude and d' scores for Noun-Adjective Gender agreement (MLAT3: r=-.022, p=.917; MLAT4: r=.286, p=.175; MLAT5: r=.262, p=.216; MLAT Total: r=.270, p=.202).

10.3.1.2 LLAMA Aptitude Tests

Scores on the LLAMA_F and LLAMA_B tests were also analyzed for correlations with d' scores for each type of agreement. The resulting Pearson correlation coefficients are presented in Table 64 and described below. No significant correlations were observed.

Table 64. Correlations between d' scores and measures of verbal aptitude on the LLAMA tests.

		S-V	N-Adj Number	N-Adj Gender
LLAMA_F	*r*	.490	.454	.360
	(*p*)	(.126)	(.161)	(.276)
LLAMA_B	*r*	.409	.350	.066
	(*p*)	(.211)	(.292)	(.847)

10.3.2 Nonverbal Aptitude

Nonverbal aptitude, as measured by the RAVEN, was not found to be significantly correlated with any behavioral measure of L2 sensitivity to agreement violations, including d' scores for Subject-Verb agreement (r=.230, p=.280), Noun-Adjective Number agreement (r=.218, p=.306), and Noun-Adjective Gender agreement (r=.135, p=.528).

10.3.3 Proficiency

Proficiency scores were found to be only marginally positively correlated with d' scores for Noun-Adjective Gender agreement (r=.362, p=.082). No relationship was found between proficiency scores and d' scores for Subject-Verb agreement (r=.323, p=.124) or Noun-Adjective Number agreement (r=.217, p=.309).

10.3.4 Summary

In summary, the above analysis uncovered significant positive correlations only between d' scores for Subject-Verb agreement and verbal aptitude as measured by the MLAT4 and MLAT5. The MLAT4, which tests language analytic ability, was also marginally correlated with N-Adj Number agreement. Proficiency scores were marginally correlated with N-Adj Gender agreement. It should be noted that the correlation for proficiency has the opposite distribution across agreement types of the correlations for the MLAT; therefore, no partial correlations were investigated. Finally, no significant correlations were found for the LLAMA tests with d' scores for any of the three types of agreement, nor was any relationship between d' scores and nonverbal aptitude uncovered. For ease of reference, all of the results of the analyses presented in this subsection are presented in Table 65.

Table 65. Correlations between d' scores and measures of aptitude and proficiency.

		S-V	N-Adj Number	N-Adj Gender
MLAT3	*r*	-.060	-.224	-.022
	(*p*)	(.780)	(.292)	(.917)
MLAT4	*r*	**.473***	**.380†**	.286
	(*p*)	**(.020)**	**(.067)**	(.175)
MLAT5	*r*	**.407***	.261	.262
	(*p*)	**(.048)**	(.217)	(.216)
MLAT Total	*r*	**.420***	.209	.270
	(*p*)	**(.041)**	(.327)	(.202)
LLAMA_F	*r*	.490	.454	.360
	(*p*)	(.126)	(.161)	(.276)
LLAMA_B	*r*	.409	.350	.066
	(*p*)	(.211)	(.292)	(.847)
Raven	*r*	.230	.218	.135
	(*p*)	(.280)	(.306)	(.528)
PROF	*r*	.323	.217	**.362†**
	(*p*)	(.124)	(.309)	**(.082)**

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

10.4 Individual Differences and ERP Responses to Agreement Violations

In order to further investigate the relationship between sensitivity to agreement violations and measures of individual differences, an analysis of correlations between aptitude/proficiency scores and ERP responses was conducted using the effect sizes (mean difference amplitudes) of P600 responses for all three agreement types in the overall P600 window (450-950ms), and the early and late P600 time windows (450-700ms, 700-950ms). Correlations for the early time windows (150-250ms, 250-450ms) are not presented here since there were no ERP effects in those time windows for learners; however, the results of those analyses can be found in

Appendix 7 for Subject-Verb agreement, Appendix 8 for N-Adj Number agreement, and Appendix 9 for N-Adj Gender agreement.

10.4.1 Verbal Aptitude and the Processing of Agreement: The MLAT

This section presents correlations between verbal aptitude, as measured by the Short Form of the Modern Languages Aptitude Test (MLAT) and its various subtests, and P600 responses to agreement violations in the form of mean difference amplitudes. Results are reported first for Subject-Verb agreement, followed by Noun-Adjective Number agreement, and then Noun-Adjective Gender agreement. An additional section reports the results of a follow-up study using the LLAMA Language Aptitude Tests (Meara, 2005) which are picture-based and therefore L1-independent.

10.4.1.1 Number Agreement on Verbs

Table 66 presents Pearson correlation coefficients representing the relationships between mean amplitude differences in the P600 time windows and the MLAT and its subtests. No significant correlations were found involving the MLAT3or MLAT Total scores. In the 450-950ms time window, only a marginal negative correlation was present between the MLAT4 and mean difference amplitudes in the Right Anterior region, r=-.380, p=.067. This marginal negative correlation was also present in the late P600 time window (700-950ms), r=-.403, p=.051. Also in that time window, MLAT5 scores exhibited a marginal positive correlation in the Left Posterior region, r=.346, p=.098.

Table 66. Correlations between verbal aptitude measures and mean difference amplitudes for Subject-Verb agreement in the P600 time windows (overall, early, late).

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
		450-950ms					
MLAT3	*r*	-.005	-.083	.000	-.101	-.098	-.113
	(*p*)	(.980)	(.699)	(.998)	(.640)	(.650)	(.600)
MLAT4	*r*	-.103	-.240	**-.380†**	.110	.037	.038
	(*p*)	(.631)	(.259)	**(.067)**	(.609)	(.863)	(.861)
MLAT5	*r*	.065	.104	.121	.286	.173	.083
	(*p*)	(.764)	(.628)	(.575)	(.175)	(.418)	(.698)
MLAT Total	*r*	-.022	-.114	-.132	.150	.056	.002
	(*p*)	(.918)	(.597)	(.538)	(.483)	(.795)	(.993)
		450-700ms					
MLAT3	*r*	-.030	-.092	-.053	-.027	.002	-.044
	(*p*)	(.888)	(.669)	(.805)	(.900)	(.993)	(.837)
MLAT4	*r*	-.099	-.194	-.287	.132	.045	.041
	(*p*)	(.646)	(.365)	(.174)	(.540)	(.836)	(.850)
MLAT5	*r*	-.048	.023	.105	.187	.133	.074
	(*p*)	(.824)	(.914)	(.625)	(.382)	(.535)	(.732)
MLAT Total	*r*	-.092	-.137	-.121	.150	.093	.035
	(*p*)	(.670)	(.524)	(.573)	(.485)	(.666)	(.871)
		700-950ms					
MLAT3	*r*	.033	-.052	.050	-.160	-.191	-.169
	(*p*)	(.879)	(.810)	(.816)	(.455)	(.372)	(.429)
MLAT4	*r*	-.095	-.242	**-.403†**	.072	.024	.031
	(*p*)	(.660)	(.254)	**(.051)**	(.737)	(.911)	(.884)
MLAT5	*r*	.227	.187	.115	**.346†**	.192	.085
	(*p*)	(.287)	(.382)	(.592)	**(.098)**	(.369)	(.692)
MLAT Total	*r*	.087	-.055	-.120	.130	.009	-.031
	(*p*)	(.687)	(.798)	(.578)	(.545)	(.967)	(.886)

† Correlation is marginally significant ($.10>p>.05$).

10.4.1.2 Number Agreement on Adjectives

The picture is quite different when correlations between MLAT measures and mean difference amplitudes for Noun-Adjective Number agreement are investigated. The resulting Pearson correlation coefficients are presented in Table 67. Significant positive correlations were found between the overall MLAT Total score and mean difference amplitudes for all posterior electrode regions in the 450-950ms time window (Left: r=.523, p=.009; Central: r=.416, p=.043,

Table 67. Correlations between verbal aptitude measures and mean difference amplitudes for Noun-Adjective Number agreement in the P600 time windows (overall, early, late).

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
		450-950ms					
MLAT3	*r*	-.027	-.031	.079	.016	.119	.342
	(*p*)	(.900)	(.885)	(.714)	(.941)	(.579)	(.102)
MLAT4	*r*	-.333	**-.359†**	**-.409***	**.523***	**.346†**	.266
	(*p*)	(.112)	**(.085)**	**(.047)**	**(.009)**	**(.098)**	(.208)
MLAT5	*r*	.111	.089	.030	**.476***	.337	.169
	(*p*)	(.605)	(.681)	(.891)	**(.019)**	(.107)	(.430)
MLAT	*r*	-.127	-.155	-.152	**.523***	**.416***	**.408***
Total	(*p*)	(.553)	(.470)	(.480)	**(.009)**	**(.043)**	**(.048)**
		450-700ms					
MLAT3	*r*	.051	.006	.078	.097	.123	.275
	(*p*)	(.813)	(.977)	(.717)	(.652)	(.568)	(.194)
MLAT4	*r*	-.163	-.136	-.140	**.552***	**.431***	**.384†**
	(*p*)	(.447)	(.527)	(.513)	**(.005)**	**(.035)**	**(.064)**
MLAT5	*r*	.246	.224	.162	**.388†**	.253	.105
	(*p*)	(.247)	(.294)	(.450)	**(.061)**	(.233)	(.625)
MLAT	*r*	.071	.050	.054	**.536***	**.418***	**.399†**
Total	(*p*)	(.740)	(.818)	(.802)	**(.007)**	**(.042)**	**(.053)**
		700-950ms					
MLAT3	*r*	-.087	-.059	.068	-.085	.100	**.362†**
	(*p*)	(.686)	(.784)	(.751)	(.694)	(.642)	**(.083)**
MLAT4	*r*	**-.409***	**-.494***	**-.566***	**.382†**	.205	.083
	(*p*)	**(.047)**	**(.014)**	**(.004)**	**(.066)**	(.337)	(.699)
MLAT5	*r*	-.024	-.051	-.080	**.486***	**.385†**	.215
	(*p*)	(.910)	(.813)	(.709)	**(.016)**	**(.063)**	(.314)
MLAT	*r*	-.269	-.311	-.295	**.401†**	**.358†**	**.349†**
Total	(*p*)	(.204)	(.139)	(.162)	**(.052)**	**(.086)**	**(.095)**

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

Right: r=.408, p=.048). These correlations were also found to be significant from 450-700ms in the Left and Central Posterior regions (Left: r=.536, p=.007; Central: r=.418, p=.042) and marginal in the Right Posterior region, r=.399, p=.053. For the 700-950ms time window, mean difference amplitudes in the posterior electrode regions were also correlated with the MLAT

Total scores, although marginally (Left: r=.401, p=.052; Central: r=.358, p=.086; Right: r=.349, p=.095).

The numerous correlations of the P600 response with the overall MLAT Total score appear to be driven by a combination of MLAT4 and MLAT5 scores in most time windows. In the 450-950ms time window, the MLAT4 was correlated significantly with the mean difference amplitudes in the Left Posterior region, r=.523, p=.009, and marginally with those for the Central Posterior region, r=.346, p=.098. Negative correlations for the MLAT4 were present marginally in the Central Anterior region, r=-.359, p=.085, and significantly in the Right Anterior region, r=-.409, p=.047. A significant correlation between the MLAT5 and mean amplitude differences in the Left Posterior region, r=.476, p=.019, also contributed to the overall correlation of verbal aptitude with the P600 response for Noun-Adjective Number agreement in the 450-950ms time window. In the early P600 time window from 450-700ms, the correlation with the MLAT5 for the Left Posterior electrodes was marginal, r=.388, p=.061, while the correlation with the MLAT4 was significant, r=.552, p=.005. The MLAT4 correlation was also significant in the Central Posterior region, r=.431, p=.035, and marginal in the Right Posterior region, r=.384, p=.064. Finally, for the late P600 time window from 700-950ms, where the correlation for the overall MLAT Total score was marginal in all posterior regions, there were a number of correlations present as well. Mean difference amplitudes in the Left Posterior region were significantly correlated with the MLAT5, r=.486, p=.016, and marginally correlated with the MLAT4, r=.382, p=.066. The MLAT 5 was also marginally correlated with mean difference amplitudes in the Central Posterior region, r=.385, p=.063. In the Right Posterior region, the only MLAT3 correlation was found, and it was marginal, r=.362, p=.083. In this late time window, the MLAT4 again produced negative correlations in the anterior electrode regions (Left: r=-.409, p=.047; Central: r=-.494, p=.014, Right: r=-.566, p=.004).

Due to overlap in significant correlations when both proficiency (see below) and MLAT scores were correlated with mean amplitude differences in the Left Posterior region in the overall and early P600 time windows, partial correlations were analyzed for the Left Posterior region in both windows. In the 450-950ms time window, when controlling for proficiency, there was still a significant positive correlation between mean difference amplitudes and the MLAT4, r=.434, p=.039, as well as a marginal correlation with the MLAT Total score, r=.363, p=.088. For the 450-700ms time window, a significant positive correlation was also still present for the MLAT4, r=.461, p=.027, as well as a marginal correlation for the MLAT Total score, r=.354, p=.098.

10.4.1.3 Gender Agreement on Adjectives

Table 68 presents Pearson correlation coefficients representing the relationship between all MLAT measures and the mean difference amplitudes for Noun-Adjective Gender agreement. There were no significant correlations present involving the MLAT3, MLAT5, or MLAT Total scores. In the 450-950ms time window, only a marginal negative correlation was present between the MLAT4 and mean difference amplitudes in the Left Anterior region, r=-.398, p=.054. This correlation with the MLAT4 was significant in the 700-950ms time window, r=-.461, p=.023, along with a marginal correlation in the Central Anterior region, r=-.397, p=.055. No other significant correlations were present in that time window, and there were no significant correlations with mean difference amplitudes from 450-700ms.

10.4.2 Verbal Aptitude and the Processing of Agreement: The LLAMA Tests

In the analysis of MLAT correlations with both behavioral and ERP data, it became clear that MLAT scores correlated more closely to sensitivity to violations involving the L1-like number feature than to gender agreement violations which occur in the L2 Spanish but not in the L1 English. Therefore, in a follow-up to the MLAT analysis, a second set of verbal aptitude tests was conducted which were not based on the L1 English, but rather on pictures: the LLAMA_F and LLAMA_B. Data from 11 participants were tested for correlation with d' scores for each type of agreement (S-V, N-Adj Number, N-Adj Gender). The mean score on the LLAMA_F was

Table 68. Correlations between verbal aptitude measures and mean difference amplitudes for Noun-Adjective Gender agreement in the P600 time windows (overall, early, late).

