An ERP study of coreference in Spanish: Semantic and grammatical gender cues

JUAN SILVA-PEREYRA, a,b EVA GUTIERREZ-SIGUT, c AND MANUEL CARREIRAS d,e

aProyecto de Neurociencias, FES Iztacala UNAM, Tlalnepantla, México
bCognitive Psychology Department, University of La Laguna, Tenerife, Spain
cCognitive Psychology Department, University of California, Davis, Davis, California, USA
dBasque Center on Cognition, Brain and Language, (BCBL), Donostia, Spain
eBasque Foundation for Science (IKERBASQUE), Bilbao, Spain
Department de Filología Vasca, Universidad del País Vasco (UPV-EHU), Bizkaia, Spain

Abstract

We report two event-related potentials (ERPs) experiments aimed to investigate the roles played by semantic and syntactic information during pronoun resolution. The first experiment was designed to show that ambiguity of the pronoun (e.g., word class ambiguity) makes an important contribution to the pattern observed in previous ERP studies. As expected, the results showed a different ERP pattern for ambiguous and nonambiguous pronouns. The second experiment analyzed pronoun resolution when gender agreement and animacy were manipulated, using only unambiguous pronouns. Results showed P600 effects at 500 to 700 ms and at 700 to 900 ms. Amplitude of the second window was significantly greater for animate than for inanimate antecedents. The modulation of the agreement effect by animacy suggests that repair processes after grammatical disagreement detection are influenced by semantics.

Descriptors: Coreference, Anaphoric pronoun resolution, Anaphora, Grammatical gender, ERP, P600

Pronominal Reference

An essential feature of discourse comprehension is the ability to resolve coreference relations and build cohesive links between sentences. One case of this coreference relationship is the link between pronouns and their antecedents. Pronouns are words with special features that allow the speaker to refer to antecedents (people, things, events, etc., previously mentioned) without having to re-enter them into the discourse. This long-distance reference is usually established according to rules of agreement.

In Spanish, gender can signal a semantic difference, such as in the case of animates, (e.g., niño–boy vs. niña–girl) indicating a biological gender difference, or can be a purely arbitrary grammatical marker (e.g., faro–lighthouse vs. mesa–table) as in the case of inanimates. It is well known that both semantic and syntactic information play roles in pronoun resolution in sentences. However, it is still unclear what the relative contribution of these sources of information is in the establishment of a coreferential relationship between the pronoun and the antecedent in combination with a local structural case constraint on the pronoun.

This research was partially supported by grants PSI2009-08889 and CONSOLIDER-INGENIO2010 CSD2008-00048 from the Spanish Ministry of Science and Innovation, and by grant ITN-LCG-GA-2009-237907 from the European Union.

Address correspondence to: Juan Silva-Pereyra, Proyecto de Neurociencias, UICSE FES-Iztacala UNAM, Ave de los Barrios 1, Los Reyes Iztacala, Tlalnepantla, Edo. de México, C.P. 54090, México. E-mail: jsilvapereyra@gmail.com

In many cases of reference resolution (especially when the referential form and its referent are in different clauses), nonstructural information appears to play a crucial role (e.g., Kehler, 2002; Wolf, Gibson, & Desmet, 2004). For instance, some researchers have argued that semantic information influences the interpretation of pronouns with cross-clausal antecedents (Arnold, 2001; Koornneef & Van Berkum, 2006; McDonald & MacWhinney, 1995; Stevenson, Crawley, & Kleinman, 1994; Stewart, Pickering, & Sanford, 2000). Moreover, Arnold (2001) found that the thematic role of an entity (e.g., agent, goal, source, experiencer) influences its likelihood of being the antecedent for a subsequent pronoun. During utterance comprehension, discourse structure as well as type of anaphor and pragmatic inference interacts in the process of integrating each new utterance with the prior discourse (Marslen-Wilson, Tyler, & Koster, 1993). Even when the anaphor is unambiguous, reference assignment may still require a concurrent inferential check on the pragmatic plausibility of the proposed antecedent (Tyler & Marslen-Wilson, 1982).

However, it has also been argued that purely formal syntactic information may alone determine the choice of antecedent for some referential forms in Romance languages. Carreiras, Garnham, and Oakhill (1993), and Garnham, Oakhill, Ehrlich, and Carreiras (1995) (see also Cacciari, Carreiras, & Cionini, 1997; Cacciari, Corradini, Padovani, & Carreiras, 2011) have provided evidence that anaphoric pronouns can be linked with antecedents by processing superficial aspects of text representation that have no semantic content, such as grammatical gender. They showed that pronouns were easier to interpret when their referents were made unambiguous by grammatical gender, even when the gender was arbitrary...
gender information is critical for understanding the meaning of sentences and for building a representation.

