SPECIAL SECTION

INTEGRATING GENDER AND NUMBER INFORMATION IN SPANISH WORD PAIRS: AN ERP STUDY

Horacio Barber and Manuel Carreiras

(Universidad de La Laguna, Tenerife, Spain)

Abstract

The aim of the current study was to explore the integration processes of gender and number morphological features, since it has been proposed that grammatical gender and number features might be associated with different strength with the word stem in lexical representation. Event related potentials (ERPs) were recorded using a 128-channel sensor net while twenty-four volunteers read Spanish word pairs and performed a syntactic judgment task. The word pairs which could agree or disagree in gender or number or in gender and number at the same time, were formed by a noun and an adjective (e.g. *faro-alto* [*lighthouse-high*]).

A negativity around 400 msec with posterior distribution, which has been related to lexical integration processes, was found in response to both gender and number violations. No differences were found between gender disagreement, number disagreement and the double disagreement. Therefore, ERPs suggest that integration of gender and number features may not be different, and that the detection of disagreement may work under a binary state, since the double disagreement condition did not differ from the others. In addition, a subsequent component (identified as P3) showed delayed latencies in the gender disagreement condition as compared to the number disagreement condition, while the double disagreement conditions showed a shorter peak latency than the other two disagreement conditions and similar to the agreement condition. The variations in the latency of the P3 component, which has been related to categorization processes, suggest that these processes are quickly triggered by the accumulation of two incongruent as compared to one disagreement features, and that reanalysis is costlier in the case of gender disagreement as compared to the number disagreement.

Key words: grammatical priming, gender and number agreement, Event Related Potentials, N400, P300

Introduction

Languages represent gender either as a conceptual characteristic or as a formal property of words. English is an example of a language with nearly exclusively conceptual gender. In English, gender is marked for some words referring to animate entities, there being a transparent relationship between the biological sex of the referents and the gender of nouns and pronouns. In contrast, romance languages have both conceptual and grammatical gender. Although some nouns with animated biological referents still keep a transparent relationship between gender and biological sex, there is no conceptual basis for the distinction of gender in many nouns (an analysis of gender in different languages can be found in Corbett, 1991). In contrast, number is always considered a conceptual feature signalling the quantity of the referent.

From a lexical point of view, number is considered a morphological marking that combines with the stem it modifies while gender seems to be a feature directly associated to the stem of the lexical representation (Ritter, 1993). Domínguez et al. (1999) gathered some evidence suggesting that grammatical gender seems to be stored in the mental lexicon, whereas number is a productive feature. They compared the effect of surface frequency in Spanish words with the same morphological stem in a lexical decision task. Thus, words in which the occurrence of the “male” form was dominant (e.g. *ciego*, blind man) were compared to words in which the “female” was dominant (e.g. *viuda*, widow woman). The same comparison was performed with the singular and plural forms of these words. The effect of surface frequency was found both for masculine dominant and for feminine dominant words. However, regarding number, while the effect of surface frequency was also observed for singulars, no effect of surface frequency was found when the plural was the dominant form. These data seem to suggest that gender information is retrieved directly from the word form, whereas number information access may be mediated by the lexical entry belonging to the single form. That is, independent representations for the single “male” and “female” forms would be stored, acting as a basis to construct the plural forms. This lexical representations feature remains as one of the most important differences between number and gender with relevant consequences for agreement processes.

Gender and number marking are generally used to compute agreement. They are important and necessary information for computing dependencies between a noun and other words in a sentence, such as determiners, adjectives, past participles, pronouns, etc., especially in richly inflected languages such as Spanish. Processing of gender and number agreement has been studied using grammatical priming (Lukatela et al., 1987; Colé and Segui, 1994; Faussart et al., 1999). The grammatical priming effect reveals that a word is recognised faster when it is preceded by another word that agrees with it, than when it is preceded by a word that does not agree. The empirical evidence on whether gender and number agreement mechanisms are similar is mixed.

Lukatela et al. (1987) compared the grammatical priming effect in pairs that did not agree in number, gender, case or in gender and case (double violation), in experiments carried out with Serbo-Croatian speakers (a language with a very rich inflectional morphology). Subjects showed longer recognition times for all disagreement conditions as compared to the agreement condition. However, there were no differences between the different disagreement conditions. The conclusion of the authors is that the syntactic processor behaves in a binary way, that is, it needs only to detect the fact of the existence or not of grammatical agreement. A similar result was obtained in French by Colé and Segui (1994). Although the main aim of their research was to study the effect of the vocabulary type on the grammatical priming, they included word pairs that disagreed in gender or number. None of the experiments in which the disagreement of gender and number was considered separately showed any differences between the two conditions, or interactions with the vocabulary type[1].

Contrary to these results, Faussart et al. (1999), after undertaking experiments in French and Spanish, found differences between number and gender processing in the auditory modality. Using the same lexical decision task in pairs of words but with auditory presentations, they found an interaction[2] between the grammatical relationship (congruent/incongruent) and the violation type (gender/number); gender violation being more disruptive than number violation. This result appeared only in the Spanish language experiment, probably due to the fact that French marks the number orthographically but not phonetically. They explain the effect within the framework of Bradley and Foster's (1987) model about 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 post-lexical processes of integration concerning the context. Grammatical agreement would take place at this last stage. According to Faussart et al. (1999), integration process failure regarding gender agreement will make the system check if the right entry was selected and then go back to the lexical identification stage, because gender is a stem inherent feature. 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 and evaluation 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.

In sum, contradictory results have been found with the grammatical priming paradigm regarding gender and number agreement. These differences may be attributable to the many procedural differences between the studies. Faussart et al.'s used French in an auditory modality whereas Lukatela et al.'s used Serbocroatian in a visual modality; in addition Faussart et al.'s design involved gender and number violations as a between subjects factor.

