Examining variability in the processing of agreement in novice learners: Evidence from event-related potentials

Article *in* Journal of Experimental Psychology Learning Memory and Cognition · January 2021

DOI: 10.1037/xlm0000983

CITATION
1

READS
286

6 authors, including:

Alison Gabriele
University of Kansas
41 PUBLICATIONS **384** CITATIONS
SEE PROFILE

José Alemán Bañón
Stockholm University
26 PUBLICATIONS **247** CITATIONS
SEE PROFILE

Lauren Covey
Montclair State University
9 PUBLICATIONS **10** CITATIONS
SEE PROFILE

Robert Fiorentino
University of Kansas
49 PUBLICATIONS **715** CITATIONS
SEE PROFILE

Examining variability in the processing of agreement in novice learners: Evidence from event-related potentials

Alison Gabriele[a], José Alemán Bañón[b], Lesa Hoffman[c], Lauren Covey[d], Amy Rossomondo[e], Robert Fiorentino[f]

[a]Second Language Acquisition Laboratory, Department of Linguistics, Dole Human Development Center, Room 4035, 1000 Sunnyside Ave., University of Kansas, Lawrence, KS 66045

[b]Centre for Research on Bilingualism, Department of Swedish Language and Multilingualism, Stockholm University, 106 91 Stockholm

[c]Department of Psychological and Quantitative Foundations, College of Education, 361 Lindquist Center, University of Iowa, Iowa City, IA 52242-1529

[d]Department of Linguistics, Schmitt Hall 240D, 1 Normal Ave., Montclair State University, Montclair, NJ 07043

[e]Department of Spanish and Portuguese, 2650 Wescoe Hall, 1445 Jayhawk Blvd., University of Kansas, Lawrence, KS 66045

[f]Neurolinguistics & Language Processing Laboratory, Department of Linguistics, Dole Human Development Center, Room 4037, 1000 Sunnyside Ave., University of Kansas, Lawrence, KS 66045

Address correspondence to
Alison Gabriele
Department of Linguistics
University of Kansas
1541 Lilac Lane
Blake Hall, Room 427
Lawrence, KS 66044-3177
Tel: +1-785-864-1014
Fax: +1-785-864-5724
E-mail: gabriele@ku.edu

In press, Journal of Experimental Psychology: Learning, Memory, and Cognition

Abstract

The present study examines both properties of the language and properties of the learner to better understand variability at the earliest stages of second language (L2) acquisition. We used event-related potentials, an oral production task, and a battery of individual differences measures to examine the processing of number and gender agreement in two groups of low-proficiency English-speaking learners of Spanish who were tested in multiple sessions. The results showed an advantage for number, the feature also instantiated in the native language, as both groups showed a native-like P600 response to subject-verb and noun-adjective number violations across sessions. The more advanced group showed larger effects for number and marginal sensitivity to gender violations. These results suggest that native-like processing of shared features is possible even for novice learners, contrary to proposals suggesting that all morphosyntactic dependencies are initially processed in a non-native manner. Working memory (WM) was a predictor of P600 effects for number and also for gender (where the effect was marginal), suggesting that similar abilities may capture variability in the processing of both shared and unique features despite differences in overall sensitivity. Furthermore, while WM predicted performance on online tasks (P600 effects/oral production), verbal aptitude predicted performance on tasks examining morphosyntactic accuracy (grammaticality judgment task/oral production). Our results show that the linguistic properties of the L2, the individual characteristics of the learner, and the nature of the task at hand all play an important role in capturing the variability often observed in the L2 processing of agreement.

Keywords: second language acquisition, event-related potentials, gender, number, agreement, individual differences

Examining variability in the processing of agreement in novice learners: Evidence from event-related potentials

While it is uncontroversial to state that a hallmark of second language (L2) acquisition is variability in performance, the factors that have been proposed to account for that variability are myriad. In his seminal paper arguing for fundamental differences between first language (L1) and L2 acquisition, Bley-Vroman (1990) argued that L2 learners who acquired the L2 late in life are constrained by the properties of their L1, and that the learning of unique properties in the L2 proceeds via general learning strategies, but not via a language-specific mechanism such as Universal Grammar. Thus, according to this proposal, success in the L2 is most likely for individuals who can best learn via explicit mechanisms, such as those with above-average, domain-general problem-solving skills. While certain aspects of this proposal were underdeveloped, particularly the nature of the problem-solving mechanism that was proposed, one key strength was that it considered both linguistic factors, such as the role of the native language in L2 acquisition, and individual-level factors, such as the problem-solving abilities that were proposed to facilitate success in the acquisition of properties unique to the L2 (see DeKeyser, 2000). Much subsequent research in L2 acquisition has largely focused on only one of these dimensions and the precise way in which properties of the language and properties of the learner interact remains underexplored.

For example, generative proposals present precise predictions with respect to the acquisition of grammatical properties of the L2 by comparing the inventory of morphosyntactic features in the L1 and L2. Both the Interpretability Hypothesis (e.g., Hawkins & Chan, 1997; Tsimpli & Dimitrakopoulou, 2007) and Full Transfer/Full Access (Schwartz & Sprouse, 1994, 1996) predict that the initial state of L2 acquisition is constrained by the L1 grammar. However, Full

Transfer/Full Access also predicts that features unique to the L2 can ultimately be acquired in the course of development while the Interpretability Hypothesis argues, in line with Bley-Vroman, that features unique to the L2 will always be processed in a qualitatively non-native-like manner. In these theoretical proposals, properties of the individual learners themselves receive less attention.

In contrast, cognitive theories emphasize the individual abilities of the learners themselves and the mechanisms associated with the learning process, with somewhat less precise attention to linguistic factors. Although the specific details of the proposals differ, cognitive theories generally propose that L2 acquisition proceeds in a series of discrete stages, from declarative to procedural or from controlled to more automatic processing, with progress being modulated by exposure and explicit practice (DeKeyser, 2000, 2007; Ullman, 2001, 2005, 2015). According to DeKeyser (2000), the most successful adult L2 learners are predicted to be those with high verbal aptitude, a set of skills related to linguistic analytic ability (e.g., grammatical sensitivity) which arguably underlie successful foreign language learning (Skehan, 1986, 2004); DeKeyser argued that these skills are relevant for late learners but not learners who acquired an L2 early in life (DeKeyser, 2000; DeKeyser et al., 2010; but see Abrahamsson & Hyltenstam, 2008). While cognitive theories acknowledge that some L1-L2 pairings may be globally more challenging than others (see DeKeyser et al., 2010), these proposals do not differentiate between precise aspects of morphosyntax that are shared between the L1 and L2 or unique to the L2. At the early stages of acquisition, it is predicted that late L2 learners will process all grammatical features in the same, non-nativelike, manner.

The goal of the present study is to focus specifically on low-proficiency learners and examine both linguistic factors and individual-level abilities in order to more comprehensively understand

the foundation on which learners build at early stages of development. The study uses event-related potentials (ERPs), a technique which allows us to examine brain responses during online processing, to investigate the processing of number and gender agreement in two groups of low-proficiency English-speaking learners of Spanish. We also include an oral production measure in order to examine whether there is a relationship between online sensitivity to grammatical violations as measured by ERPs and a learner's ability to establish morphological agreement in production in more natural contexts. Previous L2 studies of both comprehension and production suggest that variability is often observed in the acquisition and processing of agreement (e.g., Alarcón, 2011; Covey et al., 2018; Dewaele & Véronique, 2001; Franceschina, 2001, 2005; Gabriele et al., 2013; Gillon-Dowens et al., 2010; Grüter et al., 2012; Lemhöfer et al., 2014; McCarthy, 2008; Montrul et al., 2008; L. White et al., 2004) but few studies have simultaneously considered both linguistic factors and cognitive abilities.

With respect to linguistic factors, our first research question examines whether or not L2 learners build on the foundation of their L1 grammar and how processing changes over the course of L2 development. Our second research question examines whether success in the online processing of morphosyntax is related to learners' sensitivity to agreement dependencies in other contexts, such as the receptive ability to detect grammatical anomalies under less time pressure in a grammaticality judgment task (GJT) and the learners' productive ability to use morphosyntax in more meaning-focused contexts. Our third research question examines whether individual differences facilitate the acquisition and processing of morphosyntax in low-proficiency learners. The design of our study, which investigates both grammatical features that are shared between the L1 and L2 and those that are unique to the L2, allows us to examine whether the same individual differences facilitate processing for both kinds of features. Finally,

because the present study examines the acquisition of morphosyntax using three different measures (ERP, GJT, oral production), we are also able to examine whether the same individual differences impact performance across tasks that vary with respect to their focus on metalinguistic properties. We expand on each of these issues below.

Due to the complexity of the research questions examined, the paper is divided into three sections, with each research question and corresponding results presented in turn. The first section presents relevant background research and a new ERP study on L1-L2 differences. The second section further explores the ERP results presented in the first section by examining the relationship between ERP effects and both detection of agreement violations and agreement in oral production. The third section will focus on individual differences.

Examining L1-L2 Differences using Event-Related Potentials

The present study examines both number and gender agreement in L2 Spanish by native speakers of English in order to compare the processing of a feature that is shared in the L1 and L2 (number) with a feature that is unique to the L2 (gender). We use ERPs because this measure allows us to more directly test the theoretical proposals outlined above which argue that there are qualitative differences in how features which are unique to the L2 are processed (Bley-Vroman, 1990; Franceschina, 2005; Hawkins, 2001; Hawkins & Franceschina, 2004). ERPs are brain responses elicited by properties of a stimulus and are characterized by their timing, amplitude, polarity, and sensitivity to particular stimulus manipulations. A number of ERP components have been related to distinct linguistic processes (see Kaan, 2007), which thus allows for a more precise investigation of the kinds of mechanisms that learners and native speakers employ during processing and allows us to see whether those mechanisms are qualitatively similar or different.

Most relevant to the present study, the P600 is a positive-going waveform that peaks at approximately 600 ms and typically shows a centroparietal distribution. In native speakers, P600 effects for agreement violations such as mismatches in number between a subject and verb (**The boys plays*) are consistently found across languages and across different types of agreement such as number and gender (see e.g., Molinaro et al., 2011; Osterhout et al., 2004, for further discussion)[1]. P600 has also been observed in cases of syntactic violations (Friederici, et al., 1996), garden-path sentences (Osterhout & Holcomb, 1992), contexts of syntactic ambiguity (e.g., Frisch et al., 2002) and in the resolution of syntactic dependencies (Alemán Bañón et al., 2012; Felser et al., 2003; Kaan et al., 2000; Phillips et al., 2005 among others) and has thus been taken to reflect difficulty in syntactic integration. P600 has also been shown to be elicited in the context of certain kinds of semantic anomalies (e.g., Kolk et al., 2003; Kim & Osterhout, 2005). Alternative accounts have suggested P600 may reflect domain-general error detection/conflict monitoring processes (see Kaan, 2007 and Tanner et al., 2017 for discussion).

A second component, the N400, is a negative-going waveform that typically peaks around 400 ms following word onset, and has a centroparietal scalp distribution (Kutas & Federmeier,

[1] In some studies of morphosyntactic violations, another component, a Left Anterior Negativity (LAN), is also observed (Angrilli et al., 2002; Barber & Carreiras, 2005; Friederici et al., 1993; Gunter et al., 2000). However, other studies of morphosyntactic violations have not found a LAN (e.g. Gunter et al., 1997; Hagoort et al., 1993; Lau et al., 2006; see also Friederici & Weissenborn, 2007; Osterhout et al., 2004), and the reason for the mixed results is not clear (see Molinaro et al., 2011; Molinaro et al., 2015; Tanner & Van Hell, 2014, for discussion). Although some studies discuss the LAN as a metric for assessing whether L2 processing is 'native-like' (e.g., Clahsen & Felser, 2006; Steinhauer et al., 2009), the variability observed in native speakers makes it difficult to apply that metric to learners (McLaughlin et al., 2010; Tanner et al., 2013; Weber & Lavric, 2008).

2000; Lau et al., 2008 among others). N400 is typically taken to reflect semantic incongruency or the violation of lexical expectations (Kutas & Federmeier, 2011; Lau et al., 2008). While N400 is typically not elicited for agreement violations in group analyses of native speakers (but see Tanner & Van Hell, 2014), N400 has been observed in several studies examining the processing of agreement violations in L2 learners, especially those at lower levels of proficiency (Carrasco et al., 2017; McLaughlin et al., 2010; Mickan & Lemhöfer, 2020; Morgan-Short et al., 2010; Osterhout et al., 2006; Reichle et al., 2016; Tanner et al., 2013).

In a study that inspired the present investigation, Osterhout et al. (2006) used the N400 and P600 as a metric for whether or not low-proficiency learners could come to process morphosyntactic violations similarly to native speakers. This study examined the initial stages of the acquisition of L2 French by English native speakers in a longitudinal design, testing learners (n = 14) after one, four, and eight months of instruction. The study examined two different contexts focusing on agreement, subject-verb agreement (*Tu adores*/**adorez le français* 'You-2SG adore-2SG/adore-2PL French'), which is a property that is similar in the L1 and L2, and determiner-noun agreement (*Tu manges des hamburgers/*hamburger pour dîner* 'You eat some-PL hamburger-PL/hamburger-SG for dinner'), which targets a feature that is shared between English and French but is an instantiation of the number feature that is unique to the L2 French. Sentences with semantic anomalies were also included. Native speakers showed a P600 for both kinds of agreement violations and N400 for the semantic violations. For L2 learners, Osterhout et al. (2006) report that wide variability was observed, both on the GJT that accompanied the ERP experiment, and in the ERP results themselves. The results of a subset of 'fast' learners (n = 7) who performed well on the GJT were discussed; for these learners, an N400 was observed at one

month for the subject-verb agreement violations and a P600 was observed at the two later testing sessions. No effects were observed for the determiner-noun violations at any session.

Osterhout et al. argue that grammatical rules which are similar in the L1 and L2 can quickly be integrated by L2 learners, as evidenced by the P600 that emerged in the second session. However, the initial N400 suggests that this rule may not be transferred from the L1 at the initial state. The N400 may indeed reflect sensitivity to the violation, but this result also suggests that the learners are using a qualitatively different mechanism than the natives. Osterhout et al. argue that the N400 may reflect sensitivity to probabilistic dependencies between morphemes that frequently co-occur. As learners begin to grammaticalize the agreement rule in French, there is a shift from N400 to P600 (McLaughlin et al., 2010; Osterhout et al. 2006, 2008). Osterhout et al.'s proposal for these discrete stages of L2 processing is in line with the cognitive theories discussed earlier (e.g., DeKeyser, 2007; Ullman, 2001, 2005, 2015).

A similar pattern was observed in Tanner et al. (2013), a cross-sectional study of English-speaking learners of German which also focused on subject-verb agreement, a property which is similar in this L1-L2 pairing (see also Mickan & Lemhöfer, 2020). While native speakers of German and more advanced learners of German showed P600 to agreement violations, novice first-year learners showed a tendency towards a biphasic response, yielding both N400 and P600. Analyses showed that while some learners showed N400, others showed P600, and the effects within individuals were inversely related to each other. Further analyses showed that P600 effects were related to performance on the GJT associated with the ERP experiment; individuals with higher accuracy on the GJT showed larger P600s (see also Alemán-Bañón et al., 2018; Caffarra et al., 2015; Steinhauer et al., 2009; Van Hell & Tokowicz, 2010; E.J. White et al., 2012).

The results of both Osterhout et al. (2006) and Tanner et al. (2013) are interesting because they present a challenge to generative theories of transfer (Hawkins, 2001; Schwartz & Sprouse, 1994, 1996) which assume that features which are part of the L1 inventory should be available at the L2 initial state. It is important to point out that Osterhout et al. (2006, 2008) do not exclude the role of the L1 at later stages: recall that in their study, P600 was ultimately elicited for subject-verb agreement, which is similar in the L1-L2, but not determiner-noun agreement, which is unique to the L2. However, the N400 elicited at the initial stage is interpreted as a lack of transfer. These results are instead taken to show that learners pass through discrete stages in the initial learning of all morphosyntactic features regardless of L1-L2 similarity (see also Steinhauer et al., 2009).

Similar to Osterhout et al. (2006), Davidson and Indefrey (2009a) also employ a longitudinal approach and show difficulty in the processing of shared features for low-proficiency learners (see also Davidson, 2010). Their study examined Dutch learners of German in a short-term training study, examining both adjective declensions and gender agreement. The learners showed no sensitivity to either kind of violation in the pre-test, but during training and the post-test, sensitivity emerged to violations of adjective declension but not gender agreement. This is an interesting finding as both Dutch and German instantiate gender agreement. However, the learners were at a very low level of proficiency in German and the training was provided on only a single day. Thus, these results suggest that sensitivity for features such as gender, which involve both a lexical learning component (lexical gender of the nouns) and a syntactic agreement component, may take longer to emerge.

The present study builds on these previous studies, but we extend our investigation in several ways to directly address some of the open questions that remain. First, we examine a

wider range of linguistic factors, directly comparing shared and unique features in the L2, and examining the role of proficiency by testing classroom learners at two different levels of proficiency. Second, as will be discussed in the second and third sections of this paper, we further explore the factors that may underlie variability in the processing of morphosyntax. The nature of the variability that was observed in Osterhout et al. (2006) and Tanner et al. (2013) is interesting in that their results suggest that only a subset of learners can achieve native-like processing. However, the source of that variability remains unclear. While the ERP results in those studies were related to performance on the grammaticality judgment task, it is important to note that the GJT is an experiment-internal measure and thus, the relationship is not completely surprising (see Tanner et al., 2013). Our study takes a more comprehensive approach, examining both linguistic and cognitive abilities.