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
		450-950ms					
MLAT3	*r*	-.025	-.197	-.031	-.095	-.033	.049
	(*p*)	(.909)	(.357)	(.887)	(.660)	(.877)	(.819)
MLAT4	*r*	-.398	-.296	-.236	-.075	-.007	.062
	(*p*)	(.054)	(.160)	(.266)	(.726)	(.974)	(.774)
MLAT5	*r*	-.081	-.019	.089	-.009	.153	.057
	(*p*)	(.706)	(.930)	(.679)	(.967)	(.475)	(.791)
MLAT Total	*r*	-.259	-.267	-.091	-.094	.058	.088
	(*p*)	(.222)	(.207)	(.672)	(.661)	(.788)	(.683)
		450-700ms					
MLAT3	*r*	-.067	-.194	-.047	-.075	.025	.066
	(*p*)	(.756)	(.364)	.827)	(.726)	(.906)	(.761)
MLAT4	*r*	-.273	-.172	-.183	.035	.069	.048
	(*p*)	(.196)	(.420)	(.391)	(.870)	(.748)	(.824)
MLAT5	*r*	-.037	.007	.051	.011	.103	-.004
	(*p*)	(.865)	(.976)	(.812)	(.958)	(.633)	(.987)
MLAT Total	*r*	-.195	-.189	-.093	-.017	.102	.058
	(*p*)	(.361)	(.376)	(.667)	(.938)	(.634)	(.788)
		700-950ms					
MLAT3	*r*	.017	-.165	-.001	-.073	-.087	.021
	(*p*)	(.937)	(.442)	(.996)	(.734)	(.688)	(.923)
MLAT4	*r*	**-.461***	**-.397†**	-.257	-.149	-.083	.064
	(*p*)	**(.023)**	**(.055)**	(.225)	(.488)	(.698)	(.765)
MLAT5	*r*	-.111	-.047	.120	-.025	.177	.115
	(*p*)	(.604)	(.828)	(.576)	(.909)	(.409)	(.593)
MLAT Total	*r*	-.285	-.316	-.070	-.128	.002	.104
	(*p*)	(.178)	(.132)	(.746)	(.550)	(.992)	(.630)

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

56.36% (SD=27.303; Range=10-90), and the mean score on the LLAMA_B was 63.18% (SD=17.645; Range=30-85). There was a marginal correlation between the LLAMA_F and the RAVEN, r=.535, p=.090, but no significant or marginal correlations were present with other aptitude or proficiency measures. Pearson correlation coefficients are presented in Table 69 for an analysis of the relationship of the LLAMA tests to behavioral sensitivity to violations of all three types of agreement. No significant correlations were uncovered. The following sections present results of the analysis of the LLAMA tests in relation to ERP responses.

Table 69. Correlations between d' scores and L1-independent measures of verbal aptitude.

		S-V	**N-Adj Number**	**N-Adj Gender**
LLAMA_F	r	.490	.454	.360
	(p)	(.126)	(.161)	(.276)
LLAMA_B	r	.409	.350	.066
	(p)	(.211)	(.292)	(.847)

10.4.2.1 Number Agreement on Verbs

When correlation coefficients representing the relationship between the LLAMA_F and LLAMA_B tests and mean difference amplitudes for Subject-Verb agreement were calculated, no significant correlations were revealed. However, there was a marginal positive correlation between the LLAMA_B test for rapid vocabulary learning and the mean difference amplitudes from 450-950ms in the Left Anterior region, r=.543, p=.084, as well as the Central Anterior region, r=.548, p=.081. In the Left Anterior region, this correlation was driven by a marginal correlation from 700-950ms, r=.591, p=.056. Results of this analysis are presented in Table 70.

Table 70. Correlations between L1-independent verbal aptitude measures and mean difference amplitudes for Subject-Verb agreement in the P600 time windows.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
		450-950ms					
LLAMA_F	*r*	.221	.281	.121	.332	.299	.196
	(*p*)	(.514)	(.403)	(.723)	(.318)	(.372)	(.565)
LLAMA_B	*r*	**.543†**	**.548†**	.241	.193	-.189	-.134
	(*p*)	**(.084)**	**(.081)**	(.476)	(.569)	(.579)	(.694)
		450-700ms					
LLAMA_F	*r*	.215	.308	.183	.409	.443	.406
	(*p*)	(.525)	(.357)	(.590)	(.212)	(.172)	(.216)
LLAMA_B	*r*	.483	.518	.350	.295	-.115	-.058
	(*p*)	(.132)	(.102)	(.292)	(.378)	(.737)	(.865)
		700-950ms					
LLAMA_F	*r*	.207	.166	.027	.197	.106	-.007
	(*p*)	(.541)	(.627)	(.936)	(.561)	(.756)	(.983)
LLAMA_B	*r*	**.591†**	.459	.070	.048	-.247	-.188
	(*p*)	**(.056)**	(.155)	(.838)	(.889)	(.464)	(.581)

† Correlation is marginally significant (.10>*p*>.05).

Table 71. Correlations between L1-independent verbal aptitude measures and mean difference amplitudes for Noun-Adjective Number agreement in the P600 time windows.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
		450-950ms					
LLAMA_F	*r*	.029	.157	.035	.140	.334	.191
	(*p*)	(.931)	(.644)	(.919)	(.681)	(.316)	(.573)
LLAMA_B	*r*	.310	.378	.172	.265	.170	.236
	(*p*)	(.354)	(.252)	(.613)	(.431)	(.617)	(.484)
		450-700ms					
LLAMA_F	*r*	-.017	.240	.093	.185	.421	.377
	(*p*)	(.959)	(.478)	(.786)	(.587)	(.197)	(.253)
LLAMA_B	*r*	.333	.513	.400	.177	.253	.424
	(*p*)	(.317)	(.106)	(.222)	(.603)	(.452)	(.194)
		700-950ms					
LLAMA_F	*r*	.050	.048	-.026	.050	.164	-.090
	(*p*)	(.883)	(.889)	(.939)	(.884)	(.631)	(.791)
LLAMA_B	*r*	.226	.175	-.074	.325	.034	-.064
	(*p*)	(.503)	(.607)	(.829)	(.330)	(.922)	(.852)

10.4.2.2 Number Agreement on Adjectives

Correlation coefficients representing the relationship between the LLAMA_F and LLAMA_B tests and mean difference amplitudes for Noun-Adjective Number agreement are presented in Table 71. No significant correlations were evident.

10.4.2.3 Gender Agreement on Adjectives

When the LLAMA_F and LLAMA_B tests and mean difference amplitudes for Noun-Adjective Gender agreement were investigated, several significant correlations were revealed. Results of this analysis are presented in Table 72. In the 450-950ms time window, there was a significant positive correlation between the LLAMA_B and mean difference amplitudes in the Left Posterior region, *r*=.612, *p*=.045. More striking, however, were the strong positive correlations between the LLAMA_F and mean difference amplitudes in the Central Posterior

Table 72. Correlations between L1-independent verbal aptitude measures and mean difference amplitudes for Noun-Adjective Gender agreement in the P600 time windows.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
		450-950ms					
LLAMA_F	*r*	-.152	.247	.253	**.534†**	**.816***	**.774***
	(*p*)	.655	.465	.454	**.091**	**.002**	**.005**
LLAMA_B	*r*	.206	.343	.183	**.612***	.347	.328
	(*p*)	.543	.302	.590	**.045**	.295	.324
		450-700ms					
LLAMA_F	*r*	-.056	.315	.282	**.640***	**.790***	**.806***
	(*p*)	.869	.345	.400	**.034**	**.004**	**.003**
LLAMA_B	*r*	.236	.375	.297	.482	.339	.432
	(*p*)	.485	.256	.374	.133	.307	.185
		700-950ms					
LLAMA_F	*r*	-.224	.106	.150	.055	**.615***	.432
	(*p*)	.508	.756	.659	.873	**.044**	.184
LLAMA_B	*r*	.153	.238	-.021	.382	.256	.048
	(*p*)	.654	.480	.951	.246	.447	.888

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

region, r=.816, p=.002, and in the Right Posterior region, r=.774, p=.005, which were accompanied by a marginal correlation in the Left Posterior region, r=.534, p=.091. These strong correlations can account for 67% and 60% of variability in the mean difference amplitudes in those regions. They were driven by significant correlations in the early P600 time window from 450-700ms in all three posterior regions (Left: r=.640, p=.034; Central: r=.790, p=.004; Right: r=.806, p=.003), as well as a significant correlation in the Central Posterior region in the late P600 window from 700-950ms, r=.615, p=.044.

10.4.3 Nonverbal Aptitude and the Processing of Agreement

Correlation coefficients for the relationships between mean difference amplitudes for P600 responses and nonverbal aptitude were also calculated. Nonverbal aptitude was measured by scores on the Raven Advanced Progressive Matrices. Results are reported first for Subject-Verb agreement, followed by Noun-Adjective Number agreement, and then Noun-Adjective Gender agreement.

10.4.3.1 Number Agreement on Verbs

As is evident in Table 73, there were no significant correlations between RAVEN scores and the mean difference amplitudes for Subject-Verb agreement in any of the P600 time windows.

10.4.3.2 Number Agreement on Adjectives

In the 450-700ms time window, there were no significant correlations between RAVEN scores and the mean difference amplitudes for Noun-Adjective Number agreement in any of the P600 time windows, as evidenced in Table 74.

Table 73. Correlations between RAVEN scores for nonverbal aptitude and mean difference amplitudes for Subject-Verb agreement in the P600 time windows.

	Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
RAVEN	**450-950ms**					
r	.066	.119	.028	-.034	.051	-.057
(*p*)	(.759)	(.579)	(.897)	(.874)	(.813)	(.791)
	450-700ms					
r	-.062	.000	-.017	-.085	.037	-.032
(*p*)	(.774)	(.998)	(.936)	(.692)	(.863)	(.883)
	700-950ms					
r	.251	.249	.068	.024	.059	-.075
(*p*)	(.238)	(.242)	(.753)	(.913)	(.786)	(.726)

Table 74. Correlations between RAVEN scores for nonverbal aptitude measures and mean difference amplitudes for Noun-Adjective Number agreement in the P600 time windows.

	Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
RAVEN	**450-950ms**					
r	.028	.015	-.050	.065	-.073	-.130
(*p*)	(.896)	(.946)	(.818)	(.762)	(.734)	(.544)
	450-700ms					
r	.198	.171	.081	.177	-.012	-.041
(*p*)	(.354)	(.423)	(.706)	(.408)	(.955)	(.849)
	700-950ms					
r	-.120	-.131	-.147	-.085	-.131	-.213
(*p*)	(.575)	(.542)	(.492)	(.693)	(.540)	(.317)

10.4.3.3 Gender Agreement on Adjectives

Table 75 presents results of the analysis of RAVEN correlations in the 700-950ms time window, where there were still no significant correlations between RAVEN scores and the mean difference amplitudes for Noun-Adjective Gender agreement in any of the P600 time windows.

Table 75. Correlations between RAVEN scores for nonverbal aptitude measures and mean difference amplitudes for Noun-Adjective Gender agreement in the P600 time windows.

	Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
RAVEN	**450-950ms**					
r	.022	.117	.097	-.081	-.031	-.011
(*p*)	(.917)	(.586)	(.653)	(.706)	(.884)	(.958)
				450-700ms		
r	.131	.186	.108	.164	.074	.056
(*p*)	(.540)	(.385)	(.616)	(.444)	(.730)	(.795)
				700-950ms		
r	-.078	.013	.063	-.283	-.136	-.086
(*p*)	(.718)	(.951)	(.771)	(.180)	(.527)	(.691)

10.4.4 Proficiency and the Processing of Agreement

Proficiency scores on the MLA/DELE test were also tested for correlations with mean difference amplitudes for P600 responses. Results are reported first for Subject-Verb agreement, followed by Noun-Adjective Number agreement, and then Noun-Adjective Gender agreement.

10.4.4.1 Number Agreement on Verbs

The only correlation between proficiency scores and mean difference amplitudes for Subject-Verb agreement was a marginal negative correlation in the Central Anterior electrodes from 450-700ms, r=-.353, p=.091. There were no significant correlations in any time window, as can be seen in Table 76.

10.4.4.2 Number Agreement on Adjectives

In the overall P600 time window from 450-950ms, proficiency scores were significantly and positively correlated with mean difference amplitudes for Noun-Adjective Number agreement only in the Left Posterior region, r=.452, p=.027. This correlation was driven by a significant positive correlation in the early P600 time window from 450-700ms, r=.500, p=.013. Results of this analysis are presented in Table 77.

Table 76. Correlations between proficiency scores and mean difference amplitudes for Subject-Verb agreement in the P600 time windows.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
PROF		**450-950ms**					
	r	-.278	-.332	-.265	-.046	-.040	-.166
	(p)	(.188)	(.113)	(.211)	(.830)	(.854)	(.438)
		450-700ms					
	r	-.331	**-.353†**	-.304	-.116	-.073	-.197
	(p)	(.114)	**(.091)**	(.148)	(.591)	(.735)	(.356)
		700-950ms					
	r	-.155	-.222	-.179	.033	.003	-.114
	(p)	(.469)	(.296)	(.402)	(.879)	(.987)	(.594)

† Correlation is marginally significant (.10>p>.05).

Table 77. Correlations between proficiency scores and mean difference amplitudes for Noun-Adjective Number agreement in the P600 time windows.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
PROF		**450-950ms**					
	r	-.123	-.174	-.111	**.452***	.294	.053
	(p)	(.568)	(.416)	(.606)	**(.027)**	(.163)	(.807)
		450-700ms					
	r	.108	.007	.006	**.500***	.244	.007
	(p)	(.615)	(.973)	(.980)	**(.013)**	(.250)	(.973)
		700-950ms					
	r	-.291	-.305	-.187	.301	.310	.098
	(p)	(.167)	(.148)	(.381)	(.154)	(.140)	(.649)

* Correlation is significant at the p<.05 level.

10.4.4.3 Gender Agreement on Adjectives

Despite the fact that proficiency scores were correlated with behavioral sensitivity to gender violations, there were no significant correlations between proficiency scores and P600 responses to Noun-Adjective Gender agreement in any time window. Results are presented in Table 78.

Table 78. Correlations between proficiency scores and mean difference amplitudes for Noun-Adjective Gender agreement in the P600 time windows.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
PROF		**450-950ms**					
	r	-.185	-.158	.004	-.095	.106	.042
	(*p*)	(.387)	(.461)	(.984)	(.660)	(.622)	(.847)
		450-700ms					
	r	-.068	-.071	-.042	.043	.113	-.048
	(*p*)	(.751)	(.741)	(.844)	(.841)	(.601)	(.824)
		700-950ms					
	r	-.267	-.237	.064	-.188	.080	.137
	(*p*)	(.208)	(.265)	(.766)	(.380)	(.711)	(.524)

10.4.5 Summary

Correlations of aptitude and proficiency scores with measures of sensitivity to agreement violations were mixed. With regard to d' scores, verbal aptitude measures alone correlated with measures of sensitivity to number agreement, primarily in the Subject-Verb conditions (see Table 63), while proficiency scores alone were marginally correlated with d' scores for Noun-Adjective Gender agreement. The analysis of correlations between aptitude/proficiency scores and ERP responses to agreement violations also yielded stronger correlations for verbal aptitude and number agreement violations. Here, the MLAT4, MLAT5, and MLAT Total scores frequently showed significant or marginal correlations with mean difference amplitudes in regions where P600 effects were present for number violations on Adjectives, but the MLAT scores did not yield strong correlations with P600 responses to number violations on verbs despite their correlation to S-V d' scores. A few negative correlations with the MLAT4 were present in the anterior electrode regions for all three types of agreement. Interestingly, when verbal aptitude was tested using the L1-independent measures LLAMA_F and LLAMA_B, correlations were strongest for gender agreement violations. Finally, proficiency scores hardly played a role at all in ERP responses and was not correlated with sensitivity to gender violations despite the marginal correlation between proficiency and N-Adj Gender d' scores. Table 79 presents a summary of correlations between ERP responses and all aptitude and proficiency measures. Unless otherwise noted, reported correlations were significant and positive. Regions in bold indicate co-occurrence with ERP effects.

