

## Research Article

### EXAMINING THE CONTRIBUTION OF MARKEDNESS TO THE L2 PROCESSING OF SPANISH PERSON AGREEMENT AN EVENT-RELATED POTENTIALS STUDY

**José Alemán Bañón** \*

*Stockholm University*

**David Miller**

*University of Illinois at Chicago*

**Jason Rothman**

*UiT The Arctic University of Norway and Universidad Nebrija*

#### Abstract

We used event-related potentials to investigate how markedness impacts person agreement in English-speaking learners of L2-Spanish. Markedness was examined by probing agreement with both first-person (marked) and third-person (unmarked) subjects. Agreement was manipulated by crossing first-person subjects with third-person verbs and vice versa. Native speakers showed a P600 for both errors, larger for “first-person subject + third-person verb” violations. This aligns with claims that, when the first element in the dependency is marked (first person), the parser generates stronger predictions regarding upcoming agreeing elements using feature activation. Twenty-two upper-intermediate/advanced learners elicited a P600 across both errors. Learners were equally accurate detecting both errors, but the P600 was marginally reduced for “first-person subject + third-person verb” violations, suggesting that learners overused unmarked forms (third person) online. However, this asymmetry mainly characterized lower-proficiency learners. Results suggest that markedness impacts L2 agreement without constraining it, although learners are less likely to use marked features top-down.

---

The authors thank Vincent DeLuca for his help with data collection, Jon Andoni Duñabeitia for his help with the cloze probability rating, Guillermo Montero Melis for his advice with the analyses, and all participants for their time. We are also very grateful to two anonymous reviewers for their constructive comments. Any remaining errors or oversights are our own.

\* Correspondence concerning this article should be addressed to José Alemán Bañón, Centre for Research on Bilingualism, Department of Swedish and Multilingualism, Stockholm University, Stockholm, Sweden. E-mail: jose.aleman.banon@biling.su.se

© The Author(s), 2020. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

## INTRODUCTION

A central question in second language (L2) acquisition research concerns how learners represent and utilize morphosyntactic features. One generalization that emerges from this literature is that learners show variability in the comprehension and use of inflection (i.e., the morphological exponence of features), even at high-proficiency levels (e.g., Franceschina, 2005; Grüter et al., 2012; Hopp, 2010; Keating, 2009, 2010; Lardiere, 1998; McCarthy, 2008, 2012; Rossi et al., 2014; Tsimpli & Dimitrakopoulou, 2007; see Slabakova, 2018). Theoretical models have suggested distinct, non-mutually exclusive sources of variability. For example, some prominent theories attribute variability to L1-L2 (dis)similarity with respect to either feature instantiation (e.g., Hawkins, 2001; Hawkins & Chan, 1997; Schwartz & Sprouse, 1996; Tsimpli & Dimitrakopoulou, 2007) or morphological realization (e.g., Jiang, 2004, 2007; Lardiere, 2009; Tokowicz & MacWhinney, 2005). Others identify processing pressure as the locus of variability, either in production (e.g., Pienemann, 2015; Prévost & White, 2000; Rothman, 2007; White, 2011) or in online comprehension (e.g., Grüter et al., 2012; Hopp, 2010; McDonald, 2006). Additional work has shown that factors such as phonological realization (e.g., Cabrelli Amaro et al., 2018; Carrasco-Ortiz & Frenck-Mestre, 2014; Goad & White, 2006, 2019) or structural/linear distance (e.g., Gillon-Dowens et al., 2010; Keating, 2010) also account for inflectional variability.

The present study investigates the relationship between L2 inflectional variability and markedness, the observation that different feature values are asymmetrically represented. Morphological theory proposes that marked features are more complex, have more internal structure, or are less frequent than unmarked features. For *person*, the claim is that first and second person are marked, while third person is unmarked (e.g., Forchheimer, 1953; Harley & Ritter, 2002; Harris, 1995; McGinnis, 2005; Nevins, 2011).<sup>1</sup> An additional claim is that only marked features are specified. For example, first and second person are specified as *participants* in the speech act, the *speaker* and *addressee*, respectively. Third person, in contrast, is underspecified (i.e., it lacks featural information) because it is a *nonparticipant* in the speech act. Third person is considered to be the “default” person, a sort of “elsewhere” form. Our study examines whether/how this asymmetry impacts L1-English L2-Spanish learners’ processing of sentences like (1a–b), which differ with respect to person markedness. We build on previous observations that learners overuse “defaults,” that is, forms that are unmarked/underspecified (although not necessarily uninflected) and thus appear in a wider range of contexts (e.g., third-person verbal inflection emerges with personless subjects, as in *Correr relaja* “Running relax<sub>3RD-SG</sub>”) (e.g., Hawkins, 2001, 2009; McCarthy, 2008; Prévost & White, 2000).

- (1) a. Yo estudio.  
I<sub>1ST-SG</sub> study<sub>1ST-SG</sub>  
b. El médico estudia.  
the doctor<sub>3RD-SG</sub> study<sub>3RD-SG</sub>

Several L2 theoretical models claim that markedness modulates acquisition, although the specific proposals differ (e.g., Hawkins, 2001, 2009; Haznedar & Schwartz, 1997; McCarthy, 2008; Prévost & White, 2000). For example, the Missing Surface Inflection Hypothesis (MSIH) (e.g., Prévost & White, 2000) posits that learners overuse defaults

(i.e., masculine, singular, third person) in contexts that require marked features (i.e., feminine, plural, first/second person), but rarely do the reverse. The source of this variability is assumed to be a computational deficit. More specifically, it is argued that learners can represent abstract features, but have difficulty with the retrieval of inflectional morphemes and their mapping onto lexical items under computational pressure (e.g., White, 2011). For example, when establishing agreement with a first-person subject, which is marked/specified for person (*speaker*), learners might incorrectly supply third-person verbal inflection (e.g., *Yo \*habla español* “I<sub>1ST-SG</sub> speak<sub>3RD-SG</sub> Spanish”) because unmarked/underspecified forms (i.e., defaults) are easier to retrieve and function as a sort of all-purpose form. However, if the subject corresponds to the third person, learners rarely inflect verbs for first person (*Él \*hablo español* “He<sub>3RD-SG</sub> speak<sub>1ST-SG</sub> Spanish”) because they have difficulty retrieving marked/specified forms. In sum, learners tend to underspecify agreement targets such as verbs (or adjectives), but they rarely do the reverse.

Alternatively, McCarthy (2008) posits that the overuse of defaults is not due to lexical retrieval difficulty, but to the asymmetrical representation of features in the learner’s grammar. McCarthy’s account predicts qualitatively similar variability (i.e., overreliance on defaults) across tasks (i.e., comprehension, production). This contrasts with the MSIH, which predicts variability to emerge as a function of the task’s computational burden. McCarthy (2008) found that intermediate/advanced English-speaking learners of Spanish incorrectly extended masculine forms to feminine contexts in both comprehension and production, but they seldom did the reverse. Furthermore, the learners overused singular in production. With respect to person, McCarthy (2012) found that L1-English L2-Spanish learners produced third-person verbs with first-person subjects (83/111 errors), but rarely did the reverse (18/111 errors). Importantly, the intermediate group outperformed low-intermediate learners with first-person verbs (~88% vs. ~70%), but not with third-person verbs (~92% vs. ~94%, respectively), suggesting that marked forms are acquired later. Although this aligns with McCarthy’s proposal, the lack of comprehension data showing the same asymmetry precludes conclusions.

Recent investigations have examined inflectional variability in real-time comprehension (e.g., Grüter et al., 2012; Hopp, 2010; López Prego, 2015; López Prego & Gabriele, 2014). For example, López Prego and Gabriele (2014) examined number and gender agreement in Spanish with a design manipulating both markedness and task demands (untimed vs. speeded grammaticality judgment task [GJT]). Under time pressure, the intermediate and advanced learners were slower and less accurate detecting “plural noun + singular adjective” errors than the reverse configuration, suggesting some reliance on defaults (i.e., overusing singular). Crucially, the L1-Spanish controls showed a similar asymmetry for both number and gender when tested under stress, suggesting that overreliance on defaults can emerge under computational pressure in intact grammars.

Another relevant finding from López Prego and Gabriele’s study is that the intermediate/advanced learners were faster and more accurate rejecting “feminine noun + masculine adjective” errors relative to “masculine noun + feminine adjective” errors, contra McCarthy’s predictions. The authors suggest that, when the first element in the dependency is marked/specified for gender (feminine), feature activation might ease agreement resolution by allowing the parser to better anticipate the gender of the upcoming adjective (e.g., Dussias et al., 2013; Foucart et al., 2014; Hopp, 2013).

López Prego and Gabriele's proposal connecting markedness to facilitation originated in the psycholinguistics literature, where it has been argued that marked features remain longer in the focus of attention (e.g., Wagers & McElree, 2011; Wagers & Phillips, 2014) and are, thus, more likely to impact agreement operations. For example, ungrammatical strings such as *the key to the cabinets \*are* cause less disruption when the attractor noun (*cabinets*) is plural (e.g., Acuña Fariña et al., 2014; Dillon et al., 2013; Pearlmutter et al., 1999; Wagers et al., 2009). Likewise, Carminati (2005) showed that coreference between a null pronoun and a structurally dispreferred antecedent is less disruptive when the disambiguating verb is inflected for first/second person (marked) relative to third person (unmarked). Crucially, Nevins et al. (2007) posited that markedness might determine whether agreement is established predictively. Their proposal is that, upon encountering a marked feature, the parser can generate a stronger prediction regarding upcoming agreement elements.

To summarize, certain L2 theoretical models (McCarthy, 2008, 2012; Prévost & White, 2000) posit that learners have difficulty accessing marked/specified forms when establishing agreement. Consequently, they incorrectly overextend unmarked/underspecified forms (i.e., defaults) to contexts that require marked ones, due either to computational pressure (e.g., the MSIH) or to representational issues (e.g., McCarthy, 2008). Alternatively, proposals from the psycholinguistics literature (e.g., Nevins et al., 2007; Wagers & Phillips, 2014) capitalize on the predictive value of marked features. Under this proposal, when the first element in the dependency is marked, feature activation allows the parser to better resolve agreement further down the line.

### **BACKGROUND OF THE PRESENT STUDY**

In previous studies, we evaluated the previously mentioned proposals regarding how markedness impacts agreement resolution in both native (Alemán Bañón & Rothman, 2016, 2019) and nonnative speakers of Spanish (Alemán Bañón et al., 2017). In those studies, we examined online agreement resolution using event-related potentials (ERPs). In addition to providing high temporal resolution and being multidimensional, ERPs can unveil qualitative differences between different agreement dependencies. For example, Foucart and Frenck-Mestre (2012) found that the same L1-English L2-French learners elicited qualitatively different brain responses to gender errors (N400 vs. P600), depending on the syntactic configuration where they were realized. ERPs are, therefore, well suited for investigating both quantitative and qualitative differences between agreement dependencies that differ with respect to markedness.

