

# Grammatical Gender and Number Agreement in Spanish: An ERP Comparison

Horacio Barber<sup>1</sup> and Manuel Carreiras<sup>2</sup>

## Abstract

■ The role of grammatical gender and number representations in syntactic processes during reading in Spanish was studied using the event-related potentials (ERPs) technique. The electroencephalogram was recorded with a dense array of 128 electrodes while Spanish speakers read word pairs (Experiment 1) or sentences (Experiment 2) in which gender or number agreement relationships were manipulated. Disagreement in word pairs formed by a noun and an adjective (e.g., *faro-alto* [*lighthouse-high*]) produced an N400-type effect, while word pairs formed by an article and a noun (e.g., *el-piano* [*the-piano*]) showed an additional left anterior negativity effect (LAN). Agreement violations with the same words inserted in sentences (e.g., *El piano estaba viejo y desafinado* [*the m-s piano m-s was old m-s and off-key*]) resulted in a pattern of LAN-P600. This effect was found both when the violation occurred in the middle of the sentence (at the adjective), as

well as when this happened at the beginning of the sentence (at the noun), but the last segment of the P600 effect was greater for the middle sentence position, which could indicate differences in the complexity of reanalysis processes. Differences between grammatical gender and number disagreement were found in late measures. In the word pairs experiment, P3 peak latency varied across conditions, being later for gender than for number disagreement. Similarly, in the sentence experiment, the last segment of the P600 effect was greater for gender than for number violations. These event-related potentials (ERPs) effects lend support to the idea that reanalysis or repair processes after grammatical disagreement detection could involve more steps in the case of gender disagreement, as grammatical gender is a feature of the lexical representation in contrast to number, which is considered a morphological feature that combines with the stem of the word. ■

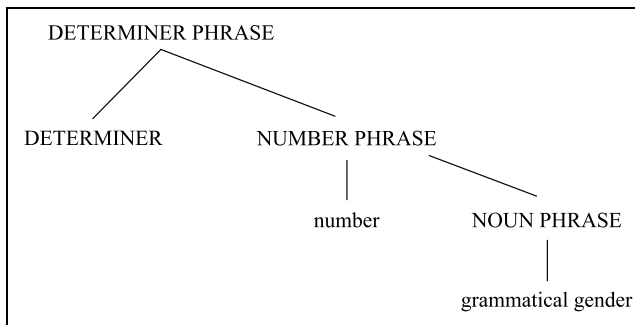
## INTRODUCTION

Agreement rules have an important role in parsing and language comprehension in general, especially in richly inflected languages. They are important and necessary information for computing grammatical dependencies between the different elements of a sentence in order to build its syntactic structure. Furthermore, agreement contributes to discourse cohesion, establishing long-distance references across sentences. Among the features used to compute agreement are the marking of gender and number, as well as others such as person or case. Gender can be a conceptual characteristic or a formal property of words. Semantic gender is marked for words referring to animate entities, there being a transparent relationship between the biological sex of the referents and the gender of nouns and pronouns. Many languages, in addition to semantic gender, also have gender as an exclusively grammatical property of many nouns. In this case, there is no conceptual basis for the distinction of gender and this can be considered as arbitrarily assigned and invariable because only one form (feminine or masculine) is assigned to each noun

(Corbett, 1991). Number, on the other hand, is always considered a conceptual feature signaling the quantity of the referent, and therefore the number morphological representation of any concept is variable because it can adopt at least two different forms, singular and plural. For this reason, number has been considered a morphological marker that combines with the stem it modifies (Ritter, 1988), whereas grammatical gender has been considered a fixed property of stem (Harris, 1991). As we will see below, this lexical feature remains as one of the most important differences between gender and number, with relevant consequences at the syntactic level.

Gender and number have been studied from both formal and processing perspectives. A concern of the formal approach is how to represent gender and number features syntactically. According to some authors (e.g., Ritter, 1988), noun phrases can be considered complements of a higher syntactic projection, the determiner phrase (see Figure 1). Thus, between the noun phrase and the determiner phrase, there could be other functional heads. If a feature is a syntactic head, that feature is available to the syntactic parser for all syntactic operations, such as movement, co-referential processes, or agreement. A feature independently represented in the lexicon must have semantic content and therefore has to be variable. Number fulfils the requisites of having

<sup>1</sup>University of California, <sup>2</sup>Universidad de La Laguna, Tenerife, Spain



**Figure 1.** Formal representation of grammatical gender and number in the determiner phrase structure (see text).

variability and semantic content and so can be considered as syntactic head. Therefore, some authors, when describing the structure of the determiner phrase, consider that number is realized as the head of an independent functional projection (namely, the number phrase) which is situated between the determiner and the noun (Ritter, 1988). On the other hand, grammatical gender, a formal feature of the noun, is not variable and cannot be selected from the lexicon independently of the rest of the noun features. Consequently, Ritter (1993) argues that gender can never be the head of its own syntactic projection. According to this proposal, under some circumstances there will be different consequences for grammatical gender and number syntactic processing.

There are language production data suggesting that gender and number agreement are probably processed or represented in different ways. For example, the rates of gender and number agreement errors within a language usually differ considerably, being greater for number than for gender, which could be interpreted as evidence for different agreement mechanisms (Vigliocco, Butterworth, & Garrett, 1996). In addition, Igoa, García-Albea, and Sánchez-Casas (1999), in an analysis of error production, found that exchange errors affected mostly number suffixes, seldom number and gender simultaneously and never gender alone. These differences in the error rates suggest that gender is retrieved directly from the lexicon and assigned to the phrase structure together with the lemma, because grammatical gender is part of the lemma, while number is derived by rule. Therefore, gender morphemes should be stranded less often than number morphemes, which are hypothesized to be assigned directly to the phrases as closed class items (Bock & Levelt, 1994; Garrett, 1982). This proposal has been recently supported by another production study in Spanish in which the number of combined gender and number errors was not greater than that expected at random, showing that these features are processed independently (Antón-Méndez, Nicol, & Garret, 2002).

Processing of gender and number agreement in comprehension has been studied using grammatical priming

with word pairs. The grammatical priming effect reveals that a word is recognized faster when preceded by another word that agrees with it than when preceded by a word that does not agree (Lukatela, Kostic, Todorovic, Carello, & Turvey, 1987). Using this procedure, the empirical evidence on whether gender and number agreement mechanisms are similar is mixed. Whereas two studies have not found differences between gender and number processing (Colé & Seguí, 1994; Lukatela et al, 1987), Faussart, Jakubowitz, and Costes (1999) reported longer recognition times when words disagreed in gender than when they disagreed in number. These authors explain this effect within the framework of classical models of lexical retrieval, according to which the target word is retrieved in three successive stages: lexical access, recognition, and integration. The first step would be the process by which the right lexical entry is located and lexical identification carried out. In the second step, the relevant lexical content of that entry would be accessed, that is, semantic information, grammatical category, or morphological information. Finally, the third stage would include all postlexical processes of integration concerning the context. Grammatical agreement would take place at this last stage. Because gender is a stem inherent feature, integration process failure regarding gender agreement would make the system go back to the lexical identification stage in order to check if the right entry had been selected. Nevertheless, number is not considered a stem inherent feature, so if number agreement is not detected, the processor would only have to check the final processes of recognition without returning to the initial processes of lexical access. The difference in access time employed by the processor when detecting a gender as compared to a number inconsistency would reflect the additional cost of going back one more step.

A different proposal that predicts differences for gender and number agreement during syntactic processing comes from the studies of De Vincenzi (1999) and De Vincenzi and Di Domenico (1999), who examined the use of morphological information in pronoun resolution mechanisms in Italian. They found that although both number and gender are used to activate an antecedent at the end of a sentence, only number information reactivates the antecedent immediately after pronouns within the sentence. They explain these differences in number and gender information relying on serial models of language processing (Frazier, 1987), according to which the construction of structural representations initially employs only information relevant to syntactic analysis. In this first step, only features with autonomous projections on syntax, such as number (following Ritter, 1993), would be used to establish coherence between pronoun and noun. On the contrary, gender information, which does not have separate identity inside the syntactic structure, would only be considered at a later stage with lexical and semantic information.

In sum, the previous studies seem to suggest that there are differences between gender and number in their representation at the lexical level. However, it is not yet clear how and when these representational differences influence agreement during syntactic processing.

### **Electrophysiological Studies**

Agreement processes have also been studied with the ERP technique, as its multidimensional character and excellent temporal resolution make this technique especially suitable for the study of when agreement relationships are made during reading. Pioneer studies have shown different effects of grammatical agreement violations. Kutas and Hillyard (1983) found an increase of negativity between 200 and 500 msec in anterior zones as a response to the reading of number agreement errors between subject and verb in a sentence. In contrast, Hagoort, Brown, and Groothusen (1993) described a P600/SPS effect (an increase of positivity from 500 msec) in response to the same type of violation. Similar effects to those described by Kutas and Hillyard have been linked to different aspects of syntactic analysis and labeled left anterior negativity (LAN) (Friederici, 1995). This anterior negativity could be reflecting, in fact, some kind of verbal working memory load, as proposed by some authors (Kluender & Kutas, 1993). The P600/SPS effect has been also linked to syntactic parsing and to reanalysis processes during syntactic processing. Various authors have differentiated at least two phases in this effect which could be related to functionally different states of the parser: the first one, between 500 and 750 msec, shows a broad distribution along the midline of the scalp, and the second one, after 750 msec, is more localized in parietal areas (Hagoort, Brown, & Osterhout, 1999).