Table 79. Summary of correlations between aptitude/proficiency measures and mean difference amplitudes for all three types of agreement in the P600 time windows.

	S-V	N-Adj Number	N-Adj Gender
	450-950ms		
MLAT3	-----	-----	-----
MLAT4	**RA(-)†**	CA(-)†, **RA(-), LP, CP†**	LA(-)†
MLAT5	-----	**LP**	-----
MLAT Total	-----	**LP, CP, RP**	-----
LLAMA_F	-----	-----	**LP†**, CP, RP
LLAMA_B	LA†, CA†	-----	**LP**
RAVEN	-----	-----	-----
PROF	-----	**LP**	-----
	450-700ms		
MLAT3	-----	-----	-----
MLAT4	-----	**LP, CP, RP†**	-----
MLAT5	-----	**LP†**	-----
MLAT Total	-----	**LP, CP, RP†**	-----
LLAMA_F	-----	-----	LP, CP, RP
LLAMA_B	-----	-----	-----
RAVEN	-----	-----	-----
PROF	CA(-)†	**LP**	-----
	700-950ms		
MLAT3	-----	-----	RP†
MLAT4	RA(-)†	LA(-), CA(-), RA(-), **LP†**	LA(-), CA(-)†
MLAT5	LP†	**LP, CP†**	-----
MLAT Total	-----	**LP†, CP†, RP†**	-----
LLAMA_F	-----	-----	CP
LLAMA_B	LA†	-----	-----
RAVEN	-----	-----	-----
PROF	-----	-----	-----

† Correlation is marginally significant (.10>*p*>.05).

CHAPTER 11:

DISCUSSION AND CONCLUSIONS

The current study was designed to investigate learner sensitivity to agreement violations in L2 Spanish, including violations of number agreement on verbs (e.g., *la viajera… descansa/*-an…* "the traveler$_{3sg}$…rest$_{3sg/*3pl}$…") and of either number or gender agreement on adjectives (e.g., *la isla es rocosa/*-o/*-as* "the island$_{fem.sg}$ is rocky$_{fem.sg/*masc.sg/*fem.pl}$"). Additionally, the role of individual differences in modulating responses to agreement violations was investigated. The following research questions were proposed:

Research Question 1:

For number agreement, do low-proficiency learners demonstrate equally native-like sensitivity to violations realized on verbs and on adjectives, despite differences in instantiation of number agreement on these categories in the L1?

Research Question 2:

Can low-proficiency learners demonstrate developing native-like sensitivity to violations of the gender feature that is unique to the L2?

Research Question 3:

Does verbal aptitude modulate sensitivity to agreement violations and/or to number and gender agreement differentially?

Research Question 4:

Does nonverbal aptitude modulate sensitivity to agreement violations and/or to number and gender agreement differentially?

This section will discuss each of these questions in terms of the results reported here, along with possible limitations in interpreting the results given the constraints of the research design and stimuli. Since the first two questions address whether learners demonstrate native-like sensitivity to agreement violations, native speakers were also tested as a control group, confirming both behavioral and neurophysiological sensitivity to all three types of agreement violations. It should be noted that no LAN effects were present for native speakers, which is not surprising given the substantial number of studies of native speakers that do not find the LAN, as discussed earlier based on comments by Osterhout et al. (2004), Molinaro et al. (2011) and Alemán Bañón et al. (2012). Critically, the P600 found for native speakers here in response to all three types of agreement violations is generally consistent with the P600 results of previous studies of native speakers, at least where violations are presented in sentences (e.g., Osterhout and Mobley, 1995; Barber and Carreiras, 2005; Alemán Bañón et al., 2012). The P600 responses to number violations on adjectives were less broadly distributed in comparison to both number violations on verbs and gender violations on adjectives, where responses in the P600 time window were found to include frontal electrodes in early stages (Friederici, 2002; Kaan and Swaab, 2003; Molinaro et al., 2011).

11.1 Learner Sensitivity to Agreement Violations

11.1.1 Number Agreement

With regard to the first research question, it was hypothesized that learners would pattern with native speakers with regard to the presence of a P600 effect for number agreement since the number feature is present in the L1 of the learners. As per the predictions of both the *Interpretability Hypothesis* (Tsimpli and Mastropavlou, 2007) and *Full Transfer/Full Access* (Schwartz and Sprouse, 1994, 1996), learners were expected to demonstrate native-like sensitivity in terms of both grammaticality judgments and ERP responses to number violations on both verbs and adjectives. However, due to the results of previous ERP studies (Tokowicz and MacWhinney, 2005; Sabourin, 2003; Foucart and Frenck-Mestre, 2010; McLaughlin et al., 2010), the question was raised as to whether or not a P600 would be present in response to number violations on adjectives, given the fact that number agreement does not occur on adjectival predicates in English.

Behavioral results for the grammaticality judgment task demonstrated learners' ability to discriminate between grammatical and ungrammatical stimuli across both types of number agreement. Learners were marginally more sensitive to number violations on adjectives versus verbs even though adjectives are not a target for agreement in the L1, but over half of the learners were within the range of the native speakers with regard to rejecting ungrammatical stimuli in both conditions. As with the native speakers, no LAN effects were present in ERP responses. Learners did exhibit a native-like P600 for both types of number agreement, including a broad distribution in the 450-700ms time window followed by a more posterior distribution in the 700-950ms time window, just as was found for native speakers for number violations on verbs. For learners, no statistical differences were observed between ERP responses to number agreement on verbs (similar to the L1) versus adjectives (different).

The results seen here are inconsistent with previous studies that found no P600 for conditions involving morphosyntactic features instantiated differently in the L1 and L2 (Tokowicz and MacWhinney, 2005; Sabourin, 2003; Foucart and Frenck-Mestre, 2010; McLaughlin et al., 2010). Stimulus design may be the primary explanation for differences across experiments or across conditions within the same experiment. The stimuli in Tokowicz and MacWhinney (2005) is a prime example, where their gender agreement condition involves a violation in sentence-final position, which is not comparable to the position of violations in other conditions. With regard to the current discussion of features that are instantiated differently in the L1 and the L2, Tokowicz and MacWhinney did not find a P600 for English-speaking learners of Spanish in response to number violations between determiners and nouns, as in (24), repeated here as (34):

(34) Los/*El *niños* están jugando.
The$_{\text{masc.pl/*masc.sg}}$ boys$_{\text{masc.pl}}$ are playing.

(Adapted from Tokowicz and MacWhinney, 2005: 178)

Critically, the sentence-initial position of the determiner-noun pair may have created a context in which these low-proficiency learners could have relied on lexical processing of the determiner and noun in the same way as native speakers in the word-pair experiment by Barber and Carreiras (2005), which found an N400 response to ungrammaticality rather than a P600. Even though Tokowicz and MacWhinney did not analyze early windows of responses for the number agreement violations, a visual inspection of the waveforms they report for the noun following the determiner demonstrates a sustained negativity for ungrammatical as opposed to grammatical sentences. It is not clear, then, that these learners are applying a non-native processing routine.

Recall that Foucart and Frenck-Mestre (2010) found a P600 in response to gender violations in a determiner-noun paradigm where the L1 also instantiates gender agreement (stimuli repeated as (35) here), but not on adjectives in a context where it does not (36):

(35) Hier la/*le *chaise* était dans le salon.
Yesterday the$_{\text{fem/*masc}}$ chair$_{\text{fem}}$ was in the living.room.
'Yesterday the chair was in the living room.'

(36) En été, les chaises *blanches/*blancs* sont dans le jardin.
In summer, the chairs$_{\text{fem}}$ white$_{\text{fem/*masc}}$ are in the garden.
'In summer, the white chairs are in the garden.'

(Adapted from Foucart and Frenck-Mestre, 2010: 12)

However, in these experiments, participants were asked to make semantic acceptability judgments. As was referenced briefly in Chapter 4 in a discussion of structural versus semantic processing in native speakers, Hagoort (2005) claims that these two types of processing are parallel, making it possible for one type of input to cancel out the other. If this is also the case for learners, it could explain the differences between experiments observed by Foucart and Frenck-Mestre. Notably, at the critical word in (35) where the P600 was observed for the morphosyntactic violaton, there is not yet enough semantic information present to trigger lexical/semantic processing, given that only an adverb, a determiner and a noun have been viewed. However, the critical word in (36) is an adjective that would necessitate lexical access at least relevant to the probability of co-occurrence with the preceding noun, particularly in light of the participants' conscious focus on making a semantic decision. (The same processing would be required in the case of nouns following pronominal adjectives in Foucart and Frenck-Mestre's third experiment, which did not produce a P600 either.) The fact that N400 effects were not reported by Foucart and Frenck-Mestre does not negate the possibility of semantic/lexical processing for the learners, since no semantic violations would have been encountered in the ungrammatical versus the grammatical conditions.

Overall, the results with regard to learner sensitivity to number agreement violations in the current study suggest that learners are able to acquire native-like processing of features even in contexts where they are instantiated differently in the L1 and L2. Here, English-speaking learners of Spanish demonstrated native-like sensitivity to number agreement both on verbs, where English also instantiates number agreement, and on adjectives, where it does not.

11.1.2 Gender Agreement

The second research question had to do with learner sensitivity to grammatical gender features. Since grammatical gender is a feature unique to the L2 Spanish, it was expected to be more difficult for learners, at least at the low level of proficiency of the learners in this study, which was confirmed by the MLA/DELE proficiency test. While the *Interpretability Hypothesis* (Tsimpli and Mastropavlou, 2007) predicts that no learners should be able to acquire the gender feature, the *Full Transfer/Full Access* theory (Schwartz and Sprouse, 1994, 1996) allows (but does not necessarily predict) that some learners at this stage may demonstrate sensitivity to gender agreement violations. The *Fundamental Difference Hypothesis* (Bley-Vroman, 1989, 1990) predicts that any learners who do show sensitivity to gender agreement violations would be exceptional learners who rely on domain-general problem-solving skills in order to attain some measure of proficiency. Given the necessity of averaging across participants, the contributions of a few exceptional learners might not be enough to produce significant effects for the group as a whole but might be reflected in tests for individual differences. It was also noted that this low-proficiency group of participants might in fact demonstrate an N400 response due to attempts to establish associations between nouns and adjectives based on orthographic regularities in the input or direct lexical associations (McLaughlin et al., 2010). The use of such a heuristic would be in line with Bley-Vroman's (2009) version of the *Fundamental Difference Hypothesis* as well as the *Shallow Structure Hypothesis* (Clahsen and Felser, 2006), although the current study does not address these theories' claims regarding the continued use of shallow processing at later stages of proficiency.

The current study finds evidence of a developing sensitivity to gender agreement on adjectives in terms of both behavioral and ERP results. On the grammaticality judgment task, almost one third of the learners performed within the range of the native speakers in rejecting ungrammatical gender, a finding that is consistent with the claims of full access theories like the FTFA (Schwartz and Sprouse, 1994, 1996), but not with the *Interpretability Hypothesis* (Tsimpli and Mastropavlou, 2007). Indeed, even though significant differences were found between sensitivity to number versus gender violations, an overall effect of grammaticality was present. With regard to learner errors, a gender assignment task demonstrated that learners were able to correctly assign gender to the nouns used in the study under the same conditions present in the grammaticality judgment task, indicating that errors were not due to faulty gender assignment.

Learners also exhibited sensitivity to gender violations in terms of ERP responses, a finding that is consistent with Tokowicz and MacWhinney (2005). The grand-averaged ERP response to gender violations on adjectives demonstrated a P600 effect that was later and more limited in distribution than those to number violations on verbs and adjectives. A region-by-region analysis showed a P600 response from 700-950ms that was significant in the Left Posterior region and marginal in the Right Posterior region. While this distribution is slightly different than the more right-hemisphere response of native speakers to these stimuli, it should be noted that both the left and right hemispheres were found to be involved in the P600 response of native speakers to the other two types of violations employed in the current study, leading to the conclusion that the left-hemisphere response in the learners cannot be classified as a non-native response. Finally, when the responses to number and gender violations on adjectives were compared, there were significant differences only in the early time window where the gender response had not yet appeared – differences between number and gender were only marginal in the late time window.

The fact that even a limited P600 response to gender violations was found is somewhat surprising given the low proficiency of the learners in the current study and the lack of P600 effects for supposedly advanced learners in Sabourin, (2003). It could be argued that learner performance with regard to gender agreement may be inflated in the current study since only nouns and adjectives demonstrating canonical gender marking were employed, making it possible for learners to have simply noticed matching patterns of *-o* endings on masculine nouns and adjectives and *-a* endings on feminine nouns and adjectives. This matching strategy might be just the type of heuristic or shallow processing that learners might rely on in L2 acquisition if the more recent version of the *Fundamental Difference Hypothesis* (Bley-Vroman, 2009) is accurate, or if results are in line with the *Shallow Structure Hypothesis* (Clahsen and Felser, 2006). However, as discussed in Chapters 4 and 5 above, it is likely that an N400 and not the P600 would have been observed if learners were making associations based on orthographic regularities. Additionally, a heuristic relying on the detection and analysis of visual patterns and the maintenance of visual material in working memory might have been expected to produce correlations between performance and RAVEN scores, given the visual nature of the RAVEN. There are a considerable number of neurocognitive studies that can serve to demonstrate the visual-perceptual nature of the RAVEN. For example, Thoma et al. (2000, 2006) found RAVEN scores to be correlated with processing speed in a test of sensory-motor control in response to visual stimuli. Prabhakaran et al. (1997) noted that a number of areas involving visual perception are activated both for RAVEN problems requiring analytical reasoning and for those requiring only figural (visual perceptual) acuity. Although negative evidence must be interpreted with caution, in the current study no correlations were present between the RAVEN and sensitivity to gender agreement violations.

