

The role of morphological markedness in the processing of number and gender agreement in Spanish: an event-related potential investigation

Article

Accepted Version

Aleman Banon, J. and 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 (10). pp. 1273-1298. ISSN 2327-3801 doi: 10.1080/23273798.2016.1218032 Available at <https://centaur.reading.ac.uk/66512/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1080/23273798.2016.1218032>

Publisher: Taylor & Francis

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

The Role of Morphological Markedness in the Processing of Number and Gender Agreement
in Spanish: An Event-Related Potential Investigation

José Alemán Bañón^a (First name: José; Last Name: Alemán Bañón)

Jason Rothman^{b,c}

^a: Basque Center on Cognition, Brain and Language

Paseo Mikeletegi 69

San Sebastián, Guipúzcoa, 20009, Spain

^b: University of Reading, School of Psychology and Clinical Language Sciences,

Harry Pitt Building, Earley Gate

Reading, Berkshire, RG6 7BE, UK

^c: The Arctic University of Norway, Department of Languages and Linguistics

Faculty of Humanities, Social Sciences, and Education

9037, Tromsø, Norway

Corresponding author: José Alemán Bañón

Email Address: j.aleman@bcbl.eu

Basque Center on Cognition, Brain and Language

Paseo Mikeletegi, 69

San Sebastián, Guipúzcoa, 20009, Spain

Phone: +44 (0) 1183788135/7466

Fax: +44 (0) 1189753365

Abstract

Current morphological theory assumes that feature values, such as masculine and feminine or singular and plural, are asymmetrically represented. That is, one member of the opposition (e.g. feminine for gender, plural for number) is assumed to be marked, and the other one, unmarked. The present study examines how these asymmetries impact agreement resolution in Spanish. Agreement was manipulated between a noun acting as head of a relative clause and an adjective located inside the relative clause (e.g. *catedral que parecía inmensa* “cathedral that looked huge”). Half of the nouns were feminine (marked) and the other half, masculine (unmarked). Half of the nouns were used in the plural (marked) and the other half, in the singular (unmarked). Twenty-seven Spanish native speakers read 240 sentences while their brain activity was recorded with EEG and performed a grammaticality judgment. Results showed that both number and gender violations elicited a central-posterior P600, a component associated with syntactic repair, and a late anterior negativity, argued to reflect working memory costs. Only the P600 was affected by markedness. It started earlier for violations where the mismatching feature was marked. Moreover, it was larger for errors where the mismatching feature was marked, although this amplitude modulation only emerged for number, possibly due to differences in how number and gender cues were realized (i.e. both masculine and feminine showed overt inflection, but singular was uninflected relative to plural). These results suggest that the parser is sensitive to markedness asymmetries in the course of online processing.

Keywords: P600; markedness; agreement; number and gender

1. Introduction

Current morphological theory posits that agreement features, such as number and gender, are organized following a markedness hierarchy (e.g. Battistella, 1990; Bonet, 1995; Cowper, 2005; Harley & Ritter, 2002; Harris, 1991). The idea is that oppositions like masculine-feminine or singular-plural are not symmetrical. Rather, one member of the opposition is unmarked or has a so-called “default status”, and the other one is marked (Corbett, 2000). These theories further assume that markedness drives the specification of features. That is, while marked feature values are fully specified, unmarked ones are assumed to be underspecified (i.e. they carry no feature information). The present study investigates how these markedness relations impact the online processing of number and gender agreement in Spanish.

Spanish has a two-way gender system that classifies nouns as either masculine or feminine. In addition, its number system distinguishes between two values, singular and plural. A number of observations suggest that, in Spanish, plural and feminine are marked for number and gender, respectively, relative to singular and masculine (Battistella, 1990; Bonet, 1995; Harris, 1991). For example, when a word with no inherent gender (e.g. the preposition *para* “for”) is modified by an agreement-bearing element, the latter can only take the masculine form (*demasiados paras en ese párrafo* “too-many-MASC fors-NO GENDER in that paragraph”) (Harris, 1991). Likewise, when nouns of different genders are conjoined, all agreement targets must take the masculine form. This suggests that masculine is unmarked, since it can modify genderless elements and even feminine ones, but the feminine value is marked since it is restricted to feminine nouns. A similar asymmetry emerges for the singular-plural opposition. For example, singular lacks overt inflection, whereas plural is almost always morphologically overt and derived from the singular (*parque-parques* “park-parks”, *mar-mares* “sea-seas”) (e.g. Battistella, 1990). In addition, singular has a

broader syntactic distribution than plural. For example, the singular dative clitic *le* can be coindexed with a plural phrase in certain contexts (*Ana le_i teme [a las brujas]_i* “Ana fears witches”), but its plural counterpart *les* cannot be coindexed with a singular phrase. This suggests that singular is unmarked for number, since it can refer to both singular and plural constituents, but plural is marked, since it is restricted to plural elements.ⁱ

Previous studies have examined how different features are retrieved and tracked online for the purposes of agreement (e.g. Alemán Bañón, Fiorentino, & Gabriele, 2012; Barber & Carreiras, 2005; Colé & Seguí, 1994; Faussart, Jakubowitz, & Costes, 1999; De Vincenzi, 1999; De Vincenzi & Di Domenico, 1999; Lukatela, Kostic, Todorovic, Carello, & Turvey, 1987; Mancini, Molinaro, Rizzi, & Carreiras, 2011; Nevins, Dillon, Malhotra, & Phillips, 2007; Osterhout & Mobley, 1995; Silva-Pereyra & Carreiras, 2007; Zawiszewski, Santesteban, & Laka, 2015). However, relatively little is known about how markedness impacts these agreement relations. One possibility, suggested by Nevins et al. (2007), is that markedness determines, at least to some extent, whether agreement engages *top-down* versus *bottom-up* processing. Under a top-down account, upon encountering an agreement trigger, predictions are generated regarding the specification of dependent elements bearing agreement features (e.g. Gibson, 1998). Alternatively, under a bottom-up approach, agreement is established backwards; after encountering an element carrying agreement features, the parser goes back to the preceding context and searches for its controller (e.g. Nicol, Forster, & Veres, 1997; Pearlmutter, Garnsey, & Bock, 1999). Nevins et al.’s suggestion is that the parser might be more likely to establish agreement predictively upon encountering an element carrying marked features, due to feature activation. This is an interesting proposal that we evaluate in our study.ⁱⁱ

In the agreement literature, one phenomenon which has provided evidence for feature value asymmetries is *attraction*. In attraction, an element carrying agreement features (e.g. a

verb) fails to agree with its controller noun and, instead, agrees with a structurally unavailable element, yielding an ungrammatical string (e.g. *the key to the cabinets *are...*; Bock & Miller, 1991). A number of studies have shown that errors are more likely to arise when the head noun (*key*) and the attractor (*cabinet*) mismatch in number and the latter is marked for plural. In contrast, when the attractor is singular, agreement errors are quite rare (e.g. production: Antón-Méndez, Nicol, & Garrett, 2002; Bock & Cutting, 1992; Bock & Eberhard, 1993; Bock & Miller, 1991; Eberhard, 1997; Hartsuiker, Antón Méndez, & van Zee, 2001; Vigliocco & Frank, 1999; comprehension: Acuña Fariña, Meseguer, & Carreiras, 2014; Nicol et al., 1997; Pearlmutter et al. 1999; Wagers, Lau, & Phillips, 2009).

Interestingly, a review of comprehension and production studies on agreement attraction reveals that this feature value asymmetry is strong and cross-linguistically reliable for number (Dutch/German: Hartsuiker, Schriefers, Bock, & Kikstra, 2003; French: Fayol, Largy, & Lemaire 1994; Spanish: Acuña Fariña et al., 2014; Antón Méndez et al., 2002), but not for grammatical gender (Acuña Fariña et al., 2014; Vigliocco & Frank 1999). Some studies examining noun-adjective agreement have reported an asymmetry between feminine and masculine in Romance languages like French (Vigliocco & Franck, 1999, 2001), Italian (Vigliocco & Franck, 2001) and Spanish (e.g. Antón-Méndez et al., 2002), but the asymmetry went in the opposite direction; that is, agreement errors were more frequent when the head noun was feminine (i.e. marked). In addition, other studies have failed to replicate this asymmetry altogether for those same languages (e.g. Acuña Fariña et al., 2014) or related ones, such as Italian (e.g. Vigliocco & Frank 1999). One possible explanation for why markedness impacts the two features differently is that grammatical gender is a lexical property of the noun (e.g. Carstens, 2000; Ritter, 1993; Eberhard et al. 2005; but see Picallo, 1991, 2008), but number is extrinsic to the noun and projects its own syntactic phrase (Ritter, 1991). This means that the same noun can be used in the singular or plural depending on the

syntactic context. In contrast, nouns from one gender category typically lack a counterpart of the opposite gender.ⁱⁱⁱ Based on these observations, it has been suggested that, from a psycholinguistic perspective, there is no markedness distinction between masculine and feminine in the noun (e.g. Vigliocco & Franck, 1999; Antón-Méndez et al. 2002) and that the higher error rate with feminine head nouns may be due to participants' overusing the default gender (masculine) on agreement targets such as adjectives. In other words, this proposal assumes an asymmetry between the two gender values, but one that emerges in syntactic categories other than the noun.

The present study examines whether feature value asymmetries also impact agreement in attractor-free contexts. Spanish being a morphologically rich language, it is possible that the abundance of agreement cues makes morphology a reliable predictor of agreement operations, such that the parser might utilize markedness information for the purposes of agreement. We focus on both number and gender, in order to examine whether markedness impacts the two agreement types differently (Acuña Fariña et al., 2014; Antón Méndez et al., 2002; Vigliocco & Frank, 1999, 2001). To address this question, we used Electroencephalography (EEG), a non-invasive electrophysiological method that provides high temporal resolution, on the order of milliseconds. EEG records brain potentials at the scalp that are time-locked to the stimulus of interest. Those brain potentials, referred to as event-related potentials (ERPs), have been found to be modulated by different linguistic processes, making this technique suitable to examine the qualitative nature of language processing. Despite the extensive literature using ERP to examine agreement phenomena, to our knowledge, no previous ERP study has examined the role of markedness on agreement by systematically comparing instances of agreement where the trigger is marked versus unmarked, in contexts other than attraction. As Molinaro, Barber, and Carreiras (2011) point out in their review of ERP studies on agreement, most previous studies have collapsed across

the values of the feature under investigation (i.e. violations involve both errors of the kind “marked trigger + unmarked target” and errors of the kind “unmarked trigger + marked target” (e.g. Alemán Bañón et al., 2012; Guajardo & Wicha, 2014; Hagoort & Brown, 1999) or by focusing on just one type of error (e.g. Alemán Bañón et al., 2012; Barber & Carreiras, 2005; Deutsch & Bentin, 2001, for number). In fact, Molinaro et al. (2011) highlight the need to bring electrophysiological evidence to bear on the question of how feature value asymmetries impact processing at the brain level, a gap in the literature that the present study addresses.

1.1 ERP Indices of Agreement

A number of studies have used ERP to examine agreement phenomena. In these studies, agreement violations consistently elicit a P600, a positive deflection in the ERP waveform between approximately 500-1000 ms that is typically captured by central-posterior electrodes of the EEG cap (e.g. Friederici, 2002; Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992). The functional significance of the P600 is still debated. It was initially linked to syntactic processing, as it was found for less preferred continuations of syntactically ambiguous sentences (e.g. Carreiras, Salillas, & Barber, 2004; Gouvea, Phillips, Kazanina, & Poeppel, 2010; Osterhout & Holcomb, 1992) and outright syntactic violations, including agreement errors (e.g. Hagoort et al., 1993; Osterhout & Mobley, 1995). It has also been argued that the P600 reflects the difficulty of integrating syntactic dependencies predicted on the basis of context, consistent with the observation that it can be found for complex but grammatical and unambiguous sentences (Kaan, Harris, Gibson, & Holcomb, 2000). More recently, a frontal positivity within the P600 time window has been linked to a failed prediction (e.g. DeLong et al., 2011; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Van Petten & Luka, 2012), and Kutas, DeLong, and Smith (2010) have suggested that this

effect may well be related to the P600. For example, Wicha, Moreno, and Kutas (2004) found a frontal positivity in the P600 time window for articles that disagreed in gender with a noun that was predicted on the basis of context, at a point when the noun was yet to be encountered (see also Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005).

Other evidence suggests that the P600 is not exclusively syntactic in nature, since it can be found for certain types of semantic anomalies (e.g. Bornkessel-Schlesewsky & Schlewsky, 2008; Brouwer, Fitz, & Hoeks, 2012; Kim & Osterhout, 2005; Kolk, Chwilla, Van Herten, & Oor, 2003; Kuperberg, 2007). Alternatively, the P600 has also been argued to reflect non language-specific processes that may be engaged during linguistic computations (e.g. Coulson, King, & Kutas, 1998; Patel, Gibson, Ratner, Besson, & Holcomb, 1998). In addition, its amplitude has been found to be modulated by salience and the probability of encountering unexpected task-relevant events (i.e. an ungrammaticality), suggesting that the P600 may not be completely independent from the P3b, an ERP component associated with the detection of unexpected non-linguistic events (Coulson et al., 1998; Sassenhagen, Schlewsky, & Bornkessel-Schlewsky, 2014; Van de Meerendonk, Kolk, Vissers, & Chwilla, 2010; but see Osterhout & Hagoort, 1999).

In the present study, we focus on the P600, which has been taken to reflect the difficulty associated with specific syntactic operations (i.e. integration, reanalysis, and repair) (Barber & Carreiras, 2005; Kaan et al., 2000) and the ease in diagnosing ungrammaticalities (Friederici, Mecklinger, Spencer, Steinhauer, & Donchin, 2001; McKinnon & Osterhout, 1996). For example, increased P600 amplitude has been interpreted as evidence that certain dependencies are more difficult to resolve than others and, therefore, require the allocation of greater resources (e.g. Barber & Carreiras, 2005; Münte, Szentkuti, Wieringa, Matzke, & Johannes, 1997). In turn, a reduced P600 has been interpreted as evidence that certain ungrammaticalities are harder to detect, such that the parser loses sensitivity to them (e.g.

Hammer, Jansma, Lamers, Münte, 2008; Kaan & Swaab, 2003; Kolk et al., 2003). In addition, since our study evaluates Nevins et al.'s proposal that marked features allow for predictive processing, we will also examine the topographical distribution of the P600, in order to evaluate current claims linking a frontal positivity in the P600 time window to prediction.