Grammatical Agreement-Related ERP Effects

ERPs technique has been also used to investigate grammatical agreement processing. Its multidimensional character and excellent temporal resolution make this technique especially suitable for the study of how and when agreement relationships are made during reading. Two pioneer studies showed different effects of grammatical subject-verb agreement violations in sentence

[1] It is worth mentioning that the most important effects for vocabulary type were obtained on experiment 4 with the shorter SOA, and precisely on that experiment gender and number conditions were not considered separately.
[2] The design included the congruency-incongruency variable as within-subject factor and the gender-number variable as between-subject factor.

context. 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 et al. (1993) in Dutch described an increase of positivity from 500 msec in response to the same type of violation[3]. Posterior studies have shown again a P600 effect for number or gender agreement violations (Ostherhout and Mobley, 1995; Gunter et al., 2000). Similar effects to those described by Kutas and Hillyard have been generally linked to different aspects of syntactic analysis and labelled as Left Anterior Negativity (LAN) (Friederici, 1995), while the P600 effect has been linked to re-analysis or syntactic repair processes (Hagoort et al., 1999).

Münte and Heinze (1994) undertook a series of experiments, of special relevance to the present research, in which the effect of grammatical agreement in word pairs was analysed. Among other types of syntactic violations, 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, English and Finnish) and with different tasks (grammatical judgement and lexical decision). Results showed that agreement errors produced anterior distribution negativities and that these effects seem to reflect relatively automatic processes. The authors point out the similitude of these negativities with those found in sentence contexts and denominated LAN effects. Inversely, the effects produced by phonological or semantic relationships produced negativities with posterior distribution, similar to the typical N400 effects found with semantic violations in sentences or with semantic or phonological priming in word pairs (reviewed in Osterhout and Holcomb, 1995).

In Spanish, Barber and Carreiras (submitted) found that either gender or number disagreement between articles and nouns inserted in sentences generated a LAN effect as well an increase in the P600 effect. No differences between gender and number disagreement were found, either in LAN or in the first stage of the P600. However, in a later stage of the P600 (between 700 and 900 msec), the gender disagreement effect was significantly higher than that of number. In a second experiment, the same article-noun word pairs were presented one after each other in a word priming task. A LAN effect was found again but accompanied by a posterior distribution effect of the N400 type. No differences between gender and number effects were observed in such time windows, but late differences were found again. The P3 component, associated with decision making (grammatical judgement), presented a higher latency in the case of gender disagreement. These latency differences may be related, according to the authors, to re-analysis processes or repairing occurring after syntactic integration failure. That is, re-analysis processes could be associated with a stronger cost because, as suggested by Faussart et al. (1999), gender checking needs more steps.

[3] In the experiment by Kutas and Hillyard (1983) only the first 600 msec were analysed, so P600 effects were not observed initially. On the other hand, Hagoort and Brown (2000) in a recent auditory replication of these experiments report a LAN effect in response to agreement violations in addition to the P600 effect, but this effect was at best marginally significant.

The Present Experiment

The behavioural data on gender and number grammatical priming show a mixed picture, which probably has to do with the fact that different languages, different procedures, and different measures have been used. The ERP studies seem to suggest that gender and number agreement processes are similar, specially at the early stages, though differences can arise at later stages. Nonetheless, the currently available data comparing directly gender and number agreement processes are scarce and restricted to a small number of agreement relations.

In the present experiment the processing of gender and number agreement is studied under the same circumstances in Spanish. We will study the ERPs associated with agreement in the context of lexical integration through the use of a priming paradigm. Grammatical priming effects are analysed in word pairs in which gender and number agreement between a noun and an adjective has been manipulated (e.g. *faro-alto* [$lighthouse_{m\text{-}sg}$-$high_{m\text{-}sg}$]). Agreement in gender and number between nouns and adjectives is mandatory in Spanish. According to previous results, a negative effect within the 300-500 window in response to the non agreement conditions is expected. In previous work (Barber and Carreiras, submitted), using article-noun word pairs, this effect was composed by two different overlapped components, one with left-anterior distribution (which was considered as syntactic) and the other one with central-posterior distribution (which was considered as lexical). The main difference between the current noun-adjective word pairs and the previously studied article-noun word pairs is that a noun plus an adjective can not be considered an autonomous syntactic unit. This way, in the present experiment we will study agreement processes in a lexical integration context where other syntactic processes may be attenuated, and we will investigate whether under such circumstances the same effect is found but only with a posterior distribution. Also, the comparison of the different kinds of non-agreement conditions will allow us to study possible lexical differences between gender and number processing without the influence of higher level syntactic process related to the construction of syntactic structures.

In addition, a double violation condition (e.g. *faro-altas* [$lighthouse_{m\text{-}sg}$-$high_{f\text{-}pl}$]) will help us to understand whether the violation of two agreement features is an additive process or if, as suggested by Lukatela and co-workers, the system works under a binary state: the word pair either agrees or disagrees, regardless of the nature of the agreement features.

Finally, in this experiment we added a phonological control condition using adjectives with common morphological marks for masculine and feminine (e.g. *poesía-triste* [poetry_{f}-$\text{sad}_{f/m}$]). These word pairs will permit us to discard an alternative explanation of the effects in terms of phonological overlap of the morphological marks in the agreement condition.

METHOD

Participants

Twenty-four undergraduates, fourteen female and ten male, participated in the experiment. 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 30 years (mean = 21.2 years). All participants were right-handed, as assessed with an abridged Spanish version of the Edinburgh Handedness Inventory (Oldfield, 1971): LQ > + 50. Six of the participants had left-handed relatives. They were volunteers and received course credit for their participation.

Stimuli

One hundred and twenty pairs of words formed by a noun (e.g. *faro [lighthouse]*) plus an adjective (e.g. *alto [high]*) were generated. In Spanish, morphological gender can be marked by several suffixes. The "-a" suffix is mostly associated with the feminine gender and the "-o" suffix mainly with the masculine gender, although there are exceptions[4]. Experimental word pairs were constructed with the canonical suffixes. Furthermore, all nouns had inanimate referents, so gender was always a strictly morphosyntactic feature without semantic significance.

The morphological marking of the adjectives was manipulated to create four different conditions:

Agreement, e.g. *Faro-alto* ($lighthouse_{m\text{-}sg}$-$high_{m\text{-}sg}$).

Gender disagreement, e.g. *Faro-alta* ($lighthouse_{m\text{-}sg}$-$high_{f\text{-}sg}$).