Research Question and Predictions

Our first research question addresses the role of the L1, examining (1) whether there are differences in sensitivity across three kinds of morphosyntactic violations of agreement which target features that are either shared in the L1 and L2 (subject-verb number, noun-adjective number) or are unique to the L2 (noun-adjective gender) and (2) whether sensitivity to both shared and unique features changes in development with increased instruction in L2 Spanish:

RQ1. Do novice learners process features that are instantiated in both the L1 and L2 differently than features unique to the L2? Does sensitivity to morphosyntactic violations for both shared and unique features change over the course of development?

By including two kinds of number agreement violations, our design further explores how to define 'similarity' between the L1 and L2. Generative theories of L2 acquisition propose that the entire feature inventory of the L1 is transferred, and thus, surface-level differences between the L1 and L2 with respect to how the feature is instantiated are not predicted to play an important role (Hawkins, 2001; Schwartz & Sprouse, 1994, 1996). For example, relevant to the present study, while both English and Spanish have a number feature in their inventory and both languages instantiate subject-verb agreement, only Spanish realizes number agreement on adjectives. Generative theories of transfer would not predict differences in the processing of these two kinds of number agreement, and moreover, would make the strong prediction that early learners would show P600s to both kinds of violations (Schwartz & Sprouse, 1994, 1996). In contrast, the cognitive theories predict that acquisition proceeds in a series of discrete stages, and thus, early learners should not show P600 at the earliest stages of processing grammatical dependencies even for features that are shared by the L1 and L2 (DeKeyser, 2007; Osterhout et al., 2006; Ullman, 2001, 2005, 2015). In addition, several ERP studies, including Osterhout et al.'s results discussed earlier, suggest the language specific instantiation of a shared feature matters (see Foucart & Frenck-Mestre, 2011; Osterhout et al., 2006; Tokowicz & MacWhinney, 2005) and thus, these proposals predict an advantage for contexts in which the shared feature is also instantiated similarly in the L1 and L2. The present study directly tests this issue by comparing the processing of number agreement in two contexts, one which is similar in the L1 and L2 (subject-verb agreement) and one which is unique to the L2 (noun-adjective agreement).

We also examine the processing of a feature unique to the L2 (gender), which allows us to test the predictions of competing theories of transfer such as Full Transfer/Full Access (Schwartz & Sprouse, 1996) and the Interpretability Hypothesis (Tsimpli & Dimitrakopoulou, 2007): while

Full Transfer/Full Access predicts that features unique to the L2 can ultimately be acquired in a native-like manner, the Interpretability Hypothesis predicts that gender would be processed in a qualitatively different manner, regardless of proficiency level. Results in the literature are quite mixed with respect to whether or not unique features are acquirable. While some ERP studies show that L2 learners do not successfully yield native-like processing profiles for agreement dependencies that are unique to the L2, even at high levels of proficiency (Chen et al., 2007; Díaz et al., 2016; Ojima et al., 2005), other studies show that native-like processing (P600) is possible (Alemán Bañón et al., 2014, 2017, 2018; Foucart & Frenck-Mestre, 2012; Gillon-Dowens et al., 2010; Gillon-Dowens et al., 2011; E.J. White at al., 2012). To our knowledge, the present study is among the first longitudinal ERP studies of novice classroom learners (in their first year of instruction) to examine a feature unique to the L2 (see Morgan-Short et al., 2010, for work on artificial language).

Our study also examines linguistic development by including an additional group of learners who have already completed the equivalent of at least a year of university-level instruction. We examine whether there is a difference between the two learner groups, who differ in the amount of Spanish instruction they have received, and whether sensitivity develops over the course of the academic year with increased instruction. This approach allows us to examine development both within and across learners.

Methods

Participants

Our study adopts both a longitudinal and cross-sectional approach, testing two groups of native English-speaking learners of Spanish (n=46) at multiple sessions across the academic year. The

first group was recruited from an introductory class sequence that was designed for learners with no previous Spanish instruction. These novice learners were recruited from Spanish classes at the '100-level' (the most basic level of instruction at the university) and thus will be referred to as the 100-level learners. A second group at a higher level of proficiency was also tested.

With respect to the novice group, the '100-level' learners, 28 participants (16 women) provided their informed consent to participate in the study, of whom 23 learners (13 women) completed all sessions of the study and were carried forward for analysis (age range at the time of testing: 18-45; mean age = 23). Only learners who remained enrolled in Spanish classes during the whole academic year were included in the study. A background questionnaire confirmed that most learners had not had any previous significant exposure to Spanish.[2] Inspired by Osterhout et al. (2006), we tested the 100-level learners in three different ERP recording sessions across the academic year. Due to the complexity of the stimuli used in the current study, the 100-level learners were tested at two months, six months, and eight months of continued Spanish instruction, while the learners in Osterhout et al. (2006) were tested after one, four, and eight months of instruction.

A supplementary grant award also allowed us to test a learner group at a somewhat higher level of proficiency. These learners were still at low levels of proficiency but were recruited from more advanced classes geared towards learners who had completed the equivalent of university classes at the 100-level. We recruited learners from '200-level' Spanish courses at the same university (Spanish 211 in the Fall/Spanish 216 in the Spring).

[2] Two of the learners reported some minor exposure to Spanish in high school. However, both learners indicated that their knowledge of Spanish before college was very limited, and they both received permission from the Spanish language coordinator at the university where the testing was carried out to enroll in a course for absolute beginners.

Twenty-nine learners were recruited from a 200-level sequence of Spanish courses of whom 23 learners (19 women) remained enrolled in Spanish classes during the whole academic year and completed all sessions of the study (age range at the time of testing: 18-30; mean age = 19). For some learners, this was their first university level Spanish class (n = 15) as they had completed introductory Spanish classes in high school. A background questionnaire confirmed that all learners had not had any significant exposure to Spanish in childhood and did not use Spanish to a significant extent outside of the classroom. Due to funding resources, the 200-level learners were tested in only two different ERP recording sessions (at two months and six months of continued instruction).

The native Spanish participant group which was reported in Bond et al. (2011) serves to establish the Spanish native response pattern for the current study. Bond et al. tested 12 native speakers of Spanish (9 women) with an age range at the time of testing of 18-41 years old (mean age = 28). With the exception of one participant who was born in the US and was then raised in Costa Rica, all of the native speakers were born and raised in a Spanish-speaking country (Peru, Bolivia, or Costa Rica). All native speakers identified Spanish as their native language and they all reported not being significantly exposed to English or any other languages until approximately age nine. Most of them were international students at a university in the US. The native speakers in Bond et al. were tested in the same laboratory as the learners reported in the current study, using the same stimulus manipulations and EEG recording parameters.

Materials

The stimuli testing subject-verb number agreement were taken from Bond et al. (2011). They consist of 80 pairs of Spanish sentences, each pair including a grammatical sentence and a

sentence with a subject-verb number violation (1a-b). Importantly, the verb is a syntactic context where both English and Spanish realize number agreement in the present tense. A sample pair is presented in (1) below, with the agreeing elements underlined; see the Appendix for a full list of stimuli for all conditions.

(1) a. El obispo católico medita en la capilla.
the bishop-3SG catholic meditate-3SG in the chapel "The catholic bishop meditates in the chapel."
b. El obispo católico *meditan en la capilla.
the bishop-3SG catholic meditate-3PL in the chapel

Agreement was realized between the subject noun (e.g., *obispo* "bishop-SG") and the verb (e.g., *medita* "meditate-SG"), which is the critical region. Only third person singular subjects were used. Ungrammatical sentences were constructed by changing the number specification of the verb from singular to plural (e.g., *meditan* "meditate-PL").

The stimuli testing number and gender agreement between nouns and adjectives consisted of 120 triplets of sentences, each triplet including a grammatical sentence, a number violation, and a gender violation, adapted from Bond et al. (2011) and a study by Alemán Bañón et al. (2012). A sample triplet is presented in (2).

(2) a. El laboratorio es moderno y el hospital también.
the lab-MASC-SG is modern-MASC-SG and the hospital too
"The lab is modern and so in the hospital"
b. El laboratorio es *modernos y el hospital también.
the lab-MASC-SG is modern-MASC-PL and the hospital too
c. El laboratorio es *moderna y el hospital también.
the lab-MASC-SG is modern-FEM-SG and the hospital too

Noun-adjective agreement was realized between the subject noun (e.g., *laboratorio* "lab-MASC-SG") and the predicative adjective (e.g., *moderno* "modern-MASC-SG"), which is the critical region. Ungrammatical sentences involved a feature mismatch between the subject

noun and the predicative adjective, either in terms of number (2b), a feature which is shared with English but not instantiated on adjectives in English, or gender (2c), a feature unique to the L2. Only singular subjects were used. An equal number of masculine and feminine subject nouns were used, all of which referred to inanimate entities, such that learners could not rely on semantic cues to establish syntactic gender agreement.

In order to minimize the effects of unfamiliar vocabulary on the learners' ability to process agreement dependencies, only lexical items that learners are likely to be familiar with, for example, cognates and words included in their Spanish textbook, were used in the study. The verb in the subject-verb conditions and the subject nouns and the adjectives in the noun-adjective conditions were used twice across the stimuli, making it possible to develop a large stimulus set with lexical items that would be familiar to learners. During recruitment, learners were also provided with a vocabulary list which included all adjectives, nouns, and verbs used in the experimental materials, together with their English translation. Learners were encouraged to study the vocabulary list before every EEG testing session. Similar to previous studies on the native and nonnative processing of agreement (Alemán Bañón et al., 2012, 2014; Barber & Carreiras, 2005; Gillon-Dowens et al., 2010, 2011), all nouns and adjectives in the noun-adjective conditions exhibited canonical number and gender marking (plural -*s*, masculine -*o*, and feminine -*a*), in order to ensure that learners were provided with overt morphophonological cues to establish agreement, a factor which has been argued to impact agreement processing in native speakers (e.g., Molinaro et al., 2011). To ensure familiarity with the vocabulary used in this experiment, learners completed a vocabulary test following their first EEG recording and a gender assignment task following their final EEG recording (see Supplementary Material 1 for methodological details regarding these tests). Learners performed

well on both the vocabulary test (100-level mean accuracy: 88.9%, 200-level mean accuracy: 95.7%) and the gender assignment test (100-level mean accuracy: 98.8%, 200-level mean accuracy: 98.2%).

Since both linear and structural distance have been shown to impact the processing of agreement in both native speakers (Alemán Bañón et al., 2012; Hammer et al., 2008) and L2 learners (Alemán Bañón et al., 2014; Foucart & Frenck-Mestre, 2012; Gillon-Dowens et al., 2010), all stimuli were controlled for linear and structural distance: in both the subject-verb and the noun-adjective agreement stimuli, agreement was realized across a verb phrase and the agreeing words were separated by one word. The critical word was the fourth word in the sentence in both the subject-verb and the noun-adjective agreement stimuli.

Three stimulus lists were created using a Latin Square design, such that participants would only see one version of each sentence per list. Each list of 240 sentences included 40 trials from each condition (subject-verb grammatical, subject-verb number violation, noun-adjective grammatical, noun-adjective number violation, and noun-adjective gender violation), as well as 40 grammatical fillers, which were simple, grammatical sentences, controlled in length to be similar to the target sentences. This yielded a 1:1 ratio of grammatical to ungrammatical sentences. Participants provided a grammaticality judgment following each sentence. No learner saw the same list more than once across the EEG recording sessions; as such, no stimuli were repeated either within or across sessions in the current study.

Procedure

All research protocols in this study were approved by the Institutional Review Board at the University of Kansas (#19586). Prior to the first EEG testing session, participants completed a

battery of individual differences measures, which will be discussed in a later section of the paper. Participants then completed a series of EEG recording sessions (three for 100-level learners, and two for 200-level learners). EEG recordings for each learner group at each testing point were always completed within a two-week period, to ensure that learners had received a comparable amount of Spanish instruction at that testing point. Following the final EEG testing session, participants completed an oral production task.

EEG Experiment Procedure

At each EEG recording session, learners were fitted with the EEG cap and seated on a comfortable chair in front of a computer monitor in a dimly-lit, sound-attenuated EEG testing room. Learners were asked to read a series of Spanish sentences and judge their grammaticality at the end of each trial by pressing one of two buttons on a response pad using the index and middle finger of their left hand. Learners were asked to favor accuracy over speed and to minimize muscle movement and blinks during the presentation of the sentences. All EEG recordings started with a practice session consisting of ten sentences, half of which were ungrammatical. None of the ungrammatical sentences involved agreement errors, and none of the content words used in the practice items appeared in the experimental stimuli. Learners received feedback for two of the practice trials. After the last practice sentence, the experiment began.

Sentences were visually presented one word at a time using RSVP (Rapid Serial Visual Presentation) with the software *Paradigm* (Tagliaferri, 2005). Each trial began with a fixation cross, which remained in the center of the screen for 500 ms. The first word of the sentence then appeared in the center of the screen. Each word in the sentence was presented for 450 ms followed by a 300 ms pause, which we have found to be a comfortable presentation rate for learners (e.g., Alemán Bañón et al., 2014, 2018; Bond et al., 2011). At the end of each sentence,

learners saw a screen with the prompts for the grammaticality judgment, the word *Bien* ("good") for grammatical sentences (on the left of the screen) and the word *Mal* ("bad") for ungrammatical sentences (on the right of the screen). The prompts remained on the screen until the participant's response, which was followed by a 500-1000 ms pause, pseudorandomly varied at 50 ms increments, after which the next trial began. Each recording consisted of six blocks of 40 sentences separated by brief breaks. Within each block, subject-verb, noun-adjective, and filler conditions were presented together in a fully randomized order. Words appeared in black text (Courier New font) on a grey background. All Spanish words exhibited the appropriate diacritics and the last word of each sentence was marked with a period. Performance on the end-of sentence GJT was quantified via acceptance rate for each condition, and via D-prime, which provides a single measure of grammatical sensitivity for each type of violation (subject-verb number, noun-adjective number, and noun-adjective gender), taking into account acceptance rates for both the grammatical and ungrammatical conditions for each violation type.[3]

The EEG was continuously recorded from 32 sintered Ag/AgCl scalp electrodes attached to an appropriately sized elastic cap (Electro-Cap International, Inc.) and placed in a modified 10-20 layout (midline: FPZ, FZ, FCZ, CZ, CPZ, PZ, OZ; lateral: FP1/2, F7/8, F3/4, FT7/8, FC3/4, T3/4, C3/4, TP7/6, CP3/4, T5/6, P3/4, O1/2). All electrodes were referenced online to the

[3] Following Bond et al. (2011), D-prime was calculated using a standard approach for false-choice experiments. We calculated H, the Hit rate for selecting *Bien* ("good") when the sentence was grammatical, and FA, the False Alarm rate for choosing *Bien* when the sentence was actually ungrammatical. D-prime was then derived using the following formula in Microsoft Excel: NORMSINV(H)-NORMSINV(FA)/SQRT(2). Hit rates of 1 (reflecting 100% acceptance rates) are corrected to 1 - 1/480, with 480 representing twice the number of items in the test). Likewise, False Alarm rates of zero were corrected to 1/480. In this analysis, chance performance would yield a D-prime score of approximately 0, with increasingly positive scores (up to 4) reflecting better performance.

left mastoid, and re-referenced offline to the average of the left and right mastoids. Electrode AFZ served as ground. Eye movements were monitored with an additional six electrodes placed on the left and right outer canthi, and above and below each eye. Impedance was maintained below 5 kΩ for scalp and mastoid electrodes, and below 10 kΩ for eye electrodes. The recordings were amplified by a Neuroscan Synamps2 amplifier (Compumedics Neuroscan, Inc.) with a bandpass filter of 0.1 to 200 Hz, and digitized continuously with a sampling rate of 1 kHz.

Since the learners in the study are at the beginning stages of L2 acquisition, and rejecting incorrectly judged trials would significantly reduce the number of trials per condition and thus the signal-to-noise ratio, we decided to include all experimental trials in the analysis regardless of accuracy in the GJT; this procedure mirrors that followed in the Osterhout et al. (2006) study which we build on here, and has also been adopted in other recent research examining beginning learners (e.g., Alemán Bañón et al., 2018; Mickan & Lemhöfer, 2020). Moreover, as pointed out by VanRullen (2011), restricting analyses to only correct trials runs the risk of introducing artifactual differences in the EEG responses across conditions. Importantly, Tanner et al. (2013) demonstrated that ERP responses were similar in analyses keeping all trials and analyses keeping only correct trials, providing further support for our decision to pursue an analysis including all trials (but see E.J. White et al., 2012 who show that an analysis of 'correct trials only' yielded larger P600s).

The continuous EEG was segmented into epochs in the interval from −300 and +1200 ms relative to the critical word (subject-verb conditions: verb; noun-adjective conditions: adjective), and trials with blinks, horizontal eye movements, excessive muscle activity or excessive alpha waves were identified via visual inspection and eliminated from analysis using Neuroscan Edit (Compumedics Neuroscan, Inc.). The number of trials kept across conditions ranged from 21-40

out of 40 trials across participants (mean by condition: Subject-Verb Grammatical: 36/40, Subject-Verb Violation: 35.9/40, Noun-Adjective Grammatical: 36/40, Noun-Adjective Number Violation: 35.7/40, Noun-Adjective Gender Violation: 35.4/40; the number of trials kept did not differ significantly across conditions; repeated-measures ANOVA; $F(4,180) = 1.745$, $p = .152$). The epoched data were then imported into EEGLAB (Delorme & Makeig, 2004) for further processing. Following interpolation of any bad channels, the epoched data were baseline-corrected relative to the 300 ms pre-stimulus interval, filtered using a 30 Hz low-pass digital filter, and averaged.