In another study, Osterhout and Mobley (1995) compared number agreement violations between subject and verb, between a reflexive pronoun and its antecedent, and between a personal pronoun and its antecedent. In the three cases, the words that infringed agreement rules produced an amplitude increase at the P600 time window. Only in the case of noun-verb agreement was an additional LAN-type effect found. In these experiments, pronouns were also manipulated to produce gender agreement violations. The effects of gender violations were never different from those of number (i.e., P600 effect), and both were different from the semantic violations. Semantic violations usually produce a significant enhancement of the N400 component, a negative waveform associated to lexical-semantic integration processes (see review in Kutas & Federmeier, 2000). Other studies have also reported the P600 effect in response to grammatical gender agreement violations in languages in which (in contrast with English) gender is not exclusively semantically based (e.g., Hagoort & Brown, 1999, in Dutch). In addition, in a German study,

the P600 effect was preceded by a LAN effect (Gunter, Friederici, & Schriefers, 2000). This LAN-P600 pattern was confirmed in Spanish when agreement violations of two types of gender (semantic and grammatical) were compared in the same experiment. Although additional differences between the semantic and the grammatical gender processing were found, the LAN-P600 response to the agreement violations was equal to both types of gender (Barber, Salillas, & Carreiras, 2004).

Münste and Heinze (1994) undertook a series of experiments in which the effect of grammatical priming in word pairs was analyzed. Among other types of syntactic relationships, gender agreement was manipulated in pairs formed by an article followed by a noun. The morphosyntactic effects were compared with other phonological and semantic effects. Moreover, the experiments were undertaken in different languages (German and Finnish) and with different tasks (grammatical judgment or lexical decision). Results showed that both agreement and verb inflection errors produced anterior distributed negativities and that these effects seem to reflect relatively automatic processes that are independent of the task. The authors point to the resemblance of these negativities with the LAN effect found in sentence contexts. Inversely, the effects produced by phonological or semantic relationships produced negativities with more posterior distributions. In a similar study, Barber and Carreiras (2003) found an effect of the N400 type in response to gender or number disagreement in adjective-noun word pairs. No differences between gender and number effects were observed in this time window. However, the P3 component that followed this effect presented a longer latency in the case of gender disagreement. The sensitiveness of the amplitude of the P3 towards characteristics such as the probability of stimulus occurrence or its subjective relevance for the task indicates that when this component is produced, the necessary processes for the categorization of the stimulus must have finalized (Donchin, 1979). Although the classic effect of this component has been located around 300 msec, several studies have shown that this latency can be retarded depending on the complexity of the stimulus (Kutas, McCarthy, & Donchin, 1977). For this reason, this component usually shows up at around 500-msec delay with linguistic stimuli. These same latency variations can occur even with the same stimuli when its categorization difficulty varies. For example, in lexical decision tasks, pseudowords produce higher latencies than words do (Bentin, McCarthy, & Wood, 1985), and repeated words have shorter latencies than those not repeated (Rugg, 1985). However, it should be borne in mind that the relationship between the P600 and the P3 is an open issue. Some authors have claimed that the P600/SPS belongs to the family of P300 effects (Coulson, King, & Kutas, 1998) although this interpretation is not shared by others (Osterhout & Hagoort, 1999). The latency differences in the experiment

of Barber and Carreiras (2003) could be related, according to the authors, to reanalysis or rechecking processes occurring after morphological integration failure. Thus, reanalysis processes after gender disagreement detection could be associated with a higher cost because, as suggested by Faussart et al. (1999), gender is a lexical feature and therefore checking of gender agreement includes both integration and lexical access processes, while number disagreement only requires checking of the integration step.

### The Present Study

In summary, the behavioral data as well as the ERP data on gender and number agreement provide a mixed picture. The behavioral data are contradictory but they indicate that there could be differences between gender and number, at least in their representation. The ERP data from different experiments seem to suggest that gender and number disagreement produce, at least, qualitatively similar effects. Nonetheless, the ERP data available at present come from different designs in different languages. As far as we know, only the study of Osterhout and Mobley (1995) has directly compared the ERP effects of gender and number agreement violations in the same experiment. Because this experiment was performed in English, only semantic gender could be manipulated. However, Spanish is a language with a rich morphology in which agreement relationships in general, and grammatical gender agreement in particular, are highly relevant in order to build the correct syntactic structure of sentences. This characteristic of the Spanish language leads us to the present research, in which the processing of both types of agreement was compared under the same circumstances with Spanish speakers. The main goal of the present experiments was to study the possible differences in the processing of gender and number, and to discriminate between the effects due to different lexical representations that could trigger different reanalysis processes (Faussart et al., 1999) and effects due to time-course differences in agreement processing during the building of a syntactic structure (De Vincenzi, 1999; De Vincenzi & Di Domenico, 1999). We compared the ERP effects of gender and number agreement violations in the context of lexical integration and syntactic processing. In the first experiment, we compared the effect of agreement in article–noun word pairs (e.g., *el-piano* [*the-piano*]) and in noun–adjective word pairs (e.g., *faro-alto* [*lighthouse-high*]), while in the second experiment the agreement between the same words was manipulated at the beginning of sentences (e.g., *El piano estaba viejo y desafinado* [*the m-s piano m-s was old and off-key*]), or in the middle of sentences (e.g., *El faro es alto y luminoso* [*the m-s lighthouse m-s is high m-s and bright*]). This way, comparing the results of both experiments, we could also trace the impact of the sentence context and reading expectations in the

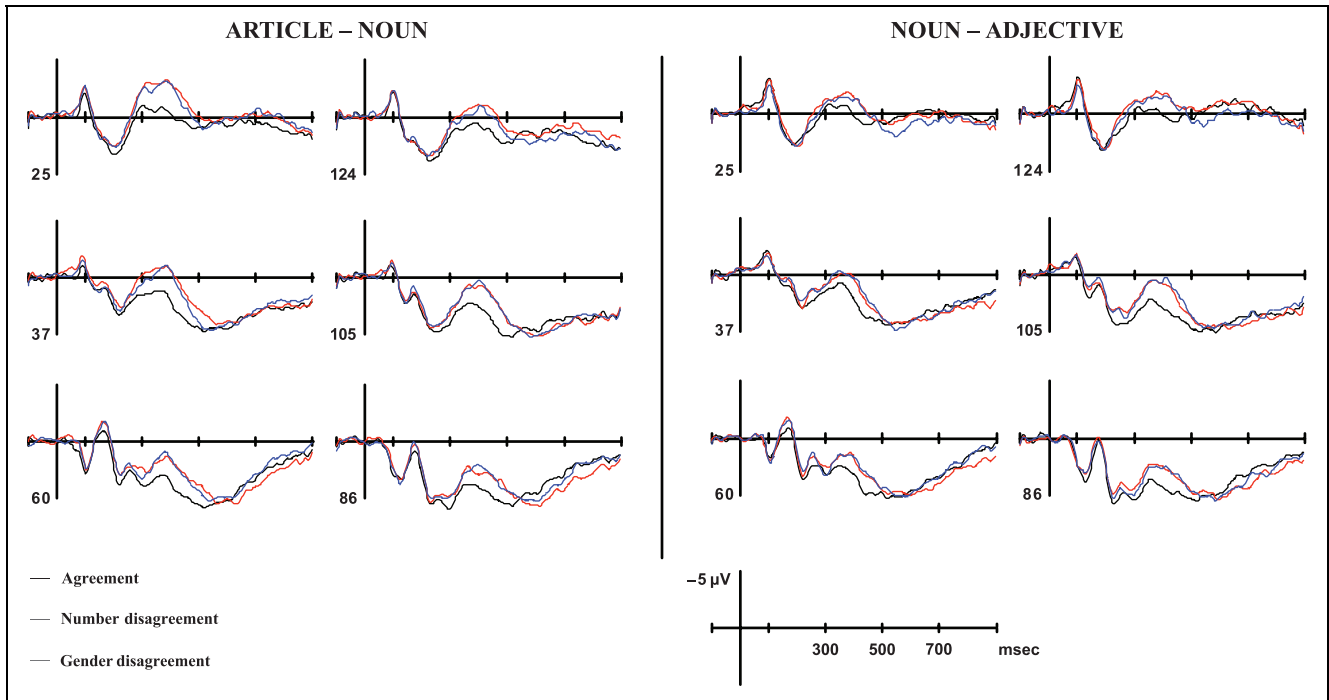
ERP components associated to syntactic processing, as well as the influence of the amount of syntactic information available during the course of sentence reading.

### EXPERIMENT 1: GENDER AND NUMBER AGREEMENT IN WORD PAIRS

Grammatical priming was analyzed in word pairs in which gender and number agreement between an article and a noun (e.g., *el-piano* [*the-piano*]), and between a noun and an adjective (e.g., *faro-alto* [*lighthouse-high*]) were manipulated. This setting provides an opportunity to investigate gender and number in the same stimuli and with the simplest syntactic context. In a previous experiment, word pairs composed by a noun and an adjective showed the same posterior negativity (N400) in response to gender and number disagreement (Barber & Carreiras, 2003). In this new experiment, we expected that the article–noun pairs (which, in contrast with adjective–noun pairs, can be considered a noun phrase) would trigger more syntactic processing and therefore a more anterior effect, similar to that found by Münte and Heinze (1994) with this type of stimuli. This way, we checked possible differences between gender and number in this very early agreement processing, as should be expected following the proposal of De Vincenzi (1999) and De Vincenzi and Di Domenico (1999). In addition, we expected to replicate previously reported late effects (i.e., P3 latency variations) that support the idea of different reanalysis costs due to differences in the lexical representation of gender and number (Barber & Carreiras, 2003).