Number disagreement, e.g. *Faro-altos* ($lighthouse_{m\text{-}sg}$-$high_{m\text{-}pl}$).

Double violation, e.g. *Faro-altas* ($lighthouse_{m\text{-}sg}$-$high_{f\text{-}pl}$).

The assignment of word pairs to conditions was counterbalanced across participants. Thus, each word pair occurred four times across subjects, once in each experimental condition, so that each subject only saw one form of each word pair during the experiment.

Forty word pairs were added as a phonological-orthographic control condition. This control was necessary because in the experimental conditions when a word pair agreed in gender, the primes and targets ended with the same letter "-a" or "-o" while in the disagreement conditions they were different for primes and target. In the control word pairs common adjectives were used. These adjectives with a neuter suffix are applied indistinctly both to masculine and feminine gender nouns (e.g. "-e"). So, those word pairs – e.g. "poesía-triste" ($poetry_{f\text{-}sg}$-$sad_{com\text{-}sg}$) – agreed in gender and number and did not present rhyme or orthographic overlap.

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). Fillers of this type were included to prevent participants from using a superficial strategy for solving the task, such as for example, attending just to the suffixes. Sixty of the filler word pairs agreed and 20 disagreed.

[4] Some feminine gender words ending in "-o" and some masculine gender words ending in "-a", as well as some other less frequent suffixes such as "dad", and even some neutral suffixes like "-e" that apply to masculine and feminine genders.

In all, each subject received 280 word pairs, half of them agreed and the other half disagreed. Within conditions there was the same number of singular, plural, masculine and feminine words in all-possible combinations. All experimental stimuli were between four, and seven letters long (2 or 3 syllables). All words have middle or high printed lexical frequency according to a Spanish standard corpus (Sebastián-Gallés et al., 2000).

Procedure

Participants were seated comfortably in a darkened sound-attenuated chamber. All stimuli were presented on a high-resolution computer that was positioned at eye level 80-90 cm in front of the participant. The words were displayed in black lower-case letters against a grey background.

Participants had to judge the grammatical agreement between words of a given pair[5]. A response button was positioned beneath each thumb. For half of the participants the right button was used to signal the “Yes” response and left button was assigned the “NO” response. For the remaining subjects the order was reversed. Thus, the assignment of buttons to hands was counterbalanced across participants.

The sequence of events in each trial is described as follows. First, a fixation point (“+”) appeared in the centre of the screen and remained there for 2800 msec. This fixation point was followed by a blank screen interval of 200 msec, then the prime word appeared for 300 milliseconds, 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 response of the participant. The inter-trial interval varied randomly between 1000 and 1500 msec. All word pairs were presented in different pseudo-random order for each participant.

Ten warm-up word pairs were provided at the beginning of the session, and they were repeated if necessary. Participants were also asked to avoid eye-movements and blinks during the interval when the fixation asterisk was not present, and they were directed to favour accuracy over speed in their responses. Each session lasted approximately 1 hr.

EEG Recording

Scalp voltages were collected from Ag/AgCl electrodes using a 128-channel Geodesic Sensor Net (Tucker, 1993). Figure 1 shows the schematic distribution of the recording sites. The vertex electrode was used as reference, and the recording was re-referred off-line to linked mastoids. Eye movements and blinks were monitored with supra- and infra-orbital electrodes and with electrodes in the external canthi. Inter-electrode impedances were kept below 30 KΩ (amplifiers input impedance < 100 MΩ). EEG was filtered with an analogue

[5] In previous works Münte and Heinze (1994) compared the grammatical priming effect in different tasks. They conclude that this effect was qualitatively equivalent for the different studied tasks, but it was bigger for the syntactic judgement task than the lexical decision task. For this reason the grammatical judgement task was choosen for this experiment.

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 experiment with a sampling rate of 250 Hz.

Analysis

Epochs of the EEG corresponding to up to 900 msec after second word onset presentation were averaged and analysed. 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, epochs with simultaneous artefacts in at least 10 channels were rejected. This operation resulted in the exclusion of approximately 10 % of the trials, which were evenly distributed along the different experimental conditions. Furthermore, electrodes with a high level of rejected trials (> 10%) were substituted by the average value of the group of nearest electrodes.

Averaging was conducted off-line using only samples recorded in trials in which correct responses had been made in a grammatical judgment task. Separate ERPs were formed for each of the experimental conditions, each of the subjects and each of the electrode sites.

Nine regions of interest were computed out of the 129 electrodes, each containing the mean of a group of electrodes. The regions were (see electrode numbers in Figure 1): midline-anterior (5, 6, 11 and 12), midline-central (7, 55, 107 and 129), midline-central-posterior (62, 68 and 73), 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), right-posterior (77, 78, 79, 80, 85, 86, 87, 92, 93, 97 and 98).

The analyses were carried out in two temporal windows: on the basis of calculations of mean amplitudes between 300 and 500 msec, and on the basis of calculations of peak latency (latency of maximum amplitude value) between 450 and 750 msec.

Different repeated measures ANOVAS for each type of measures were performed. This analysis included the grammatical relation variable as a within factor with 4 levels: agreement, gender disagreement, number disagreement, and double violation. Similar ANOVAs were performed to evaluate the effect of the orthographic overlap, in this case the grammatical factor included 2 levels: agreement and agreement with common adjectives. In mean amplitude analyses, electrode regions (anterior, central and posterior) were entered as another within subject factor in the ANOVAs. Separate analyses were carried out for the midline regions and the lateral regions. Analysis of the lateral regions included the hemisphere factor with two levels (left/right), while latency analyses only considered the posterior regions (left, central and right). A significance level of .05 was adopted for all the statistical tests. Where appropriate, critical values were adjusted using the Geisser-Greenhouse (1959) correction for violation of the assumption of sphericity. Effects related to electrode regions factor or hemisphere factor will be only reported when they interact with the experimental manipulations.