Based on the native Spanish results in Bond et al. (2011), which established P600 as the native Spanish response to violations of subject-verb number and violations of noun-adjective number and gender in our study (in line with previous ERP studies on agreement), and based on visual inspection of the grand-average waveforms for the learners, ERPs were quantified via mean amplitudes in the 500-900 ms (P600) time window (Friederici, 2002; Hagoort et al., 1993; Osterhout & Holcomb, 1992) in a broad posterior region characteristic of the P600 (TP7, CP3, CPZ, CP4, TP8, T5, P3, PZ, P4, T6, O1, OZ, and O2).[4] Averaging across sites comprising a characteristic distribution for a component of interest, as compared with approaches yielding means from multiple regions, is advantageous for studies that are not focused on topographic factors, as it allows one to quantify ERP waveforms with high signal-to-noise ratio while reducing the number of statistical comparisons that are necessary (see Luck, 2014 for discussion). We would also like to note that although the analysis reported here focuses on P600, an exploratory ANOVA analysis of ERP effects in an earlier 300-500 ms time window characteristic of the LAN and N400 revealed that neither the 100-level group nor the 200-level

[4] For a full report of the native Spanish results, we refer the readers to Bond et al. (2011), which is available with open access at http://www.lingref.com/cpp/gasla/11/paper2542.pdf.

group in the current study yielded significant negativities for either subject-verb number violations or noun-adjective number or gender violations, at any region across the scalp (left anterior, left posterior, midline anterior, midline posterior, right anterior, right posterior), for any EEG recording session. As such, this paper focuses on P600 and examining variability in this component.

ERP amplitudes from this P600 region and time window, calculated separately for each participant, condition and EEG testing session, will be utilized in the current study as outcomes within multivariate longitudinal statistical models, as discussed below.

Results

Table 1 provides descriptive statistics for the ERP amplitudes by learner group (100- versus 200-level students) and EEG testing session (after two, six, or eight months of instruction for the 100-level learners; after two or six months of instruction for 200-level learners) for each of the five conditions.

Table 1 about here

Grand-average ERP waveforms for condition and testing session are shown for representative electrode Pz in the 100-level learners (Figure 1) and 200-level learners (Figure 2); see Supplementary Material 2 for additional figures providing a larger array of electrodes for each learner group, condition, and testing session, and Supplementary Material 3 for topographic plots for each condition and for the difference between each violation condition and its corresponding grammatical condition, for each learner group and session. As shown in the bottom section of Table 1, of primary interest are the *grammaticality effects* that can be observed

in our experimental design: those elicited by the Subject-Verb Number violation condition as compared to the baseline grammatical condition, and those elicited by the Noun-Adjective Number violation condition and the Noun-Adjective Gender violation condition, as compared to the baseline grammatical Noun-Adjective condition. Accordingly, we first calculated the difference in amplitude for each violation condition relative to its grammatical counterpart, and we then analyzed these three differences as outcomes simultaneously within multivariate longitudinal models (see Hoffman, 2015). This strategy allowed us to most directly examine differences in the size of the grammaticality effects across the three conditions as posed by each research question: by EEG testing session (Session) and Learner Group (RQ1), by GJT accuracy and production abilities (RQ2), and by other sources of individual differences (RQ3), as reported in later sections of the paper.

Figure 1 about here

Figure 2 about here

Analytic Strategy

Given that the 100-level learners completed EEG testing sessions after two, six, and eight months but the 200-level learners completed EEG testing sessions only after two and six months, a traditional repeated measures analysis of variance using least squares estimation would have resulted in listwise deletion of half the sample. Instead, we estimated our multivariate linear models within SAS MIXED using restricted maximum likelihood (REML) and Kenward-Roger denominator degrees of freedom. Although equivalent to least squares given complete outcomes, REML allows cases with missing outcomes to still be included under an assumption of missing

at random (i.e., conditionally random after considering all other outcomes and model predictors). Given the complexity of the resulting doubly multivariate models, we report effect sizes for each outcome by translating the *t* or *F* test-statistics into Cohen's *d* standardized mean difference units. To ensure accurate standard errors for all hypothesis tests regarding Learner Group and Session, the adequacy of fit of the variance-covariance model across all outcomes and sessions for each Learner Group was first evaluated using likelihood ratio tests (i.e., the χ^2 for the −2LL difference between nested models given degrees of freedom equal to the number of additional model parameters). We initially estimated a fully unstructured variance-covariance matrix with separate parameters for each Learner Group (i.e., a multivariate approach for the doubly repeated measures of outcomes by sessions, resulting in 75 variance-covariance parameters). To maximize power, we then sought to reduce the number of parameters by examining alternative models. Fortunately, we found a more parsimonious model that showed no decrease in fit, $-2\Delta LL(61) = 72.7$, $p = .145$, which we used for RQ1 below. This model predicted between-subject covariance across sessions using Learner Group-specific random intercept variances (for the Noun-Adjective Gender violation condition for the 100-level learner group; for the Subject-Verb Number Violation condition for the 200-level learner group). Within subjects, the model included an unstructured matrix for the residual variances and covariances across outcomes at the same Session, whose parameters were estimated separately by Learner Group.

Research Question 1: Grammaticality Effects by Session and Learner Group

The first research question asked if learners process features that are instantiated in both the L1 and L2 differently than features unique to the L2. We also asked whether sensitivity to morphosyntactic violations changes over the course of development, examining performance in

two different learner groups (100-level, 200-level) across multiple sessions. To address RQ1, we initially estimated a saturated means model (i.e., as in a traditional analysis of variance) that included all possible effects, including main effects and interactions for Condition (violation type), Learner Group (100- vs. 200-level) and Session (after two, six, or eight months) for each outcome. However, the change between two to six months had no significant interactions with Learner Group for any outcome, $F(3,68) = 0.180$, $p = .908$, nor any significant main effects of Session, $F(3,54) = 0.710$, $p = .549$. In addition, for the 100-level learners who had one additional session, no significant change from six to eight months was found for any outcome, $F(3,54) = 0.300$, $p = .824$.

These nine nonsignificant effects related to Session were thus removed to more parsimoniously examine marginal mean differences by Learner Group for each outcome. These results are summarized in Table 2.

Table 2 about here

Both the 100-level learners and the 200-level learners showed significant P600s for Subject-Verb and Noun-Adjective Number violations; both the Subject-Verb Number effect and the Noun-Adjective Number effect were significantly larger (more positive) in the 200-level students than in the 100-level students. Further, the size of these number violation effects was equivalent across the Subject-Verb and Noun-Adjective agreement conditions for both groups. Lastly, a marginally significant positivity for the Noun-Adjective Gender violations was found only for the 200-level learners ($p = .086$).

To summarize these findings regarding sensitivity to agreement violations, the learners revealed significant P600s for number agreement violations, both in the Subject-Verb Number condition and in the Noun-Adjective Number condition, with the 200-level learners showing larger positivities than the 100-level learners. The 100-level learners showed no sensitivity to Noun-Adjective Gender violations; the marginal positivity for Noun-Adjective Gender violations in the 200-level group suggests a trend toward emerging sensitivity to gender agreement violations. Thus, learners overall showed sensitivity to agreement involving the number feature, which is shared between the L1 and L2, including number agreement on adjectives, where number is not instantiated in English. This contrasts with the findings for gender, a feature unique to the L2, where a trend toward sensitivity was only evident in the more advanced, 200-level group.

Examining the relationship between ERPs and receptive and productive knowledge of morphosyntax

While our first research question examined sensitivity to agreement dependencies at the group level, our second research question probes variability in those brain responses at the individual level. We examined whether increased sensitivity to agreement dependencies at the moment that those dependencies are computed, as measured by ERPs, is related to (1) receptive knowledge of the same dependencies when examined under less time pressure on an end-of-sentence GJT (see Alemán Bañón et al., 2018; Caffarra et al., 2015; McLaughlin et al., 2010; Steinhauer et al., 2009; Tanner et al. 2013, 2014; E.J. White et al., 2012) and (2) productive

knowledge of equivalent dependencies as measured by an oral production task (Alemán Bañón et al., 2017). Our second research question is given below:

RQ2. To what extent is sensitivity to grammatical violations (P600) within each condition related to GJT performance and oral production abilities?

The examination of performance on the GJT is motivated by research which shows a strong relationship between the online processing of agreement and GJT performance offline. We included this analysis in order to be able to compare our results to those of previous studies, which largely tested more proficient learners. In a meta-analysis of 41 studies, Caffarra et al. (2015) observed a strong relationship between proficiency, which included both performance on global standardized measures of proficiency and performance on GJTs, and the emergence of P600 for grammatical violations; proficiency was a more important predictor than a range of other factors including immersion in an L2 environment, age of acquisition, and L1-L2 similarity. In a large-scale (n=78) study examining both gender and number agreement, a study from our own lab also observed a relationship between GJT accuracy and P600 magnitude, further suggesting that neurophysiological sensitivity is related to sensitivity offline (Alemán Bañón et al., 2018). However, it is important to note that as GJT performance is an experiment-internal factor, it cannot account for individual learner variability outside of the specific experimental context (Tanner et al., 2013). In addition, grammaticality judgment tasks recruit explicit, metalinguistic knowledge and it is an open question whether online sensitivity to morphosyntactic violations, as revealed by ERPs such as the P600, is related to the ability to productively use agreement in a more natural language context.

To address this issue, the present study also examined the relationship between P600 and performance on an oral production task, which does not explicitly draw learners' attention to morphosyntax. Oral production is a domain in which even advanced learners show variability in the use of morphosyntactic agreement, and production abilities have often been argued to lag behind offline comprehension abilities (Montrul et al., 2008; Grüter et al., 2012; but see McCarthy, 2008). It is interesting to examine whether neurophysiological sensitivity as measured by P600s is related to learners' ability to establish agreement dependencies in production. The oral production task that we developed for the present study is meaning-based, focusing learners' attention on describing the differences between two related pictures. Thus, it is different from lexical gender assignment tasks, which often present words in isolation, and different from GJTs, which examine agreement dependencies in an explicitly metalinguistic context. The task is described in detail below.

Oral production task

Learners were tested on an oral production task examining the use of subject-verb and noun-adjective gender and number agreement. On each trial, learners were presented with two pictures and asked to describe what was different about them. The pictures in the subject-verb conditions showed characters involved in different activities. Crucially, one picture depicted a single character and the other picture depicted two characters. For example, one picture showed one person drinking at the beach, while the other picture showed two people drinking at a restaurant (see Figure 3). Thus, on this trial, a grammatically correct utterance would involve both an instance of plural subject-verb agreement and an instance of singular subject-verb agreement, such as "*Amy y Mary beben en el restaurante y Henry bebe en la playa*" 'Amy and Mary drink in

the restaurant and Henry drinks on the beach.' In order to minimize the effects of unfamiliar vocabulary, learners were provided with the infinitive form of the Spanish verb that they were to use in their utterance (e.g., *beber*, "to drink").

Figure 3 about here

In the noun-adjective conditions, participants were asked to describe what was different about the objects that the characters in the pictures had. For example, one picture showed a woman with one clean shirt, while the other picture showed a man with two dirty shirts (see Figure 4). As *camisa* ("shirt") is a feminine noun in Spanish, a grammatical utterance describing what is different between these two pictures would thus be expected to involve an instance of feminine, singular noun-adjective agreement marking (picture on the left), and an instance of feminine, plural noun-adjective agreement marking (picture on the right); for example, a grammatical utterance describing the pictures in Figure 4 would be "*Wendy tiene una camisa limpia y Kenny tiene unas/dos camisas sucias*" 'Wendy has a clean shirt and Kenny has some/two dirty shirts'. This part of the task included an equal number of masculine and feminine nouns; only Spanish adjectives that show the masculine-feminine distinction were included. Learners were provided with the target adjectives in English (e.g., "clean" and "dirty"), to ensure that they would not use synonymous adjectives that are not marked for gender; the adjectives were shown in English to avoid providing the adjective specified for one of the two genders. Learners were also provided with the infinitive form of the Spanish verb they were to use in their utterance (e.g., *tener*, "to have").

Figure 4 about here

The production task included 10 trials per condition, presented in blocks, plus four practice items (two targeting subject-verb agreement and two targeting noun-adjective agreement). All learners started with the subject-verb conditions. Within each block, trial presentation was counterbalanced. Learners were told that they would have limited time (approximately one minute) to describe the sentences. The experimenter controlled the presentation rate of the trials. Before their visit to the lab, learners were provided with a short vocabulary list including all nouns, adjectives, and verbs in the task to help ensure that the learners were familiar with the necessary vocabulary to perform the task. For details on the coding of the learners' productions in this task, see Supplementary Material 4.

Results

The second research question examined to what extent the ERP grammaticality effects can be predicted by the end-of-sentence GJT as well as performance on the Oral Production task. In order to evaluate the predictors' total effects, all main effects for Learner Group were removed in the analyses reported below. Preliminary analyses revealed no interactions with Session for any predictor in predicting any of the grammaticality outcomes, and thus only the predictor main effects are referred to below. Likewise, we specified a fully unstructured variance-covariance matrix across all outcomes and sessions whose parameters were combined across Learner Group given that simpler models resulted in a decrease in model fit.

The top of Table 3 provides descriptive statistics for acceptance rates on the end-of-sentence judgment task by Learner Group and Session for each of the five original conditions; these are

provided for descriptive purposes only. D-Prime values for each type of violation, which were used in the analysis, are given at the bottom of Table 3. Descriptive statistics for performance on the production task are summarized in Table 4.

Table 3 about here

Table 4 about here

The effects of D-Prime rates in predicting grammaticality effects are given in Table 5. As expected, greater D-Prime rates predicted significantly greater grammaticality effects (larger positivities) for Subject-Verb Number violations, Noun-Adjective Number violations, and Noun-Adjective Gender violations.

Table 5 about here

We then examined to what extent each outcome could be predicted by an individual's ability to correctly instantiate Subject-Verb Number agreement, Noun-Adjective Number agreement, and Noun-Adjective Gender agreement in the Oral Production task, as is also reported in Table 5. The Subject-Verb production task yielded two predictors: one for singular and one for plural. Given their moderate correlation (r = .57), prediction by singular and plural production was examined in separate models. Singular production ability did not predict any outcome, but greater P600s for Noun-Adjective Number violations were found for learners with greater ability in the Subject-Verb plural condition on the production task. Likewise, we examined production ability as measured by the Noun-Adjective conditions, which yielded four predictors: performance on Noun-Adjective singular, plural, masculine, and feminine. Given their high

correlations (*r*'s = .648 to .820), we estimated a latent factor (M = 0, VAR = 1) to represent their common variance, and then used *expected a posteriori* estimates of the factor scores as a single model predictor. As with Subject-Verb plural production, greater ability in the Noun-Adjective latent factor predicted greater P600s for Noun-Adjective Number violations.

To summarize, a clear relationship between grammatical sensitivity in the behavioral GJT task as measured by D-prime and the elicitation of P600s for violations emerged; lower acceptance rates for ungrammatical sentences (reflecting increasing sensitivity) led to significantly larger positivities for all violation types. On the Oral Production task, better performance on the production of Subject-Verb plural agreement and better performance on the production of Noun-Adjective agreement overall (as assessed by a latent factor) was associated with significantly larger P600s for Noun-Adjective Number in the ERP experiment.

Individual Differences

In the L2 psycholinguistics literature broadly, there has recently been great interest in examining whether individual differences in both linguistic and cognitive abilities are related to grammatical performance (e.g., Abrahamsson & Hyltenstam, 2008; DeKeyser, 2000; DeKeyser et al., 2010; Granena et al., 2016; Juffs & Harrington, 2011; Linck et al., 2014; Roberts, 2012; Robinson, 2001; Wen, 2014). In the ERP literature examining morphosyntax, P600 effects have been related to level of L2 use (Meulman et al., 2014) as well as motivation for learning the L2 (Tanner et al., 2014). The broader L2 ERP literature has also begun to consider individual differences in cognitive abilities, such as working memory (Dallas et al., 2013; Faretta-Stutenberg & Morgan-Short, 2017; Reichle et al., 2016), procedural learning ability (Faretta-Stutenberg & Morgan-Short, 2017), and non-verbal intelligence and reasoning (Bond et al.,

2011). Cognitive abilities such as working memory, which is responsible for the maintenance of information when confronted with processing potentially distracting information (Conway et al., 2005, 2007), are particularly interesting to examine, as working memory has been found to account for variability in language comprehension in both native speakers (see Daneman & Merikle, 1996) and L2 learners (see Linck et al., 2014).

Despite this initial progress, as many studies consider only a single skill such as language aptitude or working memory, it is difficult to determine whether one factor modulates performance in conjunction with other factors or to the exclusion of them. For example, DeKeyser and colleagues (DeKeyser, 2000; DeKeyser et al., 2010) argue for the importance of verbal aptitude in late but not early L2 learners (but see Abrahamsson & Hyltenstam, 2008, 2009), and present this evidence in support of Bley-Vroman's (1990) proposal. However, in Bley-Vroman's original proposal, he mentions "general abstract problem-solving skills" (Bley-Vroman, 1990, 4) and thus, it seems possible that *non-verbal* cognitive abilities were considered to be most important. Ideally, following recent studies of variability in the native processing literature, L2 studies would include a range of *both* verbal and non-verbal measures (e.g., Boudewyn, 2015; Van Dyke et al., 2014) so that we can evaluate whether the skills that facilitate L2 acquisition are within the language domain or whether it is general cognitive abilities that play a role, or possibly, a combination of these factors.

To our knowledge, Bond et al. (2011), a study from our own lab, is the only study to have taken such an approach. Bond et al. examined the relationship between both verbal aptitude and non-verbal intelligence and reasoning and the processing of number and gender agreement violations in English-speaking learners of Spanish at an intermediate level. The study compared the processing of subject-verb agreement, noun-adjective number agreement, and noun-adjective

gender agreement. All three violation types elicited P600 in the native speaker control group. For the learners, the three violation types also elicited P600, but the effect for noun-adjective number was larger than the effect for noun-adjective gender, suggesting a facilitative role for the L1.