### Results

Figure 2 shows grand averages of the agreement, the gender disagreement, and the number disagreement conditions for the two types of word pairs. For a clearer description of the data, in these and other figures, only one representative electrode of each analysis region is plotted. Visual inspection reveals clear differences in the responses to the agreement conditions with respect to the disagreement conditions. Targets for the disagreement word pairs start to differ from those of the well-formed pairs at 250–300 msec. Between 300 and 500 msec, disagreement waves were more negative-going than those to the agreement condition, with the largest amplitude differences at 400 msec. Furthermore, in this time window, no differences between disagreement conditions can be appreciated. As can be seen in Figure 3, the disagreement effects are broadly distributed over the scalp. The effects associated to article–noun word pairs are localized both in the frontal areas and in the central and posterior electrodes, while noun–adjective pair effects are more restricted to central and posterior areas. The subtraction of the respective topographical maps shows that differences across the



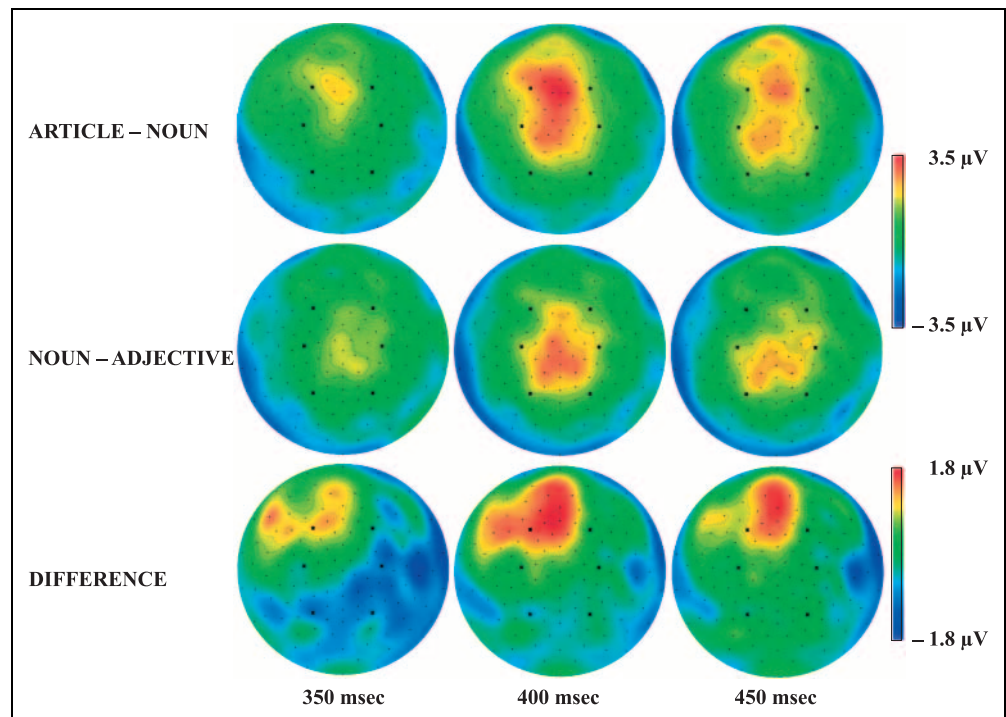
**Figure 2.** Grand averages of the target words for the agreement, gender disagreement, and number disagreement conditions, in article–noun pairs (left) and in noun–adjective pairs (right). Vertical line marks the onset of target word presentation ( $t = 0$ ). Six representative electrodes are represented (see locations in Figure 7).

two types of word pairs are localized in the left anterior area.

The N400 component was followed by the P3 component, which was present in all conditions and showed similar amplitudes across these. However, systematic peak latency variations were appreciated. This compo-

nent peaks earlier in the agreement condition than in both disagreement conditions. More detailed inspection of averaged electrode groups revealed longer peak latency for the gender disagreement condition as compared with the number disagreement condition and statistical analysis confirmed these observations.

**Figure 3.** Topographical maps obtained by interpolation from 128 electrodes at 350, 400, and 450 msec. Maps were computed from values resulting from the subtraction between agreement and disagreement waves (gender and number conditions collapsed). The first row shows the effects in article–noun word pairs, the second row in noun–adjective pairs, and the third row shows the differences between the effects of both types of word pairs.



### 300–500 msec Time Window: Amplitude Differences

Mean amplitude values corresponding to the 300–500 msec time epoch were introduced in an ANOVA with the “grammatical agreement” factor (agreement, gender disagreement, and number disagreement), the “pair type” factor (article–noun and noun–adjective), the “electrode region” factor (anterior, central, and posterior), and the “hemisphere factor” (left and right). This analysis revealed a marginally significant four-way interaction [ $F(4,18) = 2.41$ ;  $p = .08$ ;  $e = 0.64$ ], which was confirmed after the data normalization [ $F(4,18) = 2.75$ ;  $p = .05$ ;  $e = 0.64$ ]. Post hoc tests showed reliable differences between the agreement condition and the gender disagreement condition [ $F(1,21) = 33.24$ ;  $p < .001$ ] and between the agreement condition and the number disagreement condition [ $F(1,21) = 30.55$ ;  $p < .001$ ] in the article–noun pairs, as well as similar differences between the agreement condition and the gender disagreement condition [ $F(1,21) = 9.13$ ;  $p < .01$ ] and between the agreement condition and the number disagreement condition [ $F(1,21) = 7.64$ ;  $p < .01$ ] in the noun–adjective pairs. However, no differences between the disagreement conditions were found in any case ( $F < 1$ ). Table 1 shows the  $F$  values and significance levels of gender and number disagreement effects in each type of word pairs and each electrode region. Table 1 indicates different scalp distributions of the effects across both types of word pairs, while article–noun pairs show reliable effects in the left anterior electrode region that do not happen with the noun–adjective pairs.

### P3 Component: Peak Latencies Analyses

Maximum amplitude latencies between 450 and 750 msec were introduced in an ANOVA with the “grammatical agreement” factor (agreement, gender disagreement,

and number disagreement) and the “electrode region” factor (left and right). Because of the posterior distribution of the P3 component, only the posterior electrode regions were included. This analysis yielded a significant main effect of grammatical agreement [ $F(2,18) = 34.74$ ;  $p < .001$ ]. Post hoc tests showed significant differences in the article–noun pairs between the agreement condition (543 msec) as compared to both the gender disagreement condition [602 msec;  $F(1,19) = 22.58$ ;  $p < .001$ ] and the number disagreement condition [582 msec;  $F(1,19) = 13.87$ ;  $p < .01$ ], and between the gender disagreement and number disagreement conditions [ $F(1,19) = 7.56$ ;  $p < .05$ ]. Similar effects were found in the comparisons with noun–adjective pairs, between agreement (533 msec) and gender disagreement [598 msec;  $F(1,19) = 51.79$ ;  $p < .001$ ], agreement and number disagreement [576 msec;  $F(1,19) = 19.33$ ;  $p < .001$ ], and between gender disagreement and number disagreement conditions [ $F(1,19) = 9.12$ ;  $p < .01$ ].

### Discussion

The results show that grammatical disagreement for each word pair produced a remarkable effect between 300 and 500 msec. Disagreement conditions presented waves with higher negative amplitude than those of agreement. These differences were distributed over central and posterior areas in both types of word pairs, while left anterior effects are found only for the article–noun pairs.

The differences that were found in central–posterior areas fit with the classical N400 effect. This effect has been reported associated to semantic priming between words and has usually been associated with semantic integration processes (see Osterhout & Holcomb, 1995, for a review). However, the manipulation of other word

**Table 1.**  $F$  Values and Significance Levels Resulting from Post Hoc Comparisons between the Average Amplitude Values of the Experimental Conditions (Agreement versus Gender Disagreement and versus Number Disagreement), for Each Word Pair Type (Article Plus Noun and Noun Plus Adjective) and for Each Electrode Region (Anterior, Central, and Posterior) in Each Hemisphere (Left and Right)

$df = 1, 19$	Article–Noun Word Pairs				Noun–Adjective Word Pairs			
	Agreement vs. Gender Disagreement		Agreement vs. Number Disagreement		Agreement vs. Gender Disagreement		Agreement vs. Number Disagreement	
	Left H.	Right H.	Left H.	Right H.	Left H.	Right H.	Left H.	Right H.
Anterior	30.04***	15.43**	28.54***	11.55**	– <sup>a</sup>	7.22*	– <sup>a</sup>	7.05*
Central	37.00***	25.30***	27.46***	15.42**	3.60	15.80**	5.06*	18.33**
Posterior	15.21**	10.33**	15.68**	13.06**	5.86*	14.05**	3.70	8.18**

\*\*\* $p > .001$ .

\*\* $p > .01$ .

\* $p > .05$ .

<sup>a</sup> $F < 1$ .

features can result in a similar N400 effect. For instance, rhyme matching between words can result in a reduction of the N400 component, but these phonological effects, in contrast with semantic ones, have been found to be very dependent on the task that the subject has to perform, and they are only found when subjects are asked to carry out some kind of phonological judgment task (Radeau, Besson, Fonteneau, & Luis-Castro, 1998; Rugg & Barrett, 1987). Therefore, the processes associated with the N400 component should be understood as implying the integration of different types of features (e.g., lexical, conceptual, etc.) when this is requested. In the present experiment, the task that the subjects had to do (agreement judgment) could be accomplished by comparing the morphological features of the two words, so the N400 modulation could be related to the integration of these specific lexical features. However, it is worth noting that this lexical integration is neither semantic in nature nor necessarily implies syntactic parsing, as the building of a syntactic structure is not necessary.