Garrod and Sanford (1994; see also Garrod & Terras, 2000) presented a two-stage model in which the initial bonding phase involves relatively superficial information and is an automatic, more syntactic process associated with establishing a link between anaphor and a previous expression. This is followed by a later resolution phase or immediate integration stage, which involves integration of the interpretation with contextual information. The integration is responsible for evaluating the established link (i.e., by weighting syntactic and semantic discourse information).

Event-Related Potentials

Event-related potentials (ERPs) have been used to investigate pronoun or coreferential processing (see Callahan, 2008). ERPs are averages of brain electrical activity time-locked to some external or internal event and classified according to their polarity (i.e., positive or negative deflections in the waveform), the time of their peak occurrence in milliseconds, and their topographical distribution across the scalp. Some ERP studies have established that subjects may attempt to locate a correct antecedent for an anaphor as early as 280 ms after its presentation (van Berkum, Brown, & Hagoort, 1999). A left anterior negativity (LAN) has been observed for antecedent retrieval from working memory in studies of anaphora interpretation in movement constructions (King & Kutas, 1995); however, no evidence for the occurrence of this component in pronominal reference has been reported. This is probably because LAN is mainly found with structure-building syntactic processes at the sentence level (Friederici & Weissenborn, 2007). On the other hand, N400–N700 effects have also been reported in sentence contexts when readers process pronouns or closed-class words (Kutas, Van Petten, & Kluender, 2006). In addition, in discourse contexts, referential ambiguity evokes a sustained anterior negativity (Nref), which might be an electrophysiological correlate of additional retrieval processes (Nieuwland & van Berkum, 2008).

Once the appropriate antecedent is selected and retrieved, the information must be integrated with the pronoun and the surrounding context, and this is probably reflected in N400 effects (Hammer, Jansma, Lamers, & Münte, 2005; Schmitt, Lamers, & Münte, 2002). The N400, a central posterior negativity peaking approximately 300–500 ms poststimulus onset, is typically interpreted as reflecting the difficulty associated with integrating a word into a semantic or discourse representation and is taken to index semantic processing (Kutas, Federmeier, Coulson, King, & Münte, 2000; see also Molinari, Conrad, Barber, & Carreiras, 2010).

When the features (e.g., case, gender, number) of the pronoun are incompatible with the only possible antecedent, a P600 is elicited. This is a broadly distributed positivity between 400 and 1,000 ms, with a central parietal or parietal maximum at approximately 500–800 ms after stimulus onset (Hammer et al., 2005; Harris, Wexler, & Holcomb, 2000; Lamers, Jansma, Hammer, & Münte, 2006, 2008; Nieuwland & van Berkum, 2006; Osterhout, Bersick, & McLaughlin, 1997; Osterhout & Mobley, 1995; Schmitt et al., 2002). The P600 is elicited by a wide variety of linguistic contexts including, but not necessarily limited to, those involving syntactic violations or increased syntactic complexity (see Osterhout, Kim, & Kuperberg, 2007, for review). The P600 has also been seen as an index of increased processing demands during integration, arising from either syntactic or conceptual complexity (Carreiras, Salillas, & Barber, 2004; Kaan & Swaab, 2003).

ERP Studies of Pronominal Reference Resolution and Gender Agreement

One strategy to study whether syntactic, semantic, or both types of information are involved in reference resolution has been to gather ERP data on conceptual and grammatical gender agreement with an antecedent. Osterhout et al. (1997) analyzed the effects of social gender stereotypes in English. Gender incongruent pronouns elicited greater P600s than congruent pronouns. Moreover, the nonstereotyped gender condition showed a larger P600 effect at reflexive pronouns than the stereotyped gender condition (e.g., “The girl prepared herself/himself for the interview”) and “The nurse prepared herself/himself for the interview”). Violations of biological gender can also elicit an N400 effect in addition to the P600. This probably reflects semantic processing difficulty beyond the syntactic difficulty elicited by violations of syntactic gender. For example, Schmitt et al. (2002) examined the separate effects of semantic and syntactic gender violation during pronoun processing in German. They presented sentences such as #1 (below) with two clauses where the first noun phrase (NP) was the antecedent of a personal pronoun that appears in a second clause. The first NP had biological gender (male/female) while the pronoun could be either masculine, feminine, or neuter. Schmitt et al. (2002) also manipulated sentences using nondiminutives (Der Bub) and diminutives (Das Bübchen) as antecedents. This experimental manipulation is useful to show a double syntactic and semantic gender violation in the same sentence. As a consequence of the diminutive form, the syntactic gender of the noun changes from masculine or feminine to neuter but the conceptual/biological gender is kept. In German, condition 1c (Das Bübchen) can allow a neuter pronoun (es).

1. a. Der Bub/Das Bübchen masculine will schlafen und darum schaltet er masculine eine Lampe aus.
   b. Der Bub/Das Bübchen masculine will schlafen und darum schaltet sie feminine eine Lampe aus.
   c. Der Bub/Das Bübchen masculine will schlafen und darum schaltet es neuter eine Lampe aus.
   [The (little) boy wants to sleep and therefore he/she/it switches a light off.]