In Alemán Bañón and Rothman (2019), we argued in favor of Nevins et al.'s proposal (2007) that markedness allows the parser to resolve agreement top-down, at least when the dependency is sufficiently constraining (e.g., Dillon et al., 2013). In that study, we examined subject-verb person agreement with both first-person (marked/specified: *speaker*; see 2) and third-person singular subjects (unmarked/underspecified: see 3) in a group of 28 native speakers of European Spanish. We then manipulated agreement by crossing first-person subjects with third-person verbs (2b) and vice versa (3b). Both violation types yielded a P600 (500–1,000 ms), a component associated with various morphosyntactic operations (e.g., Osterhout & Holcomb, 1992), consistent with previous

studies on person agreement (e.g., Mancini et al., 2011, 2019; Nevins et al., 2007; Silva Pereyra & Carreiras, 2007; Zawiszewsky et al., 2016).<sup>2</sup>

- (2) a. Yo a menudo acaricio a los caballos.  
 I-1ST-SG often pet-1ST-SG CASE the horses  
 b. Yo a menudo \*acaricia a los caballos.  
 I-1ST-SG pet-3RD-SG
- (3) a. El cartero a menudo acaricia a los gatos.  
 the postman-3RD-SG often pet-3RD-SG CASE the cats  
 b. El cartero a menudo \*acaricio a los gatos.  
 the postman-3RD-SG pet-1ST-SG

Violations with a first-person subject (2b) yielded a larger P600 (700–900 ms) than violations with a third-person subject (3b), which we interpreted as evidence that the parser can better resolve agreement when the first element in the dependency is marked (Nevins et al., 2007). Recall that the rationale behind this proposal is that, when the first element in the dependency is marked/specified for person, feature activation allows the parser to generate a stronger prediction regarding the upcoming verb. When this prediction is unmet, the result is a larger P600. This is in line with current proposals interpreting the P600 as an index of the reanalysis processes triggered by violations of top-down expectations (e.g., Bornkessel-Schlesewsky & Schlewewsky, 2008; Kuperberg, 2007; Tanner et al., 2017; van de Meerendonk et al., 2010).<sup>3</sup>

In contrast, our investigation of noun-adjective number and gender agreement with the same native speakers (Alemán Bañón & Rothman, 2016) failed to provide similar evidence. In that study, we probed agreement with both feminine and masculine nouns (corresponding to the marked/specified and unmarked/underspecified genders, respectively), which could be used in the plural or in the singular (marked/specified vs. unmarked/underspecified number values, respectively) (see an example of a sentence with a feminine singular noun in (4)).

- (4) Carlos fotografió una catedral que parecía inmensa para una revista.  
 Carlos photographed a cathedral-FEM-SG that looked huge-FEM-SG for a magazine

Our results revealed an earlier P600 for number and gender violations realized on marked adjectives (example from the gender conditions: *coche que parecía \*cara* “car-MASC that looked expensive-FEM”) relative to the opposite error type (e.g., *catedral que parecía \*inmenso* “cathedral-FEM that looked huge-MASC”). In addition, number violations realized on plural adjectives (e.g., *coche que parecía \*caros* “car-SG that looked expensive-PL”) yielded a larger P600 than the reverse error type (e.g., *coches que parecían \*caro* “car-PL that looked expensive-SG”). In sum, native speakers were sensitive to number/gender markedness (like the Spanish natives under stress in López Prego and Gabriele’s study, 2014 and in line with McCarthy’s predictions for L2ers), but we found no evidence that the marked status of the first element in the dependency (i.e., the noun) facilitated agreement. Possibly, the structure where we examined number/gender agreement was not sufficiently constraining to allow the parser to generate predictions regarding upcoming adjectives because continuations other than an adjective

were possible (e.g., *una calle que parecía zigzaguear* “a street that seemed to zigzag”). The same is not true of the subject-verb agreement manipulation in (2–3), where the subject made it certain that a verb would appear to satisfy the sentence-building phrase structure rule (e.g., Chomsky, 1957, 1995).

In Alemán Bañón et al. (2017), we extended the examination of noun-adjective number and gender agreement to 22 upper-intermediate/advanced L1-English L2-Spanish learners, and found qualitatively similar results to the native controls. That is, the learners showed a P600 across both types of gender violations, with an earlier onset for gender violations realized on marked/specified (feminine) adjectives. In addition, they elicited a P600 across both number violation types, marginally larger for errors on marked/specified (plural) adjectives. Thus, similar to the native controls, the L2ers were sensitive to the marked status of the adjective, as opposed to the marked status of the noun (i.e., the first element in the dependency). Importantly, this sensitivity was qualitatively nativelike, which does not align with representational accounts of variability.

Herein, we investigate how the same learners resolve subject-verb person dependencies that differ with respect to markedness (first person: marked/specified as *speaker*; third person: unmarked/underspecified). We will examine whether learners overuse defaults (i.e., third-person verbal inflection), as predicted by L2 theories of morphological variability. We will further investigate whether this asymmetry is computational (e.g., Prévost & White, 2000) or representational (McCarthy, 2008). We will do so by examining the learners’ brain responses to the violating verbs (i.e., the point when the dependency is established) and their accuracy with both dependencies in the untimed GJT at the end of each sentence.

We will also examine the extent to which proficiency explains this potential overreliance on default morphology. Recall that, in McCarthy’s study (2012), intermediate learners outperformed low-intermediate learners with first-person verbs, but not with third-person verbs, suggesting that marked forms (i.e., first-person inflection) emerge later in L2 production. We will examine whether a similar asymmetry characterizes L2 comprehension, as predicted by McCarthy. Importantly, proficiency is the most reliable predictor of whether L2ers elicit a P600 for morphosyntactic errors (Caffarra et al., 2015). We will also follow recent claims for the need to dissociate global proficiency from experiential factors, such as amount of L2 instruction and immersion in L2-speaking countries (Bowden et al., 2013; Caffarra et al., 2015; DeLuca et al., 2019). A few studies have investigated the relative contribution of these factors to L2 processing, but results remain inconclusive (e.g., Alemán Bañón et al., 2018; Faretta-Stutenberg & Morgan-Short, 2018).

Alternatively, we will examine whether, similar to the native controls in Alemán Bañón and Rothman (2019), the marked status of the first element in the dependency eases person agreement resolution in the L2. If so, learners should show increased sensitivity (i.e., a larger P600) to person violations with a first-person subject (marked/specified). To our knowledge, only López Prego (2015) has examined this question in L2 comprehension. López Prego used self-paced reading to examine adjective-noun gender agreement in Spanish with a design manipulating whether the adjective showed overt gender cues (e.g., *nueva* “new-FEM”/ *verde* “green” in 5a) and whether the trigger noun was feminine (5a) or masculine (5b), corresponding to the marked/specified and unmarked/underspecified genders.

- (5) a. Como es nueva/verde, he decidido que llevaré la blusa que compré ...  
 As is new-FEM/green-NO CUE I-have decided to wear the blouse-FEM that I-bought
- b. Como es nuevo/verde, (...) el abrigo que compré ...  
 new-MASC/green-NO CUE (...) the coat-MASC

Her results showed that both native speakers and advanced English-speaking learners read the complementizer following the trigger noun (*blusa que* ...) faster when the preceding adjective was feminine (marked/specified), relative to when it was morphologically invariant (*nueva* vs. *verde* in 5a), although the effect was marginal in the learners. In contrast, no such facilitation emerged in the comparison of masculine (unmarked/underspecified) versus invariant adjectives (*nuevo* vs. *verde* in 5b). López Prego argues that marked features ease agreement resolution at the noun (although in her study facilitation emerged after the noun).

The present study examines whether other markedness-related properties, such as *speech participant* status, also facilitate L2 agreement resolution. To our knowledge, only Rossi et al. (2006) and Tanner et al. (2013) have used ERPs to examine L2 person agreement resolution, and neither study manipulated markedness. Rossi et al.'s bidirectional study found that German-Italian learners elicited a P600 for "third-person subject + first/second-person verb" errors (e.g., *Il signore...beve/\*bevo* "the man drink-3RD-SG/drink-1ST-SG"), which was modulated by proficiency. Tanner et al. (2013) found that third-year L1-English L2-German learners elicited a P600 for "first-person subject + third-person verb" errors (e.g., *Ich wohne/\*wohnt in Berlin* "I live-1ST-SG/live-3RD-SG..."), but first-year students showed a biphasic N400-P600 pattern, which they argue reflects individual differences with respect to morphosyntactic development. Thus, the question of how person markedness modulates L2 processing remains unexplored. Importantly, as Slabakova (2018) points out, studies examining potential reliance on defaults among intermediate/advanced learners are lacking. Our study fills these gaps.

## THE PRESENT STUDY

The present study investigates how markedness modulates subject-verb person agreement in Spanish, among upper-intermediate/advanced English-speaking learners. Our design manipulates both person markedness (first person: marked/specified as *speaker*; third person: unmarked/underspecified) and agreement. Our research questions "RQ" are:

*RQ1. Does markedness impact subject-verb person agreement resolution in the L2?*

McCarthy's representational account (2008, 2012) predicts reduced sensitivity to "first-person subject + third-person verb" errors relative to "third-person subject + first-person verb" errors across all metrics (accuracy in an untimed GJT, brain responses). This is because the learner's grammar allows the former error type due to representational issues. Alternatively, it is possible that such an asymmetry will only emerge in the ERP responses time-locked to the presentation of the verb because the learners' brain must detect the error exactly when the verb is encountered, whereas the end-of-the-sentence GJT is less time constrained (e.g., Prévost & White, 2000).

*RQ2. To what extent does L2 proficiency account for a potential overreliance on default morphology?*

Proficiency should explain variability with “first-person subject + third-person verb” errors across metrics, but not with the opposite error type (McCarthy, 2012). This is because the former error type involves overusing third-person inflection (a default), a representational issue that learners are only predicted to overcome at higher levels of proficiency (based on McCarthy, 2012). Alternatively, it is possible that proficiency and markedness will only interact in the ERP data (e.g., Prévost & White, 2000).

*RQ3. Do L2 learners use person markedness information to ease agreement resolution?*

If so, L2ers should elicit larger ERP responses to “first-person subject + third-person verb” errors (similar to the Spanish native speakers in Alemán Bañón and Rothman, 2019) due to the fact that feature activation at the subject (i.e., the first element in the dependency) should facilitate agreement at the verb (e.g., López Prego, 2015; Nevins et al., 2007).

## **METHODS**

### ***PARTICIPANTS***

Twenty-two L1-English L2-Spanish learners (12 females; mean age: 25; *SD*: 7.5) with a mean age of L2 acquisition of 14 years (range: 8–23) provided their informed written consent to participate in the study. Their Spanish proficiency was monitored with a 50-item test including the cloze section of the Diploma de Español como Lengua Extranjera and the vocabulary section from the MLA Cooperative Foreign Language Test (e.g., Alemán Bañón et al., 2014; McCarthy, 2008; White et al., 2004). Sixteen learners were of advanced proficiency (range: 43–50/50) and six of intermediate proficiency (33–38). Mean duration of Spanish instruction was 7.3 years (*SD*: 2.7 years; range: .5–12 years) and mean duration of immersion in Spanish-speaking countries was 15 months (*SD*: 13 months; range: 0–48 months). Only four learners had lived in L2-speaking countries for less than eight months. All learners grew up as monolingual speakers of English, with the exception of one heritage speaker of Japanese, a language that does not instantiate subject-verb person agreement. Twenty of the learners reported knowledge of other foreign languages to varying degrees of proficiency.

A group of 28 native speakers of Spanish reported in Alemán Bañón and Rothman (2019) served as the control group. All 50 participants met the standard requirements for language-related ERP studies: right-handedness (Edinburgh Handedness Questionnaire; Oldfield, 1971), normal or corrected-to-normal vision, and no history of neurological impairments. The testing took place in the United Kingdom and all participants received monetary compensation for their time.