In addition, in the article–noun word pairs, the described disagreement effect is not restricted to posterior areas, but covers anterior areas as well. The different types of vocabulary classes involved in the integration of both kinds of word pairs could trigger the activation of different types of neural networks distributed across different brain areas. The representations of content words (nouns and adjectives) could be stored in posterior areas while function word representations would be more restricted to left anterior areas (Pulvermüller, 1999). Thus, the integration of a function word plus a content word would result in the activation of both anterior and posterior areas. However, the presence of function words could be inseparable from the triggering of syntactic processing. In this way, the broader distribution of the effect associated to article–noun pairs could be the result of the overlapping of two different effects; the previously described N400 effect with central–posterior distribution and another one over the left anterior area. The anterior effect could be related to the anterior negativities that have been associated with syntactic processes, such as those reported by Münte and Heinze (1994) in response to grammatical violations between article–noun word pairs. This effect would reflect the failure in the integration of the two words in order to build a syntactic unit. It is worthy of note that, in Spanish, noun–adjective word pairs (because of the lack of the determiner) cannot constitute an autonomous syntactic unit. In contrast, article–noun word pairs can be considered as a noun phrase, and can constitute an adequate stimulus that triggers syntactic integration processes. In addition, the article–noun pair can and usually does constitute the beginning of a new sentence, so it is from the very first moment an adequate entity to open a new syntactic tree. Therefore, the presence of an effect in anterior zones in article–noun

pairs could be the result of a syntactic integration attempt that does not seem to happen in the case of noun–adjective pairs.

### *Gender versus Number*

The comparison between the gender and number disagreement effects did not reveal any reliable difference either in the size of the effects or in their distribution. However, this lack of difference could be due to the lack of a real syntactic context or due to the word pairs not having enough syntactic complexity to trigger adequate processing. In order to explore these two possibilities, Experiment 2, in which the same words were manipulated in a sentence context, was planned.

In spite of the fact that no amplitude differences between gender and number effects were found, analysis of the latency variations of the P3 peaks replicated previous results with noun–adjective word pairs (Barber & Carreiras, 2003). That is, disagreement conditions showed longer latencies than the agreement condition. In addition, although the graphics of single electrodes do not show clear differences between the peak latencies of both disagreement conditions, statistical analysis of the mean values of groups of electrodes revealed that the gender disagreement produced longer latencies than the number disagreement condition. N400 and P300 components usually present similar latencies so it is important to consider the possible effect of some partial component overlapping. This way, P300 latency variations could be the result of previous effects. However, no latency differences across conditions can be appreciated at the onsets or peaks of the previous negativities. Generally speaking, a P3 appears each time the task requires a binary-type decision (Donchin & Coles, 1988). Several studies have shown that P3 latency can be retarded depending on the complexity of the stimulus or the categorization difficulty that the subject must confront (Bentin et al, 1985; Rugg, 1985; Kutas, McCarthy, et al, 1977). Therefore, changes in the latencies of the P3 component in our data could be reflecting differences in the categorization of word pairs to prepare the response. In the case of grammatically agreeing word pairs, decision can be taken immediately after integration. However, for pairs in which such integration is impossible, the decision has to be postponed until all possible checking and integration attempts have been made. This way, differences between gender and number disagreement could be the result of the different costs of these rechecking processes. In relation to the syntactic analysis of sentences, different authors have proposed that reanalysis and repair processes are reflected in the P600 component (e.g., Friederici, 1995). If gender disagreement produces costlier reanalysis than number disagreement, greater amplitude of the P600 component in response to gender agreement violations should be expected with respect to that produced by

number agreement violation. This hypothesis was explored in the following experiment.

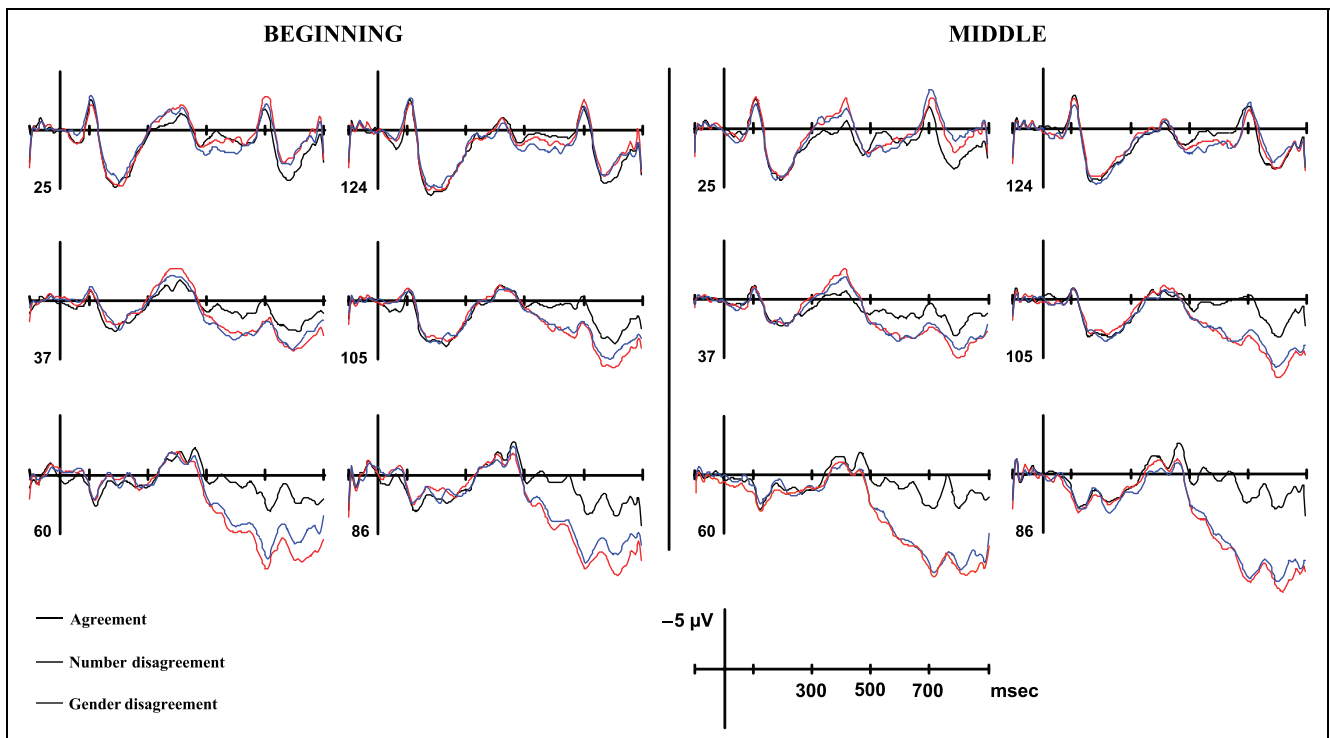
## EXPERIMENT 2: GENDER AND NUMBER AGREEMENT IN SENTENCES

In this experiment, the effects of agreement violations were analyzed in more complex syntactic structures and in a more natural reading process. The same words as in Experiment 1 were presented opening a sentence (e.g., *el piano estaba viejo y desafinado* [*the m-s piano m-s was old and off-key*]). This way, it was possible to compare the effect of the isolated word pairs of Experiment 1 (i.e., *el-piano* [*the-piano*]) with these pairs localized at the beginning of the sentence, in order to analyze the effect of the sentence context in reading. Additional sentences, made from the noun–adjective word pairs (e.g., *El faro es alto y luminoso* [*the m-s lighthouse m-s is high m-s and bright*]), presented the same type of manipulations localized after the verb (violating the noun phrase and a predicative adjective agreement), producing the violations across phrases in the middle of the sentence. Therefore, the impact of gender and number agreement violations in two different moments of sentence processing was compared. Under these circumstances, the P600 effect was expected in response to all agreement violations (Osterhout & Mobley, 1995) and a LAN effect, at least, for the gender agreement violations (Barber, Sallilas, et al., 2004). Following the models that propose that gender and number

are represented in different ways during syntactic processing (De Vincenzi & Di Domenico, 1999), a larger effect of number violation was expected than that of gender violation in the initial syntactic parsing, producing a larger effect over the LAN. However, in Experiment 1, no differences were found in the anterior negativity. Therefore, in this new experiment, it was possible to check if the lack of these differences was due to the lack of a real syntax context. In addition, considering the variations in latency of the P3 component found in Experiment 1, and that we have related these changes to rechecking processes, a larger amplitude of the P600 associated to gender violations could be hypothesized in comparison with number violations, due to the reanalysis processes being costlier when the processor has to deal with lexical features (Faussart et al, 1999). Finally, it was expected that the effect of reanalysis would be greater when the violations were localized across phrases, in the middle of the sentence, than when they were at the beginning of the sentence, within the same phrase.

## Results

Figure 4 shows the grand averaged waveforms corresponding to the agreement manipulations at the beginning and at the middle of the sentences separately. Visual inspection of both figures reveals clear differences in the responses to disagreement conditions with respect to agreement conditions. Between 300 and



**Figure 4.** Grand averages of the target words for the agreement, gender disagreement, and number disagreement conditions, at the beginning (left) and at the middle of the sentence (right).