Overall, the results showed salient P600 effects for pronouns, with violations (sentences 1b [Der Bub/Das Bübchen] and 1c [Der Bub]) more positive than nonviolations (1a [Der Bub/Das Bübchen]). Although a neuter pronoun can allow an antecedent in diminutive, there were also larger P600 effects to sentences 1c (Das Bübchen) compared to sentences 1a (Das Bübchen). Furthermore, N400 effects were observed for nondiminutives, with violations more negative (1b) than nonviolations (1a), but not for diminutives.

In order to explore whether these previous German findings could be generalized cross-linguistically to pronoun resolution for diminutives and nondiminutives in a closely related language, Lamers et al. (2008) carried out a similar study in Dutch. For sentences with a syntactically and biologically incongruent pronoun, a P600 (in absence of an N400) was obtained, which was independent of the morphological form of the referent. Additionally, in Dutch, pronoun resolution in sentences with a nondiminutive antecedent seemed to reflect processes of revision (P600 in absence of an N400), whereas for German, evidence was found for clear involvement of both conceptual/semantic processes and structure-building processes (N400/P600 complex) (Schmitt et al., 2002). Previously, Lamers et al. (2006), in both German and Dutch, manipulated gender congruency between the antecedent and the
pronoun and/or case assignment by varying the preposition. They
found that gender agreement violations showed an enlarged
broadly distributed negativity starting at approximately 350 ms
after onset of the pronoun, followed by a late positivity (N400/
P600 complex).

Hammer et al. (2005) and Hammer, Jansma, Lamers, & Münte
(2008), using a different experimental design, introduced a person
or object as an antecedent and then later referred to this by a
pronoun, which either agreed with or violated biological/syntactic
gender (as sentences in #2). The woman/The jacket (Die Frau/Die
Jacke) agrees in gender with sie but disagrees with the er pronoun.

2. a. Die Frau ist beliebt, weil sie schön ist.
   b. Die Frau ist beliebt, weil er schön ist.
   [The woman is popular because she/he is beautiful.]
   c. Die Jacke ist warm, weil sie gefüttert ist.
   d. Die Jacke ist warm, weil er gefüttert ist.
   [The jacket is warm because she/he is lined.]

Their ERP results showed a P600 effect with a central parietal
maximum for incongruent pronouns (2b and 2d) compared with
congruent pronouns (2a and 2c) in both conditions. This effect was
larger for person compared to object antecedents, and the authors
interpreted this finding as reflecting a syntactic integration process
that can be influenced by conceptual/semantic information of the
antecedent type. N400 effects were also observed after the pronoun
onset, which was greater for object antecedents than for person.
They interpreted this finding as reflecting the search for an actor
outside the sentence in the case of an incongruent pronoun.

In summary, the ERP studies carried out on pronoun compren-
hension so far have not revealed a clear pattern of results with
regard to the question of what kind of information is used to link a
pronoun with its antecedent. The current study aims to investigate
this process by taking advantage of grammatical gender agreement
in Spanish.

The Present Study

Most previous ERP studies have focused mainly on the coreferen-
tial relationship between the pronoun and the antecedent by con-
trasting biological antecedents that agree with or violate the gender
of the pronoun (Lamers et al., 2006, 2008; Schmitt et al., 2002).
These ERP studies are problematic in that they included stimuli
that are ambiguous in terms of word class and case marking. For
example, the Dutch pronoun het can be a pronoun or an article, just
like la in Spanish, so that, on seeing these words, a reader may wait
for a new noun rather than trying to establish anaphoric reference.
This is also the case for reading sentences with the German
pronoun sie, which is the same in accusative and nominative forms.
Thus, if one has interpreted sie in nominative form, one would
establish the reference as in 2c. Instead, if one has interpreted sie
in accusative, one would wait for a prepositional phrase (i.e., “Die
Jacke ist warm, weil sie in Russland gekauft hat” [The jacket is
warm because she bought it in Russia]). Therefore, it is important
to investigate pronoun resolution in situations where these ambi-
guities are not present, to see whether the agreement mechanism is
initially sensitive only to syntactic cues, as has been demonstrated
for local relations (e.g., Barber & Carreiras, 2005, see Molinaro,
Barber, & Carreiras, 2011, for a review).

The purpose of this study was to investigate gender agreement
during pronoun processing. In order to show that word class ambigu-
guity is an important factor in these kinds of studies, the first
experiment of the present study included the Spanish pronoun la,
which can also be an article. In the second experiment, we specific-
ally examine whether the agreement process is the same or dif-
f erent when antecedents are animate or inanimate nouns, using an
experimental design where the word class ambiguities were not
present. In Spanish, all nouns have syntactic gender (masculine and
feminine), but only animate nouns have semantic gender.