In some studies, the P600 is preceded by a negative component emerging between approximately 300-500 ms in left anterior areas. This Left Anterior Negativity "LAN" (e.g. Barber & Carreiras, 2005; O'Rourke & Van Petten, 2011) has been argued to index automatic morphosyntactic processing (e.g. Friederici, Hahne, & Mecklinger, 1996; Molinaro et al., 2011; Molinaro, Barber, Caffarra, & Carreiras, 2015), but a problem with this interpretation is that a number of studies examining morphosyntactic dependencies do not report it (e.g. Alemán Bañón et al., 2012; Hagoort & Brown, 1999; Kolk et al., 2003; Nevins et al., 2007; Wicha et al., 2004). Some authors have also discussed the possibility that the LAN is not completely independent from another well-known ERP component related to lexical access and semantic integration, the N400 (Guajardo & Wicha, 2014; Osterhout & Nicol, 1999; Service, Helenius, Maury, Salmelin, 2007; Tanner & Van Hell, 2014). The N400 (Kutas & Hillyard, 1980) is a negative-going wave which exhibits similar polarity and latency to the LAN, but is distinguished from it in terms of its scalp distribution, which is very broad and typically reaches its maximum in central-posterior regions, with a slight right-hemisphere bias. Since the N400 and the P600 have similar scalp distribution and can overlap to some extent in latency, their superimposition may cause the N400 to emerge in the left anterior portion of the scalp, creating the illusion of a LAN (but see Mancini et al., 2011, who found the LAN for morphosyntactic mismatches in Spanish in the absence of a P600; see also Molinaro et al., 2015 for further counterarguments). Under this view, the presence of the N400 for agreement errors may reflect individual differences with respect to processing

strategy, with some subjects relying more on lexical-semantic than syntactic processing (e.g. Tanner & Van Hell, 2014) or the difficulty with semantic integration engendered by the agreement error, especially in highly constraining contexts (e.g. Guajardo & Wicha, 2014).

In sum, studies on agreement have consistently reported the P600 for violations, and some studies have also reported an earlier negativity, although there is disagreement as to its reliability and even its identity (LAN versus N400).

1.2 ERP studies on Markedness and Feature Type

Markedness. Very few studies have used ERP to investigate the role of markedness on agreement. Deutsch and Bentin (2001) found that subject-verb gender violations in Hebrew yielded a larger P600 when the agreeing elements were marked for number (plural; “the actresses-FEM-PL *were-enchanted-MASC-PL), relative to contexts where they were unmarked (singular; “the actress-FEM-SG *was-enchanted-MASC-SG), which they relate to plural marking in Hebrew being more complex and salient than singular (e.g. Coulson et al., 1998). However, it should be taken into account that the main focus of the Deutsch and Bentin study was not markedness and, thus, their study can only speak to the asymmetry between singular and plural forms in the context of a gender agreement violation (the grammaticality of number agreement was not manipulated in their study).

Kaan (2002) examined the plural versus singular asymmetry in a context of attraction. Her study examined subject-verb number agreement in Dutch sentences and found that the P600 was larger and had an earlier onset for violations that involved a singular subject in disagreement with a plural verb, relative to “plural subject + singular verb errors”, an effect that only arose in cases where a singular object noun intervened. Although no explanation for this finding is offered, one possibility is that violations on marked elements are more salient

(consistent with Deutsch & Bentin's results, 2001), but only in contexts where there is no interference or competition from attractors.

Finally, Hoen, Deprez, and Dominey (2007) compared instances of successful agreement in French passive sentences in two conditions, one where the verbal predicate was marked for both number and gender (plural and feminine, respectively; e.g. *idioties ont été prononcées* “insanities-FEM-PL have been pronounced-FEM-PL”), and one where the verbal predicate was unmarked for both features (singular and masculine, respectively; e.g. *discours a été prononcé* “discourse-MASC-SG has been pronounced-MASC-SG”). Their results showed qualitatively different responses for marked and unmarked predicates. While marked predicates were associated with a left anterior negativity (which they refer to as the LAN), unmarked ones were associated with an N400, which the authors take as evidence that only marked values engage agreement checking, in line with previous studies on agreement comprehension (e.g. Nicol et al., 1997; Pearlmutter et al., 1999).

In sum, although the evidence is scarce, previous studies have shown that markedness impacts agreement processing, as reflected by the larger P600 effects for marked versus unmarked predicates in the study by Deutsch and Bentin (2001) and, less conclusively, in the study by Kaan (2002). In addition, it has been posited that marked and unmarked predicates engage qualitatively different processing mechanisms, with only marked forms showing evidence of feature checking (in the form of a LAN).

Feature Type. With respect to how different feature types impact agreement, the Feature Hierarchy Hypothesis (Carminati, 2005) proposes that number is more cognitively salient than gender, due to its greater prominence cross-linguistically. However, most ERP studies which have compared the two features have revealed qualitatively similar processing profiles. Barber and Carreiras (2005) examined number and gender agreement in Spanish and found that both violation types yielded a LAN and a P600. The later portion of the P600 (700-900

ms) was larger for gender than number, but this amplitude difference was not replicated in a later study by Gillon-Dowens, Vergara, Barber, and Carreiras (2010) with the same stimuli. Alemán Bañón et al. (2012) also compared number and gender agreement in Spanish and found equally robust P600 effects for both violation types, a pattern of results also reported by Nevins et al. (2007) for number and gender in Hindi. Importantly, neither Barber and Carreiras (2005) nor Alemán Bañón et al. (2012) controlled for markedness differences in the way they compared number and gender. In both studies, gender violations involved both errors of the kind “feminine noun + masculine adjective” and errors showing the opposite pattern, but number violations only included one type of error (“singular noun + plural adjective”). In addition, in Nevins et al.’s study (2007), violations of number and gender both involved an unmarked noun in disagreement with a marked verb, but the opposite error type was not examined. This means that there is no clear account in the ERP literature of how markedness impacts different agreement categories. We address this question in our study.

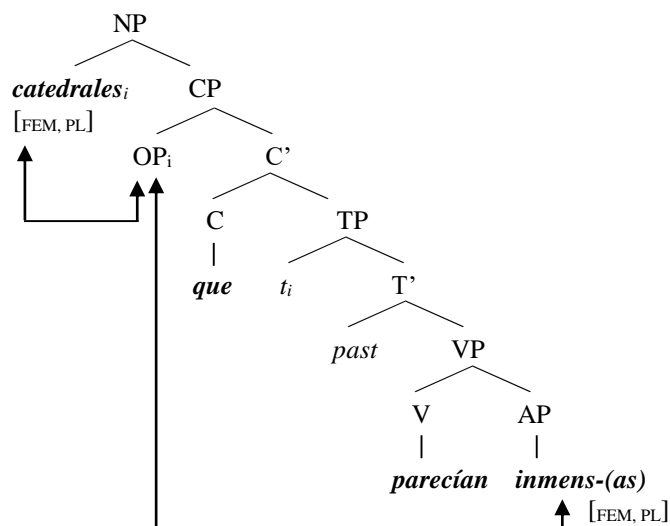
1.3 The Present Study

In the present study, we examine the role of markedness on agreement resolution in complex sentences like (1-4) below (the agreeing words are underlined for clarity):

- (1) Laura lavó un uniforme que parecía sucio con mucho detergente.
Laura washed a uniform-MASC-SG that looked dirty-MASC-SG with much detergent
- (2) Carlos fotografió una catedral que parecía inmensa para una revista.
Carlos photographed a cathedral-FEM-SG that looked huge-FEM-SG for a magazine
- (3) Laura lavó unos uniformes que parecían sucios con mucho detergente.
Laura washed some uniforms-MASC-PL that looked dirty-MASC-PL with much detergent
- (4) Carlos fotografió unas catedrales que parecían inmensas para una revista.
Carlos photographed a cathedral-FEM-PL that looked huge-FEM-PL for a magazine

The agreement relation of interest is that between the object noun of the matrix verb (e.g. *catedrales* “cathedral-FEM-PL” in 4), which acts as the head of a subject restrictive relative clause (e.g. Payne, 1997), and a predicative adjective located inside the relative clause (e.g. *inmensas* “huge-FEM-PL” in 4). A (simplified) tree of this structure is provided in (5):

(5)



One standard analysis for this type of subject relative clause is that the head NP (e.g. *catedrales* “cathedral-FEM-PL”) enters an agreement relation with an empty relative operator “OP” that raises from the specifier position of Tense Phrase “TP” (i.e. the subject position) to the specifier position of the Complementizer Phrase “CP” headed by *que* (see Rivero, 1990, 1982 for Spanish; see also Kayne, 1976 for French & Cinque, 1978, for Italian). In turn, the relative operator enters an agreement relation with the predicative adjective located inside the TP “*que parecía/parecían* ADJECTIVE”.

Within this type of complex structure, markedness was examined by systematically manipulating the number and gender specification of the head NP, such that half of the nouns are masculine, as can be seen in (1) and (3) and the other half, feminine (see 2 and 4). Likewise, half of the NPs were used in the singular, as can be seen in (1) and (2) and the other half, in the plural (see 2 and 4). Our motivation for examining agreement in a nonlocal

context is that previous ERP studies have shown that agreement resolution is impacted by both the linear (i.e. number of words) and structural distance (i.e. number of syntactic phrases) between the agreeing elements (Alemán Bañón et al., 2012; Barber & Carreiras, 2005; O'Rourke & Van Petten, 2011). For example, O'Rourke and Van Petten (2011) examined gender agreement at three levels of (a combination of linear and structural) distance and found decreasing sensitivity to gender errors (i.e. a reduced P600) with increasing distance. We, therefore, hypothesized that markedness effects (e.g. reduced sensitivity to errors where the violating feature is unmarked) might be more likely to emerge when the agreement relation is structurally nonlocal and the agreeing words are nonadjacent.

As mentioned above, the study examines the role of markedness on agreement resolution in an attractor-free context. Thus, as shown in (1-4), no nouns intervened between the agreeing elements. However, the relative clause where the critical adjective is located includes the semi-copulative verb *parecía/parecían*, which agrees in number with the trigger noun (as is the case with all Spanish verbs). Importantly, this verb, which was held constant in the design, could not cause interference, since it was always matched with the trigger noun in terms of number. In addition, the results of a study by Alemán Bañón et al. (2012) suggest that the presence of this additional target of number agreement should not impact agreement resolution on the adjective. Alemán Bañón et al. (2012) examined number and gender agreement both within a determiner phrase (e.g. *edificio muy seguro* “[DP building-MASC-SG very safe-MASC-SG]”), where the intervening adverb *muy* provided no additional cues for either number or gender, and across a verb phrase (*cuento es anónimo* “story-MASC-SG [VP is-SG anonymous-MASC-SG]”), where the intervening copulative verb *es* “is” carried number information. Their results showed similar P600 effects in the two contexts, both for number (for which there was an additional cue) and for gender (for which there was no additional

cue), suggesting that agreement resolution was similar even in the presence of an additional target of number agreement.

1.3.1 Predictions

Grammaticality

Based on previous reports, we predict that both number and gender violations overall will yield a P600 relative to grammatical sentences. It is also possible that number and gender violations will show a negativity preceding the P600, although we reserve making predictions regarding the distribution of this effect (left anterior, consistent with the LAN, or central-posterior, consistent with the N400).

Markedness

There are different ways in which markedness could modulate agreement. First, if number violations are more salient when they are realized on plural (marked) adjectives, we expect errors of the kind “singular noun + plural adjective” to yield a larger P600 than “plural noun + singular adjective” errors. This would be in line with the results by Deutsch and Bentin (2001), although recall that their study only examined how number markedness impacted the processing of gender agreement. Here we focus on how number markedness impacts number agreement. In addition, if gender markedness impacts gender agreement in a similar way, we expect violations of the kind “masculine noun + feminine adjective” to show a larger P600 than the opposite error type. Along similar lines, the parser might be less sensitive to errors where the violating feature is unmarked. This would yield the same pattern of results (a larger P600 when the violating feature is marked), although for somewhat different reasons.

Alternatively, if the parser uses marked features to predict the form of agreement targets (e.g. Nevins et al., 2007), it is possible that violations of the kind “marked noun + unmarked adjective” will yield a larger P600, due to the fact that an expectation was generated but

unmet. Another possibility is that the P600 for errors of the type “marked noun + unmarked adjective” will also span over frontal regions, consistent with previous studies linking a late frontal positivity to a failed prediction (Van Berkum et al., 2005; Wicha et al., 2004).

Feature Type

Proposals like the Feature Hierarchy Hypothesis (Carminati, 2005) predict a larger P600 for number than gender, based on the fact that number is more cognitively salient than gender. Based on previous reports, it is also possible that number and gender will show similar processing profiles overall (Alemán Bañón et al., 2012; Gillon-Dowens et al. 2010; Nevins et al., 2007). In addition, if differences in the representation of the two features make them unequally sensitive to markedness, we might expect markedness effects to emerge for number, but not gender (consistent with the literature on attraction).

The remainder of the article is structured as follows. In Section 2, we articulate the methodology of the study. We then present the results in Section 3, discuss their implications in Section 4, and provide concluding remarks in Section 5.

2. Methods

2.1 Participants

Twenty-seven native speakers of peninsular Spanish (15 females) provided their informed consent to participate in the study (age range: 18-38; mean age: 27).^{iv} All participants were right-handed, as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971), had normal or corrected-to-normal vision, and reported no history of neurological or cognitive impairments. All participants reported having been born and raised in Spain, and they all indicated that Spanish was their native language. On average, they received about 16 years of formal education in Spanish (SD: 2 years). All participants were adult language learners of English to varying degrees of proficiency, and some of them indicated having studied other

foreign languages (Chinese, French, German, Italian, and Portuguese), also to varying degrees of proficiency. In addition, four participants reported being bilingual speakers of Spanish and another one of Spain's official languages (Catalan and Galician) or Spanish Sign Language. All participants were tested in England and received monetary compensation for their participation.

2.2 Experimental Design

2.2.1 Materials

The materials for the study comprise 240 sentences of 10 words each similar to (1) above. Markedness was examined by manipulating the number and gender specification of the trigger noun, as can be seen in Table 1 below, which provides a sample of each of the 12 experimental conditions.

--Insert Table 1--

With respect to gender, the controller noun was masculine in half of the sentences (i.e. 120) and feminine in the other half (i.e. 120). This can be seen by comparing the sentences in (1-6) to those in (7-12) in Table 1. In (1-6) the noun *catedral* "cathedral" is feminine (marked) and, therefore, the agreeing adjective (*inmensa* "huge") must also be feminine. Otherwise, the string would be ungrammatical, as shown in (3) and (6). The opposite pattern is shown in (7-12), where the trigger noun (*pastel* "cake") is masculine (unmarked) and, therefore, the modifying adjective (*asqueroso* "disgusting") must also be masculine. Otherwise, the string would be ungrammatical, as shown in (9) and (12).

A similar design was used to examine number. That is, half of the sentences in the study included a singular noun and the other half, a plural one. This can be seen by comparing the

conditions in (1-3) and (7-9) to those in (4-6) and (10-12). In (4) and (10), the nouns *catedrales* “cathedrals” and *pasteles* “cakes” are marked plural and, therefore, the agreeing adjectives must also be marked plural. Otherwise, the sentence would be ungrammatical, as shown in (5) and (11). In contrast, in (1) and (7) the trigger nouns (*catedral* “cathedral”, *pastel* “cake”) are singular (unmarked) and, therefore, the predicative adjectives must also be singular. Otherwise, the string would be ungrammatical (see 2 and 8).

As can be seen in Table 1, number and gender were crossed in the present design, such that the gender manipulation involved both singular and plural nouns (equally distributed across the masculine and feminine conditions), and the number manipulation encompassed both masculine and feminine nouns (equally distributed across the singular and plural conditions). Table 2 provides a schematic of how the experimental materials were distributed across the 12 conditions above.