In addition to the preceding arguments, there is another reason why the use of canonical endings is not believed to have influenced the results of the study. The comparison of number and gender agreement in the N-Adj stimuli used here may not be equitable since the subject NPs were approximately 50% masculine and 50% feminine, but all were singular with regard to number. The use of both masculine and feminine subjects may have made processing gender more demanding than the use of all singular subjects. Additionally, the verb that intervenes between the NP subject and the adjective in these stimuli, as in the sample given above and repeated here in (37), bears a number feature but not gender, providing an additional cue for number agreement on the following adjective.

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

Finally, number features and mismatches are also present in the Subject-Verb stimuli. Thus the number of sentences exhibiting number agreement or mismatches was double the number of sentences involving gender, possibly biasing learners to focus on number features and making a P600 for gender violations less likely rather than more likely.

Overall, while the aforementioned limitations may have impacted the robustness of the response, learners did demonstrate a limited native-like sensitivity to gender violations, supporting the claims of the Full Transfer/Full Access theory (Schwartz and Sprouse, 1994, 1996). Contrary to the *Interpretability Hypothesis* (Tsimpli and Mastropavlou, 2007), learners were sensitive to the gender feature that is not present in their L1 in a context where it is uninterpretable (on the target of agreement), and several were able to perform behaviorally at native-like levels as well. Whether or not those learners are of exceptional ability, as predicted by the Fundamental Difference Hypothesis (Bley-Vroman, 1989, 1990), will be addressed below.

11.2 Individual Differences and the Processing of Agreement

11.2.1 Verbal Aptitude

Research question 3 raised the topic of whether sensitivity to specific L2 structures would be modulated by verbal aptitude. Based on the results of previous studies (DeKeyser, 2000; Abrahamsson and Hyltenstam, 2008; Harley and Hart, 1997), it was hypothesized that sensitivity to agreement violations (here, d' scores and responses in the P600 time window) would be correlated with verbal aptitude (MLAT scores). It might have been expected that the MLAT4, which tests language analytic ability, would produce the highest correlations. Given the L1-based nature of the MLAT, a follow-up study also tested verbal aptitude using the LLAMA F and B, which are based on the MLAT4 and MLAT5, respectively.

The most substantial finding of the study in terms of individual differences was the presence of correlations between verbal aptitude and measures of sensitivity to agreement violations. For the grammaticality judgment task, analyses revealed significant correlations between verbal aptitude (as measured by the MLAT4, MLAT5, and MLAT Total scores) and d' scores measuring sensitivity to number violations on verbs. There was a marginal correlation between d' scores for number agreement and the MLAT4, leaving room for the possibility that this correlation might be significant given larger numbers of participants. Furthermore, MLAT scores were positively correlated with mean amplitude differences in the P600 time windows, a finding which is a substantial contribution to further developments in the field of aptitude research. Interestingly, correlations for the P600 were present only in response to number violations and not violations involving the gender feature that is not present in the L1. Based on these results alone, it might seem that verbal aptitude differentially modulates responses to number and gender agreement, but a different picture emerges when the LLAMA results are examined below.

While the LLAMA_F and LLAMA_B might have been expected to measure the same verbal abilities as the MLAT, they produced a different pattern of correlations. Surprisingly, there were no correlations between the LLAMA tests and either d' scores or the P600 responses to number violations that demonstrated correlations with the MLAT. However, the LLAMA tests were positively correlated with mean difference amplitudes in posterior regions during responses to gender violations, indicating that verbal aptitude may also play a role in the processing of features that are unique to the L2. Neither correlations nor differences in results between the MLAT and LLAMA tests have been established in previous studies; therefore, it is difficult to interpret their differences with regard to the results of the current study. However, the one notable difference is the lack of dependence on the L1 English in the LLAMA tests. The MLAT results, then, may lend further support to the claims that adult learners rely heavily on their knowledge of their L1 in early stages of L2 acquisition, particularly with regard to features that may transfer from the L1[12]. However, the results of the LLAMA tests reported here, admittedly based on a very small sample, indicate at least a possibility that learners may also take advantage of *domain-specific* abilities that critically *do not rely on the L1* as they attempt to process unique L2 features, providing further support for theories of full access to UG. These findings show that further research with larger samples is certainly needed when it comes to correlations between aptitude tests and brain responses, particularly with regard to the nature of available aptitude batteries.

[12] The correlation of only responses to violations of the number feature, which is present in the L1, to the MLAT, which is L1-dependent, also lend support to Sagarra (2007), whose results were described in Chapter 3. Along with Service et al. (2002), Sagarra claims that L2-based tests, as opposed to L1-based tests, may be the best measure of individual differences impacting the L2.

11.2.2 Nonverbal Aptitude

The role of nonverbal aptitude, or domain-general capacities, is investigated in response to research question 4. The *Fundamental Difference Hypothesis* (Bley-Vroman, 1989, 1990, 2009) predicts such a role, while it could be argued that a role for domain-specific factors would be more strongly predicted under full-access theories. Indeed, the results of the current study find no evidence that nonverbal aptitude plays a role, even in the acquisition of unique L2 features, as seen in the lack of correlations between the RAVEN and either d' scores or ERP responses. While the matter cannot be settled on the basis of one test, it may be useful to consider here the nature of that test[13]. The RAVEN was chosen for the study due to its broad use in cognitive studies as a general measure of nonverbal intelligence and mental reasoning. Carpenter, Just, and Shell (1990) cite research into correlations between the RAVEN and other intelligence measures (Court and Raven, 1982; Jensen, 1987) suggesting that the processing that underlies the RAVEN may not be specific just to that test, but rather, general in nature and likely more central to the testing of domain-general analytic ability than a number of other tests, confirming the analysis of Snow, Kyllonen, and Marshalek (1984) that was cited previously. Carpenter et al. also report a detailed analysis of factors relevant to performance on the RAVEN. Their findings are framed in terms of two simulated models of RAVEN performance, one for median performance and one for best performance, based on the verbal protocols, eye-fixation patterns, and errors of 79 students recorded in a series of experiments involving the RAVEN test. Results indicate that the ability to decompose problems into smaller parts is central to analytic ability, and that variation in RAVEN test scores arises from individual differences in working memory and abstract reasoning. It is thus surprising that if domain-general capacities are at work, as argued by Bley-

[13] Indeed, Robinson (2002) argues that tests of individual differences will be correlated with L2 performance only where they tap the same constructs as the L2 tasks involved.

Vroman (1990, 2009), the RAVEN would not in some way reflect morphological decomposition, maintenance of features in verbal working memory, and abstraction of patterns resulting from agreement.

One other comment should be made with regard to the lack of correlation between nonverbal aptitude and P600 amplitudes, particularly in light of the correlations that were indeed found between P600 amplitudes and measures of verbal aptitude. It was mentioned earlier that Coulson et al. (1998) have argued that the P600 is one of a family of positivities that are sensitive to probability and are found in response to non-linguistic stimuli as well. Given the finding that at least one property of the P600, that is, mean amplitude, is modulated by verbal but not nonverbal aptitude, it seems likely that the P600 may at least include a domain-specific response, more in line with arguments by Osterhout (1999). This finding is only exploratory, however, and the nature of the P600 certainly merits further investigation.

Even though RAVEN scores were not found to be correlated with mean amplitude differences, caution must be exercised in ruling out altogether a role for nonverbal aptitude in L2 processing. It should be noted that mean amplitude differences are only one property of brain responses to language and may not constitute the best test of nonverbal aptitude’s relationship to L2 processing. Indeed, studies in neurocognitive research indicate that general intelligence as measured by the RAVEN may be related to processing speed. For example, Thoma et al. (2000, 2006) found RAVEN scores to be correlated with processing speed in a test of sensory-motor control in response to visual stimuli. Their study employed MEG investigation of the correlation of intelligence (RAVEN scores) with four measures of processing speed. Participants were asked to make a decision related to the location of visual stimuli on a screen and indicate their response by finger lift. The four measures that were recorded included the latency of the visual M100 (a well-documented early brain response to differences in visual stimuli), Visual Motor Integration (VMI; latency of a motor response related to readying a muscle for movement), Pre-Response Sensory-Motor (PRSM; latency of a motor dipole preceding the finger lift), and After-Response Sensory-Motor (ARSM; latency of a motor dipole immediately following finger lift). In decision conditions (versus a baseline condition in which no decision had to be made because the same finger was always lifted), the M100 latency was positively correlated with RAVEN, which was interpreted as indexing the need to record the position of stimuli in the decision conditions. VMI and ARSM were negatively correlated with RAVEN, indicating that greater intelligence and faster processing speed are linked, at least for late visual-motor processes. Further research could be useful in determining whether there is a link between nonverbal aptitude and the speed of linguistic processing, as measured by the onset latencies of responses like the P600 to agreement violations.

11.3 Conclusion

This study contributes to the growing body of literature investigating the role of morphosyntactic features in L2 processing. In particular, the study tests number and gender agreement across phrases in well-controlled contexts, allowing the comparison of featural instantiations that vary parametrically in their similarity to the L1. The development of native-like processing was evident for gender agreement on adjectives, even though grammatical gender agreement is not present in the L1. Furthermore, in contrast to earlier ERP studies (e.g., Tokowicz and MacWhinney, 2005; Frenck-Mestre et al., 2010), results showed that learners were sensitive to number agreement violations on both verbs and adjectives, despite differences between these two categories in the realization of number in the L1. These results provide evidence that adult learners even at low proficiency can exhibit development of native-like processing of (a) uninterpretable features that are not present in their L1, as well as (b) novel instantiations of features that are shared between the L1 and L2, supporting claims of full-access theories (e.g., Schwartz and Sprouse, 1994, 1996). With regard to individual differences between learners, the study does not find any evidence thus far that domain-general aptitude is predictive of success in adult L2 acquisition (contra Bley-Vroman, 1990). However, correlations were present between verbal aptitude scores and measures of sensitivity to agreement violations, providing evidence that domain-specific processes are at work in L2 acquisition, even with regard to features unique to the L2. The finding that verbal aptitude scores can be correlated to a specific property of brain responses to language opens the door to further research using ERP to investigate the role of individual differences in L2 processing.

REFERENCES

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

Alemán Bañón, J., Fiorentino, R., & Gabriele, A. (2012). The processing of number and gender agreement in Spanish: An event-related potential investigation of the effects of structural distance. *Brain Research*, *1456,* 49-63.

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

Bialystok, E. (2002). On the reliability of robustness: A reply to DeKeyser. *Studies in Second Language Acquisition, 24,* 481-488.

Bialystok, E., & Hakuta, K. (1999). Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 161–181). Mahwah, NJ: Erlbaum.

Birdsong, D. (1999). Introduction: Whys and why nots of the Critical Period Hypothesis. In D. Birdsong (Ed.), *Second language acquisition and the Critical Period Hypothesis* (pp. 1–22). Mahwah, NJ: Erlbaum.

Birdsong, D. (2005). Interpreting age effects in second language acquisition. In J. Kroll & A. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic perspectives.* New York: Cambridge University Press.

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

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

Bley-Vroman, R. (2009). The evolving context of the Fundamental Difference Hypothesis. *Studies in Second Language Acquisition, 31*, 175-198.

Bowden, H.W., Sanz, C.S., Steinhauer, K., & Ullman, M.T. (2007). An ERP study of proficiency in second language. *Journal of Cognitive Neuroscience, 170*, supplement.

Carreiras, M, Salillas, E, & Barber, H. A. (2004). Event-related potentials elicited during parsing of ambiguous relative clauses in Spanish. *Cognitive Brain Research*, *20*, 98-105.

Carroll, J. B. (1965). The prediction of success in foreign language training. In R. Glaser (Ed.), *Training, research, and education.* New York: Wiley.

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

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

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

Carstens, V. (2000). Concord in minimalist theory. *Linguistic Inquiry, 31,* 319–355.

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

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

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

Chomsky, N. (1995). *The minimalist program.* Cambridge, MA: MIT Press.

Chomsky, N. (2000). Minimalist inquiries: The framework. In R. Martin, D. Michaels, & J. Uriagereka (Eds.), Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik (pp. 89-155). Cambridge: MIT Press.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27,* 3–42.

Clahsen, H., & Muysken, P. (1986). The availability of universal grammar to adult and child learners: A study of the acquisition of German word order. *Second Language Research, 2,* 93-119.

Clahsen, H., Felser, C., Neubauer, K., Sato, M., & Silva, R. (2010). Morphological structure in native and nonnative language processing. *Language Learning, 60*, 21-43.

Corbett, G. G. (1991). *Gender*. New York: Cambridge University Press.

Coulson, S., King, J.W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes, 13,* 21-58.

Court, J. H., & Raven, J. (1982). *Manual for Raven's Progressive Matrices and Vocabulary Scales, Research Supplement No. 2* (Pt. 3, Section 7). London: H. K. Lewis.

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

Diller, K. (1981). ‘Natural methods’ of foreign language teaching: Can they exist? What criteria must they meet? In Winnitz, H. (ed.). *Native Language and Foreign Language Acquisition*. New York: The New York Academy of Sciences.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Epstein, S. D., Flynn, S., & Martohardjono, G. (1996). Second language acquisition: Theoretical and experimental issues in contemporary research. *Behavioral and Brain Sciences, 19,* 677-714.

Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics, 24*, 453-489.

Fiebach, C. J., Schlesewsky, M. & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: the processing of German WH-questions. *Journal of Memory and Language, 47*, 250-272.

Flahive, D.E. (1980). Separating the *g* factor from reading comprehension. In J. W. Oller, Jr., & K. Perkins (Eds.), *Research in Language Testing* (pp. 34-46). Rowley, MA: Newbury House.

Foucart, A., & Frenck-Mestre, C. (2010). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1-L2 syntactic similarity. *Bilingualism, Language and Cognition*.

Franceschina, F. (2001). Morphological or syntactic deficits in near-native speakers? An assessment of some current proposals. *Second Language Research, 17*, 213-247.

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

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

Gardner, R., & MacIntyre, P. (1992). A student's contributions to second language learning. Part 1: Cognitive variables. *Language Teaching, 25,* 211–220.

Gardner, R. C., & Lambert, W. E. (1965). Language aptitude, intelligence, and second language achievement. *Journal of Educational Psychology, 56,* 191-199.

Gibson, E., & Pearlmutter, N. (1998). Constraints on sentence comprehension. *Trends in Cognitive Sciences, 2*, 262-268.

Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes, 25,* 149-188.

Gunter, T.C., Friederici, A.D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience, 12,* 556-568.

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences, 9,* 416-423.

Hagoort, P. & Brown, C.M. (2000). ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia, 38*, 1531-1549.

Hagoort, P., Brown, C.M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, *8*, 439-484.

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

Hahne, A., Eckstein, K., & Friederici, A. D. (2004). Brain signatures of syntactic and semantic processes during children's language development. *Journal of Cognitive Neuroscience, 16*, 1302-1318.

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

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

Harris, J. (1991). The exponence of gender in Spanish. *Linguistic Inquiry, 22*, 27–62.

Hawkins, R. (1998). Explaining the difficulty of French gender attribution for speakers of English. Paper presented at the 8th annual EUROSLA conference, Paris.