With respect to individual differences, scores on the verbal aptitude test were related to accuracy on the GJT for subject-verb agreement and the size of the P600 effect for noun-adjective number violations. There was no relationship between verbal aptitude and dependent measures related to gender agreement and there were no significant effects for the test of nonverbal intelligence. Bond et al. (2011) interpreted these results as an argument against Bley-Vroman's (1990) proposal that it is domain-general 'problem-solving' skills that predict L2 performance, suggesting instead that verbal skills are most important. While Bond et al. provided a first step in the examination of variability in brain responses, the study is limited in that they included only one measure of non-verbal abilities (Raven, 1965) and tested learners at one level of proficiency. Thus, it is difficult to make strong conclusions with respect to the lack of correlations with non-verbal abilities.

The present study builds on Bond et al. (2011), using the same stimulus manipulations, but examining a larger group of learners at two different levels of proficiency, and including a wider range of verbal and non-verbal individual differences measures. Our third research question is given below:

RQ3. To what extent do individual-level abilities in both verbal and non-verbal skills predict sensitivity to grammatical violations (P600) of number and gender agreement? To what extent do the same individual-level abilities predict performance across three different measures of agreement dependencies (P600 effect size, GJT accuracy,

production)?

Our first goal is to examine whether individual level abilities in the verbal and non-verbal domain predict sensitivity to morphosyntactic violations as measured by ERP responses. We include two measures of verbal aptitude, a subtest of the *Modern Language Aptitude Test* (MLAT) (Carroll & Sapon, 1959) and a subtest of the *LLAMA Language Aptitude Test* (Meara, 2005a,b). The subtest of the MLAT, the MLAT-4: Words in Sentences, targets grammatical sensitivity in the learners' L1 as the test itself is in English. Our second measure, the LLAMA_F, also tests grammatical sensitivity but in a 'new' language unfamiliar to the participants.

We also included two subtests of the *Wechsler Adult Intelligence Scale IV* (*WAIS-IV*) (Wechsler, 2008), including one measure of working memory (*Letter Number Sequencing*) and one measure of perceptual reasoning (*Matrix Reasoning*). As discussed above, working memory was included because it has been shown to modulate grammatical processing in both natives and L2 learners. We included the measure of perceptual reasoning, which tests non-verbal visual intelligence and spatial ability, in order to assess the kind of non-verbal cognitive abilities or 'problem-solving' abilities that Bley-Vroman (1990) referred to. We reasoned that a relationship may emerge between perceptual reasoning and ERP effects if learners potentially rely on a heuristic strategy in which they focus on the visual co-occurrence of particular word forms, as suggested by McLaughlin et al. (2010), but would be unexpected if learners are engaged in grammatical processing.

Our second goal related to individual differences is to examine whether the same abilities predict success for features that are either similar or different in the L1 and L2. This question is interesting in light of L2 theories such as Bley-Vroman (1990) and Hawkins (2001) which

predict that a qualitatively different mechanism underlies the acquisition and processing of features unique to the L2. These proposals can be interpreted to suggest that different skills may facilitate processing of these features and thus would predict that different abilities may underlie the processing of number and gender in the present study.

Finally, we examine whether the same individual-level abilities predict performance across different measures, including an Oral Production task, as was discussed above. It has been suggested that individual differences in proficiency and cognitive abilities may influence processing only under certain experimental conditions, for example when learners are intentionally focusing on metalinguistic information (Roberts, 2012). Roberts (2012) suggests that when learners are instructed to focus on grammatical details, then they will process deeply, and individual-level abilities may play a role. If however, the task is meaning-based, learners are more likely to process shallowly (not attending to grammatical detail) and individual-level abilities will not play a role. The present study examines whether the same abilities impact processing across three different measures: ERP effect sizes, GJT accuracy, and Oral Production. The oral production measure we use focuses on meaning (a comparison of two pictures) and thus allows us to evaluate whether the individual differences that modulate performance indeed vary dependent on whether learners are asked to focus on form.

Individual differences assessments

Verbal aptitude was measured via the *Modern Language Aptitude Test, Test 4: Words in Sentences (MLAT-4)* (Carroll & Sapon, 1959) and the *LLAMA Language Aptitude Test, Test F, (LLAMA_F)* (Meara, 2005a,b). The MLAT-4 is a 45-item pencil-and-paper test. Each item presents an English sentence with several words underlined, and participants are asked to select

the one choice that plays the same grammatical function (e.g., subject of the sentence) as the underlined word in an example sentence. Learners were allowed 20 minutes to complete the task. MLAT-4 performance was quantified as number correct out of 45 questions. The LLAMA_F includes two phases: a learning phase and a test phase. During the learning phase, learners are presented with a series of pictures and sentences from an unfamiliar language describing the pictures and are given five minutes to work out as many grammatical rules for the language as possible. During the untimed test phase, learners are presented with a total of 20 pictures, each of which is accompanied by two sentences, and are asked to identify the sentence that correctly describes the picture. The test is scored automatically on a scale from 0-100%.

Learners then completed two subtests from the *WAIS-IV* (Wechsler, 2008). Participants completed the *Letter-Number Sequencing* subtest, an assessment of working memory. This subtest involves recalling intermixed strings of numbers and letters. Participants must recall the numbers first, ordering them from smallest to largest, and then the letters, ordering them alphabetically. The sets ranged in size from three to eight items. If a participant did not correctly recall any item within a particular set, the experimenter would stop the test and not move on to the next set size. The highest possible score is 30. Participants also completed the *Matrix Reasoning* subtest, which examines perceptual reasoning. Participants are asked to look at visual arrays and identify the one option that correctly completes the matrix. The highest possible score is 26. Raw scores from the Letter-Number Sequencing and Matrix Reasoning subtests were converted to scaled scores based on WAIS-IV age-based norms.

Results

Our third research question examined to what extent individual differences in verbal aptitude

and non-verbal cognitive abilities predicted performance on multiple measures of morphosyntax: ERP effects (P600), GJT accuracy rates, and performance on the Oral Production task. The statistical approach is the same as was described for the second research question. Descriptive statistics for these predictors are shown in Table 6. The results from these models are summarized below, and full results for each predictor and outcome are reported in Table 7, in which the R^2 values were computed as the square of the correlation between the outcome predicted by the model fixed effects and the observed outcome variable.

Table 6 about here

Table 7 about here

With respect to sensitivity to grammaticality violations (P600), only one individual difference predictor was uniquely significant: Learners with better working memory as assessed by the Letter-Number WAIS-subtest exhibited significantly greater ERP grammaticality effects for Noun-Adjective Number violations (p = .041), with a similar trend observed for Noun-Adjective Gender violations (p = .071). With respect to grammatical accuracy (D-prime), a different individual difference predictor was uniquely significant: learners with greater verbal aptitude as measured by the MLAT-4 had significantly higher D-prime rates for all three conditions: Subject-Verb Number (p = .016), Noun-Adjective Number (p = .009), and Noun-Adjective Gender (p = .034). Lastly, with respect to ability to establish agreement dependencies on the production task, whereas MLAT-4 again showed significant positive unique effects on all three conditions— Subject-Verb Number (p = .007), Noun-Adjective Number (p = .021), and Noun-Adjective Gender (p = .025)—higher Letter-Number scores also uniquely predicted

performance in the production task for Noun-Adjective Number (p = .005), and marginally so for Noun-Adjective Gender (p = .063)[5].

Table 8 summarizes the significant findings of our statistical models across the three different outcome measures. To summarize, working memory (Letter-Number) was the unique predictor of ERP grammaticality effects (significant for noun-adjective number/marginal for noun-adjective gender). Working memory was also a significant predictor of performance on the production task (significant for Noun-Adjective Number / marginal for Noun-Adjective Gender) but production abilities were also predicted by performance on the MLAT-4 (significant across all conditions). Performance on the GJT, as measured by D prime, was also predicted by the MLAT-4 (significant across all conditions).

Table 8 about here

General Discussion

Our first research question asked if learners process features which are shared by the L1 and L2 differently than features unique to the L2. For shared features, we further asked whether the specific morphological instantiation of the feature modulates performance. We also asked whether we would observe development both over time and with increasing exposure to the L2.

The results showed P600s for both 100-level and 200-level learners for both subject-verb agreement and noun-adjective number agreement, but we observed only a marginal positivity for gender, an effect which emerged only for the 200-level learners. This suggests that features

[5] Note that on the production task, analyses collapsed across singular/plural for number and across masculine/feminine for gender.

shared between the L1/L2 such as number, regardless of their morphological instantiation in the L1/L2, show a processing advantage as compared to features unique to the L2. These patterns are most in line with feature-based theories of transfer which propose that learners can build on the grammatical inventory of the L1 at early stages of acquisition. In addition, although the marginal effect for gender is weak, the result suggests that features unique to the L2 can potentially be acquired at more advanced levels of proficiency (Schwartz & Sprouse, 1994, 1996), in line with studies which have shown this pattern for more advanced learners (e.g., Alemán Bañón at al., 2014, 2017, 2018, Gabriele et al., 2013).

With respect to number, in contrast to the predictions of the cognitive theories (e.g., DeKeyser, 2007; Ullman, 2001, 2005, 2015), we did not observe discrete stages in which agreement dependencies were first processed as lexical associations (N400) (Carrasco et al., 2017; McLaughlin et al., 2010; Mickan & Lemhöfer, 2020; Morgan-Short et al., 2010; Osterhout et al., 2006; Reichle et al., 2016; Tanner et al., 2013). Our results did not reveal a significant effect of Session, showing similar ERP amplitudes across the different testing sessions, which suggests that the sensitivity we observed for number was in place from the first session[6]. Although we did not observe change over the course of the academic year, our group comparisons do suggest development: larger effects for both subject-verb agreement and noun-

[6] It is important to note that a direct comparison between the results of the present study and Osterhout et al.'s (2006) longitudinal study is difficult as it is unclear whether the statistical analyses in Osterhout et al. (2006) included Session as a factor, as was the case in the present study.

adjective number agreement were observed for the 200-level learners, who also performed at higher levels on the accompanying GJT[7].

While the advantage for number that emerged is most in line with theories of transfer, which argue that the entire inventory of L1 features is available at the initial state, there is an important open question. As we only tested English native speakers, it is difficult to say, without further testing, whether the processing of gender agreement dependencies is acquired later because it is a feature unique to the L2 or because it is harder due to the fact that gender agreement entails both a lexical component (assigning lexical gender to the noun) and a syntactic component (agreement on the adjective), as compared to number, which involves only syntactic agreement. Despite these differences between gender and number which may impact processing, it is important to note that previous studies have shown the two features to be processed similarly by both native speakers (e.g. Alemán-Bañón et al., 2012), and by advanced L2 learners of Spanish of varying L1 backgrounds (e.g., L1 Chinese: Gillon-Dowens et al., 2011; L1 English advanced learners: Alemán-Bañón et al., 2018). It is also important to point out that in the present study, we made a strong effort to limit the lexical burden associated with gender in order to minimize the differences between gender and number. As discussed above, learners were given a list of

[7] A reviewer points out that the learners who were at a higher level of proficiency (the 200-level learners) also had higher scores on the tests of verbal aptitude, particularly the LLAMA-F, and thus it is difficult to tell whether the larger ERP effects and higher levels of accuracy on the GJT are due to proficiency or aptitude. This is an interesting open question and the current dataset does not allow us to tease apart those two factors. It is very possible that performance on the LLAMA-F increases with increased exposure to an L2 as this aptitude test is based on learning a new grammatical system. Importantly, the LLAMA-F was not a significant predictor of any dependent measure in the study and thus, it is unlikely that verbal aptitude, as measured by the LLAMA-F, was responsible for the higher levels of performance.

vocabulary items and were tested on those items in the first session in order to ensure familiarity with the lexical items. The nouns that we selected were all canonically marked for gender (masculine nouns ending in *-o*, and feminine nouns ending in *-a*), providing strong distributional cues for lexical gender, in an attempt to focus more explicitly on agreement. As was reported above, learners performed extremely well on both the tests of vocabulary and gender assignment. Thus, while a strong claim of transfer demands testing of an additional group of novice learners whose L1 does not instantiate number or gender agreement, we believe the study took important precautions to allow for a fair comparison between number and gender.

Next, we will consider why the low-proficiency learners in our study may have been more successful than those in previous studies, yielding significant positivities for number in two conditions and showing a marginal effect for gender in the more advanced learners. We will consider specific factors related to number, gender, and task in our study.

Our results for number differ from the results for number agreement dependencies in Osterhout et al.'s (2006) longitudinal study, which inspired our own. Osterhout et al.'s (2006) results suggested discrete stages in which novice learners initially process dependencies not via grammatical means, but via lexical associations (N400). In addition, sensitivity (at any session) emerged only for the one dependency that was similar in the L1/L2. The results of our study are important because they show that it is indeed *possible* for novice learners to process grammatical agreement dependencies in a qualitatively similar way to native speakers (P600), even at very early stages of development. To our knowledge, P600s have not been observed previously in the domain of agreement for novice classroom learners. E.J. White et al. (2012) showed significant P600s for tense violations (ex. *The teacher did not start/*started the lesson*) in intermediate Chinese and Korean learners of English. Davidson and Indefrey (2009b) observed sensitivity to

violations of verb order in a longitudinal MEG study with novice German learners of Dutch, a result that they argued was due to the typological similarity of German and Dutch and the salience of the verb order violations. In the domain of agreement, Bastarrika and Davidson (2017) showed increasingly native-like patterns after explicit training in native speakers of Spanish who were exposed to a fragment of the Basque grammar and were tested on short three-word phrases targeting number violations. Number is a shared feature in Spanish/Basque, but number agreement is instantiated differently in the two languages. Bastarrika and Davidson argue that L2 processing can become increasingly 'native-like' in a matter of only hours of training, at least for a simple grammatical rule in a miniature language, tested with short phrases. With respect to our own results, the qualitatively native-like processing profile that our learners showed for number in more complex sentences was likely facilitated by the properties of the agreement system in Spanish, in which all of the agreement patterns we examined are phonologically realized, a factor which has been argued to facilitate processing (see McLaughlin et al., 2010). Our study also targeted number agreement in the 3rd person in Spanish, a context in which number agreement is overtly marked in English. This may have facilitated processing in comparison to the 2nd person context in which, for example, Osterhout et al. (2006) targeted agreement because English does not overtly mark agreement in the 2nd person. However, it is also important to note that we also observed P600s in the noun-adjective number condition in L2 Spanish, a context in which English does not mark agreement.

Notably, as compared to Osterhout et al.'s (2006) study, which included only 2nd person pronominal subjects in the subject-verb agreement condition, the present study encompassed a wide range of lexical subjects. McLaughlin et al. (2010) suggest that the N400 which Osterhout et al. (2006) observed reflects sensitivity to probabilistic dependencies between morphemes that

frequently co-occur, as in the 2nd person subject *Tu* and the inflectional ending *-s* in French. This strategy is perhaps more likely to be recruited when only one dependency is tested in a given condition. The fact that lexical subjects varied across all of our items may have made it less likely that learners would be able to rely on a strategy of 'pattern-matching' in the course of the experiment.

In addition, the specific number violation that we tested in both the subject-verb and noun-adjective conditions (plural marking in the context of a singular noun) has been found to be detected earlier by L2 learners (see McCarthy, 2008; Lopez-Prego & Gabriele, 2014) as opposed to a pattern in which singular agreement is supplied in the context of a plural noun, and thus, may be more salient. One limitation of our experimental design is that we did not include plural subject noun phrases with grammatical agreement; as such, the plural marked adjective and verb phrases may have served as an indicator of a grammatical violation. Nevertheless, the patterns that emerged in our study (better performance for number than gender and sensitivity to number agreement violations with singular nouns) are similar to those reported in other studies which included plural noun phrases in grammatical contexts (e.g., Alemán Bañón et al., 2017; Lopez-Prego & Gabriele, 2014).

With respect to gender, although the effect we observed was weak, we did observe a marginal effect in the more advanced learner group, suggesting sensitivity to gender violations. We believe this trend may have emerged due to the fact that Spanish is a language with strong distributional cues to lexical gender and our participants had strong knowledge of the lexical gender associated with the items that we tested. Many previous ERP studies which have shown native-like ERP profiles with respect to gender agreement in advanced proficiency learners have also examined L2 Spanish (see e.g., Alemán Bañón et al., 2014, 2017; Gabriele et al., 2013;

Gillon-Dowens et al., 2010, 2011)[8]. In studies examining languages such as French (Foucart & Frenck-Mestre, 2012) or Dutch (Sabourin & Stowe, 2008; Lemhöfer et al., 2014), with weaker distributional cues to lexical gender, results showing native-like processing profiles are less frequently observed and, in several studies, have been shown to be strongly related to knowledge of lexical gender (Sabourin & Stowe, 2008; Lemhöfer et al., 2014). Lemhöfer et al. (2014) showed that when responses to gender violations were examined for German learners of Dutch in accordance with the objective rules of Dutch gender assignment, L2 learners did not show native-like sensitivity. However, when the learners' subjective gender assignment (assessed in an offline task) was taken into account, and the ERP responses were re-sorted in accordance with subjective assignment, a native-like P600 response emerged. Thus, knowledge of lexical gender is arguably one of the necessary components to successfully establishing gender agreement online and learners' ability to acquire lexical gender is strongly impacted by the nature of the distributional cues available in each language (see also discussion in Grüter et al., 2012; Hopp, 2013). Following previous ERP studies, we also included a measure of lexical gender (e.g., Alemán Bañón et al., 2014, 2017; Foucart and Frenck-Mestre, 2012; Gabriele et al., 2013; Gillon-Dowens et al., 2010, 2011; Lemhöfer et al., 2014; Sabourin & Stowe, 2008) and found little variability with respect to knowledge of lexical gender even among our low proficiency learners: their target-like performance on the gender assignment task likely facilitated their processing of gender agreement dependencies online.