--Insert Table 2 here--

We created 20 items for each of the 12 conditions (i.e. a total of 240 sentences), but for the purposes of analysis, we collapsed across number when examining gender and vice versa, which yielded 40 items per condition. For example, for the examination of gender, we used 80 grammatical items (40 with a feminine noun, 40 with a masculine one) and 80 gender violations (40 with a feminine noun, 40 with a masculine one), collapsing across number. Likewise, when examining number, we used 80 grammatical items (40 with a plural noun, 40 with a singular one) and 80 number violations (40 with a plural noun, 40 with a singular one), collapsing across gender.

Together with the 240 experimental items described above, the study included another 160 sentences (80 grammatical, 80 ungrammatical) from a separate experiment which does not

manipulate number and gender agreement and which does not include any adjectives, in addition to 80 grammatical fillers. All of the fillers included adjectives, half of which do not inflect for gender (e.g. *responsable* “responsible”). No other materials were included in the study. The overall design encompassed an equal amount of grammatical and ungrammatical sentences. This is because previous studies have argued that an excessive number of ungrammatical sentences can attenuate one of the ERP components of interest, the P600 (Coulson et al., 1998; Hahne & Friederici, 1999).

These 240 sentences were counterbalanced across 6 experimental lists, such that a given participant would see 20 items per each of the 6 conditions in (1-6) (a total of 120) and 20 items per each of the 6 conditions in (7-12) (a total in 120), but no participant saw the same sentence twice. Each experimental list also included one version of each sentence type from an additional experiment, and all of the grammatical fillers.

2.2.1 Properties of the Stimuli

2.2.1.1 Nouns and Adjectives

Previous ERP studies on Spanish gender agreement have used nouns that provide strong distributional cues to gender (Alemán Bañón et al., 2012; Barber & Carreiras, 2005; Gillon-Dowens et al., 2010; O’Rourke & Van Petten, 2011; Wicha et al., 2004). This is the case for masculine nouns that end in *-o* and feminine nouns that end in *-a* (according to Teschner & Russell, 1984, approximately 99.8% of Spanish nouns ending in *-o* are masculine and 96.3% of nouns ending in *-a* are feminine). Here, we focus on nouns that do not show the *-o* versus *-a* distinction (e.g. *torre* “tower_{FEM}”, *cathedral* “cathedral_{FEM}”, *bosque* “forest_{MASC}”, *pastel* “cake_{MASC}”), whose gender cannot be as reliably determined based on the properties of the root. A number of studies (e.g. Bates, Devescovi, Pizzamiglio, D’Amico, & Hernández, 1995; Padovani, Calandra-Buonaura, Cacciari, Benuzzi, & Nichelli,

2005; Schiller, Münte, Horemans, & Jansma, 2003) have shown that, in tasks that require the retrieval of lexical gender information (e.g. lexical gender decision, article selection) participants tend to be slower and less accurate with these more opaque nouns, relative to transparent nouns. We thus hypothesized that agreement resolution (i.e. violation detection, revision/repair) would be more costly with nouns that do not provide strong distributional cues to gender.^v This additional cost in gender resolution might, in turn, make participants more sensitive to markedness asymmetries; that is, participants might show reduced sensitivity to gender errors realized on unmarked adjectives (*catedral que parecía *inmenso*) relative to the reverse type, as in the studies by Antón Méndez et al. (2012) and Vigliocco and Frank (1999). We also restrict ourselves to nouns providing no biological gender cues, to ensure that participants could not resort to semantic information to compute gender agreement.

With respect to number, we used nouns that inflect for plural with either *-s* (e.g. *torre/torres* “tower/towers”) or *-es* (*pastel/pasteles* “cake/cakes”). As for singular, it is a property of all Spanish nouns that they show no inflection in the singular.

All critical adjectives in the study show the *-o* versus *-a* distinction when inflected for gender (e.g. *sucio* vs. *sucia* “dirty-MASC vs. dirty-FEM”), and they all inflect for number by adding *-s* (e.g. *sucio* vs. *sucios* “dirty-SG vs. dirty-PL”).

2.2.1.2 Item Controls

We calculated the log count (a measure of lexical frequency) of all nouns and adjectives using the EsPal database (EsPal Written Corpus, 2012; Duchon, Perea, Sebastián Gallés, Martí, & Carreiras, 2013). The masculine and feminine nouns did not reliably differ with respect to either frequency, $t(118) = -1.471$, $p > .1$, or length, $t(118) = -1.512$, $p > .1$. The masculine and feminine versions of the adjectives (critical word where ERPs were calculated)

were also matched for frequency, $t(238) = 1.6$, $p > .1$, and their length was exactly the same. For the comparison of singular versus plural items, it was not possible to control the nouns or adjectives for either frequency or length. As expected, the singular nouns and adjectives are more frequent than their plural counterparts (Corbett, 2000, p. 281), possibly due to the fact that singular is the unmarked number value and has a wider syntactic distribution than plural. With respect to length, it was also not possible to control the singular and plural versions of the nouns or the adjectives, given that singular in Spanish is morphologically unmarked, but plural requires overt morphology ($-s/-es$).

As can be seen in Table 1, the critical word (adjective) was always located midsentence. Importantly, its position was held constant across conditions (always word #7). This is because previous studies have shown that words which are located deeper into the sentence tend to show a reduced N400 compared to earlier words (Van Petten & Kutas, 1990), which could affect later components (i.e. P600). In order to diversify the stimuli as much as possible, six different types of 3-word structures were designed that could follow the critical word.

The 120 sentences with a feminine noun encompassed 60 different critical nouns, which were used twice, but never in agreement with the same adjective. The 120 sentences with a masculine noun also involved 60 different trigger nouns, used twice with different adjectives. All critical adjectives were also used twice, once with a masculine noun (e.g. *uniforme...sucio* “uniform...dirty”) and once with a feminine noun (e.g. *calle...sucia* “street...dirty”). Since the testing involved two sessions (see *Section 2.3 Procedure*), the experimental lists were designed in such a way that participants only saw one version of each critical adjective per session, which was expected to minimize the effects of repetition on the waveforms. All of the main clause verbs in the study were transitive and they were all inflected for third person singular past tense.

Finally, all of the grammatical sentences were submitted to both a cloze probability test and a plausibility test, to ensure that the critical comparisons were not impacted by potential differences in these factors. These tests revealed that the materials were matched in terms of both cloze probability and plausibility.^{vi}

2.3 Procedure

Due to the large number of sentences in the study, the testing was carried out in two sessions (Alemán Bañón et al., 2012; O'Rourke & Van Petten, 2011), which were separated by a minimum of three days and a maximum of two weeks. Half of the experimental items were assigned to the first session, and the other half to the second session. The sessions lasted for approximately 3 hours (EEG recording itself: 1 hour). Participants sat in a comfortable chair facing a computer monitor in an electrically shielded room. They were instructed to silently read a series of sentences in Spanish and decide if they were good or bad sentences of Spanish (e.g. Alemán Bañón et al., 2012; Barber & Carreiras, 2005; Hagoort, 2003; Kaan, 2002; Nevins et al., 2007; Osterhout & Mobley, 1995; Tanner & Van Hell, 2014). Participants were asked to minimize blinks and other movements during the presentation of the sentences, and encouraged to blink between trials. Each session began with a practice set which included eight sentences, half of which were ungrammatical. None of the ungrammatical practice trials involved agreement errors. In addition, in order to avoid repetition effects, the practice sentences were designed with lexical material that did not appear in the experimental stimuli. Participants received feedback for the first three trials. Immediately after the practice, the experiment began. Each experimental session was divided into six blocks of 40 sentences, separated by five short breaks. Within each block, sentences from all experimental conditions (plus fillers and sentences from an additional experiment) were randomly intermixed. Each session involved an equal number of items per condition.

The presentation of the sentences was carried out using *Paradigm* by Perception Research Systems Inc. (Tagliaferri, 2005).

The trial structure was as follows: first, a fixation cross appeared in the center of the monitor for 500 ms. Then, the first word of the sentence was presented. Sentences were presented one word at a time using the RSVP (Rapid Serial Visual Presentation) method. Each word was presented for 450 ms and followed by a 300 ms interstimulus interval (e.g. Alemán Bañón et al., 2012). At the end of each sentence, there was a 1000 ms pause, followed by the prompts for the grammaticality judgment: the words *Bien* “good” for grammatical sentences and *Mal* “bad” for ungrammatical ones. Participants were asked to respond with their left hand (middle and index fingers, respectively) and to favor accuracy over speed. The prompts remained on the screen until the participant pressed one of the two buttons on the computer mouse. Following the behavioral response, there was an inter-trial interval ranging between 500-1000 ms, pseudo-randomly varied at 50 ms increments. Immediately after this interval, the next trial began.

At the end of the second EEG recording, participants completed a short task where they had to indicate the gender of all 120 critical nouns (60 feminine nouns, 60 masculine nouns) in the experimental materials. The purpose of this task was to ensure that all participants knew the lexical gender of the nouns where agreement was manipulated. All participants performed at 99% accuracy or above, suggesting that they knew the gender of the critical nouns in the study.

2.4 EEG Recording and Analysis

We used the Brain Vision Recorder software (BrainProducts, GmbH, Germany) to record the continuous electroencephalogram from 64 sintered Ag/AgCl electrodes buttoned into an elastic cap (Easycap, BrainProducts, GmbH, Germany). The electrodes were placed

according to the 10% System (midline: FPz, Fz, 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). Electrode AFz served as ground. The recording was referenced online to electrode FCz and re-referenced offline to the average of the right and left mastoids. An additional electrode (IO) was placed on the outer canthus of the right eye to monitor horizontal eye movements. Electrodes FP1 and FP2 (above each eyebrow) were used to monitor blinks. Electrode impedances were maintained below 10k Ω s for all electrodes. The recordings were amplified by a BrainAmp MR Plus amplifier (BrainProducts, GmbH, Germany) with a bandpass filter of .01 to 200 Hz, and digitized at a sampling rate of 1 kHz.

The EEG data analysis was conducted using the Brain Vision Analyzer 2.0 software (BrainProducts, GmbH, Germany). The EEG was first segmented into epochs relative to the critical word. Epochs started 300 ms before the critical word, which corresponds to the pre-stimulus baseline, and ended 1200 ms post-onset. Trials with artifacts (blinks, horizontal eye movements, excessive muscle artifact, and excessive alpha waves) were manually rejected from analysis. This resulted in the exclusion of approximately 7% of the data. Importantly, the mean number of artifact-free trials per condition (37-38 out of 40 across conditions) did not significantly differ across conditions (conditions 1-6 versus conditions 7-12: $F(3, 78) = 2.422, p > .05$; conditions 1-3 and 7-9 versus conditions 4-6 and 10-12: $F(3, 78) = .087, p > .05$). Data were then filtered with a low-pass filter set at 30 Hz. After filtering, epochs were baseline-corrected relative to the 300 ms pre-stimulus baseline, and averaged per condition and per participant.

Upon visual inspection of the waveforms and based on the previous literature, ERPs were quantified using mean amplitudes in two time windows of interest: the 250-450 ms time window, which includes the N400 (and the Left Anterior Negativity), and the 500-1000 ms

time window, which includes the P600. Nine regions of interest were selected for statistical analysis. Each region was calculated by averaging together the mean amplitudes of all of the 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). Analyses were carried out separately for the hemispheres and the midline (e.g. Barber & Carreiras, 2005; Deutsch & Bentin, 2001; Gillon-Dowens et al., 2010). Mean amplitudes were submitted to a repeated-measures ANOVA with Markedness (marked, unmarked), Agreement (grammatical, ungrammatical), Hemisphere (left, right) and Anterior-Posterior (anterior, central, and posterior) as repeated factors. For the analyses conducted on the midline, the only topographical factor in the ANOVA was Anterior-Posterior. Analyses were carried out separately for number and gender. This is because, in the examination of gender, it was necessary to collapse across number to achieve the number of items per condition recommended by Molinaro et al. (2011a) (and vice versa for number). The Geisser and Greenhouse correction was applied for violations of sphericity. Degrees of freedom are reported after correction (Field, 2005).

3. Results

3.1 Behavioral Results

As can be seen in Table 3 below, participants' scores in the grammaticality judgment task were very high across conditions, suggesting that they understood the task correctly and could distinguish between grammatical sentences and sentences with either a number or a gender agreement violation, regardless of markedness. No further analyses were conducted.

--Insert Table 3 here--

3.2 ERP Results

Visual inspection of the ERP waveforms reveals that, at approximately 500 ms, number and gender violations become more positive than grammatical sentences in central-posterior areas of the scalp, an effect which remains visible throughout the rest of the epoch (Figures 1 and 2 show the waveforms for the number and gender conditions, respectively; Figures 3 and 4 show topographic maps for the effects of number and gender violations, respectively). This positivity, which is consistent with the P600, appears equally robust across both levels of markedness for gender (see Figures 2 and 4). For number, violations realized on plural adjectives appear more positive than violations on singular adjectives (see Figures 1 and 3).

--Insert Figure 1 approximately here--

--Insert Figure 2 approximately here--

--Insert Figure 3 approximately here--

--Insert Figure 4 approximately here--

In approximately the same time window where the P600 emerged, all violation types also show a negative deflection relative to grammatical sentences in the anterior portion of the scalp, with a left hemisphere bias.

Preceding the P600, in the time window associated with the LAN/N400, gender violations of the kind “masculine noun + feminine adjective” and number violations of the kind “singular noun + plural adjective” appear more negative than grammatical sentences in left and mid anterior portions of the scalp, an effect that looks small (see Figures 1 and 2). For gender, this negativity does not seem to be restricted to the LAN/N400 time window, but

appears sustained. The following statistical analyses were conducted in the time window associated with the P600 (500-1000 ms) and the late frontal negativity, and in the 250-450 ms time window, which corresponds to the LAN/N400.

3.2.1 Time Window between 500-1000 ms

Gender

Hemispheres

The results of the omnibus ANOVA revealed a significant main effect of Agreement, $F(1, 26) = 22.192$, $MSE = 1.59$, $p < .001$, driven by the fact that gender violations overall yielded more positive waveforms than grammatical sentences. This effect interacted with Hemisphere, $F(1, 26) = 29.226$, $MSE = .795$, $p < .001$, and with Anterior-Posterior, $F(1.2, 31.33) = 38.664$, $MSE = 1.14$, $p < .001$. In addition, the Agreement by Hemisphere by Anterior-Posterior interaction was significant, $F(1.41, 36.78) = 5.27$, $MSE = .196$, $p < .05$. The three-way interaction suggests that gender agreement effects vary as a function of both Hemisphere and Anterior-Posterior, consistent with the pattern depicted in Figures 1 and 3, which shows a central-posterior positivity with a slight right-hemisphere bias. Post-hoc tests (collapsing across Markedness) revealed that the main effect of Agreement was significant in Right Posterior, $F(1, 26) = 39.755$, $MSE = 1.509$, $p < .001$, Left Posterior, $F(1, 26) = 42.606$, $MSE = .599$, $p < .001$, Right Medial, $F(1, 26) = 40.38$, $MSE = .787$, $p < .001$, and Left Medial, $F(1, 26) = 4.529$, $MSE = .584$, $p < .05$, and it was driven by the fact that gender violations yielded more positive waveforms than grammatical sentences. Finally, gender violations yielded more negative waveforms than grammatical sentences in Left Anterior, $F(1, 26) = 36.874$, $MSE = .764$, $p < .001$.