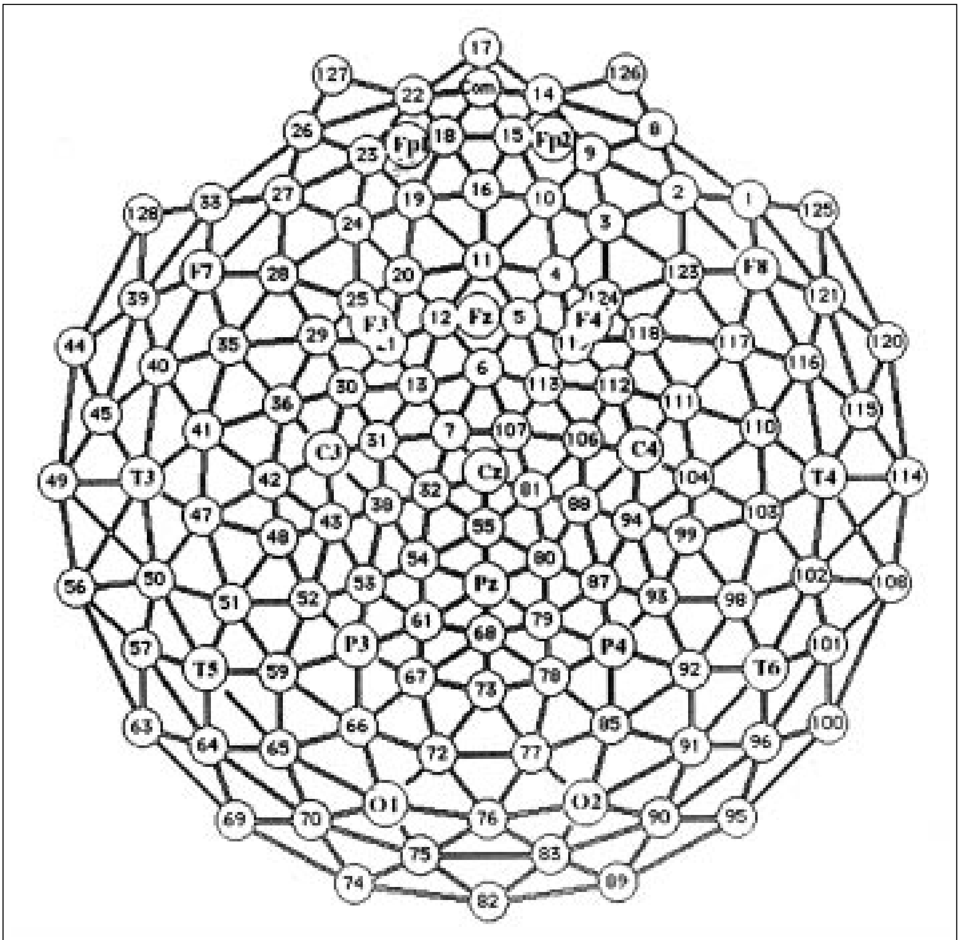

Fig. 1 – *Schematic flat representation of the 129 electrode positions from which EEG activity was recorded (front of head is at top). Channel nomenclature is by number. Approximate international 10-20 system localizations are marked.*

Results

The grand average ERPs, time-locked to the onset of the target words, are presented in Figures 2 and 3. The initial segment of all conditions showed the typical N1-P2 component. These potentials were followed by a negative going component with a peak amplitude around 400 msec which had the distribution of the classical N400. This component is followed by a positive going waveform identified as the P3 component.

Figure 2 shows the experimental agreement condition and the control agreement condition (with common adjectives) over 9 recording sites. No differences between these conditions are observed within the 300 to 500 msec window. Grand averages of the agreement, the gender disagreement, the number disagreement, and the double violation conditions are plotted in Figure 3. Visual

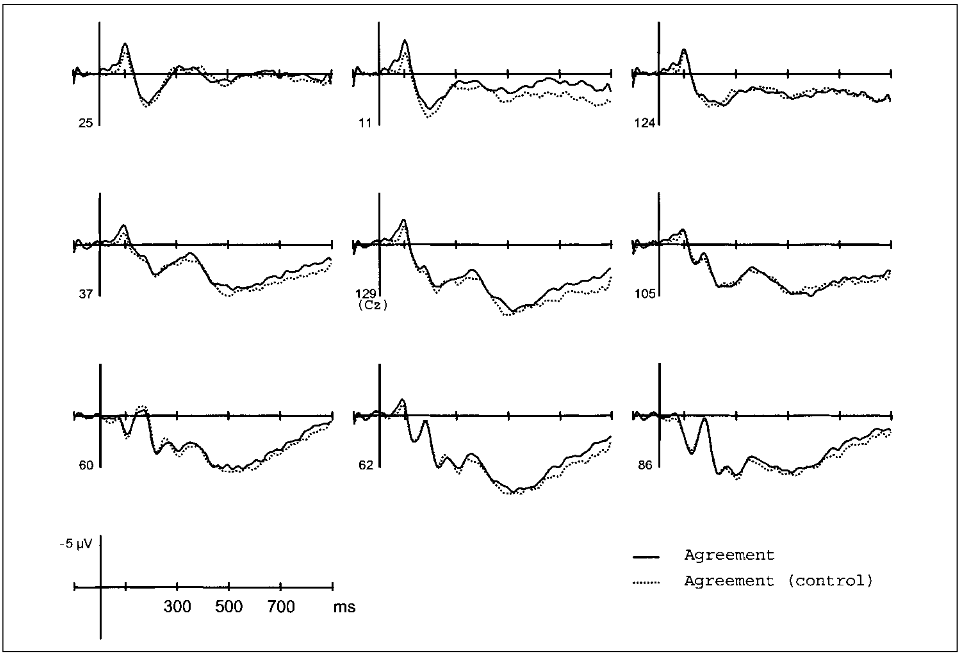

Fig. 2 – *ERPs to the target words for the experimental agreement condition and the control agreement condition. Three midline, three left and three right hemisphere electrodes sites are shown. Target onset is at 0 msec. Negative amplitude is plotted upward in this and the next figure.*