A reviewer also points out that the use of a GJT in our study may have led the participants to focus on morphosyntax, which may have led to increased overall sensitivity to the

[8] Note that Alemán Bañón et al. (2017) showed that advanced English-speaking learners of Spanish showed native-like processing profiles even when tested on sentences with nouns that do not provide strong distributional cues related to gender.

morphosyntactic violations we examined. Previous research has established that the behavioral task associated with the EEG experiment may indeed influence ERP responses, even for native speakers (Osterhout & Mobley, 1995; Nieuwland, 2014). A recent study with L2 learners found that learners showed more native-like patterns of neural processing when a GJT was included (Lewis et al. 2016). In addition, Lemhöfer et al. (in press) show, based on a study of gender agreement in Dutch, that P600 is more likely to emerge in learners when there is a reason to doubt the reliability of the input in the experiment, either due to the presence of outright grammatical violations or due to the GJT task itself. In the present study, we included a GJT in order to make the experimental task comparable to the study we were building on (Osterhout et al., 2006), as well as many other L2 studies on agreement in the literature (Mickan & Lemhöfer, 2020; Morgan-Short et al., 2010; Tanner et al., 2013, 2014; E.J. White et al., 2012, see Caffarra et al. 2015 for a review). It is important to note that differences emerged between number and gender in our experiment, and thus, the inclusion of the GJT task could not have been solely responsible for the significant effects that emerged. Future studies which compare ERP results in light of differences in the behavioral task can continue to shed light on how behavioral tasks impact brain responses in L2 learners, in line with Lewis et al. (2016) and Lemhöfer et al. (in press). A study of this kind is currently underway in our lab.

With respect to our second research question, our results were similar to the results of previous studies which also observed a significant relationship between accuracy on the GJT and size of the ERP effects (e.g., Osterhout et al., 2006; Tanner et al., 2013, 2014; E.J. White et al., 2012). This result was significant across all conditions and shows that performance related to the specific property of agreement indeed modulates processing (e.g., Alemán-Bañón et al., 2018; Steinhauer et al., 2009). However, the GJT is an experiment-internal measure requiring explicit

metalinguistic judgments and thus, we included an oral production task in order to see whether sensitivity as measured by ERPs is also related to use of agreement morphology in a more natural context. Our results showed that better performance on the production of both subject-verb plural agreement and noun-adjective agreement overall was associated with significantly larger positivities for noun-adjective number in the ERP experiment. To our knowledge, this is the first L2 study to observe a relationship between ERP responses and oral production, which is interesting because it suggests that sensitivity to grammatical violations online is indeed related to the ability to use agreement even in more natural contexts that are focused on meaning. However, as we did not observe a significant relationship between production and ERP effects for subject-verb agreement (where *p* values were just above marginal) or gender agreement, the relationship is not as robust as the one observed for GJT accuracy. One factor that may be relevant is that the production task targeted a wider range of contexts of agreement than the ERP experiment, requiring use of both number and gender agreement in plural contexts, whereas in the ERP experiment, only singular contexts were tested. The accuracy rates for the production task (Table 4) show that the task was extremely challenging for learners, particularly with respect to gender. Thus, the ability to use correct gender morphology in production, particularly across a range of contexts, may lag behind the ability to detect grammatical anomalies. Previous studies of L2 gender agreement with more advanced learners have observed similar performance in comprehension and production tasks (e.g., Alemán-Bañón et al., 2017; McCarthy, 2008) and thus, a future study should examine whether the relationship between production and sensitivity indeed emerges for gender with more advanced learners when a higher level of performance can potentially be observed.

Our third research question examined the relationship between a broad range of individual difference measures and knowledge of morphosyntax, tested via three measures: a GJT, ERP, and a production task. The statistical models we used allow us to evaluate the predictive power of one individual difference measure against the others, instead of simply testing the value of a single measure in isolation. We first asked to what extent individual-level abilities predict sensitivity to grammatical violations (P600). Note that although overall sensitivity at the group level differed for number, where P600s were observed, and gender, where only a marginal effect for more advanced learners was observed, it is still interesting to examine the role of individual differences in the processing of both features. Although the effect for gender is marginal, it suggests that sensitivity may be present in some individuals but not others. Our examination of variability allows us to examine that point further. The results of the individual difference analyses showed that learners with better working memory showed significantly larger positivities for noun-adjective number violations, with a similar trend observed for noun-adjective gender violations (p = .071). The fact that one unique predictor emerged and that it was the same for number and gender (where the result was marginal) is theoretically significant, although the effect for gender is weak and needs to be confirmed in future studies. A pattern in which the measure that best predicted responses to gender agreement violations differed from number agreement violations would lend support to a proposal such as Bley-Vroman's (1990), which argues that properties which are unique to the L2 are learned in a qualitatively different manner (see also Hawkins, 2001). Although the effect for gender is only marginal, the pattern suggests that one predictor of online sensitivity for both number and gender is working memory, a cognitive ability that has been linked to successful sentence processing in both native speakers (see Daneman & Merikle, 1996) and L2 learners at a more advanced level

of proficiency (see Linck et al., 2014; Wen et al., 2017 for a review). Our design also included a measure of perceptual reasoning (Matrix Reasoning) as we wanted to include a non-verbal measure that would tap general analytic 'problem-solving' abilities, as it was skills in this domain that Bley-Vroman argued would be most predictive of general success. We also reasoned that perceptual reasoning may be important if learners were indeed engaging in a strategy of tracking the co-occurrence of certain linguistic forms (e.g., McLaughlin et al., 2010) but this measure was not predictive of performance in any condition. Our results instead suggest it is those learners who are best able to encode and maintain the relevant linguistic features in working memory that show greatest sensitivity to grammatical anomalies and we interpret our results as tentative support for the idea that this skill underlies processing for both shared and unique features. A reviewer rightly points out that working memory may play a role in the processing of both number and gender agreement because dependency formation in language processing generally requires that linguistic features be encoded and maintained. Thus, the fact that working memory may modulate the processing of both number and gender agreement may not be surprising, and also may not indicate more broadly that the individual differences which modulate both shared and unique features will always be similar. We agree that firm conclusions with respect to whether or not similar abilities underlie the processing of features that are shared or unique to the L2 cannot be made until this question is tested across a broad array of linguistic properties, languages, and individual difference measures. As the processing of gender agreement by native speakers whose L1 does not instantiate gender agreement has been highlighted in theoretical approaches as a case in which acquisition is qualitatively different (e.g. Hawkins, 2001), we believe our findings, although they are only suggestive, make an interesting contribution to this debate.

We also asked to what extent individual-level abilities predict performance across three different measures of morphosyntactic knowledge (P600 effect size, GJT accuracy, production), allowing us to address the question of whether individual abilities play the largest role when learners are directed to use metalinguistic knowledge, as proposed by Roberts (2012). These results were summarized in Table 8. The results revealed an interesting pattern in which working memory modulated both P600 effect size and performance on the oral production task (significant for number, marginal for gender for both measures) while verbal aptitude modulated GJT accuracy (D-prime) and performance on the oral production task (significant for both number and gender on both measures). Thus, the results show that there are individual differences which modulate performance on each task, but the specific abilities that modulate performance are dependent on the nature of the task itself. Working memory plays a role when linguistic features need to be encoded and maintained during online processing. As both ERPs and the oral production task involve the online deployment of knowledge of agreement, the ability to retrieve the relevant linguistic features from memory may be particularly relevant. The measure of verbal aptitude (MLAT-4) in contrast seems to modulate variability in tasks that require attention to linguistic form. The GJT is an explicit measure of grammatical sensitivity and although the oral production task focused on content (asking participants to describe the difference between two pictures), the task was scored with respect to morphological accuracy and thus, a focus on form is crucial in both tasks.

While a significant relationship between verbal aptitude and L2 performance has been interpreted by DeKeyser (DeKeyser, 2000; DeKeyser et al., 2010) as an indicator that late L2 acquisition proceeds via qualitatively different, explicit mechanisms, we believe this conclusion may be premature. Other studies, such as Abrahamsson and Hyltenstam (2008), have observed a

relationship between verbal aptitude and GJT performance in early learners. In addition, as DeKeyser et al. (2010) themselves point out, aptitude for learning an L2 may even be related to native language proficiency in ways that have still not been extensively examined (Hulstijn & Bossers, 1992; Skehan, 1986, 1990; Sparks et al., 2012a,b). It is possible that the abilities which capture variability in the L1 and L2 are qualitatively similar, but L2 studies have generally not examined individual learners in both the L1 and L2 (e.g., Bastarrika & Davidson, 2017; Hulstijn & Bossers, 1992; see discussion in Gabriele et al., 2017) and have often not focused on linguistic domains in which variability in native speakers is likely to be observed as well; this is a direction we are currently pursuing in a new NSF-funded project. Interestingly, in the present study, while the MLAT-4 was a significant predictor of performance on the GJT and the oral production task, an additional measure of verbal aptitude, the LLAMA_F was not. While the MLAT-4 assesses verbal aptitude in the learners' native language (analyzing the grammatical roles of words in English sentences), the LLAMA_F assesses the learning of a new linguistic system. It is interesting that it was metalinguistic abilities in the native language that were predictive of L2 performance, a finding which should be further explored in future research.

In summary, our study provides evidence that several factors facilitate processing in low-proficiency learners. First, learners are able to build on the linguistic properties of the L1 in order to process the L2 in a manner qualitatively similar to native speakers, even at early stages of development. In contrast to proposals that predict discrete stages at early levels of development, our results show, at least for some morphosyntactic dependencies, that qualitatively native-like processing is possible. While our results showed facilitation for shared features as opposed to features unique to the L2, we also provided preliminary support for the idea that the metalinguistic and cognitive abilities that underlie the processing of both shared and unique

features may be similar; this is an issue that should continue to be examined in future studies to see if the patterns we observed in our data are confirmed. In addition, by looking at knowledge of morphosyntax from three different perspectives (grammatical accuracy, ERP, oral production) and examining individual differences across those different measures, we have shown that in contrast to the proposal that L2 acquisition relies on 'problem-solving' (Bley-Vroman, 1990), or any one ability (Bond et al., 2011), our results suggest that the skills that underlie the successful L2 processing of agreement are myriad, involving both language-specific abilities such as verbal aptitude, and general cognitive abilities such as working memory.

In the present study, the most successful novice learners were those with increased metalinguistic abilities and increased cognitive abilities in the domain of working memory. In more advanced learners, it is possible that language aptitude may no longer play as strong a role as it is possible that it is only those late learners with high metalinguistic abilities that go on to reach advanced proficiency levels; thus, within those learners, a skill such as verbal aptitude may not show a wide enough range of variability to be predictive (see Abrahamsson & Hyltenstam, 2008). In contrast, we would expect working memory to maintain its role even in studies of more advanced learners, because the task of establishing agreement dependencies crucially relies on the encoding of linguistic features which must be maintained in memory during processing, and the processing of agreement has been found to modulated by working memory even in native speakers (e.g. Roll et al., 2013). The fact that working memory is a significant predictor for low-proficiency L2 learners is important because it suggests that at least some of the variability in the L2 processing of agreement is qualitatively similar to the variability observed in native speakers. Relating back to theories outlined in the introduction, our results show that neither an exclusive focus on the language or the learner is sufficient. Rather, we have shown that the linguistic

properties of the L2, the individual characteristics of the learner, and the nature of the task at hand all play an important role and current theories of L2 acquisition need to take a more comprehensive approach in order to better capture the variability present in the L2 acquisition and processing of agreement.

Acknowledgements

We would like to thank Allison Edmonds, Brooke Gunter, Carla Fernández Guzmán, Maite Martínez-García, Frank Plummer, and Emma Yaffe for their help with data collection. We would like to thank Steven Politzer-Ahles as well as the Associate Editor and reviewers who provided valuable feedback on this project for which we are grateful.

Funding

This work was supported by an award from the National Science Foundation [BCS #0951900] to Robert Fiorentino and Alison Gabriele.

References

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, *30*(4), 481-509. https://doi.org/10+10170S027226310808073X

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, *59*(2), 249-306. https://doi.org/ 10.1111/j.1467-9922.2009.00507.x

Alarcón, I. V. (2011). Spanish gender agreement under complete and incomplete acquisition: Early and late bilinguals' linguistic behavior within the noun phrase. *Bilingualism: Language and Cognition*, *14*(3), 332-350. https://doi.org/ 10.1017/S1366728910000222

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. https://doi.org/10.1016/j.brainres.2012.03.057

Alemán Bañón, J., Fiorentino, R., & Gabriele, A. (2014). Morphosyntactic processing in advanced second language learners: An event related potential investigation of the effects of L1-L2 similarity and structural distance. *Second Language Research, 30*(3), 275-306.

Alemán Bañón, J., Miller, D., & Rothman, J. (2017). Morphological variability in second language learners: An examination of electrophysiological and production data. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *43*(10), 1509-1536. https://doi.org/10.1037/xlm0000394

Alemán-Bañón, J., Fiorentino, R., & Gabriele, A. (2018). Using event-related potentials to track morphosyntactic development in second language learners: The processing of number and gender agreement in Spanish. *PLoS ONE*, *13*(7), 1-35. https://doi.org/10.1371/journal.pone.0200791

Angrilli, A., Penolazzi, B., Vespignani, F., De Vincenzi, M., Job, R., Ciccarelli, L., Palomba, D., & Stegagno, L. (2002). Cortical brain responses to semantic incongruity and syntactic violation in Italian language: An event-related potential study. *Neuroscience Letters*, *322*(1), 5-8. https://doi.org/10.1016/S0304-3940(01)02528-9

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience*, *17*(1), 137-153. https://doi.org/10.1162/0898929052880101

Bastarrika, A., & Davidson, D. (2017) An event related field study of rapid grammatical plasticity in adult second-language learners. *Frontiers in Human Neuroscience*, *11*(12), 1-19. https://doi.org/10.3389/fnhum.2017.00012

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic Analysis*, *20*(1-2), 3-49.

Bond, K., Gabriele, A., Fiorentino, R., & Alemán Bañón, J. (2011). Individual differences and the role of the L1 in L2 processing: An ERP investigation. In J. Herschensohn & D. Tanner (Eds.) *Proceedings of the 11th Generative Approaches to Second Language Acquisition* (pp. 17-29). Somerville, MA: Cascadilla Proceedings Project.

Boudewyn, M. (2015). Individual differences in language processing: Electrophysiological approaches. *Language and Linguistics Compass*, *9*(10), 406-419. https://doi.org/10.1111/lnc3.12167

Caffarra, S., Molinaro, N., Davidson, D., & Carreiras, M. (2015). Second language syntactic processing revealed through event-related potentials: An empirical review. *Neuroscience and Biobehavioral Reviews*, *51*, 31-47. https://doi.org/10.1016/j.neubiorev.2015.01.010

Carrasco-Ortis, H., Herrera, A., Jackson-Maldonado, D., Ramirez, G., Pereyra, J., & Wicha, N. (2017). The role of language similarity in processing second language morphosyntax: Evidence from ERPs. *International Journal of Psychophysiology, 117*, 91-110. https://doi.org/10.1016/j.ijpsycho.2017.04.008

Carroll, J. B., & Sapon, S. M. (1959). *Modern Language Aptitude Test.* San Antonio, TX: Psychological Corporation.

Chen, L., Shu, H., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject-verb agreement in L2 learning. *Bilingualism: Language and Cognition*, *10*(2), 161-174. https://doi.org/10.1017/S136672890700291X

Clahsen, H., & Felser, C. (2006). Continuity and shallow structures in language processing. *Applied Psycholinguistics, 27*(1), 107-126. https://doi.org/10.1017/S0142716406060206

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review*, *12*(5), 769-786. https://doi.org/10.3758/BF03196772

Conway, A. R., Jarrold, C., Kane, M. J., Miyake, A., & Towse, J. N. (2007). Variation in working memory: An introduction. In *Variation in Working Memory* (pp. 3-17). Oxford, UK: Oxford University Press.

Covey, L., Gabriele, A., & Fiorentino, R. (2018). Can learners use morphosyntactic cues to facilitate processing? Evidence from a study of gender agreement in Hindi. *Language Acquisition*, 25(3), 327-337. https://doi.org/10.1080/10489223.2017.1359272

Dallas, A., DeDe, G., & Nicol, J. (2013). An event-related potential (ERP) investigation of filler-gap processing in native and second language speakers. *Language Learning*, *63*(4), 766-799. https://doi.org/10.1111/lang.12026

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin and Review*, *3*(4), 422-433. https://doi.org/10.3758/BF03214546

Davidson, D. J. (2010). Short-Term Grammatical Plasticity in Adult Language Learners. *Language Learning*, *60*(s2), 109–122. https://doi.org/10.1111/j.1467-9922.2010.00603.x

Davidson, D. J., & Indefrey, P. (2009a). An Event-related Potential Study on Changes of Violation and Error Responses during Morphosyntactic Learning. *Journal of Cognitive Neuroscience, 21*(3), 433–446. https://doi.org/10.1162/jocn.2008.21031

Davidson, D. J., & Indefrey, P. (2009b). Plasticity of grammatical recursion in German learners of Dutch. *Language and Cognitive Processes: Cognitive Neuroscience of Language*, *24*(9), 1335–1369. https://doi.org/10.1080/01690960902981883

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9-21. https://doi.org/10.1016/j.jneumeth.2003.10.009

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*(4), 499-533.

DeKeyser, R. (2007). Skill acquisition theory. In B. Van Patten and J. Williams (Eds.) *Theories in Second Language Acquisition* (pp. 97-114). Mahwah, NJ: Erlbaum.