Although Agreement and Markedness did not interact, we conducted planned comparisons in the regions where the P600 emerged, to further explore the impact of markedness on

agreement. These comparisons revealed no significant differences between the two types of gender violations.

Midline regions

The results of the omnibus ANOVA revealed a significant main effect of Agreement, $F(1, 26) = 17.443$, $MSE = 3.039$, $p < .001$, driven by the fact that gender violations overall yielded more positive waveforms than grammatical sentences. In addition, the Agreement by Anterior-Posterior interaction was significant, $F(1.24, 32.37) = 34.272$, $MSE = 1.935$, $p < .001$. Post-hoc tests (collapsing across Markedness) revealed that the main effect of Agreement was significant in both Midline Posterior, $F(1, 26) = 43.523$, $MSE = 2.107$, $p < .001$, and Midline Medial, $F(1, 26) = 19.51$, $MSE = 1.813$, $p < .001$, driven by the fact that gender violations yielded more positive waveforms than grammatical sentences. In addition, gender violations yielded more negative waveforms than grammatical sentences in Midline Anterior, $F(1, 26) = 5.554$, $MSE = 1.528$, $p < .05$.

As was the case in the hemispheres, planned comparisons revealed no significant differences between the two types of gender violations.

Number

Hemispheres

The results of the omnibus ANOVA revealed a significant main effect of Agreement, $F(1, 26) = 29.272$, $MSE = 1.352$, $p < .001$, driven by the fact that number violations overall yielded more positive waveforms than grammatical sentences. This effect interacted with Hemisphere, $F(1, 26) = 26.369$, $MSE = .947$, $p < .001$, and with Anterior-Posterior, $F(1.11, 28.91) = 30.586$, $MSE = 2.535$, $p < .001$, suggesting that the effects of number agreement varied across the scalp. In addition, the Agreement by Hemisphere by Anterior-Posterior was

marginal, $F(1.31, 34.03) = 3.243$, $MSE = .225$, $p = .07$, consistent with the pattern depicted in Figures 2 and 4, which shows a central-posterior positivity with a slight right-hemisphere bias. Importantly, the Agreement by Markedness by Hemisphere by Anterior-Posterior interaction was significant, $F(2, 52) = 6.298$, $MSE = .058$, $p < .01$. Since an interaction that involves the two linguistic factors of interest is crucial for the present discussion, we conducted planned comparisons by examining the Agreement by Markedness interaction in each one of the regions of interest. In Right Posterior, our analyses revealed a main effect of Agreement, $F(1, 26) = 38.882$, $MSE = 1.653$, $p < .001$, driven by the fact that number errors were more positive than grammatical sentences, and an Agreement by Markedness interaction, $F(1, 26) = 5.148$, $MSE = .4$, $p < .05$, driven by the fact that “singular noun + plural adjective” violations yielded more positive waveforms relative to their grammatical counterparts than violations of the kind “plural noun + singular adjective”. As shown in Figure 1, this larger positivity was caused by “singular noun + singular adjective” controls being slightly more negative than “plural noun + plural adjective controls” and also by “singular noun + plural adjective” errors being more positive than “plural noun + singular adjective” violations. A similar pattern emerged in Right Medial, although here the interaction was marginal (Agreement: $F(1, 26) = 58.156$, $MSE = .606$, $p < .001$; Agreement by Markedness: $F(1, 26) = 3.467$, $MSE = .38$, $p = .074$). In both Left Posterior and Left Medial, the only effect unveiled by the follow-ups was a main effect of Agreement, driven by the fact that number errors were more positive than grammatical sentences (Left Posterior: $F(1, 26) = 27.858$, $MSE = .91$, $p < .001$; Left Medial: $F(1, 26) = 4.455$, $MSE = .354$, $p < .05$). Although the Agreement by Markedness interaction was not significant in these regions, a numerical trend can be observed in the same direction as in the right hemisphere. Finally, number violations yielded more negative waveforms than grammatical sentences in Left Anterior, $F(1, 26) = 25.035$, $MSE = .983$, $p < .001$.

Midline regions

The results of the omnibus ANOVA revealed a significant main effect of Agreement, $F(1, 26) = 37.988$, $MSE = 2.578$, $p < .001$, driven by the fact that number violations overall yielded more positive waveforms than grammatical sentences. In addition, the Agreement by Anterior-Posterior interaction was significant, $F(1.21, 31.32) = 35.309$, $MSE = 1.893$, $p < .001$, and the Agreement by Markedness by Anterior-Posterior interaction was marginal, $F(2, 52) = 2.975$, $MSE = .351$, $p = .06$. Similar to the hemispheres, we carried out planned comparisons to examine the Agreement by Markedness interaction at different sites, but it was not significant. Analyses revealed a main effect of Agreement in Midline Posterior, $F(1, 26) = 43.054$, $MSE = 2.516$, $p < .001$, and Midline Medial, $F(1, 26) = 51.832$, $MSE = 1.309$, $p < .001$, driven by the fact that number violations were more positive than their grammatical counterparts.

500-1000 ms Time Window: Summary of Effects

Both number and gender violations yielded a positive deflection between 500-1000 ms time window relative to grammatical sentences, an effect which emerged in central-posterior areas of the EEG cap, with a slight right hemisphere bias (see Figures 1-4). The time and topographical distribution of this effect are consistent with the P600 (Barber & Carreiras, 2005; Guajardo & Wicha, 2014; Wicha et al., 2004). Interestingly, the P600 was unaffected by whether the offending adjective was feminine (i.e. marked) or masculine (i.e. unmarked). However, it was impacted by number markedness. It was larger for violations realized on plural (i.e. marked) than singular (i.e. unmarked) adjectives in the areas where the P600 reached its maximum. In this time window, violations also yielded a negative deflection in

the anterior regions (left-lateralized for number), which was significant for both number and gender across markedness conditions.

3.2.1 Time Window between 250-450 ms

Gender

Hemispheres

The only significant effect revealed by the omnibus ANOVA was an Agreement by Markedness by Anterior-Posterior interaction, $F(1.28, 33.28) = 7.105$, $MSE = .691$, $p < .01$. Planned comparisons were conducted to examine the Agreement by Markedness interaction at different sites. These analyses showed that the interaction was significant in Right Posterior, $F(1, 26) = 6.468$, $MSE = .358$, $p < .05$, and marginal in Left Posterior, $F(1, 26) = 3.491$, $MSE = .421$, $p = .073$. In both cases, the interaction was driven by the fact that violations of the type “masculine noun + feminine adjective” yielded more positive waveforms than grammatical sentences (Right Posterior: $F(1, 26) = 5.422$, $MSE = .585$, $p < .05$; Left Posterior: $F(1, 26) = 3.237$, $MSE = .585$, $p = .084$), signaling the onset of the P600, but “feminine noun + masculine adjective” errors did not differ from their grammatical counterparts. The Agreement by Markedness interaction was also significant in Left Anterior, $F(1, 26) = 8.035$, $MSE = .344$, $p < .01$, driven by the fact that “masculine noun + feminine adjective” yielded more negative waveforms than grammatical sentences, $F(1, 26) = 6.164$, $MSE = .539$, $p < .05$, but “feminine noun + masculine adjective” errors did not differ from their grammatical baseline.

Midline regions

Similar to the hemispheres, the omnibus ANOVA revealed an Agreement by Markedness by Anterior-Posterior interaction, $F(1.43, 37.21) = 4.963$, $MSE = .442$, $p < .05$, and a

marginal Agreement by Anterior-Posterior interaction, $F(1.26, 32.84) = 3.155$, $MSE = .828$, $p = .076$. Planned comparisons showed that the Agreement by Markedness interaction was only significant in Midline Posterior, $F(1, 26) = 5.379$, $MSE = .526$, $p < .05$, driven by the fact that violations of the type “masculine noun + feminine adjective” yielded marginally more positive waveforms than grammatical sentences, $F(1, 26) = 3.82$, $MSE = 1.13$, $p = .061$, but “feminine noun + masculine adjective” errors did not differ from their grammatical baseline.^{vii}

Number

Hemispheres

The omnibus ANOVA revealed a significant Agreement by Markedness by Hemisphere by Anterior-Posterior interaction, $F(1.45, 37.81) = 5.068$, $MSE = .053$, $p < .05$. In addition, the overall Agreement by Markedness interaction was marginal, $F(1, 26) = 3.723$, $MSE = .731$, $p = .065$. We conducted planned comparisons to examine the Agreement by Markedness interaction in the different regions. The Agreement by Markedness interaction was only significant in Right Posterior, $F(1, 26) = 6.92$, $MSE = .523$, $p < .05$, driven by the fact that errors of the type “singular noun + plural adjective” were more positive than their grammatical counterparts, $F(1, 26) = 5.468$, $MSE = .635$, $p < .05$, signaling the onset of the P600. In contrast, “plural noun + singular adjective” violations did not differ from grammatical sentences. The Agreement by Markedness interaction was also marginal in Right Medial, $F(1, 26) = 3.773$, $MSE = .374$, $p = .063$, also driven by the fact that “singular noun + plural adjective” violations were marginally more positive than their grammatical counterparts, $F(1, 26) = 3.522$, $MSE = .323$, $p = .072$, but “plural noun + singular adjective” errors did not differ from the grammatical baseline.

Midline regions

The results of the omnibus ANOVA revealed a significant Agreement by Markedness by Anterior-Posterior interaction, $F(1.53, 39.72) = 6.535$, $MSE = .446$, $p < .01$. Similar to the hemispheres, we conducted planned comparisons to examine the Agreement by Markedness interaction at different sites. The Agreement by Markedness interaction was marginal in Midline Posterior, $F(1, 26) = 3.215$, $MSE = .92$, $p = .085$, driven by the fact that “singular noun + plural adjective” violations were marginally more positive than their grammatical counterparts, $F(1, 26) = 3.123$, $MSE = 1.104$, $p = .089$, but the opposite error type did not differ from grammatical sentences. The interaction was also significant in Midline Anterior, $F(1, 26) = 7.008$, $MSE = .396$, $p < .05$, driven by the fact that “singular noun + plural adjective” errors yielded marginally more negative waveforms than grammatical sentences, $F(1, 26) = 3.888$, $MSE = .674$, $p = .059$, but “plural noun + singular adjective” errors did not differ from their grammatical counterparts.

250-450 ms Time Window: Summary of Effects

Analyses in this time window showed an emerging positivity for both “masculine noun + feminine adjective” violations and “singular noun + plural adjective” errors in posterior portions of the scalp, with a right-hemisphere bias. This is consistent with the beginning of the P600 effect, which becomes significant for all violation types in the following time window. The same types of violations which show this emerging P600 also show a negativity in the anterior portion of the scalp. For “masculine noun + feminine adjective” errors, this negativity remains sustained throughout the following time window.

4. Discussion

4.1 Grammaticality Effects

P600. The present study examined the role of morphological markedness in the online processing of number and gender agreement in Spanish. As predicted, our results revealed that both number and gender violations across markedness conditions yielded a robust P600 between approximately 500-1000 ms, an effect that emerged in central posterior electrodes of the EEG cap, with a slight right hemisphere bias (e.g. Barber & Carreiras, 2005; Guajardo & Wicha, 2014; Wicha et al., 2004). The finding that agreement violations yielded a P600 is consistent with an extensive body of ERP studies on the online processing of agreement in Spanish (Alemán Bañón et al., 2012; Barber & Carreiras, 2005; Bond, Gabriele, Fiorentino, & Alemán Bañón, 2011; Demestre, Meltzer, García-Albea, & Vigil, 1999; Guajardo & Wicha, 2014; Martín-Loeches, Nigbur, Casado, Hohlfeld, & Sommer, 2006; O'Rourke & Van Petten, 2011; Wicha et al., 2004; but see Martin, Nieuwland, & Carreiras, 2012) and other languages (Deutsch & Bentin, 2001; Hagoort, 2003; Nevins et al., 2007; Osterhout & Mobley, 1995). Although there is accumulating evidence that the P600 does not reflect exclusively linguistic processes, it is reliably found for morphosyntactic violations across languages, modalities, and tasks, and some authors have argued that it indexes processes of syntactic reanalysis and repair (e.g. Friederici et al., 1996; Osterhout & Holcomb, 1992). This account is compatible with the results of the present study, where all violation types (across feature types and markedness) were strongly associated with P600 effects.

Interestingly, the P600 was impacted by the markedness manipulation in two different ways. It emerged earlier for gender violations realized on feminine (i.e. marked) as opposed to masculine (i.e. unmarked) adjectives; it emerged earlier for number violations realized on plural (i.e. marked) as opposed to singular (i.e. unmarked) adjectives. In addition, the size of the P600 was modulated by markedness in the case of number, but not gender. We come back to these findings in *Section 4.2 Markedness Effects*.

(Late) Anterior Negativity. In the same time window where emerged the P600 (500-1000 ms), all violation types also elicited a late anterior negativity relative to grammatical sentences, in left-mid areas. This late anterior negativity is also consistent with previous studies on agreement (e.g. Sabourin & Stowe, 2004; Zawiszewski et al., 2015), although this effect has been found to be more broadly distributed in some reports (Alemán Bañón et al., 2012; Gillon-Dowens et al., 2010). In studies using an explicit task, such as a grammaticality judgment, this late negativity has been interpreted as a reflection of the cost of keeping the violations in working memory for the purposes of providing the grammaticality judgment at the end of the sentence. This interpretation is consistent with the results of our study, where participants showed very high accuracy rates in the judgment task, suggesting that they successfully maintained the ungrammaticalities in working memory.

In other studies, the presence of a late anterior negativity has been linked to the cost of an unfulfilled prediction. For example, Lau, Holcomb, and Kuperberg (2013) examined lexical prediction with a word-pair priming paradigm that manipulates the proportion of semantically related pairs, and found a broadly-distributed late negativity (500-800 ms) for failed predictions. The authors suggest that the late negativity might reflect the working memory costs of participants' going back to the predictive cue (i.e. the prime) to re-evaluate its semantic association with the target. Along similar lines, Otten and Van Berkum (2008) found a similar but more delayed negativity (900-1100 ms) for adjectives that disagreed in gender with a noun that was predicted on the basis of context, but not yet encountered. Given that the structure where we tested agreement made an adjective carrying agreement features quite likely to appear, one tentative interpretation is that the late anterior negativity in our study reflects the cost of a failed prediction regarding the agreement features of the target adjective. Since this late negativity has been mainly reported in agreement studies using an explicit task (Alemán Bañón et al., 2012; Gillon-Dowens et al., 2010; Sabourin & Stowe,

2004; Zawiszewski et al., 2015) this would suggest that the parser is more likely to compute agreement as a top-down process when participants' attention is drawn towards grammatical properties of the stimuli. In addition, since the late anterior negativity was not impacted by markedness, this would suggest that, contra to Nevins et al.'s proposal (2007), agreement resolution carries on as a top-down process regardless of markedness, at least as reflected by this component.