If sources of possible confusion are removed, the role of seman-
tic and syntactic information on pronoun resolution can be teased
apart. For instance, if syntactic agreement guides the bonding
process, the electrophysiological response should be similarly
modulated, at least initially, by gender agreement for animate and
inanimate antecedent-type sentences, and so a P600 should be
expected. For example, when reading sentence 3a, the antecedent
“El luchador” [The fighter masculine] is in the focus of attention and
could be easily integrated when lo is encountered.

3. a. [El luchador /La luchadora] masculine / feminine estaba intacta/a
   porque lo masculino protegieron del golpe.
   [The fighter was intact because he was protected from the
   blow.]
   b. [El foco/La bombilla] masculine / feminine estaba intacta/a porque
   lo masculino protegieron del golpe.
   [The light was intact because it was protected from the blow.]

The bonding process would rely on grammatical features such as
gender and subsequently could proceed with the integration stage.
In contrast, reading sentences 3a (“La luchadora” feminine) or 3b (“La
bombilla” feminine) where there is only one antecedent available with
incorrect gender, the system can initially detect a morphosyntactic
“dead end” during the first bonding stage; therefore, a P600 should
be expected. Thus, the P600 should be similar for animate and
inanimate antecedents.

On the other hand, if semantics plays a role in the beginning of
the pronoun resolution process, when the reader reaches the
pronoun and realizes that there is only one possible antecedent
but with incorrect gender, the most likely prediction would be a
larger P600 whether the antecedent is an animate (“La luchadora”
feminine) rather than inanimate noun (“La bombilla” feminine) due to a
mismatch in semantic and also syntactic gender on the animate
noun.

Experiment 1

In order to investigate anaphoric pronoun resolution and ascertain
whether the previous results could be modulated or contaminated
by the use of ambiguous pronouns, ERPs were used to verify
whether ambiguous and nonambiguous pronouns give rise to the
same outcome. La and lo are accusative forms of feminine and
masculine pronouns in Spanish, respectively. However, the word
“la” is also the feminine form of the determiner article. This is not
the case with lo, the masculine accusative form of the pronoun.
Therefore, in Experiment 1, we used sentences in which the gender
of antecedent was manipulated, followed by one of the two pro-
nouns la or lo. For example, participants read sentences with inani-
mate antecedents, as displayed in Table 1. If la is processed only as
a pronoun, a referential assignment can be made, and there will be
similar ERP patterns with respect to lo processing. However, if la is
also considered as an article, the processing and the ERP patterns
should be different from those of the lo pronoun. We used only
inanimate antecedents to keep the design simple within the limits
of an experimental session and because agreement with inanimates
represents a purely syntactic agreement situation.
Table 1. Example Materials for the Experiments

<table>
<thead>
<tr>
<th>Condition</th>
<th>Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>lo A</td>
<td>El foco Masculine estaba intacto porque lo Masculine protegieron del golpe.\textsuperscript{a}</td>
</tr>
<tr>
<td>lo D</td>
<td>El foco Masculine estaba intacto porque la Feminine protegieron del golpe.</td>
</tr>
<tr>
<td>la A</td>
<td>La bombilla Feminine estaba intacta porque la Feminine protegieron del golpe.</td>
</tr>
<tr>
<td>la D</td>
<td>La bombilla Feminine estaba intacto porque lo Masculine protegieron del golpe.</td>
</tr>
</tbody>
</table>

Experiment 2

| Animate A | El enfermo Masculine estaba irritado porque lo Masculine descuidaron por mucho tiempo.\textsuperscript{b} |
| Animate D | La enferma Feminine estaba irritada porque lo Masculine descuidaron por mucho tiempo. |
| Inanimate A | El sobaco Masculine estaba irritado porque lo Masculine descuidaron por mucho tiempo. |
| Inanimate D | La axila Feminine estaba irritada porque lo Masculine descuidaron por mucho tiempo. |

Notes. A = agree; D = disagree.
\textsuperscript{a}The light Masculine/Feminine was intact because it Masculine was protected from the blow.
\textsuperscript{b}The sick person Masculine/Feminine was irritated because he Masculine was neglected for a long time.

The armpit Masculine/Feminine was irritated because it Masculine was neglected for a long time.

Methods

Participants. Twenty-seven Spanish-speaking young adults (18 females; mean age = 18.9; SD = 0.6; range = 18–20) were recruited from the psychology department at the University of La Laguna. All participants were healthy (with no history of neurological illness), and they had normal or corrected-to-normal vision. All subjects included in the analyses were right-handed as assessed by an abridged Spanish version of the Edinburgh Handedness Inventory (Oldfield, 1971); LQ > +50. No subjects had a family history of left-handedness. All participants were informed of their rights and gave verbal informed consent for participation in the study. This research was carried out fulfilling ethical requirements in accordance with standard procedures at the University of La Laguna.