Hawkins, R., & Chan, C. Y.-H. (1997). The partial availability of Universal Grammar in second language acquisition: The "failed functional features hypothesis." *Second Language Research, 13*, 187–226.

Haznedar, B. (1997). L2 acquisition by a Turkish-speaking child: Evidence for L1 influence. In E. Hughes, M. Hughes, & A. Greenhill (Eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development* (pp. 245-256). Somerville, MA: Cascadilla Press.

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes, 5*, 281-312.

Ioup, G., Boustagui, E., El Tigi, M., & Moselle, M. (1994). Reexamining the critical period hypothesis: A case study of successful adult SLA in a naturalistic environment. *Studies in Second Language Acquisition, 16,* 73–98.

Jensen, A. R. (1987). The *g* beyond factor analysis. In R. R. Ronning, J. A. Glover, J. C. Conoley, & J. C. Witt (Eds.), *The influence of cognitive psychology on testing* (pp. 87-142). Hillsdale, NJ: Erlbaum.

Johnson J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21,* 60-99.

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

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

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

King, J.W., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7*, 376-395.

Kluender, R. & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience, 5*, 196-214.

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

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

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

Lau, E., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature, 9*, 920-933.

Lenneberg, E. H. (1967). *Biological Foundations of Language.* New York: Wiley.

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, **12,** 251–285.

Marinova-Todd, S. H., Marshall, D. B., & Snow, C. E. (2000). Three misconceptions about age and L2 learning. *TESOL Quarterly, 34*, 9-34.

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

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

McLaughlin, J., Tanner, D., Pitkänen, I., Frenck-Mestre, C., Inoue, K., Valentine, G., & Osterhout, L. (2010). Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning, 60 (Suppl. 2),* 123-150.

Meara, P., Milton, J., & Lorenzo-Dus, N. (2003). *Swansea language aptitude tests (LAT) v2.0*. Swansea, UK: Lognostics.

Meara, P. (2005). *LLAMA language aptitude tests.* Swansea, UK: Lognostics.

Miyake, A. & Friedman, N. F. (1998). Individual differences in second language proficiency: Working memory as language "aptitude". In A. F. Healy & L. E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training* *and retention* (pp. 339-364). Mahwah, NJ: Erlbaum.

Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex, 97*, 908-930.

Montrul, S., Foote, R., & Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: the effects of age and context of acquisition. *Language Learning, 58,* 503-553.

Montrul, S., & Slabakova, R. (2003). Competence similarities between native and near-native speakers: An investigation of the preterite/imperfect contrast in Spanish. *Studies in Second Language Acquisition, 25,* 351-398.

Nevins, A., Dillon, B., Malhotra, S., & Phillips, C. (2007). The role of feature-number and feature-type in processing Hindi verb agreement violations. *Brain Research, 1164*, 81-94.

Novoa, L., Fein, D., & Obler, L. (1988). Talent in foreign languages: A case study. In L. Obler, & D. Fein (Eds.), *The Exceptional Brain: Neuropsychology of Talent and Special Abilities.* New York: Guilford Press.

Obler, L. (1989). Exceptional second language learners. In S. Gass, C. Madden, D. Preston, & L. Selinker (Eds.), *Variation in Second Language Acquisition: Psycholinguistic Issues*. Clevedon, Avon: Multilingual Matters.

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience, 17*, 1212-28.

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

Ortega, L. (2009). *Understanding Second Language Acquisition*. London: Hodder Education.

Osterhout, L. (1999). A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King and Kutas (1998) in the P600/SPS-P300 debate. *Language and Cognitive Processes, 14,* 1-14.

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

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

Osterhout, L., McKinnon, R., Bersick, M., & Corey, V. (1996). On the language-specificity of the brain response to syntactic anomalies: Is the syntactic positive shift a member of the P300 family? *Journal of Cognitive Neuroscience*, *8*, 507–526.

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* (pp. 271–308). New York: Psychology Press.

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

Osterhout, L., & Nicol, J. (1999). On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes, 14*, 283-317.

Pakulak, E., & Neville, H. J. (2004). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience, 22*, 2728-2744.

Penfield, W. & Roberts, L. (1959). *Speech and Brain Mechanisms.* Princeton, NJ: Princeton University Press.

Phillips, C., Kazanina, N., Abada, S.H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research, 22,* 407-428.

Prabhakaran, V., Smith, J. A. L., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1997). Neural substrates of fluid reasoning: An fMRI study of neocortical activation during performance of the Raven's Progressive Matrices test. *Cognitive Psychology, 33,* 43-63.

Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research, 16,* 103-133.

Pulvermüller, F., & Schumann, J. H. (1994). Neurobiological mechanisms of language acquisition. *Language Learning* 44 (4): 681–734.

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

Renaud, C. (2010). On the nature of agreement in English-French acquisition: A processing investigation in the verbal and nominal domains. PhD. Dissertation, Indiana University.

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

Rossi, 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.

Rossi, S., Gugler, M. F., Hahne, A., & Friederici, A. D. (2005). When word category information encounters morphosyntax: An ERP study. *Neuroscience Letters, 384*, 228–233.

Rugg, M. D. (1985). The effects of semantic priming and word repetition on event-related potentials. *Psychophysiology, 22*, 642–647.

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

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

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

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

Schneiderman, E., & Desmarais, C. (1988a). The talented language learner: Some preliminary findings. *Second Language Research, 4,* 91-109.

Schneiderman, E., & Desmarais, C. (1988b). A neuropsychological substrate for talent in second language acquisition. In L. Obler & D. Fein (Eds.), *The Exceptional Brain: Neuropsychology of Talent and Special Abilities.* New York: Guilford Press.

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

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

Seliger, H. W. (1978). Implications of a multiple critical periods hypothesis for second language learning. In Ritchie, W. C. (ed.). *Second Language Acquisition Research: Issues and Implications*. New York: Academic Press.

Service, E., Simola, M., Metsanheimo, O., & Maury, S. (2002). Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology, 14,* 383–408.

Skehan, P. (1986a). The role of foreign language aptitude in a model of school learning. *Language Testing, 3*, 188–221.

Skehan, P. (1986b). Where does language aptitude come from? In P. Meara (Ed.), *Spoken Language*. London: Centre for Information on Language Teaching.

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

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

Slabakova, R. (2000). L1 transfer revisited: The L2 acquisition of telicity marking in English by Spanish and Bulgarian native speakers. *Linguistics, 38,* 739-770.

Smith, N., & Tsimpli, X. (1995). *The Mind of a Savant.* Oxford, UK: Blackwell.

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

Steinhauer, K., White, E., Cornell, S., Genesee, F., & White, L. (2006). The neural dynamics of second language acquisition: evidence from Event-Related Potentials. *Journal of Cognitive Neuroscience, 99,* supplement.

Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research, 25, 13–41*.

Sternberg, R. J. (2002). The theory of successful intelligence and its implications for language-aptitude testing. In P. Robinson (Ed.), *Individual Differences and Instructed Language Learning* (pp. 13-44). Amsterdam: John Benjamins.

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

Tanner, D., McLaughlin, J., Herschensohn, J, & Osterhout, L. (2012). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition*, 1-16. Available on CJO2012. doi:10.1017/S1366728912000302.

The *Acceso* Project. Available: http://www2.ku.edu/~spanish/acceso/.

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

Thoma, R. J., Yeo, R. A., Gangestad, S., Halgren, E., Davis, J., Paulson, K. M., & Lewine, J. D. (2006). Developmental instability and the neural dynamics of the speed-intelligence relationship. *NeuroImage, 32*, 1456-1464.

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

Thurstone, L. L., & Thurstone, T. G. (1941). *The Chicago Tests of Primary Mental Abilities, Ages 11-17, Manual*. Washington, D.C.: American Council on Education.

Thurstone, L. L., & Thurstone, T. G. (1965). *Primary Mental Abilities Test*. Chicago: Science Research Associates.

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

Tsimpli, I. M., & Dimitrakopoulou, M. (2007). The Interpretability Hypothesis: Evidence from *wh*-interrogatives in second language acquisition. *Second Language Research, 23,* 215-242.

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

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition, 18*, 380-393.

Vos, S.H., Gunter, T.C., Herman, H., Kolk, H.J., & Mulder, G. (2001). Working memory constraints on syntactic processing: An electrophysiological investigation. *Psychophysiology, 38*, 41-63.

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

Wesche, M. B., Edwards, H., & Wells, W. (1982). Foreign language aptitude and intelligence. *Applied Psycholinguistics, 3,* 127-140.

White, L. (1985). The pro-drop parameter in adult second language acquisition. *Language Learning, 35,* 47-62.

White, L. (1986). Implications of parametric variation for adult second language acquisition: An investigation of the 'pro-drop' parameter. In V. Cook (Ed.), *Experimental Approaches to Second Language Acquisition* (pp. 55-72). Oxford, UK: Pergamon Press.

White, L. (2003). *Second Language Acquisition and Universal Grammar.* Cambridge, UK: Cambridge University Press.

White, L., & Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research, 12 (3),* 233-265.

White, L., & Juffs, A. (1998). Constraints on wh-movement in two different contexts of non-native language acquisition: competence and processing. In S. Flynn, G. Martohardjono, & W. O'Neil (Eds.), *The Generative Study of Second Language Acquisition* (pp. 111-129). Mahwah, NJ: Lawrence Erlbaum.

White, L., Valenzuela, E., Kozlowska-Macgregor, M. & Leung, Y.-K.I. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics, 25,* 105–33.

Yuan, B. (1998). Interpretation of binding and orientation of the Chinese reflexive *ziji* by English and Japanese speakers. *Second Language Research, 14,* 324-340.

APPENDIX 1: Stimuli for Noun-Adjective Agreement Conditions

1. El Mediterráneo es profundo y el Adriático también.
2. El Mediterráneo es *profundos y el Adriático también.
3. El Mediterráneo es *profunda y el Adriático también.

4. El arroyo es profundo y el charco también.
5. El arroyo es *profundos y el charco también.
6. El arroyo es *profunda y el charco también.

7. El Mediterráneo es tranquilo y el Caspio también.
8. El Mediterráneo es *tranquilos y el Caspio también.
9. El Mediterráneo es *tranquila y el Caspio también.

10. La boda es tranquila y la gala también.
11. La boda es *tranquilas y la gala también.
12. La boda es *tranquilo y la gala también.

13. La naranja es redonda y la sandía también.
14. La naranja es *redondas y la sandía también.
15. La naranja es *redondo y la sandía también.

16. La Tierra es redonda y la Luna también.
17. La Tierra es *redondas y la Luna también.
18. La Tierra es *redondo y la Luna también.

19. El Atlántico es gigantesco y el Egeo también.
20. El Atlántico es *gigantescos y el Egeo también.
21. El Atlántico es *gigantesca y el Egeo también.

22. La Tierra es gigantesca y la atmósfera también.
23. La Tierra es *gigantescas y la atmósfera también.
24. La Tierra es *gigantesco y la atmósfera también.

25. El colegio es gratuito y el instituto también.
26. El colegio es *gratuitos y el instituto también.
27. El colegio es *gratuita y el instituto también.

28. La biblioteca es gratuita y la sauna también.
29. La biblioteca es *gratuitas y la sauna también.
30. La biblioteca es *gratuito y la sauna también.

31. El colegio es antiguo y el liceo también.
32. El colegio es *antiguos y el liceo también.
33. El colegio es *antigua y el liceo también.

34. La huella es antigua y la calavera también.
35. La huella es *antiguas y la calavera también.
36. La huella es *antiguo y la calavera también.

37. La isla es preciosa y la bahía también.
38. La isla es *preciosas y la bahía también.
39. La isla es *precioso y la bahía también.

40. El espejo es precioso y el mosaico también.
41. El espejo es *preciosos y el mosaico también.
42. El espejo es *preciosa y el mosaico también.

43. La isla es rocosa y la península también.
44. La isla es *rocosas y la península también.
45. La isla es *rocoso y la península también.

46. La montaña es rocosa y la sierra también.
47. La montaña es *rocosas y la sierra también.
48. La montaña es *rocoso y la sierra también.

49. La iglesia es emblemática y la abadía también.
50. La iglesia es *emblemáticas y la abadía también.
51. La iglesia es *emblemático y la abadía también.

52. El Coliseo es emblemático y el Foro también.
53. El Coliseo es *emblemáticos y el Foro también.
54. El Coliseo es *emblemática y el Foro también.

55. La pastelería es famosa y la bombonería también.
56. La pastelería es *famosas y la bombonería también.
57. La pastelería es *famoso y la bombonería también.

58. El Coliseo es famoso y el Capitolio también.
59. El Coliseo es *famosos y el Capitolio también.
60. El Coliseo es *famosa y el Capitolio también.

61. La fresa es ácida y la piña también.
62. La fresa es *ácidas y la piña también.
63. La fresa es *ácido y la piña también.

64. La naranja es ácida y la mandarina también.
65. La naranja es *ácidas y la mandarina también.
66. La naranja es *ácido y la mandarina también.

67. El Atlántico es frío y el Báltico también.
68. El Atlántico es *fríos y el Báltico también.
69. El Atlántico es *fría y el Báltico también.

70. El otoño es frío y el invierno también.
71. El otoño es *fríos y el invierno también.
72. El otoño es *fría y el invierno también.

73. La falda es florida y la blusa también.
74. La falda es *floridas y la blusa también.
75. La falda es *florido y la blusa también.

76. El trópico es florido y el prado también.
77. El trópico es *floridos y el prado también.
78. El trópico es *florida y el prado también.

79. El vestido es azulado y el manto también.
80. El vestido es *azulados y el manto también.
81. El vestido es *azulada y el manto también.

82. La falda es azulada y la corbata también.
83. La falda es *azuladas y la corbata también.
84. La falda es *azulado y la corbata también.

85. La montaña es grandiosa y la colina también.
86. La montaña es *grandiosas y la colina también.
87. La montaña es *grandioso y la colina también.

88. El faro es grandioso y el obelisco también.
89. El faro es *grandiosos y el obelisco también.
90. El faro es *grandiosa y el obelisco también.

91. El faro es luminoso y el fuego también.
92. El faro es *luminosos y el fuego también.
93. El faro es *luminosa y el fuego también.

94. La cocina es luminosa y la terraza también.
95. La cocina es *luminosas y la terraza también.
96. La cocina es *luminoso y la terraza también.

97. La pastelería es minúscula y la panadería también.
98. La pastelería es *minúsculas y la panadería también.
99. La pastelería es *minúsculo y la panadería también.

100. La casa es minúscula y la cochera también.
101. La casa es *minúsculas y la cochera también.
102. La casa es *minúsculo y la cochera también.

103. La maleta es espaciosa y la caja también.
104. La maleta es *espaciosas y la caja también.
105. La maleta es *espacioso y la caja también.

106. La casa es espaciosa y la bodega también.
107. La casa es *espaciosas y la bodega también.
108. La casa es *espacioso y la bodega también.