Yet another possibility is that the late anterior negativity reflects a phase reversal of the P600. As we mentioned earlier, the P600 in our study was most robust in posterior electrodes of the right hemisphere, consistent with a number of studies on agreement (e.g. Wicha et al., 2004; Barber & Carreiras, 2005, Tanner & Van Hell, 2014; see also Guajardo & Wicha, 2014, where the late portion of the P600 showed a slight right-hemisphere bias). Thus, it is possible that the dipole generating the electrical activity reflected by the P600 was oriented in such a way that its negative and positive ends were captured by left anterior and right posterior electrodes, respectively (e.g. see Nunez & Srinivasan, 2006). Importantly, previous studies on agreement have reported a similar polarity inversion of the P600. This is the case in the study by Barber and Carreiras (2005), who report a right-lateralized P600 in the 700-900 ms time window, and a negativity in left anterior regions in the same time window. Zawiszewski et al. (2015) report a similar effect in a study comparing person and number agreement in Basque. In their study, three types of violations (person, number, and combined "person + number") yielded both a P600 (with a slight right hemisphere bias) and a late negativity in the P600 time window. Interestingly, the conditions that showed a statistically larger P600 also showed a statistically larger late anterior negativity, suggesting that the two components might not be completely independent from one another (although as pointed out by Osterhout & Holcomb (1999), two independent components can still be sensitive to the same factor). In our study, we found that the P600 was larger for "singular + plural" errors

relative to “plural + singular” errors. If the late negativity were just a phase reversal of the P600, we would expect it to also be larger for “singular + plural” errors, relative to the reverse error type, an effect which we did not find (the Agreement by Markedness interaction was not significant in either Left Anterior or Midline Anterior). Therefore, we cautiously interpret these findings as evidence that the late anterior negativity is, at least to some extent, independent from the P600. Future research should examine the conditions under which this late anterior negativity is elicited and the factors that modulate it (e.g., working memory, failed prediction).

(Early) Anterior Negativity. Preceding the P600 (between 250-450 ms), violations where the offending feature was marked yielded a small but statistically robust anterior negativity. Thus, one possible interpretation for these results is that the feature checking process that some authors assume the LAN to index is more likely to be impacted when the violating feature is marked, due to feature activation. This would constitute evidence that abstract markedness asymmetries impact agreement, in line with current morphological theories. It would also be in line with the proposal by Hoen et al. (2007) that only marked features trigger feature checking, although our design and analyses are very different from theirs. We note, however, that this interpretation must be taken with caution for two reasons. First, the morphology of the LAN effects in the present study is atypical. For “masculine noun + feminine adjective” errors, it is left-lateralized but appears sustained, since it overlaps with the late negativity in the 500-1000 ms time window. It is, therefore, unclear whether this effect is the same one as the Left Anterior Negativity reported in previous agreement studies (see Molinaro et al., 2011). In the case of “singular noun + plural adjective” errors, the early negativity is more clearly distinguished from the late negativity in the P600 time window, but it is maximal in Midline Anterior (although a numerical trend can also be observed in Left Anterior).

Second, such a proposal is at odds with previous reports that have examined number and gender agreement with a similar design to ours. For example, Barber and Carreiras (2005) found equally-robust LAN effects for gender violations realized on both marked and unmarked elements, and for number violations that were always realized on marked items. If markedness were deterministic with respect to LAN elicitation, we would have expected number errors in Barber and Carreiras' study (2005) to show a larger LAN. Likewise, Nevins et al. (2007) do not report the LAN for either number or gender errors despite the fact that they were always realized on marked predicates.

It is, thus, unclear why certain agreement violations in the present study showed a Left Anterior Negativity and others did not, a finding that is problematic for accounts linking the LAN to automatic morphosyntactic processing (as Friederici, 2002 does). However, this within-study variability with respect to LAN elicitation is not surprising if we take into account the number of studies in the literature which do not report the LAN (Spanish: Alemán Bañón et al., 2012; Bond et al., 2011; Guajardo & Wicha, 2014; Martín-Loeches et al., 2006; Wicha et al., 2004; Dutch: Hagoort, 2003; French: Frenck-Mestre, Osterhout, McLaughlin, & Foucart, 2008; Hindi: Nevins et al., 2007). A series of recent studies have made different proposals to account for this across-study variability in LAN elicitation (Molinaro et al., 2011a; Molinaro et al., 2015; Guajardo & Wicha, 2014; Tanner & Van Hell, 2014). For example, Molinaro et al. (2011a) point out that the LAN is less likely to emerge in studies using a left hemisphere reference, which can cause a left-lateralized component like the LAN to be subtracted or severely reduced. Molinaro et al. (2015) directly tested this claim and found that the LAN for determiner-noun gender errors in Spanish was indeed significantly reduced with a left-mastoid reference, relative to an average reference. In the present study, the recordings were re-referenced offline to the average of the two mastoids and, yet, our results do not consistently show a Left Anterior Negativity for all violation

types. This suggests that, even if reference site impacts LAN elicitation, it is not the deterministic factor.

Another factor which has been argued to affect the presence of the LAN for agreement violations concerns the availability of overt agreement cues. For example, Molinaro, Vespignani, Zamparelli, and Job (2011) examined subject-verb agreement in Italian and found that number violations on the verb elicited a LAN when the subject NP was marked plural (*i ragazzi *corre* "the-PL boy-PL run-SG"), but not when the plural subject involved two coordinated singular NPs (*il ragazzo et la ragazza *corre* "the-SG boy-SG and the-SG girl-SG run-SG"). In our study, the two types of number errors showed overt number inflection (i.e. *coches barato* "car-PL cheap-SG"; *coche baratos* "car-SG cheap-PL"), although on different elements (noun vs. adjective). Interestingly, the anterior negativity only emerged when the offending adjective showed overt number inflection (*coche baratos* "car-SG cheap-PL"), which might have affected the LAN (cf. Alemán Bañón et al., 2012). Notice, however, that the type of number error which was associated with the LAN in Molinaro et al.'s study (2011b) did not show overt number inflection on the critical verb either, but on the subject (e.g. *i ragazzi *corre* "the-PL boy-PL run-SG"). In addition, the results for the gender conditions in the present study are also at odds with Molinaro et al.'s proposal (2011b). In our study, both types of gender violations showed overt gender inflection on the adjective (*coche *barata* "car-MASC cheap-FEM"; *catedral *inmensa* "cathedral-FEM huge-MASC"). However, only "masculine noun + feminine adjective" errors showed a (sustained) left anterior negativity, suggesting that the availability of inflectional cues might impact the LAN, but is also not deterministic.

Some authors have also argued that the LAN is essentially the same component as the N400, with which it shares polarity, latency, and neural generators (e.g. Guajardo & Wicha, 2014; Service et al., 2007). For example, Service et al. (2007) found similar neural activation for morphosyntactic violations and semantic violations in the left hemisphere in the N400

time window (although activation in the right hemisphere was limited to semantic incongruities). More recently, Tanner and Van Hell (2014) have argued that individual differences with respect to processing strategy can create the illusion of a LAN, with some subjects relying more on lexical semantic processing and showing an N400, and others relying more on combinatorial processing and showing a P600 for the same errors. Notice, however, that an account of individual differences cannot explain why, in the present study, the same subjects showed a LAN for certain agreement errors but not others.

4.2 Markedness Effects

An interesting finding in the present study is that markedness impacted both the onset of the P600 and the size of the P600, although the latter effect only emerged for number.

P600 Onset. The P600 emerged earlier for gender errors realized on feminine (i.e. marked) adjectives, relative to those realized on masculine adjectives. In addition, it emerged earlier for number violations realized on plural (i.e. marked) adjectives, compared to those realized on singular ones. This is consistent with the possibility that the parser can more easily detect and diagnose agreement errors when the mismatching feature is marked, in line with previous reports which have interpreted differences in ERP onset as an indication that certain ungrammaticalities are easier to detect and diagnose than others (Friederici, 1998; Friederici et al., 2001; Kaan, 2002; McKinnon & Osterhout, 1996; Nevins et al., 2007). This is the case with the study by Friederici et al. (2001), who compared the processing of object-first relative clauses and object-first complement clauses in German. Although both structures are licensed, subject-first is the preferred word order, which may lead to a temporary misparse of object-first structures in the absence of disambiguating cues. Crucially, these two object-first structures differ with respect to the relative ease in diagnosing the structural changes that are necessary to revise the sentences in case of a

misparse. Friederici et al. found that the easier to diagnose structure (i.e. object relatives) showed an earlier positivity (relative to its preferred counterpart) than the structure that was harder to diagnose.

In the domain of agreement, Nevins et al. (2007) examined four types of subject-verb violations in Hindi (gender, number, “gender + number”, and “person + gender”) and found that the P600 for violations involving the person feature exhibited an earlier onset than all other violation types. The authors interpret these findings as evidence that person violations in Hindi are easier to detect because the person feature is orthographically more marked and salient. In our study, number violations realized on plural adjectives were also orthographically more salient and morphologically more complex than the reverse type of number error. This is because the former error type involves offending inflection on the target adjective (e.g. *uniforme... *sucios* “uniform-SG...dirty-PL”), but the latter involves missing inflection (e.g. *uniformes... *sucio* “uniform-PL...dirty-SG”). Therefore, the earlier P600 that we found for number errors realized on plural adjectives is consistent with Nevins et al.’s (2007) proposal for person violations in Hindi. However, in our study, gender violations realized on feminine adjectives (e.g. *uniforme... *sucia* “uniform-MASC...dirty-FEM”) were not any more orthographically salient or morphologically complex than the opposite error type (e.g. *catedral... *inmensa* “cathedral-FEM...huge-MASC”), since the adjectives that we used require overt inflection in both the masculine and the feminine forms. Therefore, we interpret these findings as evidence that markedness impacted the onset of the P600 for reasons that go beyond orthographic salience or the presence/absence of overt inflection.

Also in the domain of agreement, Kaan (2002) found that “singular subject + plural verb” violations in Dutch had an earlier onset than “plural subject + singular verb” errors. Kaan posits that the presence of a plural subject noun in the latter construction might have triggered more taxing semantic and discourse operations, thus consuming some of the processing

resources that were necessary to diagnose the subsequent violation. While this is compatible with the results of the number conditions in our study (i.e. later P600 onset for “plural noun + singular adjective” errors), it does not account for the results of the gender conditions, since there is no clear reason why masculine and feminine noun phrases should involve different semantic/discourse operations.

One potential explanation for the earlier P600 effects for violations where the mismatching feature was marked concerns the concept of underspecification. As previously discussed, recent analytical approaches to morphological markedness argue that unmarked features are underspecified, that is, they carry no feature information (e.g. Carstairs-McCarthy, 1998; Cowper, 2005; Harley & Ritter, 2002; Noyer, 1992). For number and gender, these approaches assume that only plural and feminine are fully specified and carry number and gender information, respectively (e.g. Carstairs-McCarthy, 1998; Harley & Ritter, 2002; Noyer, 1992). In contrast, singular and masculine are assumed to act as defaults. This is consistent with the observation that syntactic contexts that are fully specified as plural can be co-indexed with singular items that lack a number specification (e.g. *Ana le_i teme [a las brujas]_i* “Ana fears witches”), but not the reverse. It is also consistent with the fact that, in certain contexts, masculine inflection can modify feminine elements (i.e. a complex phrase that involves conjoined masculine and feminine nouns requires masculine inflection on all agreement targets), but not the reverse. In other words, Spanish native speakers find evidence in the input that singular and masculine can co-occur with plural and feminine, but the reverse combination is disallowed. Thus, in the present study, it is possible that violations where the trigger noun was underspecified but a fully marked adjective was forced to agree with it (number: *uniforme... *sucios* “uniform-SG...dirty-PL; gender: e.g. *uniforme... *sucia* “uniform-MASC...dirty-FEM) were easier to diagnose, as they were more disruptive. In turn, this might have modulated the onset of the P600.

P600 Amplitude. The P600 was larger for “singular noun + plural adjective” violations, relative to the reverse type of number error. A priori, these results also are consistent with our proposal that violations where the agreement trigger is underspecified (e.g. singular) but the target is fully marked (e.g. plural) are more disruptive than the opposite error type. This interpretation is, however, slightly compromised by the fact that P600 size was similar for both types of gender errors.

One possibility is that the processes of revision and repair reflected by P600 amplitude are more impacted by markedness in the sense of orthographic salience or morphological complexity. As mentioned above, number errors where the offending feature was plural were more salient/complex than the reverse type of number error, but the same was not true for the two types of gender errors. This might explain why P600 amplitude was impacted by number but not gender markedness. This would be consistent with previous studies that have examined the role of markedness on subject-verb agreement. For example, Deutsch and Bentin (2001) found that gender violations in Hebrew yielded a larger P600 when the subject and verb were marked for number, relative to contexts where they were unmarked, which they relate to plural being more salient and morphologically complex than singular. Likewise, Kaan (2002) found a larger P600 for subject-verb number violations in Dutch realized on marked (i.e. plural) verbs, relative to unmarked ones.

This would also be consistent with previous studies which have found the P600 to be sensitive to salience. For example, the study on Hindi subject-verb agreement by Nevins et al. (2007) found that the more orthographically salient “person + gender” violations yielded a larger P600 than “gender + number” errors. Likewise, Coulson et al. (1998) compared two types of morphosyntactic violations which differed with respect to salience (case violations: *the plane took us/*we...*; agreement violations: *he mows/*mow...*) and found that the more salient case violations yielded a larger P600.^{viii} Given that our participants were asked to

judge the grammaticality of the sentences, it is possible that the larger P600 for “singular noun + plural adjective” errors results from a contribution of the P3b component for salient and task-relevant stimuli. Future studies should examine the extent to which markedness impacts agreement with less explicit tasks.

Thus, if we bring together the effects that we found on both P600 onset and P600 amplitude, our results suggest that agreement is impacted by, at least, two dimensions of markedness, one that concerns the unbalanced representation of marked and unmarked features at the level of linguistic representation, and one that concerns more superficial aspects of markedness, such as orthographic salience or morphological complexity.

The fact that our results assign a certain role to orthographic or morphological complexity raises the following question. Can markedness at the abstract level be reduced to purely distributional differences in terms of length or frequency of use (e.g. Haspelmath, 2006)? Remember that P600 amplitude was larger for number errors on plural adjectives, which were longer and less frequent than their singular counterparts. In contrast, P600 amplitude was similar for gender errors realized on masculine vs. feminine adjectives, which were of the same length and were controlled for lexical frequency. Two observations from our data speak against this possibility. First, P600 onset was modulated by markedness for both number and gender. Since the adjectives in the gender conditions were matched with respect to both length and frequency, these two factors alone cannot explain why gender errors realized on feminine adjectives were detected earlier than the reverse type of gender error. Second, these purely lexical differences should have impacted the N400 component (frequency: Neville, Mills, & Lawson, 1992; Van Petten, 1993; Van Petten & Kutas, 1990; length: Pratarelli, 1995; Osterhout et al., 2002), but not the amplitude of the P600. That is, effects of frequency and length would have resulted in plural adjectives yielding a larger N400 than singular ones in the grammatical conditions, an effect that we did not find. In fact, Figure 1 shows a trend

in the opposite direction. That is, singular adjectives were slightly more negative than plural ones, although this difference was only numerical.