DeKeyser, R., Alfi-Shabtay, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics*, *31*(3), 413-438. https://doi.org/10.1017/S0142716410000056

Dewaele, J. M., & Véronique, D. (2001). Gender assignment and gender agreement in advanced French interlanguage: A cross-sectional study. *Bilingualism: Language and Cognition*, *4*(3), 275-297. https://doi.org/10.1017/S136672890100044X

Díaz, B., Erdocia, K., de Menezes, R. F., Mueller, J. L., Sebastián-Gallés, N., & Laka, I. (2016). Electrophysiological correlates of second-language syntactic processes are related to native and second language distance regardless of age of acquisition. *Frontiers in Psychology*, *7*(133). https://doi.org/10.3389/fpsyg.2016.00133

Faretta-Stutenberg, M., & Morgan-Short, K. (2017). The interplay of individual differences and context of learning in behavioral and neurocognitive second language development. *Second Language Research*, *34*(1), 76-101. https://doi.org/10.1177/0267658316684903

Felser, C., Clahsen, H., & Münte, T. F. (2003). Storage and integration in the processing of filler-gap dependencies: An ERP study of topicalization and wh-movement in German. *Brain and Language*, *87*(3), 345-354. https://doi.org/10.1016/S0093-934X(03)00135-4

Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Bilingualism: Language and Cognition*, *14*(3), 379-399. https://doi.org/10.1017/S136672891000012X

Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language*, *66*(1), 226-248. https://doi.org/10.1016/j.jml.2011.07.007

Franceschina, F. (2001). Morphological or syntactic deficits in near-native speakers? An assessment of some current proposals. *Second Language Research*, *17*(3), 213-247.

Franceschina, F. (2005). *Fossilized Second Language Grammars: The Acquisition of Grammatical Gender*. Amsterdam: John Benjamins.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78-84. https://doi.org/10.1016/S1364-6613(00)01839-8

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1219-1248. https://doi.org/10.1037/0278-7393.22.5.1219

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological, and syntactic violations. *Cognitive Brain Research*, *1*(3), 183-192. https://doi.org/10.1016/0926-6410(93)90026-2

Friederici, A. D., & Weissenborn, J. (2007). Mapping sentence form onto meaning: The syntax-semantic interface. *Brain Research*, *1146*, 50-58. https://doi.org/10.1016/j.brainres.2006.08.038

Frisch, S., Schlesewsky, M., Saddy, D., & Alpermann, A. (2002). The P600 as an indicator of syntactic ambiguity. *Cognition*, *85*(3), B83-B92. https://doi.org/10.1016/S0010-0277(02)00126-9

Gabriele, A. Fiorentino, R., & Alemán-Bañón, J. (2013). Examining second language development using event-related potentials: a cross-sectional study on the processing of gender and number agreement. *Linguistic Approaches to Bilingualism, 3*(2), 213-232. https://doi.org/10.1075/lab.3.2.04gab

Gabriele, A., Fiorentino, R., & Covey, L. (2017). Understanding the symptoms and sources of variability in second language processing. *Bilingualism: Language and Cognition*, *20*(4), 685-686. https://doi.org/10.1017/S1366728916000961

Gillon-Dowens, M., Vergara, M., Barber, H., & Carreiras, M. (2010). Morphosyntactic processing in late second language learners. *Journal of Cognitive Neuroscience*, *22*(8), 1870–1887. https://doi.org/10.1162/jocn.2009.21304

Gillon-Dowens, M., Guo, T., Guo, J., Barber, H., & Carreiras, M. (2011). Gender and number processing in Chinese learners of Spanish – Evidence from event related potentials. *Neuropsychologia*, *49*(7), 1651-1659. https://doi.org/10.1016/j.neuropsychologia.2011.02.034

Granena, G., Jackson, D. O., & Yilmaz, Y. (2016). *Cognitive Individual Differences in Second Language Processing and Acquisition*. John Benjamins.

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, *28*(2), 191-215.

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*(4), 556-568. https://doi.org/10.1162/089892900562336

Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology*, *34*(6), 660-676. https://doi.org/10.1111/j.1469-8986.1997.tb02142.x

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*(4), 439-484. https://doi.org/10.1080/01690969308407585

Hammer, A., Jansma, B. M., Lamers, M., & Münte, T. F. (2008). Interplay of meaning, syntax and working memory during pronoun resolution investigated by ERPs. *Brain Research*, *1230*, 177-191. https://doi.org/10.1016/j.brainres.2008.07.004

Hawkins, R. (2001). The theoretical significance of Universal Grammar in SLA. *Second Language Research*, *17*(4), 345-367.

Hawkins, R., & Chan, C. Y. (1997). The partial availability of Universal Grammar in second language acquisition: The 'failed functional features hypothesis'. *Second Language Research*, *13*(3), 187-226. https://doi.org/10.1191/026765897671476153

Hawkins, R., & Franceschina, F. (2004) Explaining the acquisition and non-acquisition of determiner-noun gender control in French and Spanish. In P. Prevost & J. Paradis (Eds.), *The acquisition of French in different contexts: Focus on functional categories* (pp. 175-205). Amsterdam: Benjamins.

Hoffman, L. (2015). *Longitudinal Analysis: Modeling Within-Person Fluctuation and Change*. Routledge Taylor & Francis: New York, NY.

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33-56.

Hulstijn, J., & Bossers, B. (1992). Individual differences in L2 proficiency as a function of L1 proficiency. *European Journal of Cognitive Psychology*, *4*(4), 341-353. https://doi.org/10.1080/09541449208406192

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, *44*(2), 137-166. https://doi.org/10.1017/S0261444810000509

Kaan, E. (2007). Event-related potentials and language processing: A brief overview. *Language and Linguistics Compass*, *1*(6), 571-591. https://doi.org/10.1111/j.1749-818X.2007.00037.x

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*(2), 159-201. https://doi.org/10.1080/016909600386084

Kim, A., & Osterhout, L. (2005) The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52*(2), 205-225. https://doi.org/10.1016/j.jml.2004.10.002

Kolk, H. H., Chwilla, D. J., van Herten, M., & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language, 85*(1), 1-36. https://doi.org/10.1016/S0093-934X(02)00548-5

Kutas, M. & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science*, *4*(12), 463-470. https://doi.org/10.1016/S1364-6613(00)01560-6

Kutas, M. & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(1), 621-647. https://doi.org/10.1146/annurev.psych.093008.131123

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: [De]constructing the N400. *National Review of Neuroscience, 9*(12), 920-933. https://doi.org/10.1038/nrn2532

Lau, E., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain and Language*, *98*(1), 74-88. https://doi.org/10.1016/j.bandl.2006.02.003

Lemhöfer, K., Schriefers, H., & Indefrey, P. (2014). Idiosyncratic grammars: Syntactic processing in second language comprehension uses subjective feature representations. *Journal of Cognitive Neuroscience*, *26*(7), 1428-1444. https://doi.org/10.1162/jocn_a_00609

Lemhöfer, K., Schriefers, H., & Indefrey, P. (in press). Syntactic Processing in L2 Depends on Perceived Reliability of the Input: Evidence from P600 Responses to Correct Input. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* https://doi.org/10.1037/xlm0000895

Lewis, A. G., Lemhöfer, K., Schoffelen, J.-M., & Schriefers, H. (2016). Gender agreement violations modulate beta oscillatory dynamics during sentence comprehension: A comparison of second language learners and native speakers. *Neuropsychologia, 89*, 254–272. https://doi.org/10.1016/j.neuropsychologia.2016.06.031

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin and Review*, *21*(4), 861-883. https://doi.org/10.3758/s13423-013-0565-2

López Prego, B., & Gabriele, A. (2014). Examining the impact of task demands on morphological variability in native and non-native Spanish. *Linguistic Approaches to Bilingualism, 4*(2), 192–221. http://dx.doi.org/10 .1075/lab.4.2.03lop

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: MIT Press.

MacWhinney, B. (2005). A unified model of language acquisition. In J.F. Kroll, A.M.B. De Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 49-67). Oxford: Oxford University Press.

McCarthy, C. (2008). Morphological variability in the comprehension of agreement: An argument for representation over computation. *Second Language Research*, *24*(4), 459-486.

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*(s2), 123-150. https://doi.org/10.1111/j.1467-9922.2010.00604.x

Meara, P. (2005a). *LLAMA Language Aptitude Tests*. Swansea, UK: Lognostics.

Meara, P. (2005b). *LLAMA Language Aptitude Tests: The Manual.* Swansea, UK: Lognostics.

Meulman, N., Stowe, L. A., Sprenger, S. A., Bresser, M. & Schmid, M. S. (2014). An ERP study on L2 syntax processing: When do learners fail? *Frontiers in Psychology*, *5*:1072. https://doi.org/10.3389/fpsyg.2014.01072

Mickan, A., & Lemhöfer, K. (2020). Tracking syntactic conflict between languages over the course of L2 acquisition: A cross-sectional ERP study. *Journal of Cognitive Neuroscience, 32*(5), 822-846. https://doi.org/10.1162/jocn_a_01528

Molinaro, N., Barber, H. A., Caffarra, S., & Carreiras, M. (2015). On the left anterior negativity (LAN): The case of morphosyntactic agreement. *Cortex*, *66*, 156-159. https://doi.org/10.1016/j.cortex.2014.06.009

Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, *47*(8), 908-930. https://doi.org/10.1016/j.cortex.2011.02.019

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*(3), 503-553. https://doi.org/10.1111/j.1467-9922.2008.00449.x

Morgan-Short, K., Sanz, C., Steinhauer. K, & Ullman, M. T. (2010) Second language learning of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning*, *60*(1), 154-193. https://doi.org/10.1111/j.1467-9922.2009.00554.x

Nieuwland, M. (2014). "Who's he?" Event related potentials and unbound pronouns. *Journal of Memory and Language*, *76*(1), 1-28. https://doi.org/10.1016/j.jml.2014.06.002

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience*, *17*(8), 1212-1228. https://doi.org/10.1162/0898929055002436

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785-806. https://doi.org/10.1016/0749-596X(92)90039-Z

Osterhout, L., & Mobley, L.A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language, 34*(6), 739-773. https://doi.org/10.1006/jmla.1995.1033

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., *The On-line Study of Sentence*

Comprehension: Eyetracking, ERPs, and Beyond (pp. 271-308). New York, Hove: 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. https://doi.org/10.1111/j.1467-9922.2006.00361.x

Osterhout, L., Poliakov, A., Inoue, K., McLaughlin, J., Valentine, G., Pitkänen, I., Frenck-Mestre, C., & Hirschensohn, J. (2008). Second-language learning and changes in the brain. *Journal of Neurolinguistics, 21*(6) 509–521. https://doi.org/10.1016/j.jneuroling.2008.01.001

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, *22*(3), 407-428. https://doi.org/10.1016/j.cogbrainres.2004.09.012

Raven, J. C. (1965). *The Colored Progressive Matrices Test*. London: Lewis.

Reichle, R. V., Tremblay, A., & Coughlin, C. (2016). Working memory capacity in L2 processing. *Probus*, *28*(1), 29-55. https://doi.org/10.1515/probus-2016-0003

Roberts, L. (2012). Individual differences in second language sentence processing. *Language Learning*, *62*(s2), 172-188. https://doi.org/10.1111/j.1467-9922.2012.00711.x

Robinson, P. (2001). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research*, *17*(4), 368-392.

Roll M., Gosselke S., Lindgren M., & Horne M. (2013). Time-driven effects on processing grammatical agreement. *Frontiers in Psychology, 30*, 1004. https://doi.org/10.3389/fpsyg.2013.01004

Sabourin, L. & Stowe, L.A. (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research, 24*(3), 397-430. https://doi.org/10.11772/0267658308090186

Schwartz, B. D., & Sprouse, R. A. (1994). Word order and nominative case in nonnative language acquisition: A longitudinal study of (L1 Turkish) German interlanguage. *Language Acquisition Studies in Generative Grammar*, *31*(4), 71-89.

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, *12*(1), 40-72. https://doi.org/10.1177/026765839601200103

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

Skehan, P. (1990). The relationship between native and foreign language learning ability: Educational and linguistic factors. *Current Trends in European Second Language Acquisition Research*, 83-106.

Sparks, R. L., Patton, J., Ganschow, L., & Humbach, N. (2012a). Relationships among L1 print exposure and early L1 literacy skills, L2 aptitude, and L2 proficiency. *Reading and Writing: An Interdisciplinary Journal*, *25*(7), 1599-1634. https://doi.org/ 10.1007/s11145-011-9335-6

Sparks, R. L., Patton, J., Ganschow, L., & Humbach, N. (2012b). Do L1 reading achievement and L1 print exposure contribute to the prediction of L2 proficiency? *Language Learning*, *62*(2), 473-505. https://doi.org/10.1111/j.1467-9922.2012.00694.x

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*(1), 13-41.

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

Tanner, D., Grey, S. E., & Van Hell, J. G. (2017). Dissociating retrieval interference and reanalysis in the P600 during sentence comprehension. *Psychophysiology, 54(2), 248-259*. https://doi.org/10.1111/psyp.12788

Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition*, *17*(2), 277-293. https://doi.org/10.1017/S1366728913000370

Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition*, *16*(2), 367-382. https://doi.org/10.1017/S1366728912000302

Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia*, *56*, 289-301. https://doi.org/10.1016/j.neuropsychologia.2014.02.002

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*(2), 173-204. https://doi.org/10.1017/S0272263105050102

Tsimpli, I., & Dimitrakopoulou, M. (2007). The interpretability hypothesis: Evidence from wh-interrogatives in second language acquisition. *Second Language Research*, *23*(2), 215-242.

Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience, 2*(10), 717-726. https://doi.org/10.1038/35094573

Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In: C. Sanz C (Eds.) *Mind and context in adult second language acquisition: Methods, theory, and practice* (pp. 141-178). Washington, D.C.: Georgetown University Press.

Ullman, M.T. (2015). The declarative/procedural model: A neurobiologically motivated theory of first and second language. In: B. VanPatten and J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed.) (pp. 135-158). New York: Routledge.

Van Dyke, J. A., Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition*, *131*(3), 373-403. https://doi.org/10.1016/j.cognition.2014.01.007

Van Hell, J. G., & Tokowicz, N. (2010). Event-related brain potentials and second language learning: Syntactic processing in late L2 learners at different L2 proficiency levels. *Second Language Research*, *26*(1), 43-74.

VanRullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Frontiers in Psychology*, *2*(365). https://doi.org/ 10.3389/fpsyg.2011.00365

Weber, K., & Lavric, A. (2008). Syntactic anomaly elicits a lexico-semantic (N400) ERP effect in the second language but not the first. *Psychophysiology*, *45*(6), 920-925. https://doi.org/ 10.1111/j.1469-8986.2008.00691.x

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV)*. San Antonio, TX: Pearson.

Wen, Z. (2014). Theorizing and measuring working memory in first and second language research. *Language Teaching*, *47*(2), 174-190. https://doi.org/10.1017/S0261444813000517

Wen, Z., Biedrón, A., & Skehan, P. (2017). Foreign language aptitude theory: Yesterday, today, and tomorrow. *Language Teaching*, *50*(1), 1-31. https://doi.org/10.1017/S0261444816000276

White, E. J., Genesee, F., & Steinhauer, K. (2012). Brain responses before and after intensive second language learning: Proficiency based changes and first language background effects in adult learners. *PLoS ONE*, *7*(12), e52318. https://doi.org/10.1371/journal.pone.0052318

White, L., Valenzuela., E., Koslowska-Macgregor., M., & Leung, Y. K. I. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics*, *25*(1), 105-133. https://doi.org/10.1017/S0142716404001067

Appendix: Target stimuli

Noun-Adjective Agreement

(Items in this condition were modified from Alemán Bañón et al., 2012)

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.