Stimuli. Nouns and verbs from LEXESP (Computerized data base of Spanish language; Sebastián, Martí, Cueto, & Carreiras, 1996) were used to build 160 two-clause sentences using anaphoric pronouns at the beginning of the second clause. All nouns of antecedents were inanimate. Eighty sentences were agreement sentences divided into 40 lo-agreement and 40 la-agreement. Eighty sentences had gender disagreement between the noun subject of the first clause and the pronoun of the second clause (see Table 1). The mean frequency was 27.99 appearances/million words (SD 55.1) for lo-agreement; 25.5 (41) for lo-disagreement; 29.1 (45.8) for la-agreement; 25.4 (72.4) for la-disagreement. The frequency was controlled for agreement and disagreement conditions for la, t(318) = 0.46, p = .65, and for lo, t(318) = 0.55, p = .6. Eighty additional sentences were included as fillers in order to have 240 sentences divided into 120 grammatical and 120 ungrammatical sentences. These were sentences with different syntactic structures from an unrelated experiment. All sentences were between 9 and 10 words in length. A Latin square was used to create four lists of materials and to ensure that each sentence occurred in each of the within-materials conditions.

Procedure. A fixation point (+) appeared in the center of the screen and remained there for 700 ms. This fixation point was followed by a blank screen interval of 300 ms, then the sentence was displayed word by word. Each word appeared for 300 ms and was followed by a blank interval of 300 ms. Participants were instructed to read each sentence carefully and respond after a question mark that appeared at the end of the sentence. Their task was to decide whether the sentence was or was not correct in Spanish. The question mark was presented and remained displayed up to a maximum of 2,000 ms or until the participant responded.

ERP recording. The electroencephalogram (EEG) was recorded from 64 Ag/AgCl electrodes embedded in a standard electro-cap, each referenced online to linked earlobes. Eye movements were monitored through a bipolar recording from two electrodes placed on the outer canthi of each eye and above and under each eye. Electrode impedances were maintained below 10 kΩ. The EEG was amplified with Brain Vision amplifiers with the band-pass set from 0.1 to 100 Hz and sampled at a rate of 250 Hz. Trials with artifacts due to eye movements, excessive muscle activity, or amplifier blocking were eliminated offline before averaging—approximately 5% of the data for each of the targets (with roughly equal loss of data across conditions).

Data analysis. ERPs were computed offline using 1,100 ms epochs from each subject in each experimental condition. Epochs consisted of the 200 ms preceding the stimulus and 900 ms following the presentation of each individual critical word in each of the sentences. Automatic rejection of segments was carried out based on the following criteria: Segments with electrical activity exceeding ±100 μV and amplifier blocking for more than 50 ms at any electrode site were considered artifact, and the entire segment was rejected. Subjects with fewer than 20 artifact-free trials for each condition were excluded from the average. Further digital filtering was conducted using a band-pass of 0.1 to 20 Hz. Baseline correction was performed using the 200-ms prestimulus time window mentioned above. Averaged trials included only those with correct responses. Average number of EEG segments per condition was approximately equal (la-agreement Mean = 25.2, SD (4.2), la-disagreement 27.6 (5.3), lo-agreement 26 (5.4), la-disagreement 26.6 (5.2).

Nine regions of interest were computed out of the 54 electrodes, each containing the average of a group of 6 electrodes. The regions were (see Figure 1) left anterior (F3, F3a, F5, F7, C3a, C5a), left central (C3, C3p, C5, T3, T3l, tcp1), left posterior (C1, P3, P3p, P5, O1, T5), medial anterior (F1, F2, Fz, C1a, C2a, Cz), medial central (C1, C2, Cz, C1p, C2p, Pza), medial posterior (P1, P2, P1p, P2p, Pz, Pzp), right anterior (F4, F4a, F6, F8, C4a, C6a), right central (C4, C4p, C6, T4, T4l, tcp2), and right posterior (C2b, P4, P4p, P6, O2, T6).

A series of repeated measures ANOVAs was performed on mean amplitude values for the comparisons between disagreement and agreement sentences across experimental conditions. These analyses were done within the same ERP time windows as previous studies (Barber & Carreiras, 2005; Silva-Pereyra & Carreiras, 2007): 300 to 500 ms for the N400 component, 500 to 700 ms and 700 to 900 ms for the P600 component. Four-way repeated
measures ANOVAs were performed for each time window with Pronoun (la and lo), Agreement (agree and disagree), Anterior-Posterior (frontal, central, and parietal) and Laterality (left, medial, right) as factors. The Huynh-Feldt correction was applied to analyses with more than one degree of freedom in the numerator. The Tukey’s Honestly Significant Difference (HSD) method was used for post hoc pairwise comparisons in the repeated measures analyses. For all the reported post hoc comparisons, gender disagreement was associated with gender disagreement were smaller (i.e., less positive) than those in the agreement conditions. This effect was very similar to those found by others (Hammer et al., 2005; Lamers et al., 2006), and referred to as an N400-like effect. This was observed for the feminine pronouns and had a posterior distribution over the scalp. This N400-like effect was followed by a typical P600 effect, showing larger amplitudes for the gender disagreement conditions than agreement, but this effect was only observed for the masculine pronoun conditions. The P600 effect started around 450 ms and was maintained for more than 200 ms. Between 500 and 700 ms, the P600 effect showed a posterior distribution.