Further support against a purely length-based explanation for the markedness effects in the present study comes from the previous ERP literature comparing different agreement features. For example, Barber and Carreiras (2005) found that noun-adjective gender violations in Spanish yielded a larger P600 than number violations, in spite of the fact that (1) the latter were longer and (2) both number and gender violations involved overt offending inflection (gender: *faro es *alta* “lighthouse-MASC is high-FEM”; number: *faro es *altos* “lighthouse-SG is high-PL”). Likewise, Alemán Bañón et al. (2012) found equally robust P600s for number and gender errors in two different syntactic configurations, despite the fact that number violations were also longer (gender: *cuento es *anónima* “story-MASC is anonymous-FEM”; number: *cuento es *anónimos* “story-SG is anonymous-PL”) (see also O’Rourke & Van Petten, 2011).

Top-down vs. Bottom-up Processing. Our results are not in line with Nevins et al.’s proposal that marked feature values engage predictive (i.e. *top-down*) processing to a greater extent than unmarked ones. Under this account, upon encountering a marked noun, the parser generates a stronger prediction regarding the specification of the adjective. If this proposal is on the right track, violations of the kind “marked noun + unmarked adjective” should have yielded a larger (and more broadly distributed, spanning over frontal areas) P600 relative to the reverse error type. This is because integration of the violating adjective should have been more costly when a strong prediction was generated but proven wrong. Our results, however, revealed the opposite pattern, although in the case of P600 amplitude this was only obvious for number. One possibility is that the predictive mechanisms reflected by the P600 mainly concern syntactic category (e.g. Kaan et al., 2000). For example, Alemán Bañón et al. (2012) examined noun-adjective agreement in two configurations that differed with respect to how

predictable the critical adjective was, and found differences in the P600 time window that were not restricted to ungrammatical sentences, but also emerged for grammatical ones (consistent with Kaan et al., 2000). In the present study, the material intervening between the agreeing elements was the same across the two markedness conditions and, therefore, the syntactic category of the adjectives was not any more predictable in one configuration than the other.

An anonymous reviewer suggested that the crossed design that we used (i.e. the fact that, for example, gender errors with a masculine trigger included both singular and plural nouns) might have obscured the extent to which markedness impacted predictive processing. This is because half of the trigger nouns that were unmarked for gender were marked for number (e.g. *coche*, *coches* “car-MASC-SG”, “car-MASC-PL”) and vice versa. Our motivation for using this crossed design was that a variety of studies comparing different agreement features have provided evidence that feature checking proceeds independently for each feature. This is the case in the study by Antón Méndez et al. (2002), who compared both features in a design that also manipulates markedness (see also Igoa et al., 1999). And it is also consistent with the ERP literature examining double agreement violations, including Silva-Pereyra et al. (2007) and Zawiszewski et al. (2015), both of whom found additivity (a larger P600) in the processing of combined number + person violations, relative to single person and number violations (cf. Nevins et al., 2007), suggesting that feature checking is not necessarily a unitary process. It is therefore unlikely that the lack of predictive processing in our study is due to the crossed design that we used. Future ERP studies should examine the relation between markedness and predictive processing by examining instances of agreement where the trigger noun is doubly marked (e.g. feminine and plural) to those where it is completely unmarked for both number and gender.

Some authors have also suggested that frontal positivities in the P600 time window reflect a failed prediction (Van Berkum et al., 2005; Wicha et al., 2004; see also Van Petten & Luka, 2012), while more posterior ones reflect reprocessing (i.e. reanalysis, repair). Under this account, if marked features are used predictively, we might have expected the P600 for errors of the kind “marked noun + unmarked adjective” to show a broader scalp distribution, reflecting the summation of the effects caused by a failed prediction and the effects of reanalysis/repair, a pattern of results that we did not find.

4.3 Effects of Feature Type

Overall, number and gender showed similar processing profiles, consistent with most previous ERP studies that have compared the two features (Alemán Bañón et al., 2012; Gillon-Dowens et al., 2010; Nevins et al., 2007). They were, however, slightly differently impacted by markedness, as measured by P600 amplitude (the onset of the P600 was also impacted by markedness, but similarly for the two features). That is, number violations yielded a larger P600 when realized on plural adjectives, but no asymmetry was found for the two types of gender errors. There is a series of differences in how number and gender were tested in the present study that might account for this. First, the target adjective was always preceded by a verbal predicate carrying number, but not gender inflection. Although further analyses comparing P600 size for number and gender violations overall revealed no reliable differences, it is possible that this additional number cue facilitated the emergence of markedness effects for number, especially if we take into account that it was adjacent to the critical adjective where ERPs were calculated (e.g., *uniforme que parecía sucios* “uniform that looked-SG dirty-PL, *uniformes que parecían sucio* “uniform that looked-PL dirty-SG). Future studies comparing the impact of markedness on the two features should explore this

issue, for example, by examining agreement in a local domain, where no number-marked verbs intervene between the agreeing words.

Another difference concerns the trigger nouns that we used (e.g., *uniforme* “uniform-MASC”, *catedral* “cathedral-FEM”). These nouns do not show canonical gender marking in Spanish (i.e., masculine *-o* and feminine *-a*) but they do follow the prototypical rules for pluralization of Spanish (e.g., *uniforme* “uniform-SG” vs. *uniformes* “uniform-PL”; *catedral* “cathedral-SG vs. *catedrales* “cathedral-PL”). Thus, the absence of strong gender cues on the nouns might have resulted in similar revision processes for both types of gender errors, which might explain why P600 amplitude was unaffected by gender markedness.

Finally, another important difference in how the two features were examined is that both types of gender violations involved the presence of offending inflection on the critical adjective where ERP effects were calculated (*coche *barata* “car-MASC cheap-FEM”; *catedral *inmensa* “cathedral-FEM huge-MASC”), whereas in the case of number, this was only true for “singular noun + plural adjective” errors (*coche *baratos* “car-SG cheap-PL”). For the reverse type of number error, the adjective showed no number inflection (*coches *barato* “car-PL cheap-SG”). It is likely that the presence of incorrect inflection impacts the reanalysis/repair processes indexed by P600 amplitude to a greater extent than missing inflection, which would account for the results of the number conditions in our study. This would be consistent with the results by Vigliocco and Frank (1999), who examined the production of gender agreement in French and Italian and found a feminine-masculine asymmetry in French, where only marked (i.e. feminine) adjectives required the addition of inflection, but not in Italian, where both masculine and feminine adjectives required inflection. Future studies examining gender agreement in Spanish should compare adjectives that show inflection for both the masculine and feminine values, to those which are only inflected in the feminine (e.g. *inglés, inglesa* “English-MASC” “English-FEM”). If a feminine-masculine asymmetry emerges only for the

latter type of adjective, this would provide evidence for our proposal, and it would provide further evidence that the processing of the number and gender features is qualitatively similar. Another way to address this question would be to compare number agreement in languages like Spanish, where plural requires the addition of overt inflection, and languages like Italian, where pluralization requires different (as opposed to additional) inflection (e.g. *macchina*/*macchine* “car-SG/car-PL”), similar to how gender is morphologically realized in Spanish.^{ix}

5. Conclusion

To our knowledge, the present study is the first to use ERP to examine the unique contribution of morphological markedness to agreement processing in the nominal domain. Our results provide electrophysiological evidence that markedness impacts the establishment of agreement dependencies in several ways. First, agreement errors were detected earlier when the mismatching feature was marked (“singular noun + plural adjective”, “masculine noun + feminine adjective”), as suggested by the earlier onset of the P600, an effect that emerged for both number and gender. Along similar lines, these two error types also yielded LAN-like effect in the 250-450 ms time window, albeit without the typical morphology of the Left Anterior Negativity. In addition, number errors were found to be more disruptive when the mismatching feature was plural, relative to errors realized on singular adjectives, as suggested by P600 amplitude differences between the two types of number errors. Our results also add to a body of ERP studies showing that number and gender are processed similarly at the brain level (Alemán Bañón et al., 2012; Gillon-Dowens et al., 2010; Nevins et al., 2007). Interestingly, number and gender were somewhat differently affected by markedness, which may be due to differences in how number and gender were tested rather than qualitative differences between the two features.

Acknowledgments

The authors thank Robert Fiorentino, Alison Gabriele, and Ianthi Tsimpli for their feedback on different aspects of this project. We also thank Pilar Gray-Carlos and María de Mier for their help in recruiting participants, Dave Miller for his help with data collection, Steve Politzer-Ahles for his help in preparing the manuscript, and three anonymous reviewers for their insightful comments and suggestions. José Alemán Bañón was supported by a postdoctoral fellowship from the Spanish Ministry of Economy and Competitiveness (FPDI-2013-15813).

References

- Acuña Fariña, J. C. (2009). The linguistics and psycholinguistics of agreement: A tutorial overview. *Lingua*, 119, 389–424. doi:10.1016/j.lingua.2008.09.005
- Acuña Fariña, J. C., Meseguer, E., & Carreiras, M. (2014). Gender and number agreement in comprehension in Spanish. *Lingua*, 143, 108–128. doi: 10.1016/j.lingua.2014.01.013
- 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. doi: 10.1016/j.brainres.2012.03.057
- Antón-Méndez, I., Nicol, J. L., & Garrett, M. F. (2002). The relation between gender and number agreement processing. *Syntax*, 5, 1–25. doi: 10.1111/1467-9612.00045
- Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience*, 17, 137–153. doi: 10.1162/0898929052880101
- Bates, E., Devescovi, A., Pizzamiglio, L., D'Amico, S., & Hernandez, A. (1995). Gender and lexical access in Italian. *Perception and Psychophysics*, 57, 847–862. doi: 10.3758/BF03206800

- Battistella, E. L. (1990). *Markedness: The evaluative superstructure of language*. Albany, NY: SUNY Press.
- Bock, K., & Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language*, 31, 99–127. doi: 10.1016/0749-596X(92)90007-K
- Bock, K., & Eberhard, K. M. (1993). Meaning, sound, and syntax in English number agreement. *Language and Cognitive Processes*, 8, 57–99. doi: 10.1080/01690969308406949
- Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, 23, 45–93. doi: 10.1016/0010-0285(91)90003-7
- Bond, K., Gabriele, A., Fiorentino, R., & Alemán Bañón, J. (2011). Individual differences and the role of the L1 in L2 processing: An ERP investigation. In J. Herschensohn, & D. Tanner (Eds.), *Proceedings of the 11th Generative Approaches to Second Language Acquisition* (pp. 17–29). Somerville, MA: Cascadia Proceedings Project.
- Bonet, E. (1995). Feature structure of romance clitics. *Natural Language and Linguistic Theory*, 13, 607–47. doi: 10.1007/BF00992853
- Bornkessel-Schlesewsky, I., & Schlewsky, M. (2008). An alternative perspective on “semantic P600” effects in language comprehension. *Brain Research Reviews*, 59, 55–73. doi: 10.1016/j.brainresrev.2008.05.003
- Brouwer, H., Fitz, H., & Hoeks, J. C. J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, 1446, 127–143. doi: 10.1016/j.brainres.2012.01.055
- Caffarra, S., & Barber, H. (2015). Does the ending matter? The role of gender-to-ending consistency in sentence reading. *Brain Research*, 1605, 83–92. doi: 10.1016/j.brainres.2015.02.018.

- Carminati, M. N. (2005). Processing reflexes of hierarchy (Person>Number>Gender) and implications for linguistic theory. *Lingua*, 115, 259–285. doi: 10.1016/j.lingua.2003.10.006
- Carreiras, M., Salillas, E., & Barber, H. A. (2004). Event-related potentials elicited during parsing of ambiguous relative clauses in Spanish. *Cognitive Brain Research*, 20(1), 98–105.
- Carstairs-McCarthy, A. (1998). How lexical semantics constrains inflectional allomorphy. In G. Booij, & J. van Marle (Eds.). *Yearbook of morphology* (pp. 1–24). Dordrecht: Kluwer.
- Carstens, V. (2000). Concord in minimalist theory. *Linguistic Inquiry*, 31, 319–355. doi: 10.1162/002438900554370
- Cinque, G. (1978). La sintassi dei pronomi relativi ‘cui’ e ‘quale’ nell’italiano moderno. *Rivista di Grammatica Generativa*, 3, 31–126.
- Colé, P., & Seguí, J. (1994). Grammatical incongruency and vocabulary types. *Memory and Cognition*, 22, 387–394. doi: 10.3758/BF03200865
- Corbett, G. G. (1991). *Gender*. Cambridge, U.K.: Cambridge University Press.
- Corbett, G. G. (2000). *Number*. Cambridge, U.K.: Cambridge University Press.
- Corbett, G. G. (2006). *Agreement*. Cambridge, U.K.: Cambridge University Press.
- Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, 13, 21–58. doi: 10.1080/016909698386582
- Cowper, E. (2005). A note on number. *Linguistic Inquiry*, 36, 441–455. doi: 10.1162/0024389054396935

- DeLong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology*, 48, 1203–1207. doi: 10.1111/j.1469-8986.2011.01199.x
- De Vincenzi, M. (1999). Differences between the morphology of gender and number: evidence from establishing coreferences. *Journal of Psycholinguistic Research*, 28, 537–553. doi: 10.1023/A:1023272511427
- De Vincenzi, M., & Di Domenico, E. (1999). A distinction among phi-features: The role of gender and number in the retrieval of pronoun antecedents. *Rivista di Linguistica*, 11, 41-74.
- Demestre, J., Meltzer, S., García-Albea, J., & Vigil, A. (1999). Identifying the null subject: Evidence from event-related brain potentials. *Journal of Psycholinguistic Research*, 28, 293–312. doi: 10.1023/A:1023258215604
- Deutsch, A., & Bentin, S. (2001). Syntactic and semantic factors in processing gender agreement in Hebrew: Evidence from ERPs and eye movements. *Journal of Memory and Language*, 45, 200–224. doi: 10.1006/jmla.2000.2768
- 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. doi: 10.3758/s13428-013-0326-1
- Eberhard, K. M. (1997). The marked effect of number on subject-verb agreement. *Journal of Memory and Language*, 36, 147–164. doi: 10.1006/jmla.1996.2484
- Eberhard, K. M., Cutting, J. C., Bock, J. K. (2005). Making syntax of sense: number agreement in sentence production. *Psychological Review*, 112, 531–559. doi: 10.1037/0033-295X.112.3.531
- Faussart, C., Jakubowicz, C., & Costes, M. (1999). Gender and number processing in spoken French and Spanish. *Rivista di Linguistica*, 11, 75–101.