109. El arroyo es estrecho y el riachuelo también.
110. El arroyo es *estrechos y el riachuelo también.
111. El arroyo es *estrecha y el riachuelo también.

112. El lago es estrecho y el barranco también.
113. El lago es *estrechos y el barranco también.
114. El lago es *estrecha y el barranco también.

115. La cocina es amplia y la entrada también.
116. La cocina es *amplias y la entrada también.
117. La cocina es *amplio y la entrada también.

118. El baño es amplio y el pasillo también.
119. El baño es *amplios y el pasillo también.
120. El baño es *amplia y el pasillo también.

121. El lago es oscuro y el pozo también.
122. El lago es *oscuros y el pozo también.
123. El lago es *oscura y el pozo también.

124. El otoño es oscuro y el invierno también.
125. El otoño es *oscuros y el invierno también.
126. El otoño es *oscura y el invierno también.

127. La maleta es pesada y la cartera también.
128. La maleta es *pesadas y la cartera también.
129. La maleta es *pesado y la cartera también.

130. La mesa es pesada y la silla también.
131. La mesa es *pesadas y la silla también.
132. La mesa es *pesado y la silla también.

133. La mesa es inmensa y la cama también.
134. La mesa es *inmensas y la cama también.
135. La mesa es *inmenso y la cama también.

136. El submarino es inmenso y el pesquero también.
137. El submarino es *inmensos y el pesquero también.
138. El submarino es *inmensa y el pesquero también.

139. El baño es rosado y el dormitorio también.
140. El baño es *rosados y el dormitorio también.
141. El baño es *rosada y el dormitorio también.

142. El narciso es rosado y el gladiolo también.
143. El narciso es *rosados y el gladiolo también.
144. El narciso es *rosada y el gladiolo también.

145. El disco es clásico y el concierto también.
146. El disco es *clásicos y el concierto también.
147. El disco es *clásica y el concierto también.

148. El abrigo es clásico y el chaleco también.
149. El abrigo es *clásicos y el chaleco también.
150. El abrigo es *clásica y el chaleco también.

151. El abrigo es caluroso y el gorro también.
152. El abrigo es *calurosos y el gorro también.
153. El abrigo es *calurosa y el gorro también.

154. El trópico es caluroso y el desierto también.
155. El trópico es *calurosos y el desierto también.
156. El trópico es *calurosa y el desierto también.

157. La biblioteca es moderna y la escuela también.
158. La biblioteca es *modernas y la escuela también.
159. La biblioteca es *moderno y la escuela también.

160. La cámara es moderna y la agenda también.
161. La cámara es *modernas y la agenda también.
162. La cámara es *moderno y la agenda también.

163. La batalla es destructiva y la lucha también.
164. La batalla es *destructivas y la lucha también.
165. La batalla es *destructivo y la lucha también.

166. El terremoto es destructivo y el tornado también.
167. El terremoto es *destructivos y el tornado también.
168. El terremoto es *destructiva y el tornado también.

169. La batalla es catastrófica y la guerra también.
170. La batalla es *catastróficas y la guerra también.
171. La batalla es *catastrófico y la guerra también.

172. El terremoto es catastrófico y el maremoto también.
173. El terremoto es *catastróficos y el maremoto también.
174. El terremoto es *catastrófica y el maremoto también.

175. El plátano es beneficioso y el coco también.
176. El plátano es *beneficiosos y el coco también.
177. El plátano es *beneficiosa y el coco también.

178. La fresa es beneficiosa y la zanahoria también.
179. La fresa es *beneficiosas y la zanahoria también.
180. La fresa es *beneficioso y la zanahoria también.

181. El libro es didáctico y el periódico también.
182. El libro es *didácticos y el periódico también.
183. El libro es *didáctica y el periódico también.

184. El diccionario es didáctico y el glosario también.
185. El diccionario es *didácticos y el glosario también.
186. El diccionario es *didáctica y el glosario también.

187. El libro es anónimo y el artículo también.
188. El libro es *anónimos y el artículo también.
189. El libro es *anónima y el artículo también.

190. El cuento es anónimo y el manuscrito también.
191. El cuento es *anónimos y el manuscrito también.
192. El cuento es *anónima y el manuscrito también.

193. La guitarra es melodiosa y la flauta también.
194. La guitarra es *melodiosas y la flauta también.
195. La guitarra es *melodioso y la flauta también.

196. La samba es melodiosa y la rumba también.
197. La samba es *melodiosas y la rumba también.
198. La samba es *melodioso y la rumba también.

199. La cortina es bonita y la moqueta también.
200. La cortina es *bonitas y la moqueta también.
201. La cortina es *bonito y la moqueta también.

202. La guitarra es bonita y la armónica también.
203. La guitarra es *bonitas y la armónica también.
204. La guitarra es *bonito y la armónica también.

205. El laboratorio es conocido y el departamento también.
206. El laboratorio es *conocidos y el departamento también.
207. El laboratorio es *conocida y el departamento también.

208. El disco es conocido y el grupo también.
209. El disco es *conocidos y el grupo también.
210. El disco es *conocida y el grupo también.

211. El laboratorio es privado y el archivo también.
212. El laboratorio es *privados y el archivo también.
213. El laboratorio es *privada y el archivo también.

214. La conferencia es privada y la fiesta también.
215. La conferencia es *privadas y la fiesta también.
216. La conferencia es *privado y la fiesta también.

217. La película es pedagógica y la novela también.
218. La película es *pedagógicas y la novela también.
219. La película es *pedagógico y la novela también.

220. El diccionario es pedagógico y el tesauro también.
221. El diccionario es *pedagógicos y el tesauro también.
222. El diccionario es *pedagógica y el tesauro también.

223. El empleo es patético y el sueldo también.
224. El empleo es *patéticos y el sueldo también.
225. El empleo es *patética y el sueldo también.

226. La película es patética y la crítica también.
227. La película es *patéticas y la crítica también.
228. La película es *patético y la crítica también.

229. El narciso es decorativo y el lirio también.
230. El narciso es *decorativos y el lirio también.
231. El narciso es *decorativa y el lirio también.

232. La cortina es decorativa y la alfombra también.
233. La cortina es *decorativas y la alfombra también.
234. La cortina es *decorativo y la alfombra también.

235. El espejo es plateado y el florero también.
236. El espejo es *plateados y el florero también.
237. El espejo es *plateada y el florero también.

238. La espada es plateada y la diadema también.
239. La espada es *plateadas y la diadema también.
240. La espada es *plateado y la diadema también.

241. La cafetera es metálica y la tetera también.
242. La cafetera es *metálicas y la tetera también.
243. La cafetera es *metálico y la tetera también.

244. La ventana es metálica y la puerta también.
245. La ventana es *metálicas y la puerta también.
246. La ventana es *metálico y la puerta también.

247. El gramófono es viejo y el cronómetro también.
248. El gramófono es *viejos y el cronómetro también.
249. El gramófono es *vieja y el cronómetro también.

250. La cafetera es vieja y la tostadora también.
251. La cafetera es *viejas y la tostadora también.
252. La cafetera es *viejo y la tostadora también.

253. La ventana es amarilla y la escalera también.
254. La ventana es *amarillas y la escalera también.
255. La ventana es *amarillo y la escalera también.

256. El plátano es amarillo y el mango también.
257. El plátano es *amarillos y el mango también.
258. El plátano es *amarilla y el mango también.

259. La cámara es automática y la calculadora también.
260. La cámara es *automáticas y la calculadora también.
261. La cámara es *automático y la calculadora también.

262. El gramófono es automático y el estéreo también.
263. El gramófono es *automáticos y el estéreo también.
264. El gramófono es *automática y el estéreo también.

265. La corona es dorada y la cadena también.
266. La corona es *doradas y la cadena también.
267. La corona es *dorado y la cadena también.

268. La espada es dorada y la lanza también.
269. La espada es *doradas y la lanza también.
270. La espada es *dorado y la lanza también.

271. La corona es auténtica y la sortija también.
272. La corona es *auténticas y la sortija también.
273. La corona es *auténtico y la sortija también.

274. El cuadro es auténtico y el grabado también.
275. El cuadro es *auténticos y el grabado también.
276. El cuadro es *auténtica y el grabado también.

277. El motociclismo es costoso y el automovilismo también.
278. El motociclismo es *costosos y el automovilismo también.
279. El motociclismo es *costosa y el automovilismo también.

280. El cuadro es costoso y el retrato también.
281. El cuadro es *costosos y el retrato también.
282. El cuadro es *costosa y el retrato también.

283. El empleo es prestigioso y el título también.
284. El empleo es *prestigiosos y el título también.
285. El empleo es *prestigiosa y el título también.

286. La academia es prestigiosa y la galería también.
287. La academia es *prestigiosas y la galería también.
288. La academia es *prestigioso y la galería también.

289. La academia es pública y la guardería también.
290. La academia es *públicas y la guardería también.
291. La academia es *público y la guardería también.

292. La plaza es pública y la avenida también.
293. La plaza es *públicas y la avenida también.
294. La plaza es *público y la avenida también.

295. El cuento es entretenido y el relato también.
296. El cuento es *entretenidos y el relato también.
297. El cuento es *entretenida y el relato también.

298. El motociclismo es entretenido y el judo también.
299. El motociclismo es *entretenidos y el judo también.
300. El motociclismo es *entretenida y el judo también.

301. El camino es corto y el atajo también.
302. El camino es *cortos y el atajo también.
303. El camino es *corta y el atajo también.

304. La conferencia es corta y la entrevista también.
305. La conferencia es *cortas y la entrevista también.
306. La conferencia es *corto y la entrevista también.

307. La samba es erótica y la lambada también.
308. La samba es *eróticas y la lambada también.
309. La samba es *erótico y la lambada también.

310. El tango es erótico y el flamenco también.
311. El tango es *eróticos y el flamenco también.
312. El tango es *erótica y el flamenco también.

313. El tango es rápido y el mambo también.
314. El tango es *rápidos y el mambo también.
315. El tango es *rápida y el mambo también.

316. El submarino es rápido y el hidroplano también.
317. El submarino es *rápidos y el hidroplano también.
318. El submarino es *rápida y el hidroplano también.

319. El contrato es justo y el pago también.
320. El contrato es *justos y el pago también.
321. El contrato es *justa y el pago también.

322. La sentencia es justa y la condena también.
323. La sentencia es *justas y la condena también.
324. La sentencia es *justo y la condena también.

325. El contrato es estricto y el reglamento también.
326. El contrato es *estrictos y el reglamento también.
327. El contrato es *estricta y el reglamento también.

328. La sentencia es estricta y la pena también.
329. La sentencia es *estrictas y la pena también.
330. La sentencia es *estricto y la pena también.

331. La boda es sencilla y la ceremonia también.
332. La boda es *sencillas y la ceremonia también.
333. La boda es *sencillo y la ceremonia también.

334. El vestido es sencillo y el velo también.
335. El vestido es *sencillos y el velo también.
336. El vestido es *sencilla y el velo también.

337. La huella es misteriosa y la reliquia también.
338. La huella es *misteriosas y la reliquia también.
339. La huella es *misterioso y la reliquia también.

340. La iglesia es misteriosa y la cripta también.
341. La iglesia es *misteriosas y la cripta también.
342. La iglesia es *misterioso y la cripta también.

343. El camino es feo y el pueblo también.
344. El camino es *feos y el pueblo también.
345. El camino es *fea y el pueblo también.

346. La plaza es fea y la basílica también.
347. La plaza es *feas y la basílica también.
348. La plaza es *feo y la basílica también.

349. El gobierno es autoritario y el ejército también.
350. El gobierno es *autoritarios y el ejército también.
351. El gobierno es *autoritaria y el ejército también.

352. La aristocracia es autoritaria y la realeza también.
353. La aristocracia es *autoritarias y la realeza también.
354. La aristocracia es *autoritario y la realeza también.

355. El gobierno es poderoso y el parlamento también.
356. El gobierno es *poderosos y el parlamento también.
357. El gobierno es *poderosa y el parlamento también.

358. La aristocracia es poderosa y la burguesía también.
359. La aristocracia es *poderosas y la burguesía también.
360. La aristocracia es *poderoso y la burguesía también.

APPENDIX 2: Stimuli for Subject-Verb Agreement Conditions and Fillers

1. El mecánico astuto trabaja en el garaje.
2. El mecánico astuto *trabajan en el garaje.
3. El mecánico astuto está en el garaje.

4. La embajadora exiliada trabaja en Nueva York.
5. La embajadora exiliada *trabajan en Nueva York.
6. La embajadora exiliada está en Nueva York.

7. El sociólogo ocupado escribe en la oficina.
8. El sociólogo ocupado *escriben en la oficina.
9. El sociólogo ocupado está en la oficina.

10. La novelista prolífica escribe en la cabaña.
11. La novelista prolífica *escriben en la cabaña.
12. La novelista prolífica está en la cabaña.

13. El toro bravo corre en el campo.
14. El toro bravo *corren en el campo.
15. El toro bravo está en el campo.

16. La competidora sueca corre en la pista.
17. La competidora sueca *corren en la pista.
18. La competidora sueca está en la pista.

19. El cirujano rico almuerza en el restaurante.
20. El cirujano rico *almuerzan en el restaurante.
21. El cirujano rico está en el restaurante.

22. La empleada bancaria almuerza en la cafetería.
23. La empleada bancaria *almuerzan en la cafetería.
24. La empleada bancaria está en la cafetería.

25. El niño creativo baila en la escuela.
26. El niño creativo *bailan en la escuela.
27. El niño creativo está en la escuela.

28. La coreógrafa francesa baila en Los Ángeles.
29. La coreógrafa francesa *bailan en Los Ángeles.
30. La coreógrafa francesa está en Los Ángeles.

31. La solista asignada canta en el escenario.
32. La solista asignada *cantan en el escenario.
33. La solista asignada está en el escenario.

34. La monja bondadosa canta en el coro.
35. La monja bondadosa *cantan en el coro.
36. La monja bondadosa está en el coro.

37. El perro descuidado duerme en el garaje.
38. El perro descuidado *duermen en el garaje.
39. El perro descuidado está en el garaje.

40. La psicóloga cansada duerme en el sofá.
41. La psicóloga cansada *duermen en el sofa.
42. La psicóloga cansada está en el sofá.

43. El cocodrilo americano nada en la laguna.
44. El cocodrilo americano *nadan en la laguna.
45. El cocodrilo americano está en la laguna.

46. La ballena blanca nada en el mar.
47. La ballena blanca *nadan en el mar.
48. La ballena blanca está en el mar.

49. El loro colorado vive en la selva.
50. El loro colorado *viven en la selva.
51. El loro colorado está en la selva.

52. La doncella secuestrada vive en la torre.
53. La doncella secuestrada *viven en la torre.
54. La doncella secuestrada está en la torre.