450 msec, disagreement waves were more negative-going than those in the agreement conditions. This effect is similar for both gender and number violations and is present at anterior electrodes, especially on the left side of the scalp, fitting with the previously described LAN effect. In both figures, the negative effect is followed by a typical P600 effect, showing larger amplitudes for the disagreement conditions than for the agreement ones. The P600 effect starts at 500 msec and is maintained through more than 400 msec. Between 500 and 700 msec, the P600 effect was distributed along anterior and posterior areas. In a second stage, between 700 and 900 msec, the P600 effect is only localized over posterior areas and lateralized to the right side of the scalp. These changes in the topography of the P600 effect across time fit with the two phases previously proposed by Hagoort, Brown, and Osterhout (1999), so we will consider these two phases separately in the following analysis. The topographical distribution of the different effects is represented in Figure 5 collapsing the gender and number violations across the two positions of the sentence. The comparison between the gender and number disagreement conditions reveals differences only in the second phase of the P600 effect. In this last time window, the effect of gender violation seems to be greater as compared with number violation. As to the sentence position of the agreement violations, the LAN effect is of equivalent size in both positions of the sentence, while the last segment of the P600 effect is larger in the violations located in the middle of the sentence and especially at the posterior electrodes. Comparison of the disagreement effects (subtraction between the agreement and disagreement waves) across the two sentence positions can be found in Figure 6.

Analyses were carried out in three different temporal windows on the basis of calculations of mean amplitudes; between 300 and 450 msec for the LAN effect,

and 500–700 and 700–900 msec for the P600 effect. Omnibus ANOVAs were performed, including the “grammatical agreement” variable as a within-factor with three levels (agreement, gender disagreement, and number disagreement), the “sentence position” factor with two levels (beginning and middle), the “electrode region” factor with three levels (anterior, central, and posterior), and the “hemisphere” factor with two levels (left and right).

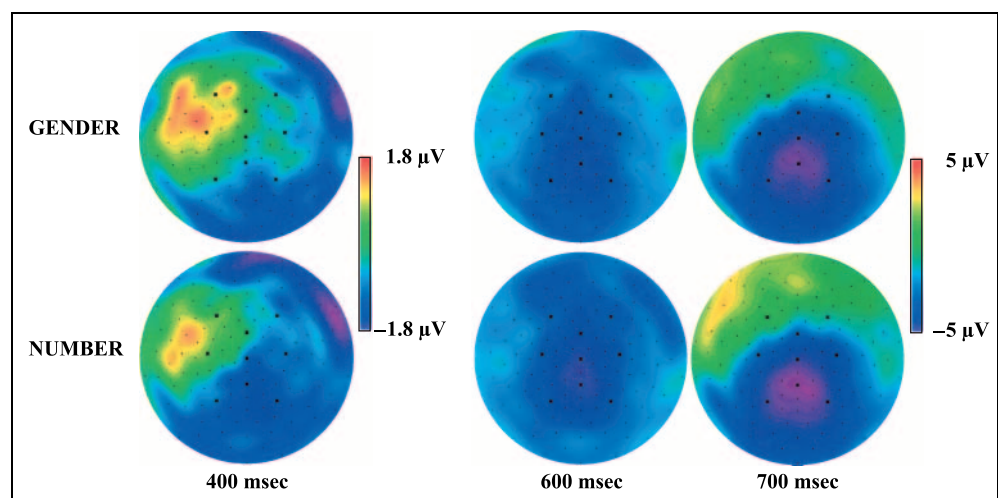
### 300–450 msec Time Window: LAN

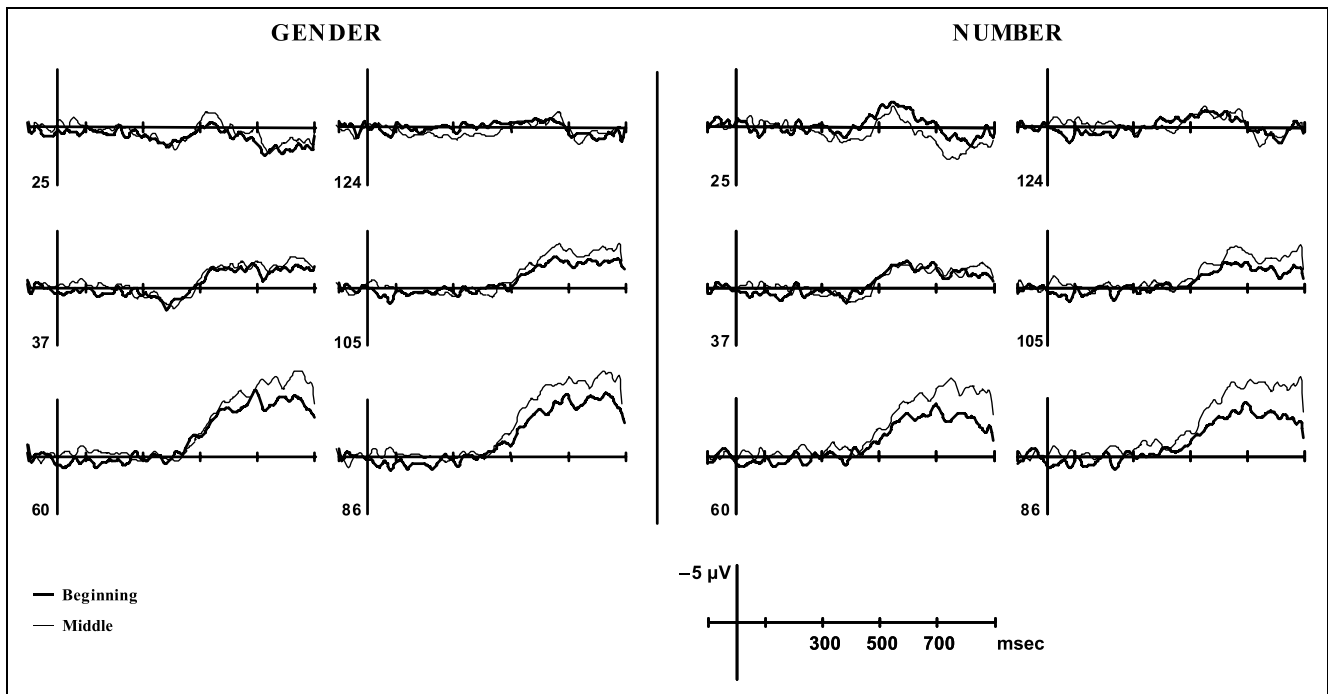
In this time window, a three-way interaction was found involving the factors “grammatical agreement,” “electrode region,” and “hemisphere” [ $F(4,20) = 6.03$ ;  $p < .01$ ;  $e = 0.55$ ] and this was confirmed after data normalization [ $F(4,20) = 4.3$ ;  $p < .05$ ;  $e = 0.63$ ]. Post hoc tests showed significant differences between the agreement condition and the gender disagreement condition just in the left anterior [ $F(1,23) = 10.15$ ;  $p < .01$ ] and the left central [ $F(1,23) = 10.37$ ;  $p < .01$ ] regions. In a similar way, differences between the agreement condition and the number disagreement condition were only present in the left anterior region [ $F(1,23) = 5.58$ ;  $p < .05$ ]. Contrasts between both disagreement conditions revealed no significant differences.

### 500–700 msec Time Window: P600a

The ANOVA revealed a three-way interaction between the factors “grammatical agreement,” “electrode region,” and “hemisphere” [ $F(4,20) = 5.8$ ;  $p < .001$ ;  $e = 0.86$ ], interaction that was maintained after the data normalization [ $F(4,20) = 4$ ;  $p < .01$ ;  $e = 0.83$ ]. However, the “grammatical agreement” variable was not modulated by the “sentence position” factor ( $F < 1$ ).  $F$  values and significance levels of the post hoc tests (comparisons between the different experimental conditions in each

**Figure 5.** Topographical maps obtained by interpolation at 400, 600, and 800 msec corresponding to LAN, P600 first phase, and P600 second phase. Maps were computed from values resulting from the subtraction between the gender disagreement and agreement condition (top row) and number disagreement less agreement condition (bottom row).





**Figure 6.** Difference waves (agreement less disagreement) comparing gender (left) and number (right) across both sentence positions.

electrode region) are listed in Table 2. It can be seen that comparisons between the disagreement conditions did not show any reliable differences in this time window.

#### 700–900 msec Time Window: P600b

As in the previous window, an ANOVA showed the same three-way interaction between the factors “grammati-

cal agreement,” “electrode region,” and “hemisphere” [ $F(4,20) = 8.44; p < .01; e = 0.62$ ; with normalized data  $F(4,20) = 8.14; p < .01; e = 0.62$ ].  $F$  values and significance levels of the post hoc tests (comparisons between the different experimental conditions in each electrode region) are listed in Table 2. This table shows how the disagreement effects are located in central and posterior regions and are greater in the right hemi-

**Table 2.**  $F$  Values and Significance Levels Resulting from Post Hoc Comparisons between the Average Amplitude Values of the Experimental Conditions (Agreement versus Gender Disagreement and versus Number Disagreement, and Gender Disagreement versus Number Disagreement), for Each Time Window of the P600 Effect (from 500 to 700 msec and from 700 to 900 msec) and for Each Electrode Region (Anterior, Central, and Posterior)

$df = 1, 23$		Agreement vs. Gender Disagreement		Agreement vs. Number Disagreement		Gender Disagreement vs. Number Disagreement	
		Left H.	Right H.	Left H.	Right H.	Left H.	Right H.
500–700 msec	Anterior	– <sup>a</sup>	3.62	– <sup>a</sup>	5.74*	– <sup>a</sup>	– <sup>a</sup>
	Central	14.61*	22.66***	17.76***	16.1***	– <sup>a</sup>	– <sup>a</sup>
	Posterior	62.87***	51.53***	39.27***	26.1***	1.25	2.35
700–900 msec	Anterior	5.85 <sup>#</sup>	– <sup>a</sup>	7.33 <sup>#</sup>	– <sup>a</sup>	– <sup>a</sup>	– <sup>a</sup>
	Central	20.29***	64.90***	17.00***	35.82***	1.42	5.84*
	Posterior	106.9***	137.0***	65.28***	70.27***	4.78*	6.05*

\*\*\* $p > .001$ .