300–500 ms time window. In this time window, there was significant Pronoun × Agreement interaction, \( F(1,26) = 9.4, p = .005 \). Tukey HSD post hoc analyses revealed smaller amplitudes for gender disagreement than for agreement conditions only on the la pronoun (mean difference, \( MD = .87 \mu V, p = .04 \)). This N400-like effect displayed a lateralized distribution (Pronoun × Agreement × Laterality interaction, \( F(2,52) = 3.13, p = .05, \epsilon = 1 \); specifically this effect was distributed over the left hemisphere (HSD \( MD = .9 \mu V, p = .014 \)).

500–700 ms time window: P600a. Repeated measures ANOVA revealed a significant Pronoun × Agreement interaction, \( F(1,26) = 4.5, p = .04 \). The la pronoun gender disagreement condition elicited a positive wave that was not significantly different from the agreement condition (\( F < 1 \)). In contrast, HSD post hoc for the lo pronoun showed that gender disagreement elicited a greater positive amplitude than the agreement condition (P600 effect; \( MD = -1.6 \mu V, p = .004 \)). This effect mainly displayed a posterior distribution (Pronoun × Gender agreement × Anterior-Posterior interaction, \( F(2,52) = 5.1, p = .03, \epsilon = .53 \)) and was observed across posterior (HSD \( MD = -2.22 \mu V, p = .0001 \)) and central regions (HSD \( MD = -1.6 \mu V, p = .003 \)), also displaying a right hemisphere bias (Pronoun × Gender agreement × Laterality, \( F(2,52) = 3.1, p = .05, \epsilon = 1 \)). The effect was mainly observed in the medial (HSD \( MD = -1.9 \mu V, p = .002 \)) and right regions (HSD \( MD = -1.6 \mu V, p = .002 \)).

700–900 ms time window: P600b. There were no significant interactions or main effects in this time window.

Discussion

The electrophysiological correlates of the la gender disagreement compared to the agreement condition showed effects only on amplitude around 400 ms with a posterior distribution, and there was no P600 effect. This is precisely what Lamers et al. (2008)
found in an ERP study, which included the neuter Dutch pronoun *het* in their experimental materials. This word is both a pronoun and the article of a neuter noun. Sentences with this word-class (pronoun/article) ambiguous pronoun *het* elicited an early negative shift, which continued in the time frame of the N400 (280–400 ms). This second part of the effect was considered to be a smaller amplitude of a positive wave corresponding to the agree condition. This pattern is similar to the one reported here for the *la* pronoun. Thus, Lamers et al.’s findings can be explained in terms of language-specific characteristics concerning the word-class ambiguous neuter pronoun *het* in combination with the difference in frequency of occurrence within sentences. In Spanish, the pronoun *la* and the article *la* are identical. It is possible that the interpretation of the pronoun *la* in the first stage is linked to the syntactic categorization of words, and the parser has two alternatives since it also considers *la* as an article or pronoun. Since there is no P600 effect in *la* gender disagreement, it is quite likely that *la* is being considered as the article of a NP because, as in Dutch, in Spanish *la* is more frequently used as an article than as a pronoun. In contrast, the response to the *lo* gender disagreement, where there is no word class ambiguity, is a positivity that appeared around 450 ms, maintained for about 300 ms, and displaying a P600 brain topography.

As Lamers et al. (2008) and the results of the present experiment have shown, there is an electrophysiological pattern in relation to word class ambiguity that is independent of the effects associated to pronoun resolution. Thus, the N400/P600 pattern found in Schmitt et al.’s and Hammer et al.’s ERP studies could be a combined effect of the manipulation of grammatical case marking, gender agreement, and animacy using German language stimuli. In sentences with unambiguous morphological case marking, the thematic properties compatible with the sentential arguments are actively used during processing, and a problem occurs when it is not possible to establish a hierarchical thematic ordering between two arguments. For example, in an ERP study, Frisch and Schlesewsky (2001) analyzed grammatical and ungrammatical sentences with two arguments (see also Schlesewsky & Bornkessel, 2004). The second argument of the sentence was rendered ungrammatical at the position of the second phrase, since both arguments are identically case marked. The difference between the two ungrammatical structures is that the argument inducing the ungrammaticality is animate in one sentence type and inanimate in the other. Each of the ungrammatical sentences elicited a P600 at the position of NP2 in comparison to its respective grammatical control sentences. Furthermore, in the ungrammatical condition with two animate arguments, a central parietal negativity (N400) was also apparent in comparison to the control condition. Frisch and Schlesewsky (2001) interpreted the N400 for sentences with two animate and identically case-marked arguments as reflecting inability to establish hierarchical thematic ordering between the

---

**Figure 2.** This figure shows ERP grand averages of all conditions across nine regions of interest. Negative is plotted up. Thicker lines represent agree conditions and thinner lines disagree conditions. Dotted gray lines represent *la* pronouns and solid black lines *lo* pronouns. *La* disagree condition elicited an amplitude effect around 400 ms with left posterior distribution and *lo* disagreement a P600 effect. Time window analyses are also marked.
arguments. By contrast, in sentences with two identically case-marked arguments that differ with regard to their animacy status, the language processing system receives an additional cue about which argument should be interpreted as thematically higher ranked and can therefore establish thematic ordering. Hence, there is no N400 effect observed.