### ***MATERIALS***

To examine the contribution of person markedness to agreement, we created 80 sentences with a first-person singular subject (Table 1: condition 1), which is marked/specified for

TABLE 1. Sample stimuli, including the conditions examining person agreement with first-person and third-person subjects (grammatical, ungrammatical), and the fillers

FIRST-PERSON SUBJECT	
Grammatical	
1. <u>Yo</u> a menudo <u>canto</u> en la ducha.	L <sub>1ST-SG</sub> often sing <sub>1ST-SG</sub> in the shower
Violation	
2. <u>Yo</u> a menudo * <u>canta</u> en la ducha.	L <sub>1ST-SG</sub> often sing <sub>3RD-SG</sub> in the shower
THIRD-PERSON SUBJECT	
Grammatical	
3. <u>La artista</u> a menudo <u>canta</u> en los festivales.	the artist <sub>3RD-SG</sub> often sing <sub>3RD-SG</sub> in the festivals
Violation	
4. <u>La artista</u> a menudo * <u>canto</u> en los festivales.	the artist <sub>3RD-SG</sub> often sing <sub>1ST-SG</sub> in the festivals
FILLERS	
Nosotros somos muy persistentes y ellos también. we <sub>1ST-PL</sub> are very persistent and they <sub>3RD-PL</sub> too.	
Ellas son más puntuales que tú. they <sub>3RD-PL</sub> are more punctual than you <sub>2ND-SG</sub>	

person (*speaker*), and 80 sentences with a third-person singular lexical subject (condition 3), which is unmarked/underspecified for person (the default person). Agreement was manipulated by pairing up first-person subjects with third-person verbs (condition 2), and vice versa (condition 4).

### *Structure of the Sentences*

All experimental sentences followed the structure in conditions 1–4. They started with the subject, followed by the adverb *a menudo* “often,” the verb, and a three-word continuation that ensured that the verb (i.e., the critical word) was not sentence-final. The adverb *a menudo* intervened between the subject and the verb to create linear distance between the agreeing elements. Previous ERP studies argued that comprehenders are more likely to exploit predictive strategies when they have sufficient time for prediction generation (e.g., Ito et al., 2017; Wlotko & Federmeier, 2015). Thus, if the L2ers herein can successfully use the subject’s person markedness to ease agreement resolution at the verb, the present setup is suitable to investigate that possibility.

### *The Subjects of the Sentences*

The conditions with a marked/specified subject involved the first-person singular pronoun *yo* (conditions 1–2). For theoretical reasons, the conditions with an unmarked/underspecified subject involved third-person singular lexical determiner phrases (DPs) (conditions 3–4). There is disagreement in the literature regarding whether third-person pronouns are underspecified for person (e.g., Harley & Ritter, 2002) or carry a *nonparticipant* person feature (e.g., Nevins, 2007). However, there is consensus that referential DPs like *el maestro* “the teacher” carry no person specification (e.g., Bianchi, 2006; Den Dikken, 2011; Nevins, 2011). Because the present study is concerned with whether the parser can better establish person agreement when the subject carries person information, relative to when it does not, we opted for lexical DPs as unmarked/underspecified subjects. Because the Spanish verb also encodes number, all subjects were used in the singular (unmarked/underspecified for number). Two measures were taken to match the two markedness conditions as much as possible. First, we submitted the sentences with lexical DP subjects, truncated at the adverb *a menudo*, to a cloze probability rating (e.g., *la artista a menudo...*) to rule out the possibility that participants could predict the target verbs based on the lexical features of the subjects. This rating, which involved 33 Spanish native speakers who did not participate in the ERP study reported in Alemán Bañón and Rothman (2019), revealed that mean cloze probability was very low (mean cloze: .03; *SD*: .1), suggesting that the target verbs were, overall, not predictable. In addition, we added 80 fillers using other (nominative case) personal pronouns in coordinated structures with ellipsis (see Table 1). In 40 of those fillers, the pronouns were used in contrastive focus. The use of other person pronouns was expected to attenuate the salience of pronoun *yo* in the experiment. Likewise, the use of contrastive focus and ellipsis was expected to improve the naturalness of overt pronouns in the sentences, given that Spanish is a null-subject language.

### *The Target Verbs*

We used the same verbs in the conditions with first- and third-person subjects. Thus, at the verb (i.e., the critical word) the two markedness conditions only differed with respect to the subject. Verbs inflected for first- and third-person singular were controlled with respect to number of characters (mean length of third-person verbs: 6.56; *SD* = 1.61; 95% CI [6.20, 6.92]; first-person verbs: 6.57; *SD* = 1.65; 95% CI [6.21, 6.94];  $t(79) = .445$ ,  $p = .658$ ;  $\eta^2 = .003$ ). However, third-person verbs were significantly more frequent than first-person ones (EsPal database; Duchon et al., 2013). Being underspecified for person, third-person verbs emerge in more syntactic contexts than first-person verbs, which results in the former being more frequent. Finally, the critical verbs were always located midsentence.

These 160 sentences were intermixed with 240 sentences from Alemán Bañón et al. (2017), a study that does not manipulate subject-verb agreement. These materials were counterbalanced across 12 lists. Across lists, all sentences occurred in their grammatical and ungrammatical versions, but no participant saw the same sentence twice. Each participant saw two different lists, administered on separate days. Each list contained an equal number of items per condition. After combining the two lists, each participant saw 80 sentences with a first-person subject (40 ungrammatical) and 80 sentences with a third-person subject (40 ungrammatical). They also saw 20 items from each of the 12 conditions in Alemán

Bañón et al. (2017) (80 grammatical, 80 number violations, 80 gender violations) and 80 grammatical fillers. The ratio of grammatical to ungrammatical sentences was 1/1. In total, participants saw 480 sentences across the two sessions. All materials associated with this study are published in Alemán Bañón and Rothman (2019). All materials from the study on number/gender agreement are published in Alemán Bañón and Rothman (2016).<sup>4</sup>

### **PROCEDURE**

The testing involved two EEG recordings, each including 240 sentences (with an equal number of items per condition, including the fillers). We used the software *Paradigm* (Perception Research Systems Inc.; Tagliaferri, 2005) for sentence presentation. Participants were instructed to read the sentences silently, without blinking, and to decide whether each was grammatical or ungrammatical in Spanish (e.g., Rossi et al., 2006; Tanner et al., 2013). They were asked to favor accuracy over speed. Each EEG recording included eight practice sentences. Four of them were ungrammatical, but none involved agreement violations or nouns/verbs from the experimental materials. Feedback was provided for the first three practice trials. The experiment began upon completion of the practice. Each recording was divided into six 40-sentence blocks, separated by five short breaks. Each EEG recording lasted approximately 1 hour.

### **Trial Structure**

First, a fixation cross was displayed in the center of the screen for 500 ms. Then, the presentation of the sentence began, one word at a time. Each word was displayed for 450 ms, followed by a 300 ms blank (e.g., Alemán Bañón et al., 2012, 2014; see Molinaro et al., 2011). At the end of the sentence came a 1,000 ms pause. Participants then saw the words *Bien* “good” or *Mal* “bad” on the screen and decided if the sentence was grammatical or ungrammatical by pressing a button (middle and index fingers of the left hand, respectively). The prompts remained on the screen until the participant provided a response. Upon the button press, an intertrial interval was added ranging between 500–1,000 ms, pseudorandomly varied at 50 ms increments.

### **EEG RECORDING AND ANALYSIS**

We used Brain Vision Recorder (Brain Products, GmbH, Germany) to record the continuous EEG from 64 sintered Ag/AgCl electrodes attached to an elastic cap (Easycap, Brain Products, GmbH, Germany). The electrodes were placed following the 10% system (midline: FPz, Fz, FCz, Cz, CPz, Pz, POz, Oz; hemispheres: FP1/2, AF3/4, AF7/8, F1/2, F3/4, F5/6, F7/8, FC1/2, FC3/4, FC5/6, FT7/8, FT9/10, C1/2, C3/4, C5/6, T7/8, CP1/2, CP3/4, CP5/6, TP7/8, TP9/10, P1/2, P3/4, P5/6, P7/8, PO3/4, PO7/8, O1/2) and referenced online to FCz (with AFz as ground). We used electrodes FP1/2 (above the eyebrows) to monitor blinks, and electrode IO (on the outer canthus of the right eye) to monitor horizontal eye movements. All electrode impedances were kept below 10 k $\Omega$ . The recordings were amplified with an online bandpass filter of .016–250 Hz (with a 12dB/octave rolloff) by a BrainAmp MR Plus amplifier (Brain Products, GmbH, Germany), and digitized at a sampling rate of 1 kHz.

We used Brain Vision Analyzer 2.0 (Brain Products, GmbH, Germany) for offline data processing. First, we re-referenced the recordings to the average of near-mastoid electrodes (TP7/8).<sup>5</sup> We then segmented the EEG into epochs from  $-300$  ms to  $1,200$  ms relative to the verb. Upon visual inspection, we rejected trials with blinks, horizontal eye movements, excessive alpha waves, or excessive muscle movement. We then discarded trials with incorrect behavioral responses. Finally, the epochs were baseline-corrected relative to the  $300$  ms prestimulus baseline, averaged per condition and per subject, and filtered with a phase-shift free Infinite Impulse Response Butterworth filter, with a high cutoff of  $30$  Hz and a  $12$  dB/octave rolloff.

Rejection of trials with artifacts or incorrect behavioral responses resulted in approximately  $16\%$  of data loss. The mean number of good trials per condition ranged between  $33$ – $35/40$  (Condition 1:  $35$ ; Condition 2:  $33$ ; Condition 3:  $34$ ; Condition 4:  $33$ ). A repeated-measures ANOVA revealed that we had retained fewer trials associated with person violations than with grammatical sentences overall,  $F(1, 21) = 4.450$ ,  $p = .047$  (we discarded incorrectly judged trials, and learners were marginally less accurate rejecting person errors overall in the GJT). As Luck (2014, supplement, chapter 8, pp. 4–5) points out, different numbers of trials per condition is not problematic when analyzing mean amplitudes. Most importantly, this difference does not affect the examination of the Markedness by Agreement interaction because we retained a comparable number of trials for each error type.

We quantified ERPs using mean amplitudes between  $250$ – $450$  ms and  $450$ – $900$ ms, corresponding to the LAN/N400 and the P600, respectively (based on Alemán Bañón et al., 2017). We used nine regions of interest (ROI) for statistical analysis, each calculated by averaging across the mean amplitudes of all electrodes in the region (Left Anterior: F1, F3, F5, FC1, FC3, FC5; Right Anterior: F2, F4, F6, FC2, FC4, FC6; Left Medial: C1, C3, C5, CP1, CP3, CP5 Right Medial: C2, C4, C6, CP2, CP4, CP6; Left Posterior: P1, P3, P5, P7, PO3, PO7; Right Posterior: P2, P4, P6, P8, PO4, PO8; Midline Anterior: Fz, FCz; Midline Medial: Cz, CPz; Midline Posterior: Pz, POz).

### *Statistical Analysis*

To evaluate RQ1 and RQ3, mean amplitudes were submitted to a repeated-measures ANOVA with Markedness (first-person subject, third-person subject), Agreement (grammatical, ungrammatical), Anterior-Posterior (anterior, medial, posterior), and Hemisphere (left, right) as the within-subjects factors. The hemisphere and midline regions were analyzed separately. For the analyses in the midline, the only topographical factor in the ANOVA was Anterior-Posterior. Because the Markedness by Agreement interaction is critical for our discussion, whenever this interaction was qualified by a topographical factor, follow-up analyses were conducted by examining the Markedness by Agreement interaction separately within the relevant ROIs. The Geisser and Greenhouse correction was applied for sphericity violations (we report corrected degrees of freedom; Field, 2005). To evaluate RQ2, we ran a series of multiple regressions with repeated measures (detailed in the following text). These analyses allowed us to examine how the linear relationship between proficiency measures (score in standardized Spanish proficiency test, amount of L2 instruction, immersion in L2-speaking countries) and measures of

TABLE 2. Learners' mean accuracy rates in the GJT ( $N = 22$ ) for the conditions examining person agreement with first-person (marked/specified) and third-person (unmarked/underspecified) subjects

	Grammatical	Violation	D-Prime Score
1st-person subject	96 (5)	92 (9)	3.43 (.8)
3rd-person subject	95 (6)	92 (13)	3.42 (1)

*Note.* Standard deviations are provided between parentheses. D-prime scores are provided in the rightmost column.

sensitivity to agreement (P600 magnitude, D-prime) varied for each level of the repeated factor (person error type) (Schneider et al., 2015).