\* $p > .05$ .

<sup>#</sup> $p > .05$  (opposite direction).

<sup>a</sup> $F < 1$ .

sphere. Even though there are significant differences in the left anterior regions, mean values indicate that they have the opposite direction to the posterior ones, so they seem to reflect the polarity inversion of the right posterior effect in the left anterior areas. Comparisons of both disagreement conditions showed significant differences in the posterior regions of both hemispheres as well as in the central region of the right side, indicating that amplitudes were larger for gender than for number. In addition to these effects, there was a significant three-way interaction between the factors “grammatical agreement,” “sentence position,” and “electrode region” [ $F(4,20) = 8.55$ ;  $p < .01$ ;  $e = 0.53$ ], an effect that also remained after the data normalization [ $F(4,20) = 4.11$ ;  $p < .05$ ;  $e = 0.53$ ], indicating the greater effect over posterior regions of the violations located in the middle of the sentence as compared with those that happened at the beginning of the sentence.

## Discussion

Agreement violations produced two effects that have been typically associated with syntactic processing; a negativity between 300 and 450 msec with a left anterior distribution that fits with the so-called LAN effect, and an inverse polarity effect that begins around 500 msec and stays for more than 400 msec, which corresponds with the P600 effect. The P600 effect was greater over the right side of the scalp and over posterior areas, especially after 700 msec.

According to previous interpretations, the LAN effect could be reflecting the detection of a mismatch between morphosyntactic features, the difficulty of integrating these characteristics in a syntactic structure (Münte, Matzke, & Johannes, 1997; Friederici, 1995), or an increase of the working memory demands implied in these processes (Kluender & Kutas, 1993). In a similar way, the P600 could also reflect the impossibility of a later integration process, different costs of reanalysis and repair processes (Hagoort, Brown, & Osterhout, 1999; Friederici, 1995) or a more general activation associated to anomaly detection (Coulson et al., 1998). In addition, some authors have proposed that different effects (with different generators) could contribute in the early and late phases of the P600 effect, the first phase being associated with diagnosis or syntactic integration difficulty and the second one with repair or reanalysis processes (Friederici, Mecklinger, Spencer, Steinhauer, & Donchin, 2001; Hagoort & Brown, 2000).

As to the position in which the agreement violation was located, while the LAN effect showed similar magnitude across positions, the second phase of the P600 effect was greater for the violations located in the middle of the sentence. The violations in the middle of the sentence involved a disagreement between a predicative adjective and a noun phrase with the verb between these and so can be considered as across-phrases dis-

agreement, whereas the article–noun disagreement at the beginning of the sentence was a within-phrase disagreement. Thus, P600 differences across positions could be reflecting the fact that mid-sentence violations involve more cost in reanalysis processes after the violation detection. The first P600 phase could reflect the detection of the violation or the failure in syntactic integration, these processes being similar in both positions. On the other hand, the second segment of the P600 would reflect the reanalysis processes which would be costlier in the case of the mid-sentence position because of the greater complexity of the agreement. However, we cannot dismiss an alternative interpretation in terms of vocabulary type, due to the fact that the target words are nouns in the beginning position and adjectives in the middle position. Although the question of the exact source of these differences is not central to the main goal of the present study, it is relevant by itself and so should be addressed in future experiments.

### *Gender versus Number*

Comparison between gender and number effects showed differences in the second phase of the P600 effect. These differences are located in the posterior areas and are larger on the right side of the scalp. In contrast, the LAN effect and the first phase of the P600 effect were similar in latency, distribution, and amplitude for gender and number violations. These data point to very late differences in the detection of gender and number disagreement during syntactic processing. Gender and number differences were found in the same time window as the effects of sentence position. If we assume that this late window is sensitive to the costs associated with reanalysis processes, then gender violations would trigger costlier reanalysis processes than number agreement violations. This idea fits with the late differences found in Experiment 1 and is considered in more depth in the general discussion.

## GENERAL DISCUSSION

In two different experiments, the effects of gender and number disagreement in visually presented words were analyzed with the ERP technique. Disagreement in article–noun word pairs resulted in a broadly distributed negativity that could be the result of the overlapping of an N400 effect and an anterior negativity, and the agreement violations during sentence reading resulted in a LAN–P600 pattern. Changes in the topographic distribution of the P600 and the different sensitivity to our manipulations (type of agreement violation and sentence position of the violation) at different times support the division of this effect into two different subeffects probably associated to different generators.

In Experiment 1, the N400 effect reflects the attempt at integration of the lexical features of the words.

Because the N400 effect is not present when subjects read the same violations in a sentence context (Experiment 2), this lexical integration could be related to the word matching that participants have to carry out in order to perform the task, rather than with real syntactic integration. In addition to the central–posterior effects, article–noun word pairs also produced disagreement effects in the left anterior area, so we have proposed the overlapping of an anterior negativity in this type of word pairs. This anterior negativity and the LAN effect of Experiment 2 could be reflecting related underlying processes. In both experiments, anterior effects can be associated with failure in the initial syntactic parsing or even with some kind of additional working memory load associated with this failure. This attempt at syntactic integration would be automatically triggered by any adequate syntactic stimuli, even when the construction of a syntactic structure is unnecessary, as in the case of the article–noun word pairs of Experiment 1.

The P600 effect was present only when participants read sentences but not when they read isolated word pairs, where only P3 latency differences were found. These differences between the experiments could be due to the fact that, in the sentence experiment, readers can categorize a sentence as wrong immediately after the reading of a violation (which would result in a P3), but in the case that they do not find a violation, they have to delay their decision (with no P3 at this point) until the end of the sentence. However, other studies have found similar P600 effects when the agreement violation (and therefore the task resolution) was always located at the end of the sentence (Hagoort, Brown, & Groothusen, 1993) or even when no task at all was requested (Osterhout & Mobley, 1995). In this respect, even when the P600 effect is only present in the sentence experiment, our data cannot resolve the question of whether the P600 and P3 are manifestations of similar underlying processes or not. An alternative interpretation can be made attending to the different complexity of the syntactic analysis across experiments. It is worth noting that in the case of the target nouns, these provided exactly the same morphosyntactic (and semantic) information when presented in a word pair (Experiment 1) as when they were presented at the beginning of a sentence (Experiment 2). Differences in the P3 latency have been explained in terms of the categorization that subjects have to do just after the reading of the second word. On the other hand, the P600 effect could be related with a more controlled process that depends on reader expectations. Thus, when the participants know that they are reading a sentence, they start to build or predict a syntactic structure right from the beginning of the sentence. This syntactic structure would be used not only to understand the sentence but also to perform the task, especially when they find an agreement violation in the middle of the sentence. The syntactic structure would not be necessary in the

word pair experiment because the task can be performed matching the lexical features of the words.

### **Gender versus Number**

As we described above, whereas grammatical gender is usually considered a lexical feature of the representation of some words, number is considered an autonomous feature that combines with the stem of the word. For this reason, it has been proposed that only number features have autonomous representations during syntactic analysis and therefore only number could be considered in early syntactic processing. On the other hand, the different lexical representations of gender and number features could result in different reanalysis processes after the detection of agreement violations.

Models that propose different roles of gender and number features during the early stages of syntactic processing (De Vincenzi, 1999; De Vincenzi & Di Domenico, 1999) predict different effects of gender and number disagreement, especially in relation to anterior negativities, as these have been related with first pass syntactic processing (Friederici, 1995). Thus, gender disagreement might produce weaker negative effects or none at all. The results of these two experiments cannot support this hypothesis, as no differences were found between the negative effects related to the two types of disagreement, either in the word pairs experiment or in the sentence experiment. In both experiments, the early effects of gender and number agreement violations produced effects with the same amplitude, latency, and distribution, indicating that the detection of such violations could involve at least some common process. Obviously, equivalent effects on the surface of the scalp could be a result of changes originating in different neural generators and additional processes were possibly not detected by our technique. However, the existence of early effects derived from gender disagreement indicates that gender information is used in the first stages of syntactic processing.

The P600 has also been considered a manifestation of syntactic integration difficulty (Hagoort & Brown, 2000; Kaan, Harris, Gibson, & Holcomb, 2000). However, in our data, gender and number differences were found at a very late stage of this effect, while the early phase was similar in both gender and number violations. Although differences between gender and number are predicted in initial syntactic processing, our effects after 700 msec seem to be too late to be reflecting this initial syntactic stage.

In both experiments, differences in the effects associated to gender and number violations are found at late time windows. In the word pairs experiment, P3 latency is longer to gender as compared to number disagreement, indicating that the response-related decision (the categorization of a word pair as wrong) takes place later

for gender disagreement. In a similar way, in the sentence experiment, the size of the effect of gender agreement violation is greater than the number effect after 700 msec, indicating a greater impact of gender violation on the later syntactic processes. Following the proposal by Faussart et al. (1999), when an inconsistency is detected in agreement relations, in the case of gender failure, the processor would have to check both syntactic integration processes and lexical access (because gender is a lexical feature). In contrast, in the case of number information posing a problem, the processor would only have to check the syntactic integration processes. Therefore, differences in the reanalysis processes would explain the higher latencies in the case of gender disagreement as compared to number disagreement in the word pairs experiment, as well as the greater amplitude of the second phase of the P600 effect in the sentence experiment. It is worthy of note that these late differences would depend on the way in which gender and number representations affect the reanalysis processes, rather on the role that these representations play in the early construction of the syntactic structure.