55. El ajo sabroso crece en el jardín.
56. El ajo sabroso *crecen en el jardín.
57. El ajo sabroso está en el jardín.

58. La planta nutritiva crece en la selva.
59. La planta nutritiva *crecen en la selva.
60. La planta nutritiva está en la selva.

61. El anillo lustroso brilla en el estante.
62. El anillo lustroso *brillan en el estante.
63. El anillo lustroso está en el estante.

64. La joya egipcia brilla en el museo.
65. La joya egipcia *brillan en el museo.
66. La joya egipcia está en el museo.

67. El leopardo africano caza en la sabana.
68. El leopardo africano *cazan en la sabana.
69. El leopardo africano está en la sabana.

70. La pantera negra caza en el valle.
71. La pantera negra *cazan en el valle.
72. La pantera negra está en el valle.

73. El académico venezolano estudia en la universidad.
74. El académico venezolano *estudian en la universidad.
75. El académico venezolano está en la universidad.

76. La investigadora meticulosa estudia en el archivo.
77. La investigadora meticulosa *estudian en el archivo.
78. La investigadora meticulosa está en el archivo.

79. El neurólogo pediátrico enseña en la universidad.
80. El neurólogo pediátrico *enseñan en la universidad.
81. El neurólogo pediátrico está en la universidad.

82. La profesora simpática enseña en el instituto.
83. La profesora simpática *enseñan en el instituto.
84. La profesora simpática está en el instituto.

85. El panadero perezoso lee en la cafetería.
86. El panadero perezoso *leen en la cafetería.
87. El panadero perezoso está en la cafetería.

88. La bibliotecaria desocupada lee en la sala.
89. La bibliotecaria desocupada *leen en la sala.
90. La bibliotecaria desocupada está en la sala.

91. El caballero heróico lucha en el castillo.
92. El caballero heróico *luchan en el castillo.
93. El caballero heróico está en el castillo.

94. La armada británica lucha en el Pacífico.
95. La armada británica *luchan en el Pacífico.
96. La armada británica está en el Pacífico.

97. El obispo católico medita en la capilla.
98. El obispo católico *meditan en la capilla.
99. El obispo católico está en la capilla.

100. La abuela piadosa medita en la catedral.
101. La abuela piadosa *meditan en la catedral.
102. La abuela piadosa está en la catedral.

103. El muchacho travieso grita en la calle.
104. El muchacho travieso *gritan en la calle.
105. El muchacho travieso está en la calle.

106. La maestra enojada grita en el pasillo.
107. La maestra enojada *gritan en el pasillo.
108. La maestra enojada está en el pasillo.

109. El soldado herido descansa en el hospital.
110. El soldado herido *descansan en el hospital.
111. El soldado herido está en el hospital.

112. La turista agotada descansa en el hotel.
113. La turista agotada *descansan en el hotel.
114. La turista agotada está en el hotel.

115. El genio artístico pinta en la galería.
116. El genio artístico *pintan en la galería.
117. El genio artístico está en la galería.

118. La retratista boliviana pinta en el estudio.
119. La retratista boliviana *pintan en el estudio.
120. La retratista boliviana está en el estudio.

121. El alumno aburrido dibuja en su cuaderno.
122. El alumno aburrido *dibujan en su cuaderno.
123. El alumno aburrido está en su escritorio.

124. La artista talentosa dibuja en el parque.
125. La artista talentosa *dibujan en el parque.
126. La artista talentosa está en el parque.

127. El equipo colombiano juega en el estadio.
128. El equipo colombiano *juegan en el estadio.
129. El equipo colombiano está en el estadio.

130. La tenista japonesa juega en Nueva York.
131. La tenista japonesa *juegan en Nueva York.
132. La tenista japonesa está en Nueva York.

133. El músico nervioso fuma en el camerino.
134. El músico nervioso *fuman en el camerino.
135. El músico nervioso está en el camerino.

136. La administradora estresada fuma en su oficina.
137. La administradora estresada *fuman en su oficina.
138. La administradora estresada está en su oficina.

139. El novio abandonado llora en su habitación.
140. El novio abandonado *lloran en su habitación.
141. El novio abandonado está en su habitación.

142. La viuda deprimida llora en el cementerio.
143. La viuda deprimida *lloran en el cementerio.
144. La viuda deprimida está en el cementerio.

145. El aficionado expulsado protesta en la entrada.
146. El aficionado expulsado *protestan en la entrada.
147. El aficionado expulsado está en la entrada.

148. La traidora sentenciada protesta en la corte.
149. La traidora sentenciada *protestan en la corte.
150. La traidora sentenciada está en la corte.

151. El aventurero chileno esquía en los Andes.
152. El aventurero chileno *esquían en los Andes.
153. El aventurero chileno está en los Andes.

154. La deportista suiza esquía en los Alpes.
155. La deportista suiza *esquían en los Alpes.
156. La deportista suiza está en los Alpes.

157. El abogado australiano practica en Buenos Aires.
158. El abogado australiano *practican en Buenos Aires.
159. El abogado australiano está en Buenos Aires.

160. La oradora inquieta practica en el auditorio.
161. La oradora inquieta *practican en el auditorio.
162. La oradora inquieta está en el auditorio.

163. El auto nuevo ruge en la pista.
164. El auto nuevo *rugen en la pista.
165. El auto nuevo está en la pista.

166. La turbina anticuada ruge en el hangar.
167. La turbina anticuada *rugen en el hangar.
168. La turbina anticuada está en el hangar.

169. El farmacéutico argentino reside en Buenos Aires.
170. El farmacéutico argentino *residen en Buenos Aires.
171. El farmacéutico argentino está en Buenos Aires.

172. La directora excéntrica reside en Los Ángeles.
173. La directora excéntrica *residen en Los Ángeles.
174. La directora excéntrica está en Los Ángeles.

175. El asesino contratado espera en el vestíbulo.
176. El asesino contratado *esperan en el vestíbulo.
177. El asesino contratado está en el vestíbulo.

178. La prisionera condenada espera en su celda.
179. La prisionera condenada *esperan en su celda.
180. La prisionera condenada está en su celda.

181. El barco vacío flota en la laguna.
182. El barco vacío *flotan en la laguna.
183. El barco vacío está en la laguna.

184. La balsa perdida flota en el río.
185. La balsa perdida *flotan en el río.
186. La balsa perdida está en el río.

187. El paso peligroso comienza en el cañón.
188. El paso peligroso *comienzan en el cañón.
189. El paso peligroso está en el cañón.

190. La carretera montañosa comienza en los Pirineos.
191. La carretera montañosa *comienzan en los Pirineos.
192. La carretera montañosa está en los Pirineos.

193. El sendero romántico termina en el jardín.
194. El sendero romántico *terminan en el jardín.
195. El sendero romántico está en el jardín.

196. La ruta arriesgada termina en la sierra.
197. La ruta arriesgada *terminan en la sierra.
198. La ruta arriesgada está en la sierra.

199. El teléfono ruidoso suena en el vestíbulo.
200. El teléfono ruidoso *suenan en el vestíbulo.
201. El teléfono ruidoso está en el vestíbulo.

202. La campana hermosa suena en el campanario.
203. La campana hermosa *suenan en el campanario.
204. La campana hermosa está en el campanario.

205. El espectro espantoso aparece en la torre.
206. El espectro espantoso *aparecen en la torre.
207. El espectro espantoso está en la torre.

208. La tabla informativa aparece en el apéndice.
209. La tabla informativa *aparecen en el apéndice.
210. La tabla informativa está en el apéndice.

211. El ogro adormilado ronca en la cueva.
212. El ogro adormilado *roncan en la cueva.
213. El ogro adormilado está en la cueva.

214. La oficinista dormida ronca en su asiento.
215. La oficinista dormida *roncan en su asiento.
216. La oficinista dormida está en su asiento.

217. La arqueóloga fanática acampa en el Sáhara.
218. La arqueóloga fanática *acampan en el Sáhara.
219. La arqueóloga fanática está en el Sáhara.

220. El montañero peruano acampa en los Andes.
221. El montañero peruano *acampan en los Andes.
222. El montañero peruano está en los Andes.

223. El rabino sincero ora en la sinagoga.
224. El rabino sincero *oran en la sinagoga.
225. El rabino sincero está en la sinagoga.

226. La pecadora arrepentida ora en la catedral.
227. La pecadora arrepentida *oran en la catedral.
228. La pecadora arrepentida está en la catedral.

229. El científico marino bucea en las Bahamas.
230. El científico marino *bucean en las Bahamas.
231. El científico marino está en las Bahamas.

232. La bióloga mexicana bucea en el Caribe.
233. La bióloga mexicana *bucean en el Caribe.
234. La bióloga mexicana está en el Caribe.

235. El oso hambriento pesca en el río.
236. El oso hambriento *pescan en el río.
237. El oso hambriento está en el río.

238. La financiera retirada pesca en las Bahamas.
239. La financiera retirada *pescan en las Bahamas.
240. La financiera retirada está en las Bahamas.

241. El extranjero testarudo regatea en el mercado.
242. El extranjero testarudo *regatean en el mercado.
243. El extranjero testarudo está en el mercado.

244. La vendedora sabia regatea en el quiosco.
245. La vendedora sabia *regatean en el quiosco.
246. La vendedora sabia está en el quiosco.

247. El cocinero renombrado cocina en el restaurante.
248. El cocinero renombrado *cocinan en el restaurante.
249. El cocinero renombrado está en el restaurante.

250. La criada italiana cocina en el hotel.
251. La criada italiana *cocinan en el hotel.
252. La criada italiana está en el hotel.

253. El ejecutivo obeso desayuna en su despacho.
254. El ejecutivo obeso *desayunan en su despacho.
255. El ejecutivo obeso está en su despacho.

256. La conductora apresurada desayuna en su vehículo.
257. La conductora apresurada *desayunan en su vehículo.
258. La conductora apresurada está en su vehículo.

259. El canguro capturado boxea en su jaula.
260. El canguro capturado *boxean en su jaula.
261. El canguro capturado está en su jaula.

262. La campeona cubana boxea en el gimnasio.
263. La campeona cubana *boxean en el gimnasio.
264. La campeona cubana está en el gimnasio.

265. El caballo enfermo convalece en el establo.
266. El caballo enfermo *convalecen en el establo.
267. El caballo enfermo está en el establo.

268. La víctima asustada convalece en el hospital.
269. La víctima asustada *convalecen en el hospital.
270. La víctima asustada está en el hospital.

271. El enfermero infectado estornuda en la clínica.
272. El enfermero infectado *estornudan en la clínica.
273. El enfermero infectado está en la clínica.

274. La señora delgada estornuda en el aeropuerto.
275. La señora delgada *estornudan en el aeropuerto.
276. La señora delgada está en el aeropuerto.

277. El político protegido desembarca en el aeropuerto.
278. El político protegido *desembarcan en el aeropuerto.
279. El político protegido está en el aeropuerto.

280. La pasajera brasileña desembarca en San Francisco.
281. La pasajera brasileña *desembarcan en San Francisco.
282. La pasajera brasileña está en San Francisco.

283. El monstruo acuático reaparece en la bahía.
284. El monstruo acuático *reaparecen en la bahía.
285. El monstruo acuático está en la bahía.

286. La enemiga vengativa reaparece en la secuela.
287. La enemiga vengativa *reaparecen en la secuela.
288. La enemiga vengativa está en la cárcel.

289. El tirano malvado reina en el archipiélago.
290. El tirano malvado *reinan en el archipiélago.
291. El tirano malvado está en el archipiélago.

292. La soberana tiránica reina en el castillo.
293. La soberana tiránica *reinan en el castillo.
294. La soberana tiránica está en el castillo.

295. El destacamento especializado ataca en el este.
296. El destacamento especializado *atacan en el este.
297. El destacamento especializado está en el este.

298. La brigada rusa ataca en el norte.
299. La brigada rusa *atacan en el norte.
300. La brigada rusa está en el norte.

301. El anciano respetado habla en el coloquio.
302. El anciano respetado *hablan en el coloquio.
303. El anciano respetado está en el coloquio.

304. La arquitecta exitosa habla en el auditorio.
305. La arquitecta exitosa *hablan en el auditorio.
306. La arquitecta exitosa está en el auditorio.

307. El cautivo torturado sufre en su celda.
308. El cautivo torturado *sufren en su celda.
309. El cautivo turturado está en su celda.

310. La huérfana traumatizada sufre en el orfanato.
311. La huérfana traumatizada *sufren en el orfanato.
312. La huérfana traumatizada está en el orfanato.

313. El juego entero cabe en la caja.
314. El juego entero *caben en la caja.
315. El juego entero está en la caja.

316. La bolsa llena cabe en el baúl.
317. La bolsa llena *caben en el baúl.
318. La bolsa llena está en el baúl.

319. El geranio híbrido florece en el patio.
320. El geranio híbrido *florecen en el patio.
321. El geranio híbrido está en el patio.

322. La orquídea púrpura florece en la ribera.
323. La orquídea púrpura *florecen en la ribera.
324. La orquídea púrpura está en la ribera.

325. El candidato próspero triunfa en la elección.
326. El candidato próspero *triunfan en la elección.
327. El candidato próspero está en San Francisco.

328. La modista imaginativa triunfa en la pasarela.
329. La modista imaginativa *triunfan en la pasarela.
330. La modista imaginativa está en la pasarela.

331. El funcionario designado participa en la investigación.
332. El funcionario designado *participan en la investigación.
333. El funcionario designado está en la capital.

334. La secretaria discreta participa en la transacción.
335. La secretaria discreta *participan en la transacción.
336. La secretaria discreta está en el banco.

337. El cartero maltratado persevera en el trabajo.
338. El cartero maltratado *perseveran en el trabajo.
339. El cartero maltratado está en la camioneta.

340. La carpintera trabajadora persevera en el taller.
341. La carpintera trabajadora *perseveran en el taller.
342. La carpintera trabajadora está en el taller.

343. El fotógrafo experto colabora en el proyecto.
344. El fotógrafo experto *colaboran en el proyecto.
345. El fotógrafo experto está en la gala.

346. La reportera inquisitiva colabora en la investigación.
347. La reportera inquisitiva *colaboran en la investigación.
348. La reportera inquisitiva está en la capital.

349. El paramédico musculoso compite en el torneo.
350. El paramédico musculoso *compiten en el torneo.
351. El paramédico musculoso está en la ambulancia.

352. La poetisa española compite en el concurso.
353. La poetisa española *compiten en el concurso.
354. La poetisa española está en la librería.

355. El árbitro frustrado intercede en la pelea.
356. El árbitro frustrado *interceden en la pelea.
357. El árbitro frustrado está en el vestuario.

358. La consejera respetuosa intercede en el conflicto.
359. La consejera respetuosa *interceden en el conflicto.
360. La consejera respetuosa está en la sala.

APPENDIX 3: Individual Mean Acceptability Rates of Native Speakers for All Three Types of Agreement

Table 80. Individual mean acceptability rates of native speakers for Subject-Verb agreement, N-Adj Number agreement, and N-Adj Gender agreement, along with average, standard deviation, and range.