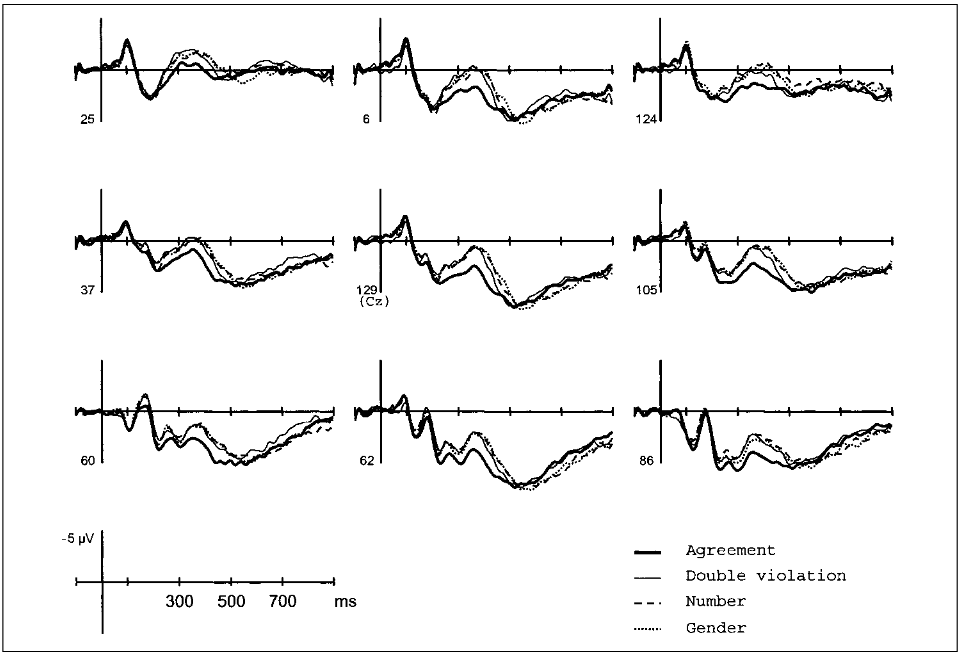

Fig. 3 – *ERPs to the target words for the experimental agreement, gender disagreement, number disagreement and double violation conditions.*

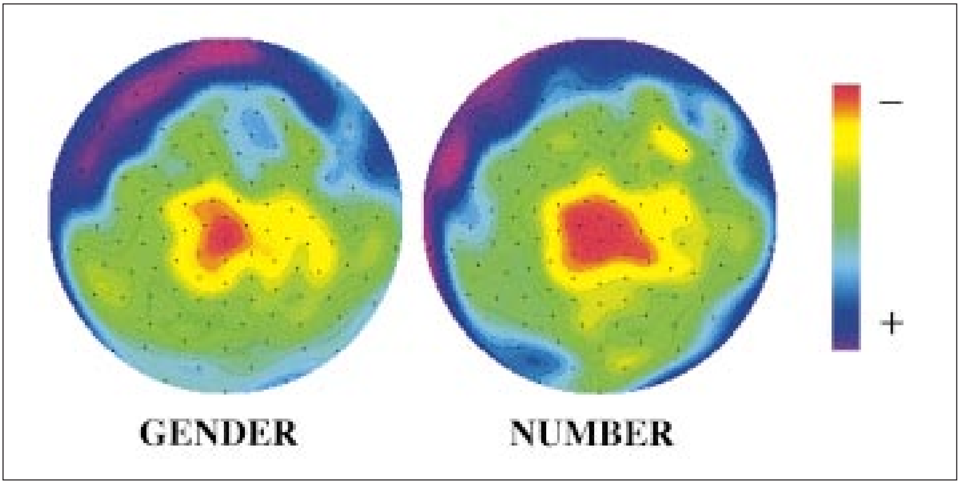

Fig. 4 – *Topographical maps obtained by interpolation. Maps were computed for difference waves from gender disagreement (left) and number disagreement (right) conditions use relative scaling to show distribution of negativities at 400 msec.*

inspection reveals clear differences in the responses to the agreement condition with respect to all disagreement conditions. Targets for the disagreement word pairs start to differ from the well-formed pairs at 250-300 msec. Between 300 and 500 msec disagreement waves were more negative going than those in the agreement condition, with largest amplitude differences at 400 msec (Figure 4 shows the topographical distribution of these effects). Furthermore, in this time window no differences between each disagreement condition are appreciated. Following these effects, the P3 component is present in all conditions showing peak latency differences across conditions. This component peaks earlier in the agreement condition and the double violation condition than in the gender disagreement condition and the number disagreement condition. Furthermore, the gender disagreement latency is slightly longer than the number disagreement condition. Mean values of peak latencies over central-posterior region can be found in the Figure 5.

The ANOVA for the time epoch 300-500 msec with the factor grammatical relation (agreement, gender disagreement, number disagreement and double violation) and the factor electrode regions revealed a significant main effect of grammatical relation both in the midline analysis [$F(3, 21) = 5.27$; $p < 0.01$; $\varepsilon = 0.91$] and the lateral regions analysis [$F(3, 21) = 5.16$; $p < 0.01$; $\varepsilon = 0.92$]. No interactions between grammatical relation and electrode regions (or hemisphere) were found in these analyses. *Post-hoc* tests resulted in a significantly less prominent amplitude for the agreement condition compared to the gender disagreement [$F(23, 1) = 9.35$; $p < 0.01$], number disagreement [$F(23, 1) = 8.89$; $p < 0.01$] and double violation [$F(23, 1) = 9.14$; $p < 0.01$] conditions, but the disagreement conditions did not differ from each other ($F < 1$).