In sum, manipulations of gender and number agreement resulted in different ERP effects. Anterior negativities were found in both experiments and have been related with the failure of automatic syntactic integration. This effect was accompanied by a more posterior N400 effect in the word pair presentation and a P600 effect in the sentence context. While the N400 effect seems to reflect difficulty in the integration of the morphological features of the words, the P600 effect would be associated to syntactic analysis of the sentence and presented two different phases. The late phase was related with reanalysis processes and was sensitive to the position in which the violation was presented, being greater when the violation was located in the middle of the sentence. The comparison between gender and number agreement processes showed differences in late measures, that is, gender disagreement produced longer latencies of the P3 component in the word pair experiment and greater amplitudes of the P600 second phase in the sentence experiment. These late effects are congruent with the proposal that reanalysis processes are costlier after detection of gender violations because of the lexical nature of grammatical gender.

## METHODS

### Participants

Twenty-two undergraduates (11 women and 11 men) participated in Experiment 1, and 24 (18 women and 6 men) participated in Experiment 2. All of them were native Spanish speakers, with no history of neurological or psychiatric impairment, and with normal or corrected-to-normal vision. Ages ranged from 18 to 26 years (mean = 20.3 years) in Experiment 1, and from 18 to

29 years (mean = 22.7 years) in Experiment 2. All participants were right-handed, as assessed by an abridged Spanish version of the Edinburgh Handedness Inventory (Oldfield, 1971): LQ > + 50. Seven of the participants in Experiment 1 and six in Experiment 2 reported having left-handed relatives. Subjects were volunteers and received course credit for their participation. None of the subjects of Experiment 2 had participated in Experiment 1.

### EEG Recording and Analysis

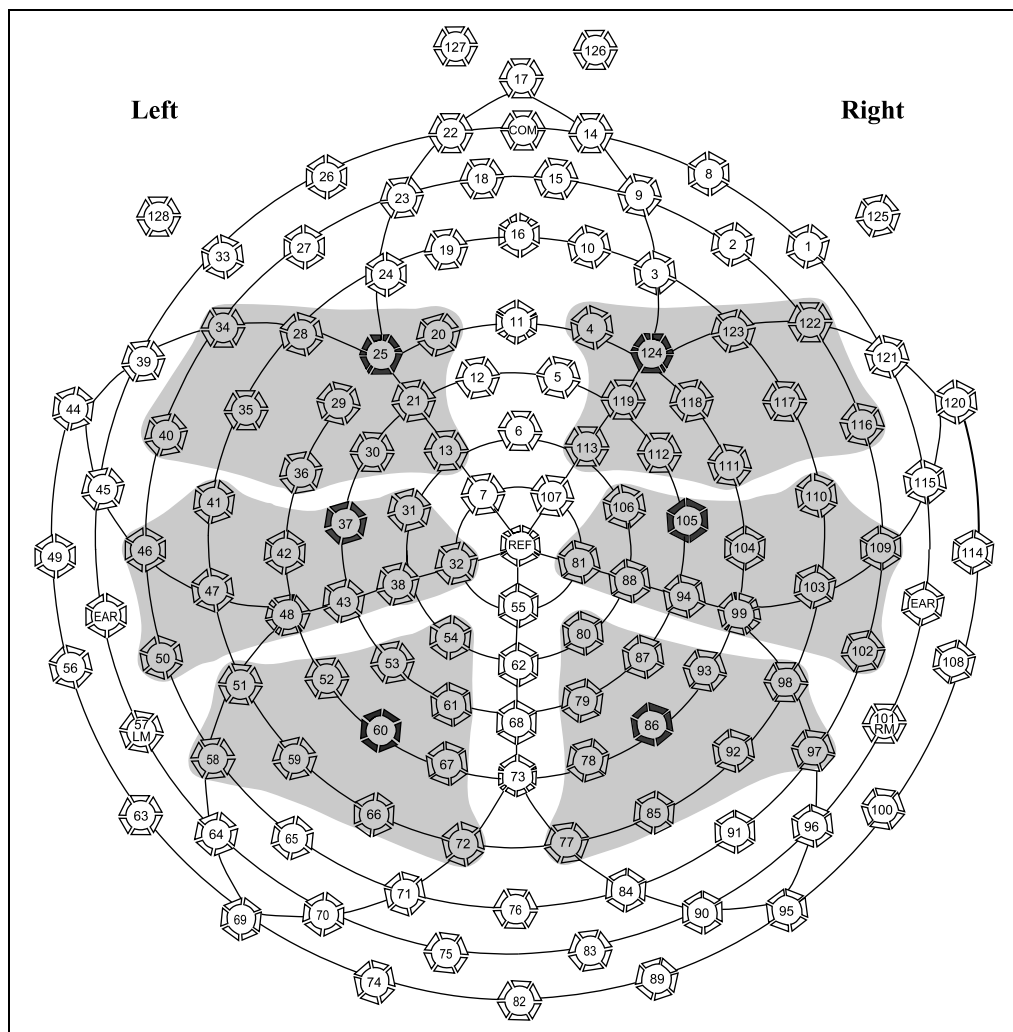
Scalp voltages were collected from Ag/AgCl electrodes using a 128-channel Geodesic Sensor Net (see Figure 7). The vertex electrode was used as reference (REF), and the recording was re-referred off-line to linked mastoids (57 and 101). Eye movements and blinks were monitored with supra- and infraorbital electrodes and with electrodes in the external canthi (14, 22 and 125–128). Interelectrode impedances were kept below 30 k $\Omega$  (amplifiers input impedance >200 M $\Omega$ ). EEG was filtered with an analogue bandpass filter of 0.01–100 Hz (50 Hz notch filter) and a digital 35-Hz low-pass filter was applied before analysis. The signals were sampled continuously throughout the sessions with a sampling rate of 250 Hz.

Epochs from the continuous EEG in the interval between –100 and +900 msec with respect to the onset of the target were averaged and analyzed. Baseline correction was performed using the average EEG activity in the 100 msec preceding the onset of the target word as a reference signal value. Following baseline correction, trials with artifacts were rejected. Less than 15% of the trials were excluded by this operation and these were evenly distributed across the different experimental conditions. Furthermore, electrodes with a high level of artifacts (>10%) were substituted by the average value of the group of nearest electrodes. Separate ERPs were formed for each of the experimental conditions, each of the subjects and each of the electrode sites.

Six regions of interest were computed out of the 129 electrodes, each containing the mean of a group of 11 electrodes. The regions were (see electrode numbers in Figure 7): left anterior (13, 20, 21, 25, 28, 29, 30, 34, 35, 36, and 40), left central (31, 32, 37, 38, 41, 42, 43, 46, 47, 48, and 50), left posterior (51, 52, 53, 54, 58, 59, 60, 61, 66, 67, and 72), right anterior (4, 111, 112, 113, 116, 117, 118, 119, 122, 123, and 124), right central (81, 88, 94, 99, 102, 103, 104, 105, 106, 109, and 110), and right posterior (77, 78, 79, 80, 85, 86, 87, 92, 93, 97, and 98). Additional analyses were carried out for other electrode groups, including frontal and occipital regions, and midline locations. As these revealed similar results they are not considered here.

Different repeated-measures ANOVAs were performed introducing these groups of electrodes as the different levels of the electrode region factor (anterior, central,

**Figure 7.** Schematic flat representation of the 129 electrode positions from which EEG activity was recorded (front of head is at top). Grouped electrodes in the analysis (six electrode regions) and six specific positions represented in Figures 2, 4, and 6 are marked.



and posterior) and the hemisphere factor (left and right), and the experimental variables as additional factors. Effects related with electrode region factor or hemisphere factor will only be reported when they interact with the experimental manipulations. In cases of interaction of any experimental factor with the electrode region or hemisphere factors, data were normalized following the vectorial scaled procedure recommended by McCarthy and Wood (1985). Where appropriate, critical values were adjusted using the Geisser and Greenhouse (1959) correction for violation of the assumption of sphericity.

Behavioral measures are not reported because speed was not required of the subjects so reaction times showed great variability.

## Stimuli

### Experiment 1

Two lists of 120 experimental word pairs each were generated. In the first list, word pairs were made up of a determiner article followed by a noun, whereas in the

second list, word pairs included a noun followed by an adjective. The combination of the different gender and number forms of articles and adjectives resulted in the different experimental conditions:

Article–noun word pairs:

- (a) Agreement, e.g., *El piano (the m-s piano m-s)*.
- (b) Gender disagreement, e.g., *La piano (the f-s piano m-s)*.
- (c) Number disagreement, e.g., *Los piano (the m-p piano m-s)*.

Noun–adjective word pairs:

- (d) Agreement, e.g., *Faro alto (lighthouse m-s high m-s)*.
- (e) Gender disagreement, e.g., *faro alta (lighthouse m-s high f-s)*.
- (f) Number disagreement, e.g., *faro altos (lighthouse m-s high m-p)*.

Assignment of word pairs to conditions in each list was counterbalanced across participants. Thus, each pair occurred three times across subjects, once in each condition, so that each subject only saw one form of

each pair during the experiment. In Spanish, it is mandatory that articles, nouns, and adjectives agree in gender and in number. In the present stimuli, gender was always a strictly morphosyntactic feature without semantic significance. All nouns and adjectives were morphologically marked in gender and number, that is, they ended with the canonical suffixes in Spanish for gender (“-o” for masculine and “-a” for feminine) and number (“-s” or “-es” for plural).

In addition, a list of 80 filler trials was introduced. Some fillers had opaque gender (e.g., the word “reloj” [clock] lacks any explicit morphological mark) and some other fillers were irregular words (e.g., “mano” [hand] ends with the letter “-o” but is feminine). This type of fillers was included to prevent participants using a superficial strategy for solving the task such as, for example, attending only to the suffixes. All the filler word pairs agreed in gender and number.