A False Discovery Rate correction (Benjamini & Hochberg, 1995) was applied to all follow-up tests, to control for Type I error. For all follow-up tests, we provide both the raw  $p$  value and the adjusted significance level ( $q^*$ ). We consider effects where  $p$  is below .05 as significant and effects where  $p$  is between .05 and .1 as marginal.

## RESULTS

### BEHAVIORAL

The percentage of accurate responses in the GJT is provided in Table 2 for all experimental conditions. For each person dependency type, the rightmost column of the table provides D-Prime Scores (i.e., a single measure of sensitivity to each person dependency type that controls for response bias). The learners were very accurate across the board (above 90% across conditions), suggesting that they understood the task and were able to complete it. A repeated-measures ANOVA with Markedness (first-person subject, third-person subject) and Agreement (grammatical, ungrammatical) as the repeated factors revealed a marginal main effect of Agreement,  $F(1, 21) = 3.641$ ,  $p = .07$ ;  $\eta^2 = .148$ , driven by the fact that learners were less accurate rejecting ungrammatical sentences overall ( $M = 92$ ;  $SD = 11$ ; 95% CI [87, 97]) than accepting grammatical ones ( $M = 96$ ;  $SD = 5$ ; 95% CI [93, 98]).

### ERP EFFECTS

Both types of person errors yielded more positive waveforms than their grammatical counterparts (Figures 1–2). This positivity starts ~450 ms upon presentation of the violating verb, does not go back to baseline by the end of the epoch (1,200 ms), and shows a central-posterior distribution. Overall, this is consistent with the P600. The positivity appears less robust for “first-person subject + third-person verb” errors, as can be seen in Figure 3, which shows the magnitude of the effects. In the same time window where the P600 emerged (~450 ms until the end of the epoch), violations yielded more negative waveforms than their grammatical counterparts in Left Anterior. This negativity appears equally robust for both error types.

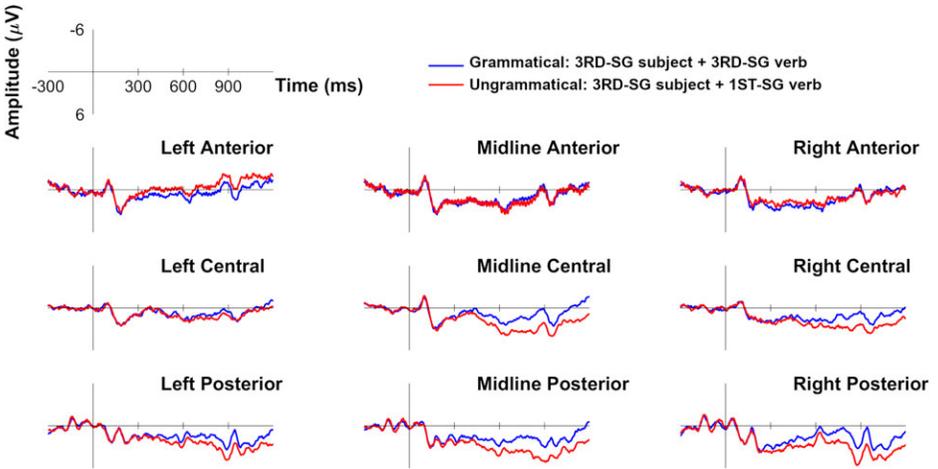


FIGURE 1. Grand average ERP waveforms for the conditions examining person agreement with third-person (unmarked/underspecified) subjects: third-person subject + third-person verb (grammatical) and third-person subject + first-person verb (ungrammatical).

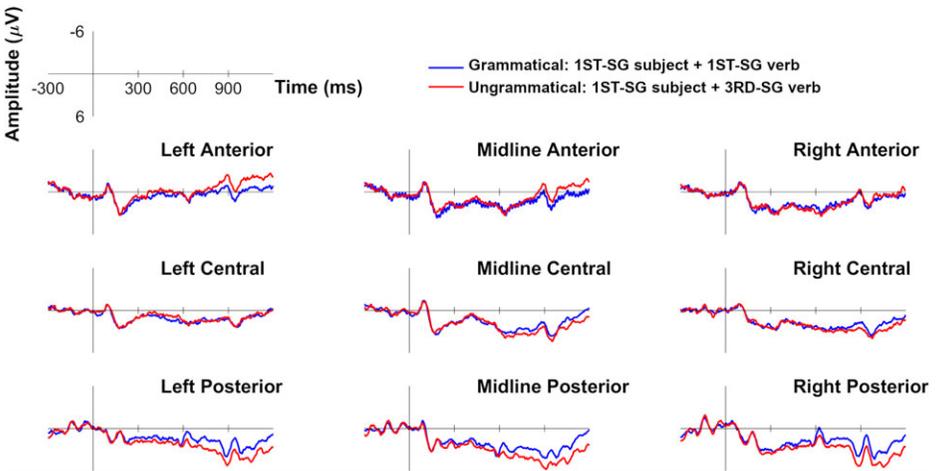


FIGURE 2. Grand average ERP waveforms for the conditions examining person agreement with first-person (marked/specified) subjects: first-person subject + first-person verb (grammatical) and first-person subject + third-person verb (ungrammatical).

**250–450 ms Time Window (LAN/N400)**

Table 3 (middle column) summarizes the results of the omnibus ANOVA in the 250–450 ms time window. In the hemispheres, the Markedness by Agreement by Anterior-Posterior by Hemisphere interaction was significant. As a follow-up, we examined the Markedness by Agreement interaction within each ROI, but no significant interactions emerged in any ROI.

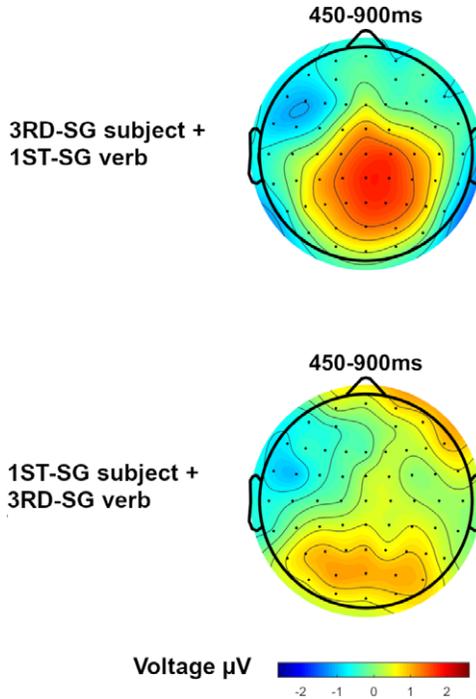


FIGURE 3. Topographic plots for “third-person subject + first-person verb” violations and for “first-person subject + third-person verb” violations in the 450–900 ms time window.

*Note.* Plots were computed by subtracting the grammatical sentence from the violation condition. First and third person correspond to marked/specified and unmarked/underspecified values, respectively.

In the hemispheres, the omnibus ANOVA also revealed a significant Agreement by Anterior-Posterior interaction (Table 3). Therefore, we examined the main effect of Agreement within each level of Anterior-Posterior. The main effect of Agreement was only significant in Posterior,  $F(1, 21) = 10.748, p = .004, q^* = .017; \eta^2 = .339$ , driven by the fact that person violations overall yielded more positive waveforms ( $M = 1.85 \mu\text{V}; SD = 2.11; 95\% \text{ CI } [.92, 2.79]$ ) than grammatical sentences ( $M = 1.41 \mu\text{V}; SD = 2.27; 95\% \text{ CI } [.40, 2.42]$ ). The omnibus ANOVA also revealed a marginal Agreement by Hemisphere interaction, but follow-up tests revealed no effects of Agreement in either hemisphere. Finally, follow-up tests to the Markedness by Anterior-Posterior by Hemisphere interaction revealed no effects of Markedness in any ROI.

In the midline, the omnibus ANOVA revealed a significant Agreement by Anterior-Posterior interaction (Table 3). Follow-up tests revealed a main effect of Agreement in Midline Posterior,  $F(1, 21) = 9.376, p = .006, q^* = .017; \eta^2 = .309$ , driven by the fact that person violations overall were more positive ( $M = 2.23 \mu\text{V}; SD = 2.50; 95\% \text{ CI } [1.13, 3.34]$ ) than grammatical sentences ( $M = 1.67 \mu\text{V}; SD = 2.58; 95\% \text{ CI } [.53, 2.81]$ ).

To summarize, the analyses conducted between 250–450 ms revealed no reliable LAN or N400 effects for either person violation type (see Figures 1–2). These

TABLE 3. Results of the omnibus ANOVA in the 250–450 ms and 450–900 ms time windows

LATERAL REGIONS	250–450 ms	450–950 ms
Marked	$F(1, 21) = .301$ , $p = .589$ ; $\eta p^2 = .014$	$F(1, 21) = .977$ , $p = .334$ ; $\eta p^2 = .044$
Agree	$F(1, 21) = 1.128$ , $p = .300$ ; $\eta p^2 = .051$	$F(1, 21) = 7.242$ , $p = .014$ ; $\eta p^2 = .256$
Marked × Agree	$F(1, 21) = .093$ , $p = .764$ ; $\eta p^2 = .004$	$F(1, 21) = .282$ , $p = .601$ ; $\eta p^2 = .013$
Marked × Ant-Post	$F(1.23, 25.88) = 1.526$ , $p = .233$ ; $\eta p^2 = .068$	$F(1.24, 25.97) = .863$ , $p = .384$ ; $\eta p^2 = .039$
Agree × Ant-Post	$F(1.29, 27.22) = 9.965$ , $p = .002$ ; $\eta p^2 = .322$	$F(1.41, 29.63) = 12.519$ , $p = .0004$ ; $\eta p^2 = .373$
Marked × Agree × Ant-Post	$F(1.58, 33.17) = 1.135$ , $p = .322$ ; $\eta p^2 = .051$	$F(1.56, 32.71) = 2.125$ , $p = .145$ ; $\eta p^2 = .092$
Marked × Hemis	$F(1, 21) = .081$ , $p = .779$ ; $\eta p^2 = .004$	$F(1, 21) = 1.674$ , $p = .21$ ; $\eta p^2 = .074$
Agree × Hemis	$F(1, 21) = 3.327$ , $p = .082$ ; $\eta p^2 = .137$	$F(1, 21) = 7.029$ , $p = .015$ ; $\eta p^2 = .251$
Marked × Agree × Hemis	$F(1, 21) = .110$ , $p = .744$ ; $\eta p^2 = .005$	$F(1, 21) = .160$ , $p = .693$ ; $\eta p^2 = .008$
Marked × Ant-Post × Hemis	$F(2, 42) = 3.671$ , $p = .034$ ; $\eta p^2 = .149$	$F(2, 42) = 1.002$ , $p = .376$ ; $\eta p^2 = .046$
Agree × Ant-Post × Hemis	$F(1.19, 24.92) = .953$ , $p = .354$ ; $\eta p^2 = .043$	$F(1.31, 27.54) = 5.052$ , $p = .024$ ; $\eta p^2 = .194$
Marked × Agree × Ant-Post × Hemis	$F(2, 42) = 8.611$ , $p = .001$ ; $\eta p^2 = .291$	$F(2, 42) = 1.926$ , $p = .158$ ; $\eta p^2 = .084$
MIDLINE REGIONS		
Marked	$F(1, 21) = 1.126$ , $p = .301$ ; $\eta p^2 = .051$	$F(1, 21) = 2.686$ , $p = .116$ ; $\eta p^2 = .113$
Agree	$F(1, 21) = 2.091$ , $p = .163$ ; $\eta p^2 = .091$	$F(1, 21) = 16.965$ , $p = .0004$ ; $\eta p^2 = .447$
Marked × Agree	$F(1, 21) = .304$ , $p = .587$ ; $\eta p^2 = .014$	$F(1, 21) = 3.638$ , $p = .07$ ; $\eta p^2 = .148$
Marked × Ant-Post	$F(1.29, 27.16) = 1.572$ , $p = .225$ ; $\eta p^2 = .07$	$F(1.58, 33.12) = 2.742$ , $p = .09$ ; $\eta p^2 = .115$
Agree × Ant-Post	$F(2, 42) = 6.004$ , $p = .005$ ; $\eta p^2 = .222$	$F(1.29, 27.03) = 6.383$ , $p = .012$ ; $\eta p^2 = .233$
Marked × Agree × Ant-Post	$F(1.31, 27.57) = .360$ , $p = .612$ ; $\eta p^2 = .017$	$F(1.33, 27.98) = 1.202$ , $p = .298$ ; $\eta p^2 = .054$

Marked: Markedness; Agree: Agreement; Ant-Post: Anterior-Posterior; Hemis: Hemisphere  
 Note. Significant/marginal effects are shaded.

analyses do reveal that, by 450 ms, the P600 has already emerged for person errors overall.