P3 peaks latencies (maximum amplitude latencies between 450 and 750 msec) were introduced in an ANOVA with the grammatical relation factor (agreement, gender disagreement, number disagreement and double violation)

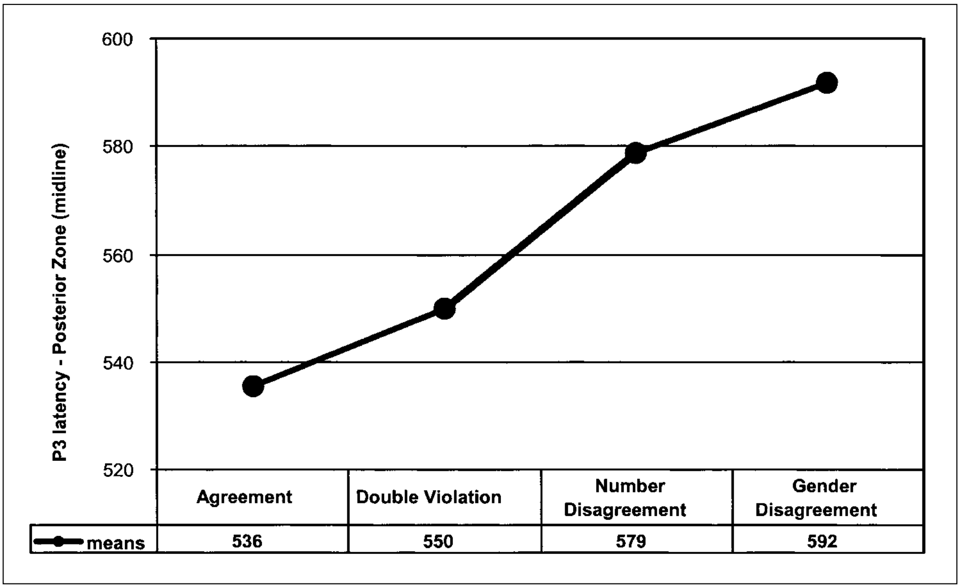

Fig. 5 – *Mean values in milliseconds after target onset of P3 peak latencies for all experimental conditions over midline posterior electrodes region (averaged).*

and the electrode region factor. Because of the posterior distribution of the P3 component only the posterior regions data (left, central and right) were included. This analysis yielded a significant main effect of the grammatical relation [F (3, 21) = 18.73; p < 0.001; ε = 0.79]. *Post-hoc* tests showed shorter latencies for the agreement condition as compared to both gender disagreement condition [F (1, 23) = 69.79; p < 0.001] and number disagreement condition [F (1, 23) = 26.54; p < 0.001], and significant differences between gender and number disagreement conditions [F (1, 23) = 11.02; p < 0.01]. Moreover, the tests did not show significant differences between the agreement condition and the double violation condition (F < 1), but the double violation latencies were significantly shorter than those of gender disagreement [F (1, 23) = 18.86; p < 0.001] and number disagreement [F (1, 23) = 5.95; p < 0.05] conditions.

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 a higher negative amplitude than those of agreement. In this latency range, N400 and LAN components are typically described. These two effects, even if they appear on the same temporal frame, have different distributions over the scalp and have been found to be sensitive to different experimental manipulations. The N400 component typically presents a central-posterior distribution, while LAN is described as a effect with an anterior left hemisphere localization. Our effects present a distribution that reaches central and posterior zones of the middle line, which is usually regarded as an N400 effect. Furthermore, part of the posterior distribution of the component could be eclipsed by the P3 overlapped later component, which also has a central and posterior distribution. Thus, the distribution and latency of the effect we have found match those of the classic N400 effect described in semantic priming experiments.

Generally speaking, the N400 is considered a reflection of word integration processes to build higher order representations. These processes would be influenced by the information available in working memory (contextual information) and in long-term memory, which depend on the lexical-semantic system and are widely distributed over different nets (Kutas and Ferdermeier, 2000). Given this functional interpretation of N400, this component has been mainly linked to a semantic processor. However, it is clearly the case that our data cannot be explained by such an account, especially considering that adjective morphemes lack of any referential dimension. Gender and number morphemes do not mark the lexical content of the adjective, but are just signallers of the gender and number of the noun to which they apply.

Several psycholinguistic models propose that lexical-semantic and morphosyntactic characteristics be analysed separately, without mutual influence (Forster, 1979). This autonomy is limited to the first lexical retrieval processes. In a second step, the information from both modules – lexical semantic and morphosyntactic – is used in post-lexical integration processes during which it is also contrasted with contextual information. Several studies have shown that grammatical priming effects are originated during these post-lexical processes (Seidenberg et al., 1984), in the same way as effects in N400 are linked to post-lexical integration processes (Kutas and Federmeier, 2000). If we assume that integration processes use information from the several modules, which contain different information and perform different analysis on the input signal, the concept of lexical-semantic integration that modulates N400 can be understood in a wide sense. N400 could reflect integration processes of the different characteristics associated with lexical representations, one of them being their morphologic features.

In previous studies with word pairs, grammatical disagreement has shown an effect with the same polarity as well as in the same time window, but with an anterior distribution (Münte and Heinze, 1994). In addition, Barber and Carreiras (submitted) found a LAN effect in Spanish when participants were considering whether article-noun pairs agreed in gender or in number, either in isolation or embedded in sentences. When words were presented in word pairs, an additional N400 effect, similar to that of this experiment, was also found. This being so, another aspect that must be considered in depth is the absence of effects in the anterior electrodes in our data.

The processes related to the establishment of grammatical agreement or to its detection can be shared by those that perform syntactic analysis of more complex structures, because agreement plays an important role in the construction of the syntactic tree. Thus, it would be expected that the effect produced by the lack of agreement in word pairs would be located in the anterior zones (i.e., a LAN effect). Word pairs of an article followed by a noun are noun phrases, a minimal syntactic unit, while the adjective-noun pairs used in this experiment need at least one determiner to form such a syntactic unit. Thus, it may not be difficult to accept that article-noun pairs can trigger the syntactic integration processes, while noun-adjective pairs, although susceptible to such integration, may not. The determiner-noun pair can and usually does constitute the beginning of a new sentence. Keeping in mind that the building – at least provisionally – of the syntactic structure is a process that takes place earlier than the integration of morphological characteristics (Frazier, 1987), it could also be a previous requisite for the agreement features integration to happen in the determiner-noun pairs. In contrast, noun-adjective pairs are not found in initial sentence positions, so do not constitute from the very first moment an entity to open a new syntactic tree and so syntactic integration processes may not take place. Therefore, the presence of an effect in anterior zones of the left hemisphere 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.

Finally, an alternative interpretation for the N400 effects observed could be entertained in terms of orthographic or phonological processing. The letter ending of each word pair end is the same in the agreement conditions, but it is different in the case of disagreement conditions. Thus, a word rhyme effect could be brought into play to account for the effects. However, the effect found in the disagreement conditions should also be present in the agreement condition with common adjectives (poesía-triste [poetry-sad]), in which final letters did not overlap. As has been shown, the common adjective condition did not differ from the agreement condition in which there was an overlap of the final letters. Thus, the effects cannot be accounted for by orthographic or phonological processes. Besides, the existence of filling pairs with different endings made it impossible for the task to be resolved attending only to phonological overlapping.