### Experiment 2

Two lists of 120 experimental sentences each were generated in order to manipulate the agreement at the beginning and in the middle of the sentence. The first list manipulated the agreement relationships at the beginning of the sentence using the article–noun word pairs list from Experiment 1, while the second list manipulated the agreement relationships in the middle of the sentence using the noun–adjective word pairs list from Experiment 1:

Manipulation at the beginning of the sentence:

- (a) Agreement, e.g., *El piano estaba viejo y desafinado* (the  $m-s$  piano  $m-s$  was old and off-key).
- (b) Gender disagreement, e.g., *La piano estaba viejo y desafinado* (the  $f-s$  piano  $m-s$  was old and off-key).
- (c) Number disagreement, e.g., *Los piano estaba viejo y desafinado* (the  $m-p$  piano  $m-s$  was old and off-key).

Manipulation in the middle of the sentence:

- (d) Agreement, e.g., *El faro es alto y luminoso* (the  $m-s$  lighthouse  $m-s$  is high  $m-s$  and bright).
- (e) Gender disagreement, e.g., *El faro es alta y luminoso* (the  $m-s$  lighthouse  $m-s$  is high  $f-s$  and bright).
- (f) Number disagreement, e.g., *El faro es altos y luminoso* (the  $m-s$  lighthouse  $m-s$  is high  $m-p$  and bright).

Between two and four filler words were included after the target adjective in order to avoid wrap-up effects on the target words. As in Experiment 1, assignment of sentences to conditions in each list was counterbalanced across participants. Thus, each sentence occurred three times across subjects, once in each condition, so that each subject only saw one form of each sentence during the experiment.

In addition, a list of 160 well-formed filler sentences was introduced. Fillers included nouns with opaque or irregular gender, and adjectives with neuter gender.

### Procedure

Participants were seated comfortably in a darkened sound-attenuated chamber. Stimuli were displayed in black letters against a gray background. All trials were presented in different pseudorandom order for each participant. Participants had to judge the grammatical agreement between words of a given pair (Experiment 1) or to what extent a sentence was grammatically correct (Experiment 2). Participants in Experiment 1 had to make the decision after reading the second word, but they were not encouraged to respond quickly. In Experiment 2, the response was delayed until the reading of the last word of the sentence. The grammatical judgment task was chosen taking into consideration previous studies, both with word pairs (Münte & Heinze, 1994) and with sentences (Osterhout & Mobley, 1995), which compared the effects of agreement violations with or without task. In these studies, the same qualitative effects were found but they were quantitatively greater with the grammatical judgment task. A response button was positioned beneath each thumb. For half of the participants, the right button was used to signal the “Yes” response and the left button was assigned to the “No” response. For the remaining subjects, the order was reversed. Participants were also asked to avoid eye movements and blinks during the interval when a fixation asterisk was not present.

The sequence of events in each trial in each experiment is described as follows:

### Experiment 1

First, a fixation point (“+”) appeared in the center of the screen and remained there for 2700 msec. This fixation point was followed by a blank screen interval of 300 msec, then the prime word appeared for 300 msec, which was followed by a 500-msec blank interval, and finally the target word appeared and remained there up to a maximum of 2000 msec or until the participant’s response. The intertrial interval varied randomly between 1000 and 1500 msec.

### Experiment 2

First, a fixation point (“+”) appeared in the center of the screen and remained there for 2700 msec. This fixation point was followed by a blank screen interval of 300 msec, then the sentence was displayed word by word. Each word appeared for 300 msec and was followed by a 300-msec blank interval. Participants were instructed to respond after the last word of the sentence. At that moment, a question mark was presented and remained there up to a maximum of 2000 msec or until the participant’s response. The intertrial interval varied randomly between 1000 and 1500 msec.

## Acknowledgments

This research was partially supported by fellowship EX2003-0012 (Secretaría de Estado de Educación y Universidades of the Spanish Government and the European Social Fund) to the first author, and grant BSO2003-01135 (Spanish Ministry of Science and Technology) to the second author. We thank Lee Osterhout and two anonymous reviewers for helpful comments, and Margaret Gillon Dowens for her revision of the manuscript.

Reprint requests should be sent to Horacio Barber, PhD, University of California, San Diego, Department of Cognitive Science, 9500 Gilman Drive, La Jolla, CA 92093-0515, or via e-mail: hbarber@cogsci.ucsd.edu.

## REFERENCES

- Antón-Méndez, I., Nicol, J. L., & Garret, M. F. (2002). The relation between gender and number agreement processing. *Syntax*, *5*, 1–25.
- Barber, H., & Carreiras, M. (2003). Integrating gender and number information in Spanish word pairs: An ERP study. *Cortex*, *39*, 465–482.
- Barber, H., Salillas, E., & Carreiras, M. (2004). Gender or genders agreement? In M. Carreiras & C. Clifton (Eds.), *On-line study of sentence comprehension; eye-tracking, ERP and beyond*. Brighton, UK: Psychology Press.
- Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology*, *60*, 343–355.
- Bock, K., & Levelt, W. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). San Diego, CA: Academic Press.
- Colé, P., & Seguí, J. (1994). Grammatical incongruity and vocabulary types. *Memory and Cognition*, *22*, 387–394.
- Corbett, G. G. (1991). *Gender*. Cambridge: Cambridge University Press.
- Coulson, S., King, J., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, *13*, 21–58.
- De Vincenzi, M. (1999). Differences between the morphology of gender and number: Evidence from establishing coreferences. *Journal of Psycholinguistic Research*, *28*, 5, 537–553.
- 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.
- Donchin, E. (1979). Event-related brain potentials: A tool in the study of human information processing. In H. Begleiter (Ed.), *Evoked potentials and behaviour*. New York: Plenum.
- Donchin, E., & Coles, M. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*, 357–374.
- Faussart, C., Jakubowicz, C., & Costes, M. (1999). Gender and number processing in spoken French and Spanish. *Rivista di Linguistica*, *11*, 75–101.
- Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance*. Hillsdale, NJ: Erlbaum.
- Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and Language*, *50*, 259–281.
- 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.
- Garrett, M. F. (1982). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production* (pp. 170–220). London: Academic Press.
- Geisser, S., & Greenhouse, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*, 95–112.
- Gunter, T. C., Friederici, A. D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience*, *12*, 556–568.
- Hagoort, P., & Brown, C. (1999). Gender electrified: ERP evidence on the syntactic nature of gender processing. *Journal of Psycholinguistic Research*, *28*, 6, 715–728.
- Hagoort, P., & Brown, C. (2000). ERP effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia*, *38*, 1531–1549.
- Hagoort, P., Brown, C., & Grootjens, J. (1993). The syntactic positive shift (SPS) as an ERP-measure of syntactic processing. *Language and Cognitive Processes*, *8*, 439–483.
- Hagoort, P., Brown, C., & Osterhout, L. (1999). The neurocognition of syntactic processing. In C. Brown & P. Hagoort (Eds.), *Neurocognition of language* (pp. 273–316). Oxford: Oxford University Press.
- Harris, J. W. (1991). The exponence of gender in Spanish. *Linguistic Inquiry*, *22*, 27–62.
- Igoa, J. M., García-Albea, J. M., & Sánchez-Casas, R. (1999). Gender–number dissociation in sentence production in Spanish. *Rivista di Linguistica*, *11*, 163–193.
- Kaan, E., Harris, A., Gibson, G., & Holcomb, P. J. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*, 159–201.
- Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, *5*, 196–214.
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*, 463–470.
- Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory and Cognition*, *11*, 539–550.
- Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science*, *197*, 792–795.
- 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.
- McCarthy, G., & Wood, C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology*, *62*, 203–208.
- Münte, T. F., & Heinze, H. J. (1994). ERP negativities during syntactic processing of written words. In H.-J. Heinze, T. F. Münte, & G. R. Mangun (Eds.), *Cognitive electrophysiology* (pp. 211–238). Boston: Birkhäuser.
- Münte, T. F., Matzke, M., & Johannes, S. (1997). Brain activity associated with syntactic incongruencies in words and pseudo-words. *Journal of Cognitive Neuroscience*, *9*, 318–329.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, *9*, 97–113.



- 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.
- Osterhout, L., & Holcomb, P. J. (1995). Event related potentials and language comprehension. In M. D. Rugg & M. Coles (Eds.), *Electrophysiology of mind, event related brain potentials and cognition*. Oxford: Oxford University Press.
- Osterhout, L., & Mobley, L. A. (1995). Event-related potentials elicited by failure to agree. *Journal of Memory and Language, 34*, 739-773.
- Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences, 22*, 253-336.
- Radeau, M., Besson, M., Fonteneau, E., & Luis Castro, S. (1998). Semantic, repetition and rime priming between spoken words: Behavioral and electrophysiological evidence. *Biological Psychology, 48*, 183-204.
- Ritter, E. (1988). A head-movement approach to construct state noun phrases. *Linguistics, 26*, 909-929.
- Ritter, E. (1993). Where is gender? *Linguistic Inquiry, 24*, 795-803.
- Rugg, M. D. (1985). The effects of semantic priming and word repetition on event related potentials. *Psychophysiology, 22*, 642-647.
- Rugg, M. D., & Barrett, S. E. (1987). Event-related potentials and the interaction between orthographic and phonological information in a rhyme-judgment task. *Brain and Language, 32*, 336-361.
- Vigliocco, G., Butterworth, B., & Garrett, M. F. (1996). Subject-verb in Spanish and English: Differences in the role of conceptual factors. *Cognition, 51*, 261-298.