- Fayol, M., Largy, P., & Lemaire, P. (1994). When cognitive overload enhances subject-verb agreement errors. A study in French written language. *Quarterly Journal of Experimental Psychology*, 47, 437–464. doi: 10.1080/14640749408401119
- Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, 1146, 75–84. doi: 10.1016/j.brainres.2006.06.101
- Field, A. (2005). *Discovering Statistics Using SPSS (2nd ed)*. London: SAGE Publications.
- Frenck-Mestre C., Osterhout, L., McLaughlin, J., & Foucart, A. (2008). The effect of phonological realization of inflectional morphology on verbal agreement in French: Evidence from ERPs. *Acta Psychologica*, 128, 528–536. doi: 10.1016/j.actpsy.2007.12.007
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6, 78–84. doi: 10.1016/S1364-6613(00)01839-8
- Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1219–1248. doi: 10.1037/0278-7393.22.5.1219
- Friederici, A. D., Mecklinger, A., Spencer, K. M., Steinhauer, K., & Donchin, E. (2001). Syntactic parsing preferences and their on-line revisions: A spatio-temporal analysis of event-related brain potentials. *Cognitive Brain Research*, 11, 305–323. doi: 10.1016/S0926-6410(00)00065-3
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1–76. doi: 10.1016/S0010-0277(98)00034-1
- Gillon Dowens, M., Vergara, M., Barber, H., & Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience*, 22, 1870–1887. doi: 10.1162/jocn.2009.2134

- Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes*, 25, 149–188. doi: 10.1080/01690960902965951
- Guajardo, L. F., & Wicha, N. Y. Y. (2014). Morphosyntax can modulate the N400 component: Event-related potentials to gender-marked post-nominal adjectives. *NeuroImage*, 91, 262–272. doi: 10.1016/j.neuroimage.2013.09.077
- Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, 15, 883–899. doi: 10.1162/089892903322370807
- Hagoort, P., & Brown, C. M. (1999). Gender electrified: ERP evidence on the syntactic nature of gender processing. *Journal of Psycholinguistic Research*, 28, 715–728. doi: 10.1023/A:1023277213129
- Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8, 439–484. doi: 10.1080/01690969308407585
- Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, 11, 194–205. doi: 10.1162/089892999563328
- Hammer, A., Jansma, B. M., Lamers, M., & Münte, T. F. (2008). Interplay of meaning, syntax and working memory during pronoun resolution investigated by ERPs. *Brain Research*, 1230, 177–191. doi: 10.1016/j.brainres.2008.07.004
- Harley, H., & Ritter, E. (2002). Person and number in pronouns: a feature geometric analysis. *Language*, 78, 482–526. doi: 10.1353/lan.2002.0158
- Harris, J. (1991). The exponence of gender in Spanish. *Linguistic Inquiry*, 22, 27–62.

- Hartsuiker, R. J., Antón Méndez, I., & van Zee, M. (2001). Object attraction in subject-verb agreement construction. *Journal of Memory and Language*, 45, 546–573. doi: 10.1006/jmla.2000.2787
- Hartsuiker, R. J., Schriefers, H. J., Bock, K., & Kikstra, G. M. (2003). Morphophonological influences on the construction of subject-verb agreement. *Memory & Cognition*, 31, 1316–1326. doi: 10.3758/BF03195814
- Haspelmath, M. (2006). Against markedness and what to replace it with. *Journal of Linguistics*, 42, 25–70. doi: 10.1017/S0022226705003683
- Hoen, M., Deprez, V., & Dominey, P. F. (2007). Do you agree? Electrophysiological characterization of online agreement checking during the comprehension of correct French passive sentences. *Journal of Neurolinguistics*, 20, 395–421. doi: 10.1016/j.jneuroling.2007.03.001
- Igoa, J. M., García Albea, J. E., & Sánchez Casas, R. (1999). Gender-number dissociations in sentence production in Spanish. *Rivista de Lingüística*, 11, 165–198. doi: 10.1.1.175.2275
- Kaan, E. (2002). Investigating the effects of distance and number interference in processing subject-verb dependencies: An ERP Study. *Journal of Psycholinguistic Research*, 31, 165–193. doi: 10.1023/A:1014978917769
- Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, 15, 159–201. doi: 10.1080/016909600386084
- Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience*, 15, 98–110. doi: 10.1162/089892903321107855

- Kayne, R. S. (1976). French relative “que”. In F. Hensey, & M. Luján (Eds.), *Current Studies in Romance Linguistics* (pp. 255–299). Washington D.C.: Georgetown University Press
- Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52, 205–255. doi: 10.1016/j.jml.2004.10.002
- Kolk, H. H. J., Chwilla, D. J., Van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, 85, 1–36. doi: 10.1016/S0093-934X(02)00548-5
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146, 23–49. doi: 10.1016/j.brainres.2006.12.063
- Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the Brain: Using Our Past to Generate a Future* (pp. 190–207). Oxford University Press.
- Kutas, M., & Hillyard, S. A. (1980). Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science*, 207, 203–205. doi: 10.1126/science.7350657
- Lau, E. F., Holcomb, P., & Kuperberg, G. (2013). Dissociating N400 effect of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience*, 25, 484–502. doi: 10.1162/jocn_a_00328
- Lukatela, G., Kostic, A., Todorovic, D., Carello, C., & Turvey, M. T. (1987). Type and number of violations and the grammatical congruency effect in lexical decision. *Psychological Research*, 49, 37–43. doi: 10.1007/BF00309201
- 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. doi: 10.1016/j.brainres.2011.06.055

- Martin, A., Nieuwland, M., & Carreiras, M. (2012). Event-related brain potentials index cue-based retrieval interference during sentence comprehension. *NeuroImage*, *59*, 1859–1869. doi: 10.1016/j.neuroimage.2011.08.057
- Martín-Loeches, M., Nigbur, R., Casado, P., Hohlfeld, A., & Sommer, W. (2006). Semantics prevalence over syntax during sentence processing: A brain potential study of noun-adjective agreement in Spanish. *Brain Research*, *1093*, 178–189. doi: 10.1016/j.brainres.2006.03.094
- McKinnon, R., & Osterhout, L. (1996). Constraints on movement phenomena in sentence processing: Evidence from event-related brain potentials. *Language and Cognitive Processes*, *11*, 495–523. doi: 10.1080/016909696387132
- Molinaro, N., Barber, H. A., & Carreiras, M. (2011a). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, *47*, 908–930. doi: 10.1016/j.cortex.2011.02.019
- Molinaro, N., Barber, H. A., Caffarra, S., & Carreiras, M. (2015). On the left anterior negativity (LAN): The case of morphosyntactic agreement. *Cortex*, *66*, 156–159. doi: 10.1016/j.cortex.2014.06.009.
- Molinaro, N., Vespignani, F., Zamparelli, R., & Job, R. (2011b). Why brother and sister are not just siblings: Repair processes in agreement computation. *Journal of Memory and Language*, *64*, 211–232. doi: 10.1016/j.jml.2010.12.002
- Münte, T. F., Szentkuti, A., Wieringa, B. M., Matzke, M., & Johannes, S. (1997). Human brain potentials to reading syntactic errors in sentences of different complexity. *Neuroscience Letters*, *235*, 105–108. doi: 10.1016/S0304-3940(97)00719-2
- Neville, H. J., Mills, D. L., & Lawson, D. S. (1992). Fractionating language: Different neural subsystems with different sensitive periods. *Cerebral Cortex*, *2*, 244–258. doi: 10.1093/cercor/2.3.244

- 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. doi: 10.1016/j.brainres.2007.05.058
- Nicol, J. L., Forster, K. I., & Veres, C. (1997). Subject–verb agreement processes in comprehension. *Journal of Memory and Language, 36*, 569–587. doi: 10.1006/jmla.1996.2497
- Noyer, R. (1992). Features, projections and affixes in autonomous morphological structure. Ph.D. dissertation, MIT.
- Nunez, P. L., & Srinivasan, R. (2006). *Electric fields of the brain: The neurophysics of EEG* (2nd ed). Oxford University Press.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia, 9*, 97–113. doi: 10.1016/0028-3932(71)90067-4
- O'Rourke, P. L., & Van Petten, C. (2011). Morphological agreement at a distance: Dissociation between early and late components of the event-related brain potential. *Brain Research, 1392*, 62–79. doi: 10.1016/j.brainres.2011.03.071
- Osterhout, L., Allen, M., & McLaughlin, J. (2002). Words in the brain: Lexical determinants of word-induced brain activity. *Journal of Neurolinguistics, 15*, 171–187. doi: 10.1016/S0911-6044(01)00036-7
- Osterhout, L., & Hagoort, P. (1999). A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King and Kutas (1998) in the P600/SPS-P300 debate. *Language and Cognitive Processes, 14*, 1–14. doi: 10.1080/016909699386356
- Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785–806. doi: 10.1016/0749-596X(92)90039-Z

- Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, 34, 739–773. doi: 10.1006/jmla.1995.1033
- Osterhout, L., Nicol, J. L. (1999). On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes*, 14, 283–317. doi: 10.1080/016909699386310
- Otten, M., & van Berkum, J. J. A. (2008). Discourse-based word anticipation during language processing: prediction or priming? *Discourse Processes*, 45, 464–496. doi: 10.1080/01638530802356463
- Padovani, R., Calandra-Buonaura, G., Cacciari, C., Benuzzi, F., & Nichelli, P. (2005). Grammatical gender in the brain: evidence from an fMRI study on Italian. *Brain Research Bulletin*, 65 (4), 301–308. doi: 10.1016/j.brainresbull.2004.11.025.
- Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: an event-related potential study. *Journal of Cognitive Neuroscience*, 10, 717–733. doi: 10.1162/089892998563121
- Payne, T. E. (1997). *Describing morphosyntax: A guide for field linguists*. Cambridge, U.K.: Cambridge University Press.
- Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, 41, 427–456. doi: 10.1006/jmla.1999.2653
- Picallo, M. C. (1991). Nominals and nominalizations in Catalan. *Probus*, 3, 279–316. doi: 10.1515/prbs.1991.3.3.279
- Picallo, M. C. (2008). Gender and number in Romance. *Lingue e Linguaggio*, 7, 47–66. doi: 10.1418/26709

- Pollard, C., & Sag, I. A. (1988). An information-based theory of agreement. In: G. Brentari, G. Larson, & L. McLeod (Eds.), *Papers from the 24th Annual Regional Meeting of the Chicago Linguistics Society* (pp. 236-257). Chicago: Chicago Linguistics Society.
- Pratarelli, M. E. (1995). Modulation of semantic processing using word length and complexity: An ERP study. *International Journal of Psychophysiology*, 19, 233–246. doi: 10.1016/0167-8760(95)00015-K
- Ritter, E. (1991). Two functional categories in noun phrases: Evidence from Modern Hebrew. In S. D. Rothstein (Ed.), *Syntax and semantics 25: Perspectives on phrase structure: Heads and licensing* (pp. 37–62). New York: Academic Press.
- Ritter, E. (1993). Where is gender? *Linguistic Inquiry*, 24, 795–803. doi: 35400002376664.0090
- Rivero, M. L. (1980). That-relatives and deletion in COMP in Spanish. *Cahiers linguistiques d'Ottawa*, 9, 383–399.
- Rivero, M. L. (1982). Restrictive relatives with que. *Nueva Revista de Filología Hispánica*, 31, 195–234.
- Sabourin, L., & Stowe, L. A. (2004). Memory effects in syntactic ERP tasks. *Brain and Cognition*, 55, 392–395. doi: 10.1016/j.bandc.2004.02.056
- Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, 137, 29-39. doi: 10.1016/j.bandl.2014.07.010
- Schiller, N. O., Munte, T. F., Horemans, I., & Jansma, B. M. (2003). The influence of semantic and phonological factors on syntactic decisions: An event-related brain potential study. *Psychophysiology*, 40, 869-877. doi: 10.1111/1469- 8986.00105.

- Service, E., Helenius, P., Maury, S., & Salmelin, R. (2007). Localization of syntactic and semantic brain responses using magnetoencephalography. *Journal of Cognitive Neuroscience*, 19, 1193–1205. doi: 10.1162/jocn.2007.19.7.1193
- Shieber, S. (1986). *An Introduction to Unification-based Approaches to Grammar* (CSLI Lecture Notes 4). Stanford: CSLI.
- Silva-Pereyra, J. F., & Carreiras, M. (2007). An ERP study of agreement features in Spanish. *Brain Research*, 1185, 201–211. doi: 10.1016/j.brainres.2007.09.029
- Tagliaferri, B. (2005). *Paradigm*. Perception Research Systems, Inc.
www.perceptionresearchsystems.com
- Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia*, 56, 289–301. doi: 10.1016/j.neuropsychologia.2014.02.002
- Teschner, R. V., & Russell, W. M. (1984). The gender patterns of Spanish nouns: An inverse dictionary-based analysis. *Hispanic Linguistics*, 1(1), 115–32.
- Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 443–467. doi: 10.1037/0278-7393.31.3.443
- Van de Meerendonk, N., Kolk, H. J., Vissers, C. Th. W. M., & Chwilla, D. J. (2010). Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, 22, 67–82. doi: 10.1162/jocn.2008.21170
- Van Petten, C. (1993). A comparison of lexical and sentence-level context effects in event-related potentials. *Language and Cognitive Processes*, 8, 485–532. doi: 10.1080/01690969308407586

- Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83, 176–190. doi: 10.1016/j.ijpsycho.2011.09.015
- Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition*, 18, 380–393. doi: 10.3758/BF03197127
- Vigliocco, G., & Franck, J. (1999). When sex and syntax go hand in hand: Gender agreement in language production. *Journal of Memory and Language*, 40, 455–478. doi: 10.1006/jmla.1998.2624
- Vigliocco, G., & Franck, J. (2001). When sex affects syntax: Contextual influences in sentence production. *Journal of Memory and Language*, 45, 368–390. doi: 10.1006/jmla.2000.2774
- Wagers, M. W., Lau, E., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61, 206–237. doi: 10.1016/j.jml.2009.04.002
- Wechsler, S., & Zlatic, L. (2003). *The Many Faces of Agreement*. Stanford: CSLI.
- Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, 16, 1272–1288. doi: 10.1162/0898929041920487
- Zawiszewski, A., Santesteban, M., & Laka, I. (2015). Phi-features reloaded: An ERP study on person and number agreement processing. *Applied Psycholinguistics*. doi: 10.1017/S014271641500017

ⁱ Harley and Ritter (2002) and Corbett (2000, p. 154) suggest that, for languages with a two-way number distinction, this pattern (i.e. plural being marked with respect to singular) holds true cross-linguistically. For

grammatical gender, the issue is more complicated, since there are languages with a two-way gender system where nouns are not classified along the masculine vs. feminine divide. Corbett (1991, p. 206-207, 220) discusses at least three languages with a masculine-feminine system where the feminine value appears to be the unmarked one (Khasi, Maasai, and Zayse), suggesting that markedness asymmetries for gender may be language-specific.

ⁱⁱ For a review of different theoretical approaches to agreement, the reader is referred to Acuña-Fariña (2009).

See also Corbett (2006); Pollard and Sag (1988); Shieber (1986); Wechsler & Zlatić (2003).

ⁱⁱⁱ This is the case for nouns where no gender-based biological distinction is possible.

^{iv} Data from one additional participant who performed more than two standard deviations below the mean in all ungrammatical conditions of the judgment task were excluded from analysis.

^v Caffarra and Barber (2015) did not manipulate markedness, suggesting that there is no clear account of how markedness impacts gender agreement when distributional cues are absent. In addition, Caffarra and Barber (2015) only tested determiner-noun gender agreement (i.e. linearly and structurally local), but it is possible that the lack of distributional cues impacts gender agreement when the agreeing elements are not adjacent and belong to different phrases.