#### 450–900 ms Time Window (P600)

**Table 3** (right column) summarizes the results of the omnibus ANOVA between 450–900 ms. In the hemispheres, the Agreement by Hemisphere by Anterior-Posterior interaction was significant. Follow-up tests revealed that the main effect of Agreement was significant in Right Posterior,  $F(1, 21) = 13.590, p = .001, q^* = .006, \eta^2 = .393$ ; Left Posterior,  $F(1, 21) = 12.988, p = .002, q^* = .011, \eta^2 = .382$ ; Right Medial,  $F(1, 21) = 10.972, p = .003, q^* = .017, \eta^2 = .343$ ; and Left Medial,  $F(1, 21) = 6.796, p = .016, q^* = .027, \eta^2 = .244$ . In all central-posterior regions, person violations overall elicited more positive waveforms than grammatical sentences (Right Posterior, violation:  $M = 1.45 \mu\text{V}; SD = 1.66; 95\% \text{ CI } [.71, 2.18]$ ; grammatical:  $M = .70 \mu\text{V}; SD = 1.63; 95\% \text{ CI } [-.02, 1.43]$ ; Left Posterior, violation:  $M = 1.06 \mu\text{V}; SD = 2.10; 95\% \text{ CI } [.13, 1.99]$ ; grammatical:  $M = .40 \mu\text{V}; SD = 1.82; 95\% \text{ CI } [-.40, 1.21]$ ; Right Medial, violation:  $M = 2.45 \mu\text{V}; SD = 1.61; 95\% \text{ CI } [1.73, 3.17]$ ; grammatical:  $M = 1.70 \mu\text{V}; SD = 1.57; 95\% \text{ CI } [1.01, 2.39]$ ; Left Medial, violation:  $M = 1.70 \mu\text{V}; SD = 1.61; 95\% \text{ CI } [.99, 2.42]$ ; grammatical:  $M = 1.34 \mu\text{V}; SD = 1.70; 95\% \text{ CI } [.58, 2.09]$ ). In Left Anterior, the main effect of Agreement was also significant,  $F(1, 21) = 9.021, p = .007, q^* = .022, \eta^2 = .30$ , but here person violations overall were more negative ( $M = .21 \mu\text{V}; SD = 1.93; 95\% \text{ CI } [-.64, 1.07]$ ) than grammatical sentences ( $M = .73 \mu\text{V}; SD = 1.62; 95\% \text{ CI } [.01, 1.45]$ ), which partly explains the three-way interaction.

In the midline, the omnibus ANOVA revealed a marginal Markedness by Agreement interaction ( $p = .07$ ) (**Table 3**). Because this effect is critical to our discussion, we followed up on it by examining the main effect of Agreement within each level of Markedness. The interaction was driven by the fact that “third-person subject + first-person verb” errors yielded more positive waveforms ( $M = 2.84 \mu\text{V}; SD = 2.11; 95\% \text{ CI } [1.91, 3.78]$ ) than grammatical sentences ( $M = 1.54 \mu\text{V}; SD = 1.86; 95\% \text{ CI } [.71, 2.36]$ ), a difference that was significant,  $F(1, 21) = 13.691, p = .001, q^* = .025, \eta^2 = .395$ . For the reverse error type, violations were also more positive ( $M = 2.73 \mu\text{V}; SD = 1.79; 95\% \text{ CI } [1.94, 3.53]$ ) than grammatical sentences ( $M = 2.51 \mu\text{V}; SD = 2.00; 95\% \text{ CI } [1.36, 3.14]$ ), but this difference was only marginal,  $F(1, 21) = 3.687, p = .069, q^* = .05, \eta^2 = .149$ . This is visible in **Figure 3**.

In the midline, the omnibus ANOVA also revealed a significant main effect of Agreement, which was qualified by an interaction with Anterior-Posterior (see **Table 3**). Follow-up tests showed a main effect of Agreement in Midline Posterior,  $F(1, 21) = 19.939, p = .0002, q^* = .017, \eta^2 = .487$  (violation:  $M = 3.24 \mu\text{V}; SD = 1.62; 95\% \text{ CI } [2.52, 3.95]$ ; grammatical:  $M = 1.99 \mu\text{V}; SD = 1.97; 95\% \text{ CI } [1.11, 2.86]$ ) and Midline Medial,  $F(1, 21) = 16.368, p = .001, q^* = .033, \eta^2 = .438$  (violation:  $M = 3.08 \mu\text{V}; SD = 2.28; 95\% \text{ CI } [2.07, 4.09]$ ; grammatical:  $M = 2.02 \mu\text{V}; SD = 2.47; 95\% \text{ CI } [.93, 3.12]$ ).

To summarize, person violations overall yielded a P600 with a central-posterior distribution. Importantly, the P600 was reduced for “first-person subject + third-person verb” violations (in the midline), although the interaction remained marginal. In this time

TABLE 4. Zero-order correlations between the learners' ( $N = 22$ ) overall proficiency in and experience with the L2 (Global\_Proficiency, Instruction, Months\_Abroad) and measures of sensitivity to person agreement (D-prime\_Score, P600\_Magnitude)

	Global_Proficiency	Instruction	Months_Abroad	D-prime Score
D-prime_Score 3rd-person subject + 1st-person verb	$r = .559$ $p = .007$	$r = .144$ $p = .523$	$r = -.429$ $p = .047$	
D-prime_Score 1st-person subject + 3rd-person verb	$r = .598$ $p = .003$	$r = -.023$ $p = .919$	$r = -.296$ $p = .181$	
P600_Magnitude 3rd-person subject + 1st-person verb	$r = -.031$ $p = .890$	$r = -.088$ $p = .697$	$r = .533$ $p = .011$	$r = -.095$ $p = .673$
P600_Magnitude 1st-person subject + 3rd-person verb	$r = .550$ $p = .008$	$r = .125$ $p = .580$	$r = .012$ $p = .959$	$r = .263$ $p = .236$

Global\_Proficiency: Score in standardized Spanish proficiency test;

Instruction: Years of instruction in L2 Spanish;

Months\_Abroad: Months spent in Spanish-speaking countries.

Note. Significant correlations are shaded.

window, a negativity emerged across person violations relative to grammatical sentences in Left Anterior. The P600 was already apparent in the preceding time window (250–450 ms).

## REGRESSION ANALYSIS

### *P600 Magnitude*

We used multiple regression with repeated measures to examine whether overreliance on defaults (reduced sensitivity to “first-person subject + third-person verb” errors) was accounted for by variables related to the learners’ proficiency in and experience with their L2. The dependent variable was P600 magnitude, corresponding to the mean amplitude between 450–900 ms in the ungrammatical minus the grammatical condition in a ROI including all regions where the P600 emerged (Left Medial, Right Medial, Left Posterior, Right Posterior, Midline Medial, and Midline Posterior). P600 magnitude was calculated separately for each person violation type. The two error types correspond to the two levels of the within-subjects predictor, Error\_Type. The between-subjects predictors were Global\_Proficiency (score in a standardized Spanish proficiency test), Instruction (years of instruction in L2 Spanish), and Months\_Abroad (months spent in a Spanish-speaking country). Table 4 shows all zero-order correlations between the variables of interest.

The assumptions of multiple regression (evaluated using the residuals) were met. Table 5 summarizes the results of the regression. The significant Global\_Proficiency by Error\_Type interaction was driven by the fact that Global\_Proficiency significantly predicted P600 magnitude for “first-person subject + third-person verb” violations,  $t(18) = 2.84$ ,  $p = .011$ ,  $q^* = .017$ ,  $\eta^2 = .309$ , an instance of overreliance on defaults, but not for the reverse error (Figure 4, Plots A–B). Examination of the unstandardized

TABLE 5. Results of the multiple regression analysis examining the relationship between the learners' proficiency in and experience with the L2 (Global\_Proficiency, Instruction, Months\_Abroad) and measures of sensitivity to person agreement (P600\_Magnitude, D-Prime\_Score)

Effects	P600_Magnitude	D-Prime_Score
Error_Type	$F(1, 18) = 1.598$ , $p = .222$ , $\eta^2 = .082$	$F(1, 18) = .013$ , $p = .910$ , $\eta^2 = .001$
Global_Proficiency	$F(1, 18) = 2.805$ , $p = .111$ , $\eta^2 = .135$	$F(1, 18) = 15.086$ , $p = .001$ , $\eta^2 = .456$
Instruction	$F(1, 18) = 1.041$ , $p = .321$ , $\eta^2 = .055$	$F(1, 18) = .346$ , $p = .564$ , $\eta^2 = .019$
Months_Abroad	$F(1, 18) = 4.842$ , $p = .041$ , $\eta^2 = .212$	$F(1, 18) = 6.567$ , $p = .020$ , $\eta^2 = .267$
Error_Type by Global_Proficiency	$F(1, 18) = 4.559$ , $p = .047$ , $\eta^2 = .202$	$F(1, 18) = .430$ , $p = .520$ , $\eta^2 = .023$
Error_Type by Instruction	$F(1, 18) = .003$ , $p = .956$ , $\eta^2 = .0002$	$F(1, 18) = .443$ , $p = .514$ , $\eta^2 = .024$
Error_Type by Months_Abroad	$F(1, 18) = 3.747$ , $p = .069$ , $\eta^2 = .172$	$F(1, 18) = 1.057$ , $p = .318$ , $\eta^2 = .055$

Global\_Proficiency: Score in standardized Spanish proficiency test;

Instruction: Years of instruction in L2 Spanish;

Months\_Abroad: Months spent in Spanish-speaking countries;

Error\_Type: Person violation type.

Note. Significant/marginal effects are shaded.

regression coefficient suggests that a one standard deviation increase in Global\_Proficiency (i.e., a 5-point score increase in the proficiency test) results in an estimated .57  $\mu\text{V}$  increase in P600 magnitude (95% CI [.15, .98]) for “first-person subject + third-person verb” errors. Notice also that a few learners have negative values, consistent with the possibility that their sensitivity to the errors was qualitatively different. Thus, the positive value of the regression coefficient suggests development toward increasingly nativelike processing for “first-person subject + third-person verb” errors, both qualitatively and quantitatively.

The main effect of Months\_Abroad was also significant, driven by the fact that P600 magnitude overall tended to be larger for learners with longer immersion time. The estimated increase in P600 magnitude for every one standard deviation increase in Months\_Abroad (i.e., a 13-month increase in immersion time) was .37  $\mu\text{V}$  (95% CI [.02, .73]). However, this effect was qualified by a marginal interaction with Error\_Type (Table 5). As Plots A and B of Figure 5 reveal, this interaction was driven by the fact that Months\_Abroad predicted P600 magnitude for “third-person subject + first-person verb” errors,  $t(18) = 2.782$ ,  $p = .012$ ,  $q^* = .017$ ,  $\eta^2 = .301$  (estimated increase in P600 magnitude for every one standard deviation increase in Months\_Abroad = .68  $\mu\text{V}$ ; 95% CI [.17, 1.19]), but not for the other error.