In sum, while the ERPs linked to the processing of the lo pronoun seem to reflect only pronoun comprehension, the ERPs associated with the interpretation of the la word (pronoun/article) could reflect several other processes, including those linked to word class ambiguity, and so are inappropriate for investigating pronoun resolution. As we predicted, the electrophysiological pattern is different when word ambiguity is present, and previous ERP results could be challenged by the present findings. The N400/P600 pattern could vanish if word class ambiguity is controlled.

**Experiment 2**

Experiment 2 investigated the resolution of pronouns with animate and inanimate antecedents in an experimental design avoiding word class ambiguities. In this experiment, sentences with the pronoun la were removed and the lo trials were retained from Experiment 1. In addition, we added sentences with animate antecedents. It is important to notice that using the same pronoun (i.e., lo) and only manipulating the gender agreement of the antecedent leaves a cleaner comparison (see Table 1), and as results from Experiment 1 have shown, the N400/P600 complex might not be present during pronoun resolution, even with animate antecedents.

Based on previous findings regarding biological/semantic and syntactic gender agreement on pronoun resolution, we predicted two possible outcomes for this experiment: (1) According to the two-stage bonding model, syntactic information is used in a first stage and semantics in a second during pronoun resolution, so a P600 effect should be expected with similar size for the two types of antecedents. (2) Alternatively, if semantic information is used from the very beginning during the bonding and integration processes, a P600 effect should be expected, and this effect should be larger for animates than inanimates.

**Methods**

**Participants.** Twenty-nine Spanish-speaking young adults (19 females; mean age = 18.6, SD = 1.2; range = 19–21) were recruited from psychology undergraduate courses at the University of La Laguna. All participants were healthy (with no history of neurological illness), and they had normal or corrected-to-normal vision. All subjects included in the analyses were right-handed as assessed by an abridged Spanish version of the Edinburgh Handedness Inventory (Oldfield, 1971): LQ > +50. All subjects had no family history of left-handedness. This research was carried out fulfilling ethical requirements in accordance with standard procedures at the University of La Laguna.

**Stimuli.** Nouns and verbs from LEXESP were used to build 160 two-clause sentences using anaphoric pronouns at the beginning of the second clauses. Care was taken to match word frequencies for persons (25.2 ± 60) and things (32.8 ± 67.3, 76.9 ± 377; two-tailed = -1.53, p = .13). Eighty were control gender agreement sentences divided into 40 animate antecedent-type sentences and 40 inanimate antecedent-type sentences. Eighty were experimental gender disagreement sentences also divided in 40 animate antecedent-type and 40 inanimate antecedent-type sentences (see Table 1). All sentences were between 9 and 10 words in length and contained a noun phrase as an object. Fifty percent of the sentences were correct Spanish sentences, and the other half were ungrammatical sentences. One hundred additional sentences were included as filler sentences in order to have 130 grammatical and 130 ungrammatical sentences. These fillers were sentences with the same structure but using la as a pronoun, to avoid participants engaging in strategic reading. A Latin square was used to create four lists of materials and to ensure that each sentence occurred in each of the within-materials conditions.

**Procedure.** This experiment followed the same procedure regarding timing of stimulus presentation and ERP recordings as Experiment 1.

**Data analysis.** Data analyses for this experiment were the same as in Experiment 1, but in this experiment ERPs were computed offline from 1,000-ms epochs for each subject in each experimental condition. Epochs consisted of the 100 ms preceding and 900 ms following the presentation of each individual critical word in each of the sentences. Baseline correction was performed using the 100-ms prestimulus time window mentioned above. Averaged trials include only those with correct responses, and the average number of EEG segments per experimental condition was approximately equal (animate-agreement 28.2 (4.3), animate-disagreement 27.1 (3.5), inanimate-agreement 27.4 (5.7), inanimate-disagreement 27 (4.6).

The statistical analyses were the same as the first experiment, but in this one a series of four-way repeated measures ANOVAs was performed for each time window with Animacy (animate and inanimate), Agreement (agree and disagree) Anterior-Posterior (frontal, central, and parietal), and Laterality (left, medial, right) as factors. For all the reported post hoc comparisons, gender disagreement versus agreement conditions were compared. Only interactions that involved Animacy and Agreement are reported.