Gender versus Number

The main goal of this research was to study the possible differences between gender and number agreement; in other words, the effects of their agreement violation. Gender and number information can be represented differently in the lexicon. Gender seems to be a feature directly associated with the stem, while number seems to be a trait added to the stem during the retrieval process. This difference may not only affect storage and retrieval processes but could also have an impact on agreement processes. As has been pointed out before, the N400 component has been associated with lexical-semantic integration processes. Therefore, if gender is a feature associated with the morphological stem but number is not, a higher effect of N400 for gender disagreement as compared to number disagreement could be expected. Our data do not seem to support these predictions, nor the idea that gender and number information are processed through different mechanisms when integration takes place. Gender agreement violation produced the same N400 effect as number agreement violation did. No differences in amplitude, latency or distribution of the N400 were found for gender and number agreement violations. Furthermore, the simultaneous violation of gender and number was also studied in the present experiment. The same N400 effect was obtained again for the conjoined violation of gender and number as compared to for each type of violation separately. This result points again towards the idea of gender and number agreement sharing a common agreement mechanism, which acts as a binary switch when detecting the incongruence of one or both grammatical features.

Certainly, the absence of differences between both types of agreement can not be considered as proof against models that propose differences in gender and number syntactic processing, because null results may have multiples causes. In the same way, although it is tempting to consider that the presence of identical effects may imply the existence of shared processes in the integration of both types of information, this idea should be approached with caution because equivalent effects on the surface of the scalp could be a result of changes originated in different neural generators.

The negative component produced in the 300-500 msec window is followed by a positive component identified as P3. The identification of this component was made taking into account its polarity, latency, distribution, and sensitivity towards experimental manipulations. This positive wave is usually present in priming experiments using word pairs that require an immediate response (e.g. Bentin et al., 1985) and does not appear when the subject is simply required to read the words (Brown et al., 2000). In general, P3 appears each time the task requires a binary type decision (Donchin and Coles, 1988). Thus, P3 is associated with the type of task and with the response that participants have to execute just after the reading of the critical word. Although the polarity, latency and distribution of this component coincide with those of the P600/SPS effect, it must be pointed out that that the P600 effect only appears in syntactic transgressions or anomalies, while the positivity under discussion is also present in the agreement condition. Besides, the morphology of these effects (P3 vs. P600/SPS) differ slightly. The P600 is typically maintained for several hundreds of milliseconds and does not usually present a defined peak, although it should be borne in mind that the relationship between the P600 and the P3 is an open issue[6].

The sensitivity 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 categorisation of the stimulus must have finalized (Donchin, 1979). Although other processes may take place between categorisation and the visibility of the component, the P3 latency varies as a function of the difficulty of the stimulus evaluation (Kutas et al., 1977). On the other hand, later processes such as motor preparation of the answer do not seem to affect the latency of the P3. In experiments in which stimulus-response incompatibility was manipulated, effects on reaction times have been observed, but not on the P3 latency (McCarthy and Donchin, 1981).

[6] Some authors have claimed that the P600/SPS belongs to the family of P300 effects (Coulson et al.,1998; Gunter et al., 1997) although others authors do not share this interpretation (Osterhout and Hagoort, 1999).

Although there were no amplitude differences in the P3, its latency varies accordingly to the different experimental conditions. Gender and number disagreement conditions present higher latencies than agreement conditions. Importantly, the double violation condition caused a similar latency to that of agreement. Moreover, the gender disagreement condition produced a longer latency than that of number disagreement. The differences in latencies between conditions seem to be reflecting processes that are posterior to lexical analysis and word integration. Although the classic effect of this component has been located around the 300 msec, several studies have shown that this latency can be retarded depending on the complexity of the stimulus or the categorization difficulty that the subject must confront (Kutas et al., 1977). For this reason, this component usually shows up after a delay of around 500 msec delay with linguistic stimuli or faces. 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 (Rugg, 1983; Bentin et al., 1985) and repeated words have shorter latencies than those not repeated (Rugg, 1985).

Therefore, could changes in the latencies of the P3 component in our data be reflecting differences in the categorization of word pairs to prepare the response or could it be that the beginning of the categorisation process is occurring later in some cases[7]. In the case of grammatically agreeing word pairs, the 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. 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 syntactical integration processes and lexical access. 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 re-analysis processes would explain the higher latencies in the case of gender disagreement as compared to number disagreement. This conclusion is also consistent with some of our previous findings (Barber and Carreiras, submitted). In a sentence reading task we have found higher amplitudes for gender disagreement as compared to number disagreement in the final segment of the P600, while with article-noun word pairs, the latency of the P3 component was also higher for gender disagreement than for number. Thus, the results of both experiments suggest a higher cost in the last integration or closure processes, indicating that re-analysis processes seem to be more complex for gender information.

Regarding the pairs with a double gender-number violation, these items include two "incorrect" cues, so the decision can be taken quickly without any further checking. As a result, pairs that include double violation present similar latency to that of grammatically agreeing word pairs, and lower than that of each

[7] Peak latency was used in the analysis as a measure of the P3 component latency, although we could have taken the latency of the onset of the component. This measure was chosen due to there is the possibility of a N400 and P3 overlapping, so determination of the onset of the component point in some cases can be complex and comparisons among conditions may result in confusion. In any case, some works have showed that there is a close correlation between peak latency and onset latency in the P3 (Scheffers et al., 1991).

of the violations separately. It seems that at later stages, after integration failure, the combination of two features triggers the response mechanism more rapidly.

In summary, the manipulation of grammatical agreement produced effects in a negative component with central-posterior distribution that is named N400 and is usually linked with lexical integration processes. This component could be reflecting the detection of the impossibility of grammatical integration in a word pair or the activity that underlies the integration attempt itself. These processes are affected in the same way when the integration-impeding trait is gender, when it is number or both at the same time. In any case, even if the beginning of these processes is initiated and reaches maximal effect at the same moment in all disagreement experimental conditions, it may require additional checking. During this last stage, participants seem to use two tracks simultaneously with double violations pairs which allow them a quick decision, while in other cases of disagreement they need to pursue the additional checking, which is more complex for gender agreement rules. Therefore, although we did not find any differences between gender and number processing in their first stages, our data indicate that re-analysis processes are costlier in the case of gender disagreement detection.