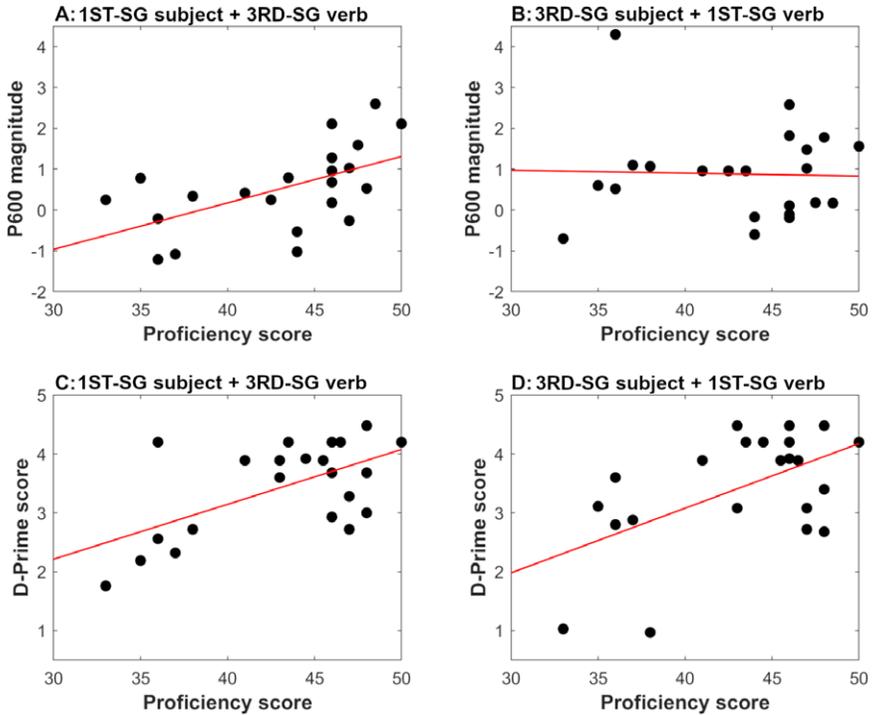


FIGURE 4. Scatterplots showing the relation between the learners’ global proficiency (score in a standardized Spanish proficiency test) and their sensitivity to person agreement both in terms of P600 magnitude (Plots A and B) and in terms of behavioral accuracy (Plots C and D).  
*Note.* P600 effect size was calculated by subtracting the grammatical from the ungrammatical condition. Effects were averaged across all regions where P600 effects emerged for both types of person errors. Behavioral accuracy was operationalized as D-prime Score for each type of person dependency in the GJT. Each dot represents a data point from a single learner. The dashed line represents the best-fit regression line. Minimal jitter has been added to make learners with identical or near identical values visible. Marked subject = first-person; Unmarked subject = third-person.

**D-Prime Scores**

A similar approach was undertaken to examine the relationship between the learners’ behavioral sensitivity to person agreement and their proficiency in and experience with L2 Spanish. The dependent variable was D-Prime Score. All other aspects of the analysis were held constant. Zero-order correlations and a summary of the effects are provided in Tables 4 and 5, respectively. The assumptions of multiple regression were met. The main effect of Global\_Proficiency was significant, with more proficient L2ers showing higher D-Prime scores overall (Figure 4, Plots C–D). The estimated increase in D-Prime\_Score for every one standard deviation increase in Global\_Proficiency (i.e., a 5-point score increase in the proficiency test) was .52 (95% CI [.24, .80]). The main effect of Months\_Abroad was also significant. Learners with longer immersion time showed lower D-Prime scores (i.e., lower accuracy) (Figure 5, Plots C–D). The estimated decrease in

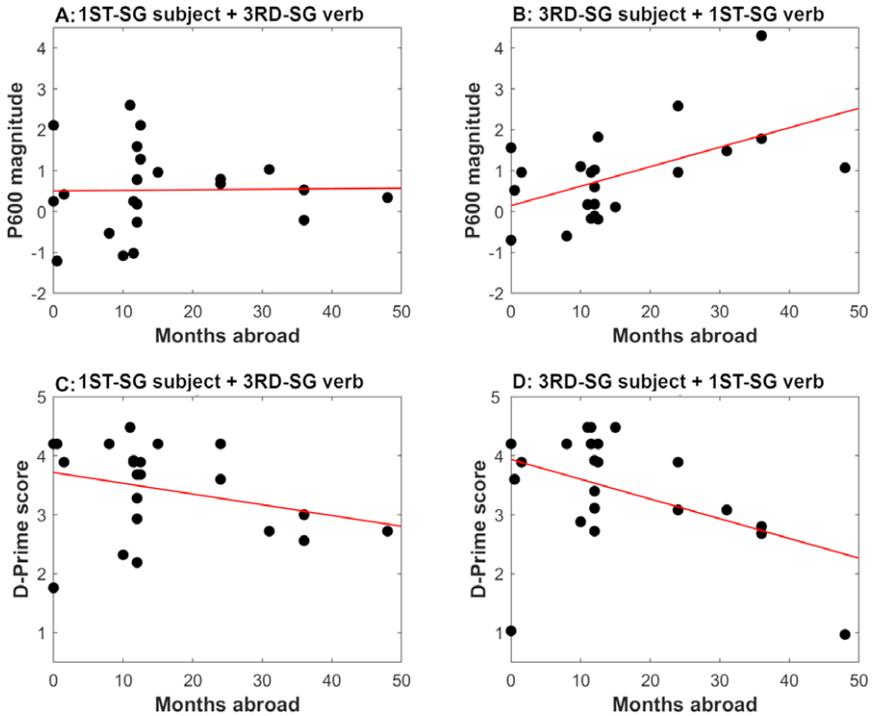


FIGURE 5. Scatterplots showing the relation between the learners' immersion time (number of months spent in Spanish-speaking countries) and their sensitivity to person agreement both in terms of P600 magnitude (Plots A and B) and in terms of behavioral accuracy (Plots C and D).

*Note.* P600 effect size was calculated by subtracting the grammatical from the ungrammatical condition. Effects were averaged across all regions where P600 effects emerged for both types of person errors. Behavioral accuracy was operationalized as D-prime Score for each type of person dependency in the GJT. Each dot represents a data point from a single learner. The dashed line represents the best-fit regression line. Minimal jitter has been added to make learners with identical or near identical values visible. Marked subject = first-person; Unmarked subject = third-person.

D-Prime\_Score for every one standard deviation increase in Months\_Abroad (i.e., a 13-month increase in immersion time) was  $-.38$  (95% CI  $[-.68, -.07]$ ). We come back to this surprising finding in the "Discussion" section.

## DISCUSSION

The present study investigated the role of markedness in the processing of subject-verb person agreement in Spanish by upper-intermediate/advanced English-speaking learners. Recall that our main aim was to adjudicate between different proposals regarding how markedness impacts L2 agreement resolution. L2 theoretical models posit that learners overuse default morphology (e.g., third-person verbal inflection) in contexts that require marked/specified forms (e.g., first-person subjects) due either to computational pressure (Prévost & White, 2000) or to the asymmetrical representation of features in the learner's

grammar (e.g., McCarthy, 2008, 2012). Alternative proposals from the psycholinguistics literature suggest that marked features remain longer in the focus of attention and, thus, facilitate agreement operations, for example, by recruiting top-down mechanisms to resolve agreement (Nevins et al., 2007; Wagers & McElree, 2011; see López Prego, 2015). Under the latter proposal, feature activation upon encountering a marked/specified feature (e.g., a first-person subject) allows the parser to better resolve agreement.

To that end, we probed subject-verb person agreement with both first-person (marked/specified as *speaker*) and third-person (unmarked/underspecified) subjects (e.g., Bianchi, 2006; Den Dikken, 2011; Harley & Ritter, 2002; Harris, 1995; Jakobson, 1971; McGinnis, 2005; Nevins, 2011). By crossing each subject type with verbs inflected for the opposite person, we created two error types for which the previously mentioned proposals make different predictions. The learners' brain responses revealed reduced sensitivity (i.e., a reduced P600) to "first-person subject + third-person verb" errors, relative to the opposite error type, although this effect remained marginal in the ANOVA. Recall however that, in the regression analyses, Proficiency interacted with Error\_Type (i.e., P600 magnitude as a function of error type). This analysis showed that the reduced sensitivity to "first-person subject + third-person verb" errors characterized the less proficient learners in the sample. These data suggest that, in the course of online processing, learners tolerated unmarked/underspecified forms (third-person inflection) in contexts that required marked/specified ones (first-person subject), mainly at intermediate levels of proficiency. Importantly, the fact that no asymmetry emerged in the end-of-the-sentence GJT, for which learners took as much time as they needed, indicates that such overreliance on defaults most likely results from computational burden (e.g., Alemán Bañón et al., 2017; Hopp, 2010; López Prego & Gabriele, 2014; McDonald, 2006; Prévost & White, 2000). That is, while the learners' brain might not have detected "first-person subject + third-person verb" errors exactly at the time when the violating verb was presented, it is possible that the learners detected the agreement error by the time they provided the grammaticality judgment. The learners also elicited a Late Anterior Negativity (similar to the native speakers) across both error types. This component has been argued to reflect the costs associated with keeping the violations in working memory for the GJT (e.g., Alemán Bañón et al., 2012; Gillon-Dowens et al., 2010; Sabourin & Stowe, 2004). This might explain why this component was unimpacted by markedness because the learners were equally accurate detecting both error types.

Recall that the same learners were sensitive to markedness asymmetries in the processing of noun-adjective number and gender agreement (Alemán Bañón et al., 2017). In that study, number violations on plural adjectives (marked for number) yielded a marginally larger P600 than number errors on singular adjectives. In addition, the P600 emerged earlier for gender errors realized on feminine (marked for gender) relative to masculine adjectives. Crucially, similar effects emerged in the L1-Spanish controls (Alemán Bañón & Rothman, 2016), suggesting that the learners' processing profile was qualitatively nativelike, at least for noun-adjective number/gender. The overall picture that emerges from these studies is that learners are sensitive to markedness asymmetries at the point when the agreement dependency is resolved (the adjective for noun-adjective agreement; the verb for subject-verb agreement), and that such sensitivity can also characterize native speaker processing (e.g., Alemán Bañón & Rothman, 2016; López Prego & Gabriele, 2014).<sup>6</sup>

We also asked whether learners eventually abandon their reliance on default morphology with development. Our regression analysis suggests so. As Figure 4 shows, more proficient learners showed larger P600 effects for “first-person subject + third-person verb” violations (Plot A), but not for “third-person subject + first-person verb” errors (Plot B). One possibility is that, at the upper-intermediate level of proficiency (i.e., the lower bound of the proficiency range examined herein), the learners’ grammar still allows first-person subjects with third-person inflection, which still functions as a sort of all-purpose form. However, the learners’ grammar already rules out “third-person subject + first-person verb” configurations. This aligns with McCarthy’s analysis (2012) of corpus production data, which showed that less proficient learners supplied third-person verbs with first-person subjects, but rarely did the reverse. Plots A–B from Figure 4 reveal that P600 effects were larger for “third-person subject + first-person verb” errors overall (Plot B). Twelve learners show a P600 of approximately 1  $\mu$ V or larger for this error type, and these learners are scattered across the proficiency spectrum examined. In contrast, fewer learners (seven) show P600 effects of approximately similar size for “first-person subject + third-person verb” violations (Plot A), and all of them scored in the upper range of the advanced level. In addition, although we see negative effects for both error types, negativities tend to be larger for “first-person subject + third-person verb” errors, and the largest negativities for this error type are associated with learners near the lower end of the proficiency range examined. It is thus possible that lower proficiency learners rely on qualitatively different processing mechanisms for harder dependencies (e.g., Carrasco-Ortiz et al., 2017; Osterhout et al., 2006; Tanner et al., 2014) although the small number of negative responders in our sample precludes strong conclusions. That global proficiency interacted with error type in the ERP data is consistent with the claim we made in Alemán Bañón et al. (2017) that markedness impacts L2 processing without constraining it. Overreliance on default forms (i.e., reduced sensitivity to “first-person subject + third-person verb” errors) might characterize the intermediate levels of proficiency, but is progressively abandoned at higher ones.