Subject-Verb Agreement

1. El mecánico astuto trabaja en el garaje.
2. El mecánico astuto *trabajan en el garaje.

3. La embajadora exiliada trabaja en Nueva York.
4. La embajadora exiliada *trabajan en Nueva York.

5. El sociólogo ocupado escribe en la oficina.
6. El sociólogo ocupado *escriben en la oficina.

7. La novelista prolífica escribe en la cabaña.
8. La novelista prolífica *escriben en la cabaña.

9. El toro bravo corre en el campo.
10. El toro bravo *corren en el campo.

11. La competidora sueca corre en la pista.
12. La competidora sueca *corren en la pista.

13. El cirujano rico almuerza en el restaurante.
14. El cirujano rico *almuerzan en el restaurante.

15. La empleada bancaria almuerza en la cafetería.
16. La empleada bancaria *almuerzan en la cafetería.

17. El niño creativo baila en la escuela.
18. El niño creativo *bailan en la escuela.

19. La coreógrafa francesa baila en Los Ángeles.
20. La coreógrafa francesa *bailan en Los Ángeles.

21. La solista asignada canta en el escenario.
22. La solista asignada *cantan en el escenario.

23. La monja bondadosa canta en el coro.
24. La monja bondadosa *cantan en el coro.

25. El perro descuidado duerme en el garaje.
26. El perro descuidado *duermen en el garaje.

27. La psicóloga cansada duerme en el sofá.
28. La psicóloga cansada *duermen en el sofa.

29. El cocodrilo americano nada en la laguna.
30. El cocodrilo americano *nadan en la laguna.

31. La ballena blanca nada en el mar.
32. La ballena blanca *nadan en el mar.

33. El loro colorado vive en la selva.
34. El loro colorado *viven en la selva.

35. La doncella secuestrada vive en la torre.
36. La doncella secuestrada *viven en la torre.

37. El ajo sabroso crece en el jardín.
38. El ajo sabroso *crecen en el jardín.

39. La planta nutritiva crece en la selva.
40. La planta nutritiva *crecen en la selva.

41. El anillo lustroso brilla en el estante.
42. El anillo lustroso *brillan en el estante.

43. La joya egipcia brilla en el museo.
44. La joya egipcia *brillan en el museo.

45. El leopardo africano caza en la sabana.
46. El leopardo africano *cazan en la sabana.

47. La pantera negra caza en el valle.
48. La pantera negra *cazan en el valle.

49. El académico venezolano estudia en la universidad.
50. El académico venezolano *estudian en la universidad.

51. La investigadora meticulosa estudia en el archivo.
52. La investigadora meticulosa *estudian en el archivo.

53. El neurólogo pediátrico enseña en la universidad.
54. El neurólogo pediátrico *enseñan en la universidad.

55. La profesora simpática enseña en el instituto.
56. La profesora simpática *enseñan en el instituto.

57. El panadero perezoso lee en la cafetería.
58. El panadero perezoso *leen en la cafetería.

59. La bibliotecaria desocupada lee en la sala.
60. La bibliotecaria desocupada *leen en la sala.

61. El caballero heróico lucha en el castillo.
62. El caballero heróico *luchan en el castillo.

63. La armada británica lucha en el Pacífico.
64. La armada británica *luchan en el Pacífico.

65. El obispo católico medita en la capilla.
66. El obispo católico *meditan en la capilla.

67. La abuela piadosa medita en la catedral.
68. La abuela piadosa *meditan en la catedral.

69. El muchacho travieso grita en la calle.
70. El muchacho travieso *gritan en la calle.

71. La maestra enojada grita en el pasillo.
72. La maestra enojada *gritan en el pasillo.

73. El soldado herido descansa en el hospital.
74. El soldado herido *descansan en el hospital.

75. La turista agotada descansa en el hotel.
76. La turista agotada *descansan en el hotel.

77. El genio artístico pinta en la galería.
78. El genio artístico *pintan en la galería.

79. La retratista boliviana pinta en el estudio.
80. La retratista boliviana *pintan en el estudio.

81. El alumno aburrido dibuja en su cuaderno.
82. El alumno aburrido *dibujan en su cuaderno.

83. La artista talentosa dibuja en el parque.
84. La artista talentosa *dibujan en el parque.

85. El equipo colombiano juega en el estadio.
86. El equipo colombiano *juegan en el estadio.

87. La tenista japonesa juega en Nueva York.
88. La tenista japonesa *juegan en Nueva York.

89. El músico nervioso fuma en el camerino.
90. El músico nervioso *fuman en el camerino.

91. La administradora estresada fuma en su oficina.
92. La administradora estresada *fuman en su oficina.

93. El novio abandonado llora en su habitación.
94. El novio abandonado *lloran en su habitación.

95. La viuda deprimida llora en el cementerio.
96. La viuda deprimida *lloran en el cementerio.

97. El aficionado expulsado protesta en la entrada.
98. El aficionado expulsado *protestan en la entrada.

99. La traidora sentenciada protesta en la corte.
100. La traidora sentenciada *protestan en la corte.

101. El aventurero chileno esquía en los Andes.
102. El aventurero chileno *esquían en los Andes.

103. La deportista suiza esquía en los Alpes.
104. La deportista suiza *esquían en los Alpes.

105. El abogado australiano practica en Buenos Aires.
106. El abogado australiano *practican en Buenos Aires.

107. La oradora inquieta practica en el auditorio.
108. La oradora inquieta *practican en el auditorio.

109. El auto nuevo ruge en la pista.
110. El auto nuevo *rugen en la pista.

111. La turbina anticuada ruge en el hangar.
112. La turbina anticuada *rugen en el hangar.

113. El farmacéutico argentino reside en Buenos Aires.
114. El farmacéutico argentino *residen en Buenos Aires.

115. La directora excéntrica reside en Los Ángeles.
116. La directora excéntrica *residen en Los Ángeles.

117. El asesino contratado espera en el vestíbulo.
118. El asesino contratado *esperan en el vestíbulo.

119. La prisionera condenada espera en su celda.
120. La prisionera condenada *esperan en su celda.

121. El barco vacío flota en la laguna.
122. El barco vacío *flotan en la laguna.

123. La balsa perdida flota en el río.
124. La balsa perdida *flotan en el río.

125. El paso peligroso comienza en el cañón.
126. El paso peligroso *comienzan en el cañón.

127. La carretera montañosa comienza en los Pirineos.
128. La carretera montañosa *comienzan en los Pirineos.

129. El sendero romántico termina en el jardín.
130. El sendero romántico *terminan en el jardín.

131. La ruta arriesgada termina en la sierra.
132. La ruta arriesgada *terminan en la sierra.

133. El teléfono ruidoso suena en el vestíbulo.
134. El teléfono ruidoso *suenan en el vestíbulo.

135. La campana hermosa suena en el campanario.
136. La campana hermosa *suenan en el campanario.

137. El espectro espantoso aparece en la torre.
138. El espectro espantoso *aparecen en la torre.

139. La tabla informativa aparece en el apéndice.
140. La tabla informativa *aparecen en el apéndice.

141. El ogro adormilado ronca en la cueva.
142. El ogro adormilado *roncan en la cueva.

143. La oficinista dormida ronca en su asiento.
144. La oficinista dormida *roncan en su asiento.

145. La arqueóloga fanática acampa en el Sáhara.
146. La arqueóloga fanática *acampan en el Sáhara.

147. El montañero peruano acampa en los Andes.
148. El montañero peruano *acampan en los Andes.

149. El rabino sincero ora en la sinagoga.
150. El rabino sincero *oran en la sinagoga.

151. La pecadora arrepentida ora en la catedral.
152. La pecadora arrepentida *oran en la catedral.

153. El científico marino bucea en las Bahamas.
154. El científico marino *bucean en las Bahamas.

155. La bióloga mexicana bucea en el Caribe.
156. La bióloga mexicana *bucean en el Caribe.

157. El oso hambriento pesca en el río.
158. El oso hambriento *pescan en el río.

159. La financiera retirada pesca en las Bahamas.
160. La financiera retirada *pescan en las Bahamas.

161. El extranjero testarudo regatea en el mercado.
162. El extranjero testarudo *regatean en el mercado.

163. La vendedora sabia regatea en el quiosco.
164. La vendedora sabia *regatean en el quiosco.

165. El cocinero renombrado cocina en el restaurante.
166. El cocinero renombrado *cocinan en el restaurante.

167. La criada italiana cocina en el hotel.
168. La criada italiana *cocinan en el hotel.

169. El ejecutivo obeso desayuna en su despacho.
170. El ejecutivo obeso *desayunan en su despacho.

171. La conductora apresurada desayuna en su vehículo.
172. La conductora apresurada *desayunan en su vehículo.

173. El canguro capturado boxea en su jaula.
174. El canguro capturado *boxean en su jaula.

175. La campeona cubana boxea en el gimnasio.
176. La campeona cubana *boxean en el gimnasio.

177. El caballo enfermo convalece en el establo.
178. El caballo enfermo *convalecen en el establo.

179. La víctima asustada convalece en el hospital.
180. La víctima asustada *convalecen en el hospital.

181. El enfermero infectado estornuda en la clínica.
182. El enfermero infectado *estornudan en la clínica.

183. La señora delgada estornuda en el aeropuerto.
184. La señora delgada *estornudan en el aeropuerto.

185. El político protegido desembarca en el aeropuerto.
186. El político protegido *desembarcan en el aeropuerto.

187. La pasajera brasileña desembarca en San Francisco.
188. La pasajera brasileña *desembarcan en San Francisco.

189. El monstruo acuático reaparece en la bahía.
190. El monstruo acuático *reaparecen en la bahía.

191. La enemiga vengativa reaparece en la secuela.
192. La enemiga vengativa *reaparecen en la secuela.

193. El tirano malvado reina en el archipiélago.
194. El tirano malvado *reinan en el archipiélago.

195. La soberana tiránica reina en el castillo.
196. La soberana tiránica *reinan en el castillo.

197. El destacamento especializado ataca en el este.
198. El destacamento especializado *atacan en el este.

199. La brigada rusa ataca en el norte.
200. La brigada rusa *atacan en el norte.

201. El anciano respetado habla en el coloquio.
202. El anciano respetado *hablan en el coloquio.

203. La arquitecta exitosa habla en el auditorio.
204. La arquitecta exitosa *hablan en el auditorio.

205. El cautivo torturado sufre en su celda.
206. El cautivo torturado *sufren en su celda.

207. La huérfana traumatizada sufre en el orfanato.
208. La huérfana traumatizada *sufren en el orfanato.

209. El juego entero cabe en la caja.
210. El juego entero *caben en la caja.

211. La bolsa llena cabe en el baúl.
212. La bolsa llena *caben en el baúl.

213. El geranio híbrido florece en el patio.
214. El geranio híbrido *florecen en el patio.

215. La orquídea púrpura florece en la ribera.
216. La orquídea púrpura *florecen en la ribera.

217. El candidato próspero triunfa en la elección.
218. El candidato próspero *triunfan en la elección.

219. La modista imaginativa triunfa en la pasarela.
220. La modista imaginativa *triunfan en la pasarela.

221. El funcionario designado participa en la investigación.
222. El funcionario designado *participan en la investigación.

223. La secretaria discreta participa en la transacción.
224. La secretaria discreta *participan en la transacción.

225. El cartero maltratado persevera en el trabajo.
226. El cartero maltratado *perseveran en el trabajo.

227. La carpintera trabajadora persevera en el taller.
228. La carpintera trabajadora *perseveran en el taller.

229. El fotógrafo experto colabora en el proyecto.
230. El fotógrafo experto *colaboran en el proyecto.

231. La reportera inquisitiva colabora en la investigación.
232. La reportera inquisitiva *colaboran en la investigación.

233. El paramédico musculoso compite en el torneo.
234. El paramédico musculoso *compiten en el torneo.

235. La poetisa española compite en el concurso.
236. La poetisa española *compiten en el concurso.

237. El árbitro frustrado intercede en la pelea.
238. El árbitro frustrado *interceden en la pelea.

239. La consejera respetuosa intercede en el conflicto.
240. La consejera respetuosa *interceden en el conflicto.

Table 1

Descriptive Statistics for ERP Amplitudes by Condition, Learner Group, and Session

Amplitudes by Condition	Session	100-Level				200-Level			
		Mean	SD	Min	Max	Mean	SD	Min	Max
Subject-Verb Grammatical: No Violation	2 months	−0.02	1.73	−4.22	3.29	−0.65	2.19	−3.01	6.56
	6 months	−1.60	1.36	−4.95	0.97	−0.61	1.77	−3.48	2.75
	8 months	0.12	1.71	−3.56	3.47				
	Overall	−0.50	1.77	−4.95	3.47	−0.63	1.97	−3.48	6.56
Subject-Verb Ungrammatical: Number Violation	2 months	0.33	2.06	−3.81	5.74	0.88	3.21	−4.21	7.80
	6 months	−0.61	1.97	−3.30	3.34	2.00	3.47	−3.06	11.85
	8 months	0.78	1.72	−2.09	4.78				
	Overall	0.17	1.98	−3.81	5.74	1.44	3.35	−4.21	11.85
Noun-Adjective Grammatical: No Violation	2 months	2.07	2.00	−2.58	5.16	2.55	2.52	−2.81	7.01
	6 months	1.55	2.14	−1.75	6.35	2.59	2.17	−0.75	6.88
	8 months	2.19	2.10	−1.98	5.88				
	Overall	1.94	2.07	−2.58	6.35	2.57	2.32	−2.81	7.01
Noun-Adjective Ungrammatical: Number Violation	2 months	2.50	2.39	−1.34	9.03	4.69	3.19	0.09	13.41
	6 months	2.20	2.53	−3.23	9.01	4.65	3.34	−3.69	11.69
	8 months	2.76	2.80	−1.29	8.92				
	Overall	2.49	2.55	−3.23	9.03	4.67	3.23	−3.69	13.41
Noun-Adjective Ungrammatical: Gender Violation	2 months	1.80	1.77	−1.72	5.90	2.78	2.47	−3.94	7.26
	6 months	1.64	1.90	−1.09	6.14	3.46	3.20	−0.88	13.60
	8 months	2.49	2.19	−1.19	5.96				
	Overall	1.97	1.97	−1.72	6.14	3.12	2.85	−3.94	13.60
Violation Effect Size: Subject-Verb Number	2 months	0.35	2.12	−3.29	6.33	1.53	2.36	−2.24	6.42
	6 months	0.99	1.60	−1.64	4.99	2.61	3.04	−1.45	10.00
	8 months	0.66	1.51	−3.68	3.44				
	Overall	0.67	1.76	−3.68	6.33	2.07	2.75	−2.24	10.00
Violation Effect Size: Noun-Adjective Number	2 months	0.44	1.47	−2.24	3.87	2.15	2.62	−2.65	7.91
	6 months	0.65	1.88	−1.77	5.25	2.06	3.02	−6.42	7.40
	8 months	0.57	1.80	−3.57	3.56				
	Overall	0.55	1.70	−3.57	5.25	2.10	2.79	−6.42	7.91
Violation Effect Size: Noun-Adjective Gender	2 months	−0.27	1.50	−3.56	2.70	0.23	1.66	−2.85	3.47
	6 months	0.09	1.60	−4.29	3.22	0.87	2.51	−2.53	7.23
	8 months	0.29	1.88	−5.29	3.34				
	Overall	0.04	1.66	−5.29	3.34	0.55	2.13	−2.85	7.23

Note: SD = Standard Deviation, Min = Minimum, Max = Maximum.

Table 2

Simplified Means Model Results for Amplitude Grammaticality Effects by Condition and Learner Group

Fixed Effects by Outcome	Est	SE	Lower CI	Upper CI	Cohen's *d*	*p* <
Subject-Verb (SV) Number Violation Difference						
100-Class Marginal Mean	0.67	0.21	0.25	1.09	0.77	**.002**
200-Class Marginal Mean	2.07	0.44	1.16	2.97	1.68	**.001**
Class Mean Difference	1.40	0.49	0.41	2.39	0.85	**.007**
Noun-Adjective (NA) Number Violation Difference						
100-Class Marginal Mean	0.55	0.21	0.14	0.96	0.65	**.009**
200-Class Marginal Mean	2.10	0.41	1.27	2.93	1.52	**.001**
Class Mean Difference	1.55	0.46	0.63	2.47	0.82	**.001**
Noun-Adjective (NA) Gender Violation Difference						
100-Class Marginal Mean	0.04	0.26	−0.51	0.58	0.06	.888
200-Class Marginal Mean	0.55	0.31	−0.08	1.18	0.52	.086
Class Mean Difference	0.51	0.41	−0.30	1.33	0.31	.214
Number Violation Difference: SV vs. NA						
100-Class Marginal Mean Difference	−0.12	0.28	−0.67	0.44	0.10	.676
200-Class Marginal Mean Difference	0.03	0.52	−1.02	1.09	0.02	.948
Class Mean Difference	0.15	0.59	−1.03	1.33	0.07	.800
Noun-Adjective Violations: Number vs. Gender						
100-Class Marginal Mean Difference	−0.51	0.31	−1.14	0.11	0.57	.103
200-Class Marginal Mean Difference	−1.55	0.43	−2.42	−0.68	1.07	**.001**
Class Mean Difference	−1.04	0.53	−2.09	0.02	0.45	.055

Note: EST = Estimate, SE = Standard Error, CI = 95% Confidence Interval. Bold values indicate effects with *p*'s < .05.

Table 3

Acceptance Rates and D-Prime Descriptive Statistics by Condition, Learner Group, and Session

Acceptance Rates by Condition	Session	100-Level				200-Level			
		Mean	SD	Min	Max	Mean	SD	Min	Max
Subject-Verb Grammatical: No Violation	2 months	0.65	0.18	0.40	1.00	0.84	0.17	0.45	1.00
	6 months	0.70	0.20	0.38	1.00	0.85	0.17	0.45	1.00
	8 months	0.70	0.20	0.38	1.00				
	Overall	0.68	0.19	0.38	1.00	0.84	0.17	0.45	1.00
Subject-Verb Ungrammatical: Number Violation	2 months	0.58	0.17	0.13	0.88	0.25	0.26	0.00	0.83
	6 months	0.45	0.27	0.00	0.93	0.22	0.27	0.00	0.83
	8 months	0.41	0.25	0.00	0.80				
	Overall	0.48	0.24	0.00	0.93	0.24	0.26	0.00	0.83
Noun-Adjective Grammatical: No Violation	2 months	0.68	0.18	0.33	0.98	0.87	0.16	0.45	1.00
	6 months	0.69	0.23	0.10	1.00	0.88	0.16	0.45	1.00
	8 months	0.74	0.23	0.18	1.00				
	Overall	0.70	0.21	0.10	1.00	0.87	0.16	0.45	1.00
Noun-Adjective Ungrammatical: Number Violation	2 months	0.57	0.20	0.05	0.95	0.22	0.24	0.00	0.78
	6 months	0.49	0.25	0.00	0.95	0.19	0.24	0.00	0.78
	8 months	0.44	0.26	0.00	0.85				
	Overall	0.50	0.24	0.00	0.95	0.21	0.24	0.00	0.78
Noun-Adjective Ungrammatical: Gender Violation	2 months	0.63	0.17	0.15	0.93	0.48	0.30	0.03	1.00
	6 months	0.57	0.27	0.00	0.98	0.43	0.33	0.03	1.00
	8 months	0.58	0.28	0.05	0.98				
	Overall	0.59	0.24	0.00	0.98	0.45	0.32	0.03	1.00
D-Prime: Subject-Verb Number Violation	2 months	0.22	0.69	−0.37	2.84	1.66	1.29	−0.57	3.41
	6 months	0.62	1.02	−0.54	4.05	2.08	1.01	0.09	4.05
	8 months	0.79	1.17	−0.40	3.41				
	Overall	0.54	0.99	−0.54	4.05	1.87	1.17	−0.57	4.05
D-Prime: Noun-Adjective Number Violation	2 months	0.27	0.57	−0.37	2.33	1.91	1.29	−0.09	3.41
	6 months	0.54	1.00	−0.44	4.05	2.46	1.28	0.04	4.05
	8 months	0.93	1.27	−0.14	4.05				
	Overall	0.58	1.01	−0.44	4.05	2.18	1.30	−0.09	4.05
D-Prime: Noun-Adjective Gender Violation	2 months	0.15	0.42	−0.27	1.90	1.06	1.20	−0.86	2.84
	6 months	0.36	0.93	−0.37	4.05	1.60	1.32	−1.01	3.41
	8 months	0.56	1.06	−0.25	3.04				
	Overall	0.36	0.85	−0.37	4.05	1.33	1.28	−1.01	3.41

Note: SD = Standard Deviation, Min = Minimum, Max = Maximum.