^{vi} These tests were carried out as follow-ups, and we thank an anonymous reviewer for this suggestion. To measure cloze probability, the sentences were truncated right before the critical word and participants were asked to complete the sentences with the first adjective that came to their mind. Results from 40 native speakers of Spanish who did not participate in the EEG experiment revealed that the cloze probability of the sentences ranged from 0 to .55, and that the mean cloze per condition was low (for gender, conditions 1 and 4: .022; conditions 7 and 10: .025; for number, conditions 1 and 7: .024; conditions 4 and 10: .022) mainly because the sentences were not designed to be semantically constraining. Crucially, the difference across conditions was not significant ($p > .2$ for both comparisons). In order to measure the plausibility of the sentences, they were truncated at the critical adjective. A total of 40 native speakers of Spanish who did not participate either in the cloze test or in the EEG task rated how plausible the sentences were on a Likert scale ranging from 1 (completely implausible) to 7 (completely plausible). The mean plausibility per condition was above 5.5 for all conditions (for gender, conditions 1 and 4: 5.8; conditions 7 and 10: 5.55; for number, conditions 1 and 7: 5.71; conditions 4 and 10: 5.62) and the difference between conditions was not significant ($p > .3$ for both comparisons).

^{vii} Follow-ups to the marginal Agreement by Anterior-Posterior interaction showed no effects.

^{viii} A related finding in Coulson et al.’s study is that P600 amplitude was reduced when the proportion of morphosyntactic violations was higher. In this respect, our design included an equal number of “singular noun + plural adjective” and “plural noun + singular adjective” violations, suggesting that—in terms of proportion—one violation type was not any more salient/probable than the other.

^{ix} We thank an anonymous reviewer for this suggestion.

Table 1: Sample stimuli for the experimental conditions

FEMININE SINGULAR NOUN	
<i>Grammatical</i>	
1. Carlos fotografió una <u>catedral</u> que parecía <u>inmensa</u> para una revista. Carlos photographed a cathedral-FEM-SG that looked huge-FEM-SG for a magazine	
<i>Number Violation</i>	
2. Carlos fotografió una <u>catedral</u> que parecía <u>*inmensas</u> para una revista. Carlos photographed a cathedral-FEM-SG that looked huge-FEM-PL for a magazine	
<i>Gender Violation</i>	
3. Carlos fotografió una <u>catedral</u> que parecía <u>*inmenso</u> para una revista. Carlos photographed a cathedral-FEM-SG that looked huge-MASC-SG for a magazine	
FEMININE PLURAL NOUN	
<i>Grammatical</i>	
4. Carlos fotografió unas <u>catedrales</u> que parecían <u>inmensas</u> para una revista. Carlos photographed some cathedral-FEM-PL that looked huge-FEM-PL for a magazine	
<i>Number Violation</i>	
5. Carlos fotografió unas <u>catedrales</u> que parecían <u>*inmensa</u> para una revista. Carlos photographed some cathedral-FEM-PL that looked huge-FEM-SG for a magazine	
<i>Gender Violation</i>	
6. Carlos fotografió unas <u>catedrales</u> que parecían <u>*inmenso</u> para una revista. Carlos photographed some cathedral-FEM-PL that looked huge-MASC-PL for a magazine	
MASCULINE SINGULAR NOUN	
<i>Grammatical</i>	
7. Antonio hizo un <u>pastel</u> que parecía <u>asqueroso</u> para el desayuno. Antonio made a cake-MASC-SG that looked disgusting-MASC-SG for the breakfast	
<i>Number Violation</i>	
8. Antonio hizo un <u>pastel</u> que parecía <u>*asquerosos</u> para el desayuno. Antonio made a cake-MASC-SG that looked disgusting-MASC-PL for the breakfast	
<i>Gender Violation</i>	

9. Antonio hizo un <u>pastel</u> que parecía <u>*asquerosa</u> para el desayuno. Antonio made a cake-MASC-SG that looked disgusting-FEM-SG for the breakfast
MASCULINE PLURAL NOUN
<i>Grammatical</i>
10. Antonio hizo unos <u>pasteles</u> que parecían <u>asquerosos</u> para el desayuno. Antonio made some cake-MASC-PL that looked disgusting-MASC-PL for the breakfast
<i>Number Violation</i>
11. Antonio hizo unos <u>pasteles</u> que parecían <u>*asqueroso</u> para el desayuno. Antonio made some cake-MASC-PL that looked disgusting-MASC-SG for the breakfast
<i>Gender Violation</i>
12. Antonio hizo unos <u>pasteles</u> que parecían <u>*asquerosas</u> para el desayuno. Antonio made some cake-MASC-PL that looked disgusting-FEM-PL for the breakfast

Table 2. Distribution of the experimental materials across conditions and number of items per condition
(between parentheses)

	Feminine Noun	Masculine Noun
Singular	<u>Condition 1</u>	<u>Condition 7</u>
	GR (20)	GR (20)
	<u>Condition 2</u>	<u>Condition 8</u>
	*NUM (20)	*NUM (20)
	<u>Condition 3</u>	<u>Condition 9</u>
	*GEN (20)	*GEN (20)
Plural	<u>Condition 4</u>	<u>Condition 10</u>
	GR (20)	GR (20)
	<u>Condition 5</u>	<u>Condition 11</u>
	*NUM (20)	*NUM (20)
	<u>Condition 6</u>	<u>Condition 12</u>
	*GEN (20)	*GEN (20)

GR: grammatical; *NUM: number violation; *GEN: gender violation

Table 3. Mean accuracy rates in the grammaticality judgment task for the conditions examining agreement with marked versus unmarked nouns. Results are reported separately for number and gender (GR: grammatical; UN: ungrammatical).

Gender				Number			
Marked (FEM) Noun		Unmarked (MASC) Noun		Marked (PL) Noun		Unmarked (SG) Noun	
GR	UN	GR	UN	GR	UN	GR	UN
93.9	95	94.5	96.6	93.7	93.7	94.6	93.9

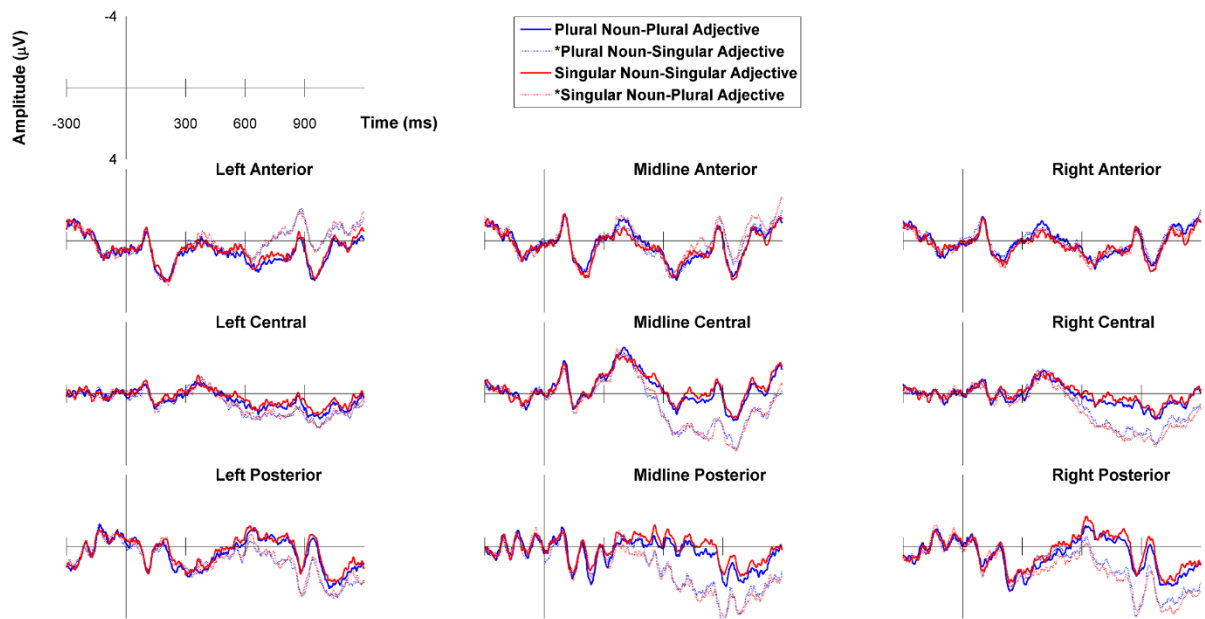
Figure Captions

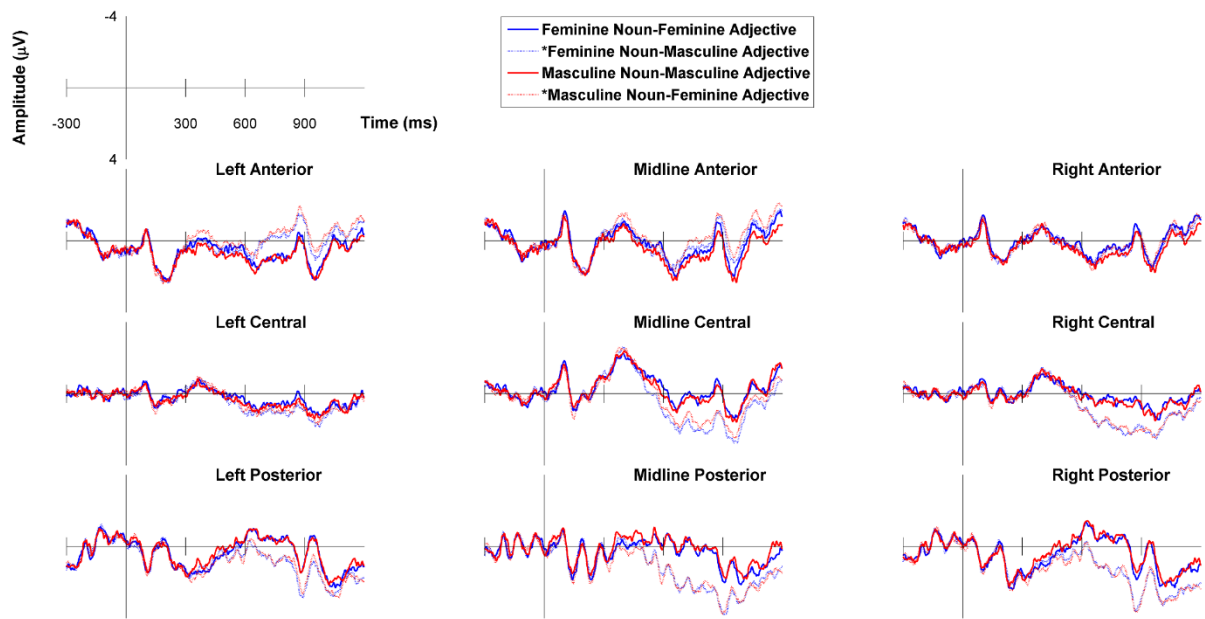
Figure 1. Grand average ERP waveforms for the conditions examining number agreement with plural (marked) versus singular (unmarked) nouns: plural noun-plural adjective (grammatical), plural noun-singular adjective (ungrammatical), singular noun-singular adjective (grammatical), singular noun-plural adjective (ungrammatical). ERPs are plotted for the average of all electrodes within each region of interest.

Figure 2. Grand average ERP waveforms for the conditions examining gender agreement with feminine (marked) versus masculine (unmarked) nouns: feminine noun-feminine adjective (grammatical), feminine noun-masculine adjective (ungrammatical), masculine noun-masculine adjective (grammatical), masculine noun-feminine adjective (ungrammatical). ERPs are plotted for the average of all electrodes within each region of interest.

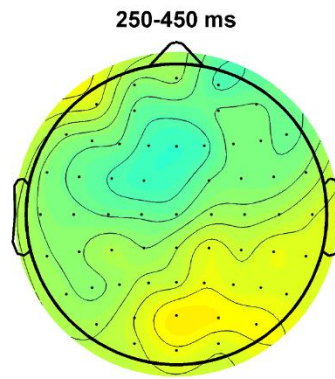
Figure 3. Topographic plots of the number effects in the plural (marked) versus singular (unmarked) noun conditions in the 250-450 ms and 500-1000 ms time window. Plots were computed by subtracting the grammatical sentence from the violation condition.

Figure 4. Topographic plots of the gender effects in the feminine (marked) versus masculine (unmarked) noun conditions in the 250-450 ms and 500-1000 ms time window. Plots were computed by subtracting the grammatical sentence from the violation condition.

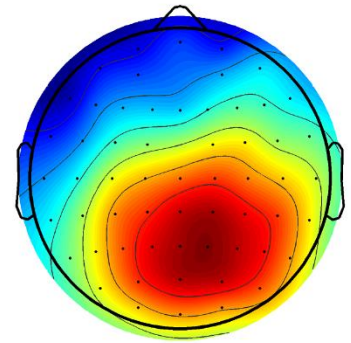




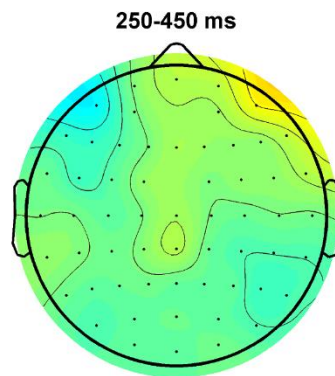
SINGULAR NOUN
PLURAL ADJECTIVE



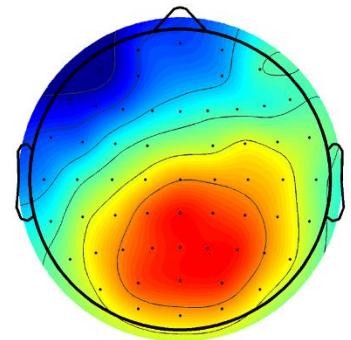
500-1000 ms



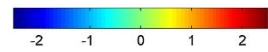
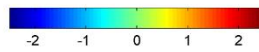
PLURAL NOUN
SINGULAR ADJECTIVE



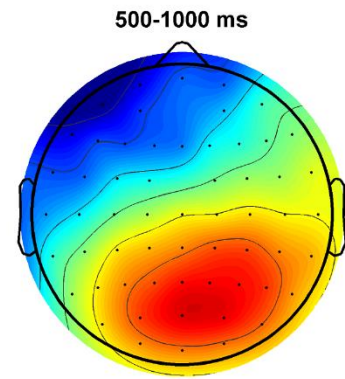
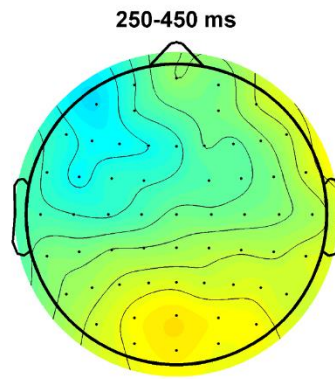
500-1000 ms



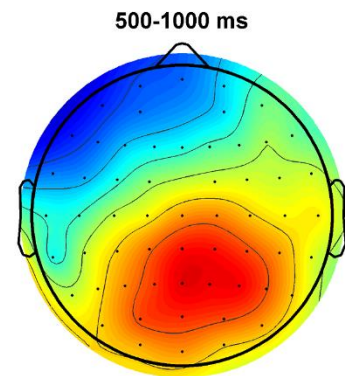
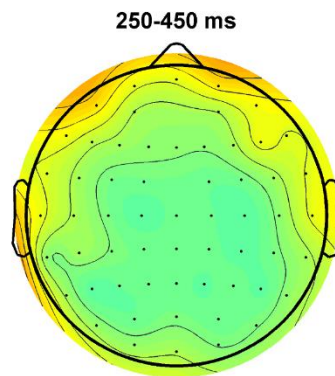
Voltage μV



MASCULINE NOUN
FEMININE ADJECTIVE



FEMININE NOUN
MASCULINE ADJECTIVE



Voltage μV