Figure 4 (Plots C–D) shows that higher proficiency learners also tended to show higher D-Prime scores for both person violation types, providing additional evidence for development (but not markedness) (e.g., Alemán Bañón et al., 2018). The reader might wonder why global proficiency interacted with markedness in the ERP data, but not in the D-Prime data. The two metrics probably tap into different types of sensitivity. While P600 magnitude is a measure of brain sensitivity to person dependencies exactly at the time when they are established, D-Prime scores provide a measure of sensitivity to the same dependencies once the learner has read the whole sentence. That D-Prime scores and P600 magnitude did not significantly correlate for either type of person dependency is consistent with this possibility (see Table 4). It is thus conceivable that proficiency modulated both types of sensitivity differently. Given what is required of a judgment task (i.e., detecting the ungrammaticality, maintaining decisions about grammaticality in working memory until the end of the sentence, wrapping up sentence meaning), a lower level of proficiency might have impacted accuracy with both types of person dependencies.

Therefore, the response to RQ1 and RQ2 (*Does markedness impact subject-verb person agreement resolution in the L2? To what extent does L2 proficiency account for a potential overreliance on default morphology?*) is that L2 processing is not constrained

by markedness, but is sensitive to it, particularly among less advanced learners. Because markedness did not impact the learners' accuracy while judging the sentences at a later point and with no time pressure, sensitivity to markedness is more likely to be computational.

We now turn to the question of why these learners, who were qualitatively nativelike with noun-adjective number and gender agreement, showed a qualitatively different processing profile from native speakers for subject-verb person agreement.<sup>7</sup> One possibility is that, although the learners could successfully resolve person agreement at the verb, the markedness of the subject (i.e., the first element in the dependency) did not facilitate agreement. The question of whether L2 learners can use linguistic cues to facilitate integration of the bottom-up input has played a central role in recent L2 processing research (see Kaan, 2014 for a review). While some studies have argued that L2 learners, even advanced ones, fail to use lexical, morphosyntactic, syntactic, or discourse cues predictively (e.g., Grüter et al., 2012, 2017; Kaan et al., 2016; Martin et al., 2013), others have claimed that predictive processing is similar in the L1 and L2 (e.g., Kaan, 2014), but is modulated by proficiency (e.g., Dussias et al., 2013), the strength of lexical representations (e.g., Hopp, 2013), L1-L2 similarity (e.g., Foucart et al., 2014; Hopp & Lemmerth, 2018), and individual differences in cognitive factors (e.g., Hopp, 2013). It is still unclear, however, which cues learners can use predictively and which linguistic representations they can activate because most of the evidence supporting prediction in the L2 comes from studies manipulating overt gender cues (Dussias et al., 2013; Foucart et al., 2014; Hopp, 2013; López Prego, 2015). Thus, it is possible that, unlike gender cues, *speech participant* status was insufficient to facilitate agreement at the verb. So, the answer to RQ3 (*Do L2 learners use person markedness information to ease agreement resolution?*) is preliminarily “no,” based on the evidence provided herein.

It could be argued that the larger P600 for “third-person subject + first-person verb” errors does not reflect sensitivity to markedness, but rather facilitation from L1 English, which instantiates agreement with third-person singular subjects (e.g., Jiang, 2004, 2007; Lardiere, 2009; Tokowicz & MacWhinney, 2005). First, we point out that using plural subjects would not have worked because third-person plural DPs in Spanish agree with first- and second-person plural verbs, a process called *unagreement* (e.g., *Las viudas lloramos/lloráis* “[we/you-2ND-PL] the widows cry-1ST-PL/cry-2ND-PL”) (e.g., Höhn, 2016; Hurtado, 1985), and previous work by Mancini et al. (2011, 2019) has shown that native speakers treat these sentences differently from outright person violations (i.e., they do not elicit a P600). Most importantly, while previous ERP studies manipulating L1-L2 similarity have consistently shown an advantage for L2 features instantiated in the L1 (e.g., Alemán Bañón et al., 2014; Bond et al., 2011; Gabriele et al., *in press*; Gillon-Dowens et al., 2010, 2011) they have consistently found no advantage for shared morphological instantiations of common features (e.g., Alemán Bañón et al., 2014; Bond et al., 2011; Gabriele et al., *in press*). For example, Bond et al. (2011) and Gabriele et al. (*in press*) examined L1-English L2-Spanish learners' brain responses to subject-verb number agreement with third-person singular subjects (instantiated in English) and noun-adjective number agreement (unique to Spanish), and found no facilitation for the former. Other studies have even found a disadvantage for contexts where the L1 overtly marks agreement. For example, Alemán Bañón et al. (2014) examined demonstrative-noun and noun-adjective number agreement in Spanish (e.g., *este apartamento/\*apartamentos*

“this apartment/apartments”; *órgano muy complejo/\*complejos* “organ-SG very complex-SG/complex-PL”) among advanced English-speaking learners. Crucially, English instantiates number on demonstratives, but not on adjectives. Their results showed a larger P600 for number violations on adjectives, an effect that can be explained by markedness (i.e., *este* was unmarked, but *complejos* was marked) or differences in syntactic category, but crucially not by transfer (e.g., Tokowicz & MacWhinney, 2005). Thus, it is unlikely that the effects herein reflect L1 facilitation, especially because the same learners showed enhanced sensitivity to gender errors realized on marked adjectives (Alemán Bañón et al., 2017), an effect that cannot come from the L1.

Before concluding, we address the role of immersion in L2 morphosyntactic development, which showed inconsistency across measures. While Plot B of Figure 5 suggests that longer immersion results in greater brain sensitivity to “third-person subject + first-person verb” errors, Plots C–D reveal lower D-Prime scores (i.e., less sensitivity) to both error types for learners with similar immersion time. We can think of no reason why immersion in an L2-speaking environment would result in poorer ability to resolve any type of morphosyntactic dependency. Importantly, immersion represents an indirect measure of proficiency. Learners with longer immersion time are assumed to have benefited from richer input, alongside increased opportunities for output. In turn, this is expected to promote morphosyntactic development. However, this might simply not be the case. In fact, previous studies on morphosyntactic development do not consistently report advantages for learners in study-abroad programs relative to “at home” learners (Faretta-Stutenberg & Morgan-Short, 2018, p. 5). In our study, the two learners with the longest immersion time (48 and 36 months, respectively) had mainly lived in English-speaking communities and they both scored in the intermediate range in the standardized test (38 and 36, respectively). Both of them showed D-Prime scores roughly one or two standard deviations below the mean. Likewise, one of the learners with the highest D-Prime scores (approximately one standard deviation above the mean) had never resided in Spanish-speaking countries. However, this learner had lived immersed in a Spanish-speaking community for two years in the United States and benefited from native speaker input. It is, therefore, possible that, being an indirect measure of proficiency, immersion time is a noisier predictor of grammatical development. Another possibility is that there was not sufficient variability in the sample for reliable relations to emerge because half of the learners had an immersion time of ~10 months.

## CONCLUSION

The present study found that a group of 22 L1-English L2-Spanish learners showed brain sensitivity (i.e., a P600) across two types of subject-verb person dependencies that differed with respect to markedness. The learners were marginally less sensitive to violations where the subject corresponded to a marked/specified person (first person: *speaker*) and the verb was unmarked/underspecified for person (third person), suggesting some overreliance on default morphology (i.e., third-person inflection). Markedness did not impact the learners’ ability to detect the same violations in an end-of-the-sentence GJT, for which no time constraints were imposed, suggesting that overreliance on defaults is computational. Regression analyses showed that L2ers gradually abandon this overreliance on defaults with increased proficiency. Finally, unlike what we found in the

Spanish controls (Alemán Bañón & Rothman, 2019), the speech participant status of the subject (i.e., *speaker* vs. the default person) did not ease agreement resolution among learners.

## NOTES

<sup>1</sup>See Battistella (1990) for a list of criteria to determine markedness asymmetries.

<sup>2</sup>Between ~300–500 ms, the P600 is sometimes preceded by a Left Anterior Negativity (e.g., Rossi et al., 2005), associated with morphosyntactic processing, or an N400 (e.g., Mancini et al., 2011, 2019; Zawiszewsky et al., 2016), associated with lexical-semantic operations. Because no negativities emerged in Alemán Bañón and Rothman (2019) for either person error type, we focus on the P600.

<sup>3</sup>Because we examined agreement violations, our study cannot dissociate prediction from integration effects (see Kuperberg & Jaeger, 2016). We reasoned, however, that only “first-person subject + third-person verb” violations would contribute to top-down processing of person information because third-person subjects lack person information. In Alemán Bañón and Rothman (2019), we did not argue that the larger P600 for violations with first-person subjects reflected prediction disconfirmation, but rather the reanalysis cost engendered by an unmet prediction.

<sup>4</sup>The 12 counterbalanced lists and the instructions to the task can be downloaded from the IRIS database at <https://www.iris-database.org/iris/app/home/index>

<sup>5</sup>Because electrode TP10 was noisy in several recordings, we re-referenced them to the average of TP7/TP8 (e.g., Simor et al., 2019).

<sup>6</sup>Although here we focused on comprehension, our design includes an elicited production task, which revealed virtually no variability.

<sup>7</sup>Because our learners’ proficiency ranged from intermediate to advanced, we analyzed them separately from native speakers. An ANOVA with Error\_Type as the repeated factor and Group (natives, learners) as the between-subjects factor showed a significant Error\_Type by Group interaction,  $F(1, 48) = 4.359, p = .042, \eta^2 = .083$ , with learners showing a reduced P600 for “first-person subject + third-person verb” errors relative to native speakers,  $F(1, 48) = 6.672, p = .013, q^* = .033, \eta^2 = .122$ . P600 was calculated for one region including 16 central-posterior electrodes, between 500–1,000 ms for natives and 450–900 ms for L2ers).

## REFERENCES

- Acuña Fariña, J. C., Meseguer, E., & Carreiras, M. (2014). Gender and number agreement in comprehension in Spanish. *Lingua, 143*, 108–128.
- 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.
- Alemán Bañón, J., Fiorentino, R., & Gabriele, A. (2014). Morphosyntactic processing in advanced second language (L2) learners: An event-related potential investigation of the effects of L1-L2 similarity and structural distance. *Second Language Research, 30*, 275–306.
- 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*, e0200791. <https://doi.org/10.1371/journal.pone.0200791>.
- 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*, 1509–1536.
- Alemán Bañón, J., & Rothman, J. (2016). The role of morphological markedness in the processing of number and gender agreement in Spanish: An event-related potential investigation. *Language, Cognition and Neuroscience, 31*, 1273–1298.
- Alemán Bañón, J., & Rothman, J. (2019). Being a *participant* matters: Event-related potentials show that markedness modulates person agreement in Spanish. *Frontiers in Psychology, 10*, 746. <https://doi.org/10.3389/fpsyg.2019.00746>.
- Battistella, E. (1990). *Markedness: The evaluative superstructure of language*. SUNY Press.

- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57, 289–300.
- Bianchi, V. (2006). On the syntax of personal arguments. *Lingua*, 116, 2023–2067.
- 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 D. Tanner & J. Herschensohn (Eds.), *Proceedings of the 11th generative approaches to second language acquisition conference* (pp. 17–29). Cascadilla Press.
- Bornkessel-Schlesewsky, I., & Schlewsky, M. (2008). An alternative perspective on “semantic P600” effects in language comprehension. *Brain Research Reviews*, 59, 55–73.
- Bowden, H. W., Steinhauer, K., Sanz, C., & Ullman, M. T. (2013). Native-like brain processing of syntax can be attained by university foreign language learners. *Neuropsychologia*, 51, 2492–2511.
- Cabrelli Amaro, J., Campos, G., & Rothman, J. (2018). The role of L1 phonology in L2 morphological production: L2 English past tense agreement from Spanish, Mandarin, and Japanese speakers. *Studies in Second Language Acquisition*, 40, 503–527.
- 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.
- Carminati, M. N. (2005). Processing reflexes of the feature hierarchy (Person > Number > Gender) and implications for linguistic theory. *Lingua*, 115, 259–285.
- Carrasco-Ortiz, H., & Frenck-Mestre, C. (2014). Phonological and orthographic cues enhance the processing of inflectional morphology. ERP evidence from L1 and L2 French. *Frontiers in Psychology*, 5, 888. <https://doi.org/10.3389/fpsyg.2014.00888>.
- Carrasco-Ortiz, H., Velázquez Herrera, A., Jackson-Maldonado, D., Avecilla Ramírez, G. N., Silva Pereyra, J., & Wicha, N. Y. Y. (2017). The role of language similarity in processing second language morphosyntax: Evidence from ERPs. *International Journal of Psychophysiology*, 117, 91–110.
- Chomsky, N. (1957). *Syntactic structures*. Mouton.
- Chomsky, N. (1995). *The minimalist program*. The MIT Press.
- DeLuca, V., Rothman, J., & Pliatsikas, C. (2019). Linguistic immersion and structural effects on the bilingual brain: A longitudinal study. *Bilingualism: Language and Cognition*, 22, 1160–1175.
- Den Dikken, M. (2011). Phi-feature inflection and agreement: An introduction. *Natural Language & Linguistic Theory*, 29, 857–874.
- Dillon, B., Mischler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, 69, 85–103.
- Duchon, A., Perea, M., Sebastián Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods*, 45, 1246–1258.
- Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition*, 35, 353–387.
- Faretta-Stutenberg, M., & Morgan-Short, K. (2018). The interplay of individual differences and context of learning in behavioral and neurocognitive second language development. *Second Language Research*, 34, 67–101.
- Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). SAGE Publications.
- Forchheimer, P. (1953). *The category of person in language*. Walter de Gruyter.
- 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, 226–248.
- Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 1461–1469.
- Franceschina, F. (2005). *Fossilized second language grammars: The acquisition of grammatical gender*. John Benjamins.
- Gabriele, A., Alemán Bañón, J., Hoffman, L., Covey, L., Rossomondo, A., & Fiorentino, F. (in press). Linguistic and cognitive abilities facilitate the processing of agreement in novice learners: Evidence from event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Gillon-Downs, 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, 1651–1659.

- Gillon-Dowens, M., Vergara, M., Barber, H., & Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience*, 22, 1870–1887.
- Goad, H., & White, L. (2006). Ultimate attainment in interlanguage grammars: A prosodic approach. *Second Language Research*, 22, 243–268.
- Goad, H., & White, L. (2019). Prosodic effects on L2 grammars. *Linguistic Approaches to Bilingualism*, 9, 769–808.
- Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, 28, 191–215.
- Grüter, T., Rohde, H., & Schafer, A. (2017). Coreference and discourse coherence in L2: The roles of grammatical aspect and referential form. *Linguistic Approaches to Bilingualism*, 7, 199–229.
- Harley, H., & Ritter, E. (2002). Person and number in pronouns: A feature-geometric analysis. *Language*, 78, 482–526.
- Harris, J. (1995). The morphology of Spanish clitics. In H. Campos (Ed.), *Evolution and revolution in linguistic theory: Essays in honor of Carlos Otero* (pp. 168–197). Georgetown University Press.
- Hawkins, R. (2001). The theoretical significance of universal grammar in SLA. *Second Language Research*, 17, 345–367.
- Hawkins, R. (2009). Statistical learning and innate knowledge in the development of second language proficiency: Evidence from the acquisition of gender concord. In A. G. Benati (ed.), *Issues in second language proficiency* (pp. 63–78). Continuum.
- Hawkins, R., & Chan, C. (1997). The partial availability of universal grammar in second language acquisition: The “failed functional features hypothesis.” *Second Language Research*, 13, 187–226.
- Haznedar, B., & Schwartz, B. D. (1997). Are there optional infinitives in child L2 acquisition? In E. Hughes, M. Hughes, & A. Greenhill (Eds.), *Proceedings of the 21st annual Boston University conference on language development* (pp. 257–268). Cascadilla Press.
- Höhn, G. F. (2016). Unagreement is an illusion: Apparent person mismatches and nominal structure. *Natural Language & Linguistic Theory*, 34, 543–592.
- Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, 120, 901–931.
- Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, 29, 33–56.
- Hopp, H., & Lemmerth, N. (2018). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition*, 40, 171–199.
- Hurtado, A. (1985). The unagreement hypothesis. In L. King & C. Maley (Eds.), *Selected papers from the Thirteenth Linguistic Symposium on Romance Languages* (pp. 187–211). John Benjamins.
- Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). How robust are prediction effects in language comprehension? Failure to replicate article-elicited N400 effects. *Language, Cognition and Neuroscience*, 32, 954–965.
- Jakobson, R. (1971). Shifters, verbal categories, and the Russian verb. In R. Jakobson (ed.), *Selected writings: Word and language* (Vol. 2, pp. 130–147). Mouton.
- Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, 25, 603–634.
- Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning*, 57, 1–33.
- Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism*, 4, 257–282.
- Kaan, E., Kirkham, J., & Wijnen, F. (2016). Prediction and integration in native and second-language processing of elliptical structures. *Bilingualism: Language and Cognition*, 19, 1–18.
- Keating, G. D. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning*, 59, 503–535.
- Keating, G. D. (2010). The effects of linear distance and working memory on the processing of gender agreement in Spanish. In B. Van Patten & J. Jegerski (Eds.), *Research in second language processing and parsing* (pp. 113–134). John Benjamins.
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146, 23–49.
- Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, cognition and neuroscience*, 31, 32–59.

- Lardiere, D. (1998). Case and tense in a fossilized steady state. *Second Language Research*, 14, 1–26.
- Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research*, 25, 173–227.
- López Prego, B. (2015). *The online use of markedness information in L1 and L2 Spanish gender agreement* (PhD dissertation). University of Kansas.
- 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, 192–221.
- Luck, S. (2014). *An introduction to the event-related potential technique* (2nd ed.). MIT Press.
- Mancini, S., Massol, S., Duñabeitia, J. A., Carreiras, M., & Molinaro, N. (2019). Agreement and illusion of disagreement: An ERP study on Basque. *Cortex*, 116, 154–167.
- Mancini, S., Molinaro, N., Rizzi, L., & Carreiras, M. (2011). A person is not a number: Discourse involvement in subject-verb agreement computation. *Brain Research*, 1410, 64–76.
- Martin, C. D., Thierry, G., Kuipers, J. R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, 69, 574–588.
- McCarthy, C. (2008). Morphological variability in the comprehension of agreement: An argument for representation over computation. *Second Language Research*, 24, 459–486.
- McCarthy, C. (2012). Modeling morphological variation and development. *Linguistic approaches to Bilingualism*, 2, 25–53.
- McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, 55, 381–401.
- McGinnis, M. (2005). On markedness asymmetries in person and number. *Language*, 81, 699–718.
- Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, 47, 908–930.
- Nevins, A. (2011). Multiple agree with clitics: Person complementarity vs. omnivorous number. *Natural Language & Linguistic Theory*, 25, 939–971.
- Nevins, A., Dillon, B., Malhotra, S., & Phillips, C. (2007). The role of feature-number and feature-type in processing Hindi verb agreement violations. *Brain Research*, 1164, 81–94.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9, 97–113.
- Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785–806.
- 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.
- Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, 41, 427–456.
- Pienemann, M. (2015). An outline of processability theory and its relationship to other approaches to SLA. *Language Learning*, 65, 123–151.
- Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, 16, 103–133.
- Rossi, E., Kroll, J. F., & Dussias, P. E. (2014). Clitic pronouns reveal the time course of processing gender and number in a second language. *Neuropsychologia*, 62, 11–25.
- Rossi, S., Gugler, M. F., Friederici, A., & Hahne, A. (2006). The impact of proficiency on syntactic second language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, 18, 2030–2048.
- Rossi, S., Gugler, M. F., Hahne, A., & Friederici, A. D. (2005). When word category information encounters morphosyntax: An ERP study. *Neuroscience Letters*, 384, 228–233.
- Rothman, J. (2007). Sometimes they use it, sometimes they don't: An epistemological discussion of L2 morphological production and its use as a competence measurement. *Applied Linguistics*, 28, 609–614.
- Sabourin, L., & Stowe, L. A. (2004). Memory effects in syntactic ERP tasks. *Brain and Cognition*, 55, 392–395.
- Schneider, B. A., Avivi-Reich, M., & Mindaugas, M. (2015). A cautionary note on the use of the analysis of covariance (ANCOVA) in classification designs with and without within-subject factors. *Frontiers in Psychology*, 6, 474. <https://doi.org/10.3389/fpsyg.2015.00474>.

- Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, 12, 40–72.
- Silva Pereyra, J. F., & Carreiras, M. (2007). An ERP study of agreement features in Spanish. *Brain Research*, 1185, 201–211.
- Simor, P., van Der Wijk, G., Gombos, F., & Kovács, I. (2019). The paradox of rapid eye movement sleep in the light of oscillatory activity and functional synchronization during phasic and tonic microstates. *NeuroImage*, 202, 116066. <https://doi.org/10.1016/j.neuroimage.2019.116066>
- Slabakova, R. (2018). Inflectional morphology. In P. Malovrh & A. Benati (Eds.), *The Handbook of advanced proficiency in second language acquisition* (1st ed., pp. 381–400). John Wiley & Sons.
- Tagliaferri, B. (2005). *Paradigm*. Perception Research Systems. [www.perceptionresearchsystems.com](http://www.perceptionresearchsystems.com)
- Tanner, D., Grey, S., & van Hell, J. G. (2017). Dissociating retrieval interference and reanalysis in the P600 during sentence comprehension. *Psychophysiology*, 54, 248–259.
- Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition*, 17, 277–293.
- Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition*, 16, 367–382.
- Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition*, 27, 173–204.
- Tsimpli, I. M., & Dimitrakopoulou, M. (2007). The interpretability hypothesis: Evidence from *wh*-interrogatives in second language acquisition. *Second Language Research*, 23, 215–242.
- van de Meerendonk, N., Kolk, H. H., Vissers, C. T., & Chwilla, D. J. (2010). Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, 22, 67–82.
- Wagers, M. W., Lau, E., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61, 206–237.
- Wagers, M. W., & McElree, B. (2011). *Unpublished manuscript*. University of California Santa Cruz and New York University.
- Wagers, M. W., & Phillips, C. (2014). Going the distance: Memory and control processes in active dependency construction. *Quarterly Journal of Experimental Psychology*, 67, 1274–1304.
- White, L. (2011). Second language acquisition at the interfaces. *Lingua*, 121, 577–590.
- White, L., Valenzuela, E., Kozłowska Macgregor, M., & Leung, Y. K. I. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics*, 25, 105–133.
- Wlotko, E. W., & Federmeier, K. D. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex*, 68, 20–32.
- Zawiszewsky, A., Santesteban, M., & Laka, I. (2016). Phi-features reloaded: An event-related potential study on person and number agreement processing. *Applied Psycholinguistics*, 37, 601–626.