Acknowledgments. This research was supported by grants from the Ministry of Science and Technology (BSO2000-0862 and BSO2001-3492-C04-03). We are grateful to David Corina for helpful comments on earlier drafts of this manuscript.

References

Barber H and Carreiras M. Gender and number representation during syntactic processing: An ERP study. Submitted.

Bentin S, McCarthy G and Wood CC. Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology, 60:* 343-355, 1985.

Bradley DC and Forster KI. A reader's view of listening. In UH Frauenfelder and LK Tyler (Eds), *Spoken word recognition*. Cambridge: MIT Press, 1987, pp. 103-133.

Brown C, Hagoort P and Chwilla D. An even-related brain potential analysis of visual word priming effects. *Brain and Language, 72:* 158-190, 2000.

Colé P and Segui J. Grammatical incongruency and vocabulary types. *Memory and Cognition, 22:* 387-394, 1994.

Corbett GG. *Gender*. Cambridge: University Press, 1991.

Coulson S, King J and Kutas M. ERPs and domain specificity: Beating a straw horse. *Language and Cognitive Processes, 13:* 653-672, 1998.

Domínguez A, Cuetos F and Segui J. The processing of grammatical gender and number in Spanish. *Journal of Psycholinguistic Research, 28, 5:* 485-497, 1999.

Donchin E. Event-related brain potentials: A tool in the study of human information processing. In H Begleiter (Ed), *Evoked potentials and behavior*. New York: Plenum, 1979.

Donchin E and Coles M. Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences, 11:* 357-374, 1988.

Faussart C, Jakubowicz C and Costes M. Gender and number processing in spoken French and Spanish. In E Di Domenico and M De Vincenzi (Eds), Gender and number in normal and impaired language processing. Special Issue of *Rivista di linguistica, 11:* 75-101, 1999.

Forster KI. Levels of processing and the structure of language processor. In WE Cooper and EC Walker (Eds), *Sentence processing*. Hillsdale: Erlbaum, 1979, pp. 27-85.

Frazier L. Sentence processing: A tutorial review. In M Coltheart (Ed), *Attention and performance*. Hillsdale: Erlbaum, 1987.

Friederici AD. The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and Language, 50:* 259-281, 1995.

Geisser S and Greenhouse S. On methods in the analysis of profile data. *Psychometrika, 24:* 95-112, 1959.

Gunter TC, Friederici AD and Schriefers H. Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience, 12:* 556-568, 2000.

GUNTER TC, STOWE LA and MULDER G. When syntax meets semantics. *Psychophysiology, 34:* 660-676, 1997.

HAGOORT P and BROWN C. 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, 2000.

HAGOORT P, BROWN C and GROOTHUSEN J. The syntactic positive shift (SPS) as an ERP-measure of syntactic processing. *Language and Cognitive Processes, 8:* 439-483, 1993.

HAGOORT P, BROWN C and OSTERHOUT L. The neurocognition of syntactic processing. In C Brown and P Hagoort (Eds), *Neurocognition of Language*. Oxford: Oxford University Press, 1999.

KUTAS M and FEDERMEIER KD. Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4:* 463-470, 2000.

KUTAS M and HILLYARD SA. Event-related brain potentials to grammatical errors and semantic anomalies. *Memory and Cognition, 11:* 539-550, 1983.

KUTAS M, MCCARTHY G and DONCHIN E. Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science, 197:* 792-795, 1977.

LUKATELA G, KOSTIC A, TODOROVIC D, CARELLO C and TURVEY MT. Type and number of violations and the grammatical congruency effect in lexical decision. *Psychological Research, 49:* 37-43, 1987.

MCCARTHY G and DONCHIN E. A metric of thought: A comparison of P300 latency and reaction time. *Science, 211:* 77-80, 1981.

MÜNTE TF and HEINZE HJ. ERP negativities during syntactic processing of written words. In HJ Heinze, TF Münte and GR Mangun (Eds), *Cognitive Electrophysiology*. Birkäuser, 1994.

OLDFIELD RC. The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia, 9:* 97-113, 1971.

OSTERHOUT L and HAGOORT P. 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, 1999.

OSTERHOUT L and HOLCOMB PJ. Event related potentials and language comprehension. In MD Rugg and M Coles (Eds), *Electrophysiology of mind, Event related brain potentials and cognition.* Oxford: Oxford University Press, 1995.

OSTERHOUT L and MOBLEY LA. Event-related potentials elicited by failure to agree. *Journal of Memory and Language, 34:* 739-773, 1995.

RITTER E. Where is gender? *Linguistic Inquiry, 24:* 795-803, 1993.

RUGG MD. Further study of the electrophysiological correlates of lexical decision. *Brain and Language, 19:* 142-152, 1983.

RUGG MD. The effects of semantic priming and word repetition on event related potentials. *Psychophysiology, 22:* 642-47, 1985.

SCHEFFERS MK, JOHNSON R and RUCHKIN DS. P300 in patients with unilateral temporal lobectomies: The effects of reduced stimulus quality. *Psychophysiology, 28:* 274-284, 1991.

SEBASTIÁN-GALLÉS N, MARTÍ MA, CARREIRAS M and CUETOS F. LEXESP: Una base de datos informatizada del español. Universitat de Barcelona, España, 2000.

SEIDENBERG MS, WALTERS GS, SANDERS M and LANGER P. Pre- and postlexical loci of contextual effects on word recognition. *Memory and Cognition, 12:* 315-328, 1984.

TUCKER DM. Spatial sampling of head electrical fields: The geodesic sensor net. *Electroencephalography and Clinical Neurophysiology, 87:* 154-163, 1993.

Horacio Barber, Departamento de Psicología Cognitiva, Universidad de La Laguna, 38205 S/C de Tenerife, Spain.
e-mail: hbarber@ull.es