Table 4

Descriptive Statistics for Production Measures by Learner Group

Production Tests (% Correct)	100-Class				200-Class			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Subject-Verb Number Singular	73.9	36.9	0	100	90.9	23.1	0	100
Subject-Verb Number Plural	62.6	43.3	0	100	90.0	22.3	0	100
Noun-Adjective Number Singular	60.4	21.8	20	100	77.6	17.3	40	100
Noun-Adjective Number Plural	20.4	22.9	0	80	58.3	26.4	0	100
Noun-Adjective Gender Masculine	50.0	24.7	10	100	72.6	20.9	30	100
Noun-Adjective Gender Feminine	27.5	25.4	0	100	49.6	25.0	10	90

Table 5

Results for D-Prime and Performance on the Production Task Predicting ERP Grammaticality Effects by Condition

Predictor Effects by Condition	Est	SE	Lower CI	Upper CI	Cohen's *d*	*p* <
D-Prime Predicting Violation Differences:						
Subject-Verb (SV) Number	0.44	0.19	0.07	0.82	0.67	**.021**
Noun-Adjective (NA) Number	0.69	0.11	0.48	0.91	1.74	**.001**
Noun-Adjective (NA) Gender	0.36	0.17	0.01	0.70	0.48	**.044**
Singular Production Predicting Violation Differences:						
Subject-Verb (SV) Number	0.01	0.01	−0.01	0.03	0.59	.284
Noun-Adjective (NA) Number	0.01	0.01	−0.00	0.02	0.77	.129
Noun-Adjective (NA) Gender	−0.00	0.01	−0.02	0.01	0.05	.881
Plural Production Predicting Violation Differences:						
Subject-Verb (SV) Number	0.01	0.01	−0.00	0.03	0.86	.102
Noun-Adjective (NA) Number	0.01	0.01	0.00	0.02	1.34	**.011**
Noun-Adjective (NA) Gender	−0.00	0.01	−0.02	0.01	0.24	.460
Noun-Adjective Production Predicting Violation Differences:						
Subject-Verb (SV) Number	0.40	0.26	−0.14	0.94	0.70	.134
Noun-Adjective (NA) Number	0.57	0.19	0.11	0.92	1.12	**.014**
Noun-Adjective (NA) Gender	0.22	0.22	−0.22	0.65	0.31	.316

Note: EST = Estimate, SE = Standard Error, CI = 95% Confidence Interval. Bold values indicate effects with *p*'s < .05.

Table 6

Descriptive Statistics for Individual Difference Measures by Learner Group

Production and ID Measures	100-Level				200-Level			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Verbal Aptitude Measures:								
MLAT-4	19.1	5.7	11	31	23.1	5.5	11	31
LLAMA_F	44.3	28.1	0	90	57.8	20.7	0	80
WAIS Subtests*:								
Matrix Reasoning	9.8	2.6	4	14	10.1	2.7	5	16
Letter Number (Working Memory)	8.9	2.0	6	14	10.4	1.8	8	16

Note: SD = Standard Deviation, Min = Minimum, Max = Maximum.
*The WAIS-IV subtests target perceptual reasoning (Matrix Reasoning) and working memory (Letter Number). According to WAIS-IV age-based norms, scaled scores of 1-7 indicate weakness (below average scores), scores of 8-12 indicate average performance, and scores of 13-19 indicate above average performance.

Table 7

Results for Individual Differences Predicting Grammaticality Effects by Outcome

Outcome	Predictor	Estimate	SE	Lower CI	Upper CI	$p <$
Predicting Sensitivity to Grammatical Violations (P600):						
Subject-Verb Number Violation Difference ($R^2 = .041$; $p = .457$)	Intercept	1.149	0.296	0.538	1.761	.001
	MLAT-4	0.000	0.057	−0.119	0.120	.995
	LLAMA_F	0.019	0.013	−0.008	0.047	.156
	Letter Number	−0.061	0.152	−0.377	0.254	.691
	Matrix Reasoning	0.062	0.119	−0.186	0.310	.606
Noun-Adjective Number Violation Difference ($R^2 = .124$; $p = .038$)	Intercept	1.180	0.210	0.750	1.609	.001
	MLAT-4	0.045	0.041	−0.039	0.129	.278
	LLAMA_F	0.008	0.009	−0.011	0.028	.400
	Letter Number	0.233	0.109	0.011	0.454	**.041**
	Matrix Reasoning	−0.020	0.085	−0.195	0.156	.819
Noun-Adjective Gender Violation Difference ($R^2 = .079$; $p = .235$)	Intercept	0.187	0.212	−0.241	0.615	.382
	MLAT-4	0.018	0.042	−0.066	0.102	.666
	LLAMA_F	0.005	0.010	−0.014	0.025	.580
	Letter Number	0.203	0.109	−0.018	0.424	.071
	Matrix Reasoning	−0.019	0.087	−0.195	0.156	.824
Predicting Grammatical Accuracy (D-prime):						
Subject-Verb Number Violation Difference ($R^2 = .030$; $p = .009$)	Intercept	1.101	0.156	0.785	1.417	.001
	MLAT-4	0.077	0.031	0.015	0.140	**.016**
	LLAMA_F	0.008	0.007	−0.006	0.023	.249
	Letter Number	0.025	0.065	−0.106	0.156	.702
	Matrix Reasoning	0.028	0.081	−0.135	0.192	.728
Noun-Adjective Number Violation Difference ($R^2 = .096$; $p = .007$)	Intercept	1.230	0.172	0.881	1.579	.001
	MLAT-4	0.093	0.034	0.024	0.162	**.009**
	LLAMA_F	0.008	0.008	−0.008	0.024	.302
	Letter Number	0.024	0.089	−0.157	0.204	.793
	Matrix Reasoning	0.041	0.055	−0.078	0.161	.470

Table 7 continued

Results for Individual Differences Predicting Grammaticality Effects by Outcome

Outcome	Predictor	Estimate	SE	Lower CI	Upper CI	$p <$
Predicting Grammatical Accuracy (D-prime) continued:						
Noun-Adjective Gender Violation Difference ($R^2 = .024$; $p = .044$)	Intercept	0.578	0.139	0.282	0.874	.001
	MLAT-4	0.063	0.027	0.006	0.121	**.034**
	LLAMA_F	0.004	0.006	−0.009	0.018	.487
	Letter Number	0.005	0.072	−0.148	0.157	.951
	Matrix Reasoning	0.023	0.071	−0.122	0.167	.752
Predicting Ability to Use Agreement Morphology (Production Task):						
Subject-Verb Number Violation Difference ($R^2 = .032$; $p = .001$)	Intercept	79.266	3.910	71.374	87.158	.001
	MLAT-4	2.194	0.767	0.645	3.744	**.007**
	LLAMA_F	0.338	0.180	−0.026	0.702	.068
	Letter Number	−0.478	2.016	−4.549	3.593	.814
	Matrix Reasoning	0.725	1.607	−2.520	3.970	.654
Noun-Adjective Number Violation Difference ($R^2 = .118$; $p = .001$)	Intercept	54.776	2.707	49.334	60.217	.001
	MLAT-4	1.269	0.532	0.200	2.338	**.021**
	LLAMA_F	0.031	0.124	−0.220	0.281	.807
	Letter Number	4.124	1.400	1.309	6.939	**.005**
	Matrix Reasoning	0.805	1.114	−1.434	3.044	.473
Noun-Adjective Gender Violation Difference ($R^2 = .066$; $p = .001$)	Intercept	50.484	2.780	44.894	56.075	.001
	MLAT-4	1.263	0.546	0.165	2.361	**.025**
	LLAMA_F	0.179	0.128	−0.078	0.436	.167
	Letter Number	2.738	1.438	−0.154	5.630	.063
	Matrix Reasoning	−0.167	1.143	−2.467	2.132	.884

Note: EST = Estimate, SE = Standard Error, CI = 95% Confidence Interval. Bold values indicate effects with p's < .05.

Table 8

Summary of significant results by condition for analyses of individual differences predicting performance on three measures (P600 effect size, GJT accuracy, oral production)

	P600 effect size	GJT Accuracy: D prime	Oral Production
MLAT-4: Verbal aptitude		-subject-verb number -noun-adj. number -noun-adj. gender	-subject-verb number -noun-adj. number -noun-adj. gender
Letter-Number: Working Memory	-noun-adj. number -noun-adj. gender (p = .071)		-noun-adj. number -noun-adj. gender (p = .063)

Figure 1: Grand-average ERP waveforms for each agreement type and testing session (100-level learners). Grand average ERP waveforms are plotted at representative electrode PZ for the 100-level learners for each agreement type (Subject-Verb Number, Noun-Adjective Number, Noun-Adjective Gender) and EEG testing session (Session 1, Session 2, Session 3). Waveforms for grammatical conditions are plotted in blue, and those for ungrammatical sentences (agreement violations) are plotted in red.

Figure 2: Grand-average ERP waveforms for each agreement type and testing session (200-level learners). Grand average ERP waveforms are plotted at representative electrode PZ for the 200-level learners for each agreement type (Subject-Verb Number, Noun-Adjective Number, Noun-Adjective Gender) and EEG testing session (Session 1, Session 2). Waveforms for grammatical conditions are plotted in blue, and those for ungrammatical sentences (agreement violations) are plotted in red.

Figure 3: Sample trial from the subject-verb agreement condition in the Oral Production task. The picture on the left depicts two people drinking in a restaurant, and the picture on the right depicts one person drinking on the beach.

Figure 4: Sample trial from the noun-adjective agreement condition in the Oral Production task. The picture on the left shows a person with one clean shirt and the one on the right a person with two dirty shirts.

Figure 1

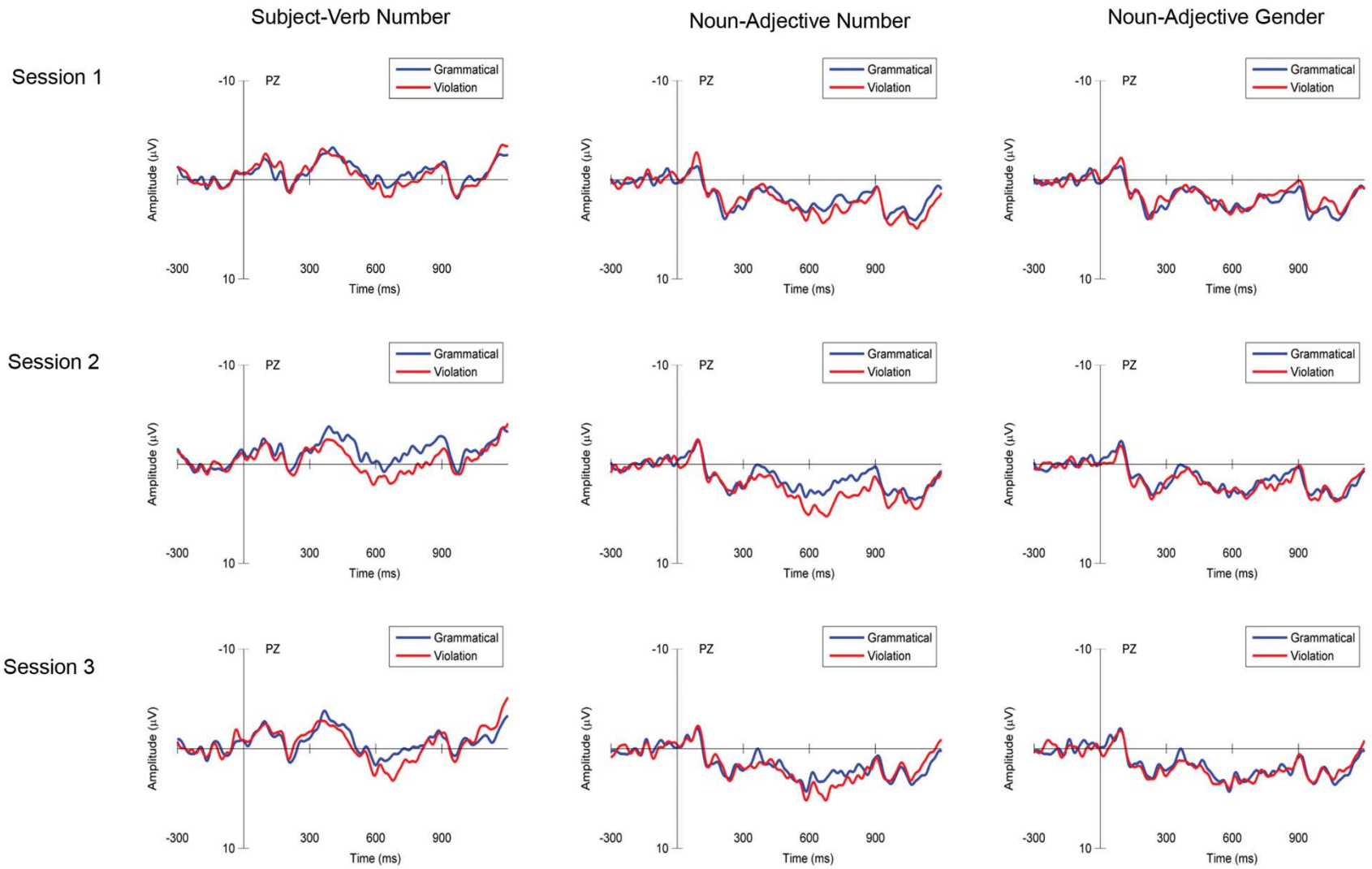
100-Level Learners
Subject-Verb Number
Noun-Adjective Number
Noun-Adjective Gender
Session 1
Session 2
Session 3
PZ
Grammatical
Violation
Amplitude (μV)
-10
10
-300
300
600
900
Time (ms)

Figure 2

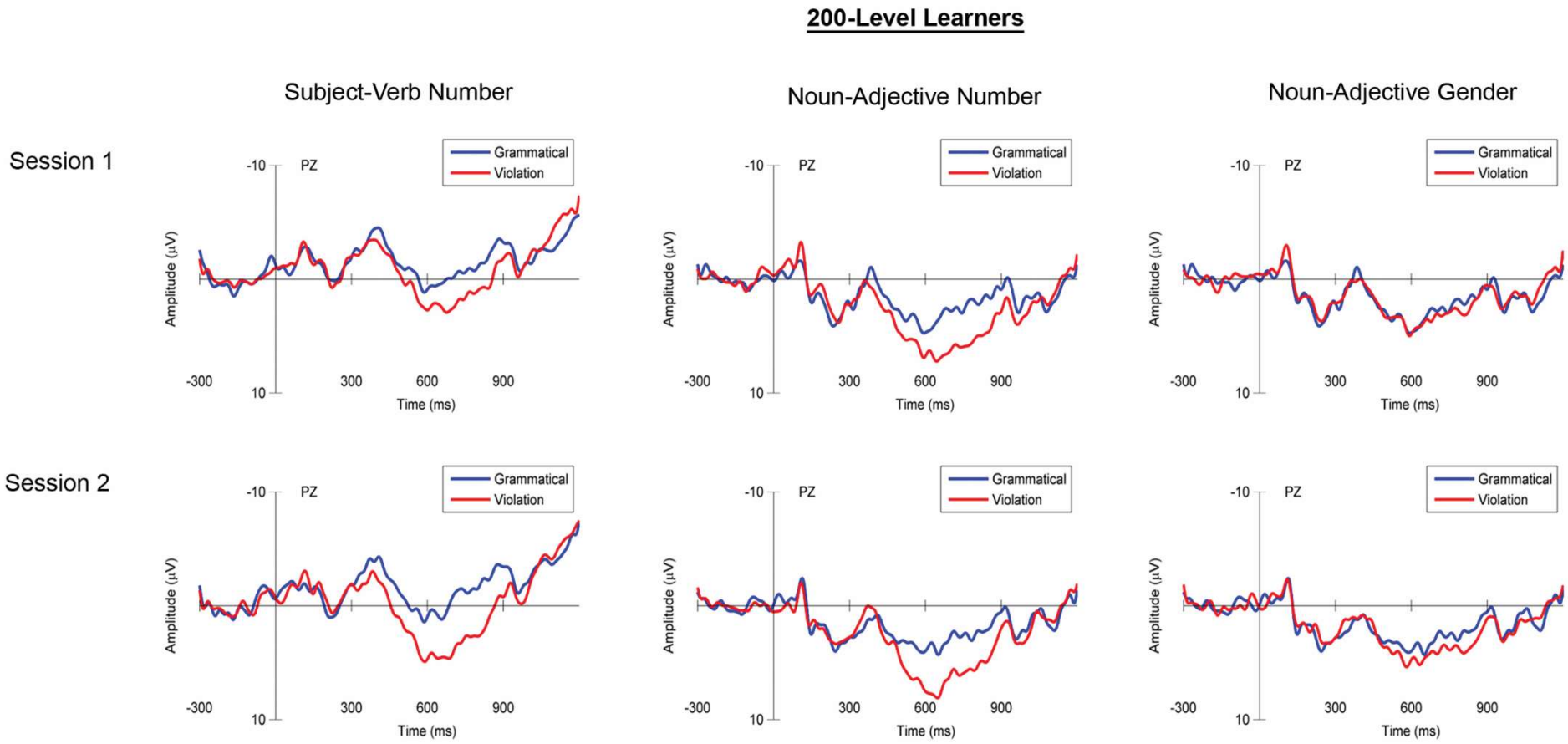
200-Level Learners
Subject-Verb Number
Noun-Adjective Number
Noun-Adjective Gender
Session 1
Session 2
PZ
Grammatical
Violation
Amplitude (μV)
Time (ms)
-10
10
-300
300
600
900

Figure 3

Amy y Mary

Henry

beber

Figure 4

Wendy

Kenny

tener

View publication stats