		S-V	*S-V	N-Adj	*N-Adj Number	*N-Adj Gender
Participant	NS001	100.00	0.00	97.50	0.00	0.00
	NS003	92.50	2.50	100.00	5.00	2.50
	NS005	97.50	0.00	97.50	0.00	0.00
	NS006	95.00	5.00	97.50	2.50	0.00
	NS008	92.50	0.00	100.00	2.50	2.50
	NS009	100.00	7.50	100.00	5.00	10.00
	NS011	90.00	5.00	97.50	2.50	7.50
	NS012	100.00	0.00	97.50	0.00	5.00
	NS014	92.50	2.50	90.00	2.50	2.50
	NS015	97.50	0.00	97.50	0.00	10.00
	NS017	100.00	2.50	95.00	0.00	2.50
	NS018	97.50	10.00	90.00	10.00	0.00
	Average	96.25	2.92	96.67	2.50	3.54
	(SD)	(3.615)	(3.343)	(3.427)	(3.015)	(3.763)
	(Range)	(90-100)	(0-10)	(90-100)	(0-10)	(0-10)

APPENDIX 4: Native Speakers' d' Scores for All Three Types of Agreement

Table 81. Native speakers' d' scores for Subject-Verb agreement, N-Adj Number agreement, and N-Adj Gender agreement, along with average and standard deviation.

		S-V	**N-Adj Number**	**N-Adj Gender**
Participant	NS001	4.052	3.412	3.412
	NS003	2.404	3.189	3.412
	NS005	3.412	3.412	3.412
	NS006	2.326	2.772	3.412
	NS008	3.044	3.412	3.412
	NS009	3.044	3.189	2.932
	NS011	2.069	2.772	2.404
	NS012	4.052	3.412	2.549
	NS014	2.404	2.292	2.292
	NS015	3.412	3.412	2.292
	NS017	3.412	3.189	2.549
	NS018	2.292	1.812	2.932
	Average	2.994	3.023	2.918
	(SD)	(.691)	(.519)	(.481)

APPENDIX 5: Individual Mean Acceptability Rates of Learners for All Three Types of Agreement

Table 82. Individual mean acceptability rates of learners for Subject-Verb agreement, N-Adj Number agreement, and N-Adj Gender agreement, along with average, standard deviation, and range.

		S-V	***S-V**	**N-Adj**	***N-Adj Number**	***N-Adj Gender**
Participant	L2003	97.50	15.00	100.00	2.50	67.50
	L2004	85.00	10.00	87.50	15.00	45.00
	L2006	90.00	12.50	95.00	7.50	7.50
	L2007	85.00	45.00	77.50	57.50	87.50
	L2008	95.00	0.00	97.50	5.00	15.00
	L2009	82.50	40.00	92.50	92.50	55.00
	L2010	90.00	10.00	97.50	15.00	17.50
	L2011	55.00	45.00	75.00	47.50	57.50
	L2012	90.00	5.00	95.00	5.00	7.50
	L2014	85.00	15.00	92.50	10.00	22.50
	L2016	97.50	5.00	100.00	0.00	5.00
	L2019	97.50	0.00	97.50	0.00	7.50
	L2022	75.00	5.00	92.50	22.50	55.00
	L2023	95.00	15.00	97.50	10.00	42.50
	L2024	92.50	10.00	97.50	10.00	72.50
	L2025	70.00	77.50	60.00	50.00	47.50
	L2026	100.00	7.50	92.50	0.00	27.50
	L2028	97.50	0.00	90.00	0.00	2.50
	L2029	100.00	40.00	97.50	12.50	50.00
	L2030	97.50	5.00	100.00	0.00	7.50
	L2031	77.50	35.00	65.00	60.00	62.50
	L2032	87.50	12.50	95.00	7.50	17.50
	L2037	87.50	2.50	92.50	0.00	7.50
	L2038	90.00	2.50	95.00	0.00	37.50
	Average	88.33	17.29	90.94	17.92	34.38
	(SD)	(10.675)	(19.518)	(10.754)	(24.756)	(25.348)
	(Range)	(55-100)	(0-77.5)	(60-100)	(0-92.5)	(2.5-87.5)

APPENDIX 6: Learners' d' Scores for All Three Types of Agreement

Table 83. Learners' d' scores for Subject-Verb agreement, N-Adj Number agreement, and N-Adj Gender agreement, along with average and standard deviation.

		S-V	N-Adj Number	N-Adj Gender
Participant	L2003	2.119	3.412	1.705
	L2004	1.639	1.546	0.902
	L2006	1.720	2.181	2.181
	L2007	0.822	0.400	-0.279
	L2008	3.189	2.549	2.119
	L2009	0.840	0.000	0.929
	L2010	1.812	2.119	2.047
	L2011	0.178	0.521	0.343
	L2012	2.069	2.326	2.181
	L2014	1.466	1.924	1.552
	L2016	2.549	4.052	3.189
	L2019	3.412	3.412	2.404
	L2022	1.640	1.552	0.929
	L2023	1.896	2.292	1.520
	L2024	1.924	2.292	0.963
	L2025	-0.163	0.179	0.223
	L2026	3.044	3.044	1.441
	L2028	3.412	2.932	2.292
	L2029	2.205	2.199	1.386
	L2030	2.549	4.052	3.044
	L2031	0.807	0.093	0.047
	L2032	1.627	2.181	1.824
	L2037	2.199	3.044	2.036
	L2038	2.292	3.189	1.388
	Average	1.885	2.146	1.515
	(SD)	(.931)	(1.201)	(.892)

APPENDIX 7: Correlations Between Aptitude and Proficiency Measures and Early ERP Amplitudes in Response to Subject-Verb Agreement

Table 84. Correlations between aptitude and proficiency measures and mean difference amplitudes for Subject-Verb agreement in the early time windows.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
		150-250ms					
MLAT3	*r*	- .060	- .282	**-.437***	**- .403†**	- .083	- .059
	(*p*)	(.780)	(.182)	**(.033)**	**(.051)**	(.700)	(.783)
MLAT4	*r*	**.473***	- .142	- .134	- .156	.063	- .003
	(*p*)	**(.020)**	(.509)	(.534)	(.465)	(.769)	(.990)
MLAT5	*r*	**.407***	- .208	- .026	.088	- .027	- .021
	(*p*)	**(.048)**	(.328)	(.902)	(.684)	(.900)	(.921)
MLAT Total	*r*	**.420***	- .332	- .318	- .252	- .026	- .044
	(*p*)	**(.041)**	(.113)	(.131)	(.236)	(.902)	(.837)
LLAMA_F	*r*	.409	.489	.311	.128	.512	.139
	(*p*)	(.211)	(.127)	(.351)	(.708)	(.107)	(.683)
LLAMA_B	*r*	.490	.028	- .153	- .272	.129	- .120
	(*p*)	(.126)	(.936)	(.654)	(.418)	(.704)	(.726)
RAVEN	*r*	.230	.085	.081	.049	- .099	- .085
	(*p*)	(.280)	(.692)	(.708)	(.820)	(.646)	(.695)
PROF	*r*	.323	- .318	- .145	- .094	**-.424***	**- .351†**
	(*p*)	(.124)	(.130)	(.500)	(.661)	**(.039)**	**(.092)**
		250-450ms					
MLAT3	*r*	.013	-.140	-.152	.031	.026	-.103
	(*p*)	(.950)	(.515)	(.478)	(.887)	(.904)	(.632)
MLAT4	*r*	-.123	-.132	-.139	.190	.091	.064
	(*p*)	(.566)	(.539)	(.516)	(.373)	(.673)	(.765)
MLAT5	*r*	-.107	.049	.226	.017	.019	.005
	(*p*)	(.618)	(.818)	(.289)	(.938)	(.929)	(.982)
MLAT Total	*r*	-.111	-.117	-.037	.123	.071	-.020
	(*p*)	(.605)	(.586)	(.865)	(.568)	(.743)	(.926)
LLAMA_F	*r*	.448	.367	.222	**.524†**	.066	.151
	(*p*)	(.167)	(.267)	(.512)	**(.098)**	(.848)	(.658)
LLAMA_B	*r*	.408	.500	.421	.424	.427	.492
	(*p*)	(.213)	(.117)	(.197)	(.193)	(.190)	(.124)
RAVEN	*r*	.082	.182	.191	.079	.179	.167
	(*p*)	(.703)	(.396)	(.372)	(.712)	(.401)	(.436)
PROF	*r*	-.222	-.080	.061	-.172	-.108	-.168
	(*p*)	(.296)	(.712)	(.778)	(.423)	(.614)	(.434)

* Correlation is significant ($p<.05$).
† Correlation is marginally significant ($.10>p>.05$).

APPENDIX 8: Correlations Between Aptitude and Proficiency Measures and Early ERP Amplitudes in Response to Noun-Adjective Number Agreement

Table 85. Correlations between aptitude and proficiency measures and mean difference amplitudes for Noun-Adjective Number agreement in the early time windows.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
		150-250ms					
MLAT3	*r*	-.224	-.114	-.142	-.102	**-.393†**	-.214
	(*p*)	(.292)	(.596)	(.508)	(.634)	**(.057)**	(.316)
MLAT4	*r*	**.380†**	.058	.017	-.105	.191	.159
	(*p*)	**(.067)**	(.786)	(.937)	(.624)	(.371)	(.458)
MLAT5	*r*	.261	**.394†**	**.358†**	.153	.205	.124
	(*p*)	(.217)	**(.057)**	**(.086)**	(.474)	(.337)	(.565)
MLAT Total	*r*	.209	.172	.118	-.030	-.008	.030
	(*p*)	(.327)	(.420)	(.584)	(.891)	(.970)	(.889)
LLAMA_F	*r*	.350	.207	.289	.272	**.563†**	.456
	(*p*)	(.292)	(.542)	(.389)	(.419)	**(.071)**	(.158)
LLAMA_B	*r*	.454	.055	.108	.073	.392	.279
	(*p*)	(.161)	(.873)	(.753)	(.830)	(.233)	(.407)
RAVEN	*r*	.218	.249	.271	.256	.118	.213
	(*p*)	(.306)	(.241)	(.201)	(.227)	(.583)	(.317)
PROF	*r*	.217	.323	.263	.112	-.028	.204
	(*p*)	(.309)	(.124)	(.214)	(.603)	(.896)	(.338)
		250-450ms					
MLAT3	*r*	-.298	-.198	-.041	-.291	-.192	-.060
	(*p*)	(.158)	(.355)	(.848)	(.167)	(.370)	(.781)
MLAT4	*r*	-.208	-.045	-.077	-.050	-.053	-.180
	(*p*)	(.330)	(.836)	(.722)	(.817)	(.805)	(.401)
MLAT5	*r*	.183	.268	.077	.190	-.028	-.306
	(*p*)	(.392)	(.206)	(.719)	(.373)	(.898)	(.146)
MLAT Total	*r*	-.172	.009	-.021	-.084	-.145	-.283
	(*p*)	(.421)	(.965)	(.921)	(.697)	(.500)	(.181)
LLAMA_F	*r*	.325	.491	.188	.506	.504	.394
	(*p*)	(.329)	(.125)	(.580)	(.112)	(.114)	(.231)
LLAMA_B	*r*	.089	.215	.073	.458	**.567†**	.448
	(*p*)	(.795)	(.526)	(.832)	(.156)	**(.069)**	(.167)
RAVEN	*r*	.137	.193	.213	.175	.160	.212
	(*p*)	(.524)	(.366)	(.317)	(.414)	(.456)	(.320)
PROF	*r*	-.207	-.112	-.009	-.085	-.204	**-.358†**
	(*p*)	(.331)	(.602)	(.965)	(.694)	(.340)	**(.086)**

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

APPENDIX 9: Correlations Between Aptitude and Proficiency Measures and Early ERP Amplitudes in Response to Noun-Adjective Gender Agreement

Table 86. Correlations between aptitude and proficiency measures and mean difference amplitudes for Noun-Adjective Gender agreement in the early time windows.

		Left Anterior	Central Anterior	Right Anterior	Left Posterior	Central Posterior	Right Posterior
		150-250ms					
MLAT3	*r*	-.022	-.216	-.316	**-.405***	**-.396†**	-.193
	(*p*)	(.917)	(.311)	(.133)	**(.050)**	**(.055)**	(.366)
MLAT4	*r*	.286	-.134	.028	-.010	-.015	.064
	(*p*)	(.175)	(.533)	(.897)	(.961)	(.946)	(.766)
MLAT5	*r*	.262	.019	.095	.040	.106	.299
	(*p*)	(.216)	(.930)	(.660)	(.854)	(.620)	(.155)
MLAT Total	*r*	.270	-.175	-.107	-.203	-.166	.084
	(*p*)	(.202)	(.414)	(.619)	(.342)	(.439)	(.697)
LLAMA_F	*r*	.066	.182	.354	.367	**.530†**	.229
	(*p*)	(.847)	(.592)	(.285)	(.266)	**(.093)**	(.497)
LLAMA_B	*r*	.360	-.234	.061	.124	.145	.227
	(*p*)	(.276)	(.489)	(.859)	(.716)	(.671)	(.503)
RAVEN	*r*	.135	-.066	.039	.090	.137	.225
	(*p*)	(.528)	(.758)	(.856)	(.676)	(.524)	(.291)
PROF	*r*	**.362†**	-.070	-.003	-.128	-.218	.069
	(*p*)	**(.082)**	(.746)	(.988)	(.552)	(.306)	(.748)
		250-450ms					
MLAT3	*r*	-.284	**-.417***	**-.354†**	-.272	-.194	-.333
	(*p*)	(.179)	**(.043)**	**(.089)**	(.198)	(.365)	(.112)
MLAT4	*r*	-.192	.054	.033	-.287	-.195	-.237
	(*p*)	(.369)	(.802)	(.879)	(.174)	(.362)	(.264)
MLAT5	*r*	-.119	.020	.009	.059	.160	.006
	(*p*)	(.580)	(.925)	(.966)	(.783)	(.455)	(.977)
MLAT Total	*r*	-.313	-.187	-.169	-.263	-.121	-.298
	(*p*)	(.137)	(.382)	(.429)	(.215)	(.573)	(.158)
LLAMA_F	*r*	.219	.425	.384	**.540†**	.350	.416
	(*p*)	(.517)	(.192)	(.244)	**(.086)**	(.291)	(.203)
LLAMA_B	*r*	-.132	.230	.276	.403	**.545†**	**.562†**
	(*p*)	(.699)	(.496)	(.411)	(.219)	**(.083)**	**(.072)**
RAVEN	*r*	.047	.174	.155	.216	.314	**.350†**
	(*p*)	(.827)	(.416)	(.469)	(.310)	(.135)	**(.094)**
PROF	*r*	-.205	-.061	-.054	-.217	-.020	-.172
	(*p*)	(.337)	(.776)	(.802)	(.307)	(.927)	(.422)

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