**Results**

Percentages of correct responses were transformed using ARCSIN [SQRT (percentage/100)]. Two-way repeated measures ANOVAs were performed for transformed percentage and reaction times data with Animacy (animate and inanimate) and Agreement (agree and disagree) as factors. There were no significant main effects or Animacy × Agreement interaction on reaction times (F < 1). However, in terms of percentage of correct responses, there was a marginal Animacy × Agreement interaction, F(1,28) = 3.5, p = .07, which means a higher percentage of correct responses for the agree condition in inanimates antecedents (HSD MD = .2, p = .001). There were also significant Animacy, F(1,28) = 16.3, p = .0001, and Agreement main effects, F(1,28) = 10.4, p = .003. These results indicate a higher percentage of correct responses to animate than to inanimate conditions and to gender agreement than to disagreement conditions, as can be seen in Table 3.

Figure 3 shows grand average ERPs for gender disagreement and agreement sentences at the onset of the lo pronoun. Visual inspection reveals differences in the responses to the disagreement and the agreement condition relative to animate antecedent and inanimate antecedent sentences. The P600 effect starts around 450 ms and is maintained for more than 300 ms. The P600 effect seems to be distributed along posterior areas between 500 and 700 ms.
300–500 ms time window. Disagreeing pronouns following animate antecedents elicit somewhat more positive ERPs in this latency range, but this effect was the beginning of a P600 effect that attained a larger (significant) amplitude in the 700–900 ms latency range.

500–700 ms time window: P600a. Repeated measures ANOVA in this time window revealed a greater amplitude positive wave elicited by gender disagreement than in the agreement condition (Agreement main effect, $F(1,28) = 10.8, p = .003$). Agreement by Anterior-posterior interaction was significant, $F(2,56) = 3.9, p = .05$, epsilon = .65, which means that this P600 effect was significant at posterior (HSD MD = −1.5 μV, $p = .002$) and central sites (HSD MD = −1.2 μV, $p = .004$). The interaction Animacy × Agreement interaction was not significant, $F(1,28) = 1.2, p = .3$. Nonetheless, the P600 effect seems to be greater (i.e., larger P600 amplitudes for disagreement than agreement condition) for the animate antecedent than the inanimate antecedent condition (Animacy × Agreement × Anterior-posterior interaction, $F(1,28) = 10.8, p = .003$). Agreement by Anterior-posterior interaction was significant, $F(2,56) = 3.9, p = .05$, epsilon = .66, which means that this P600 effect was significant at posterior (HSD MD = −1.5 μV, $p = .002$) and central sites (HSD MD = −1.2 μV, $p = .004$). The interaction Animacy × Agreement interaction was not significant, $F(1,28) = 1.2, p = .3$.

700–900 ms time window: P600b. Statistical analyses showed a significant Animacy × Agreement interaction, $F(1,28) = 5.6, p = .025$, which indicates a greater amplitude positive wave elicited by the gender disagreement than the agreement condition on animate antecedent type sentences (HSD MD = −1.2 μV, $p = .003$). The P600 effect on animate antecedent sentences was mainly distributed over posterior regions (Animacy × Agreement × Anterior-posterior interaction, $F(2,56) = 6, p = .02$, epsilon = .54), and this was largest over posterior (HSD MD = −2.04 μV, $p = .0001$) and central (HSD MD = −1.2 μV, $p = .005$) regions.

### Table 3. Mean RTs and Percentage of Correct Responses (29 Subjects), and the Corresponding SDs for Animacy and Gender Agreement

<table>
<thead>
<tr>
<th></th>
<th>Gender agreement</th>
<th>RT ms</th>
<th>% Correct responses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Animacy Agree</td>
<td>541.8</td>
<td>179.7</td>
<td>95.3</td>
</tr>
<tr>
<td>Animacy Disagree</td>
<td>532.2</td>
<td>175.6</td>
<td>91.5</td>
</tr>
<tr>
<td>Inanimate Agree</td>
<td>554.3</td>
<td>168.3</td>
<td>94.4</td>
</tr>
<tr>
<td>Inanimate Disagree</td>
<td>530.1</td>
<td>173.4</td>
<td>84.2</td>
</tr>
</tbody>
</table>

Notes. RT = reaction time; SD = standard deviation.

Figure 3. ERP grand averages of all conditions across nine regions of interest. Negative is plotted up. Thicker lines represent agree conditions and thinner lines disagree conditions. Dotted gray lines represent inanimate pronouns and solid black lines animate pronouns. Disagreements elicited a P600 effect. Time window analyses are also marked.
Discussion

The aim of this experiment was to investigate the processing of gender agreement of pronouns when the antecedent is an animate or inanimate noun by using an ERP experimental design that avoids word class ambiguity. A P600 of the same size for animate and inanimate antecedents would be expected in agreement with the two-stage model (Garrod & Sanford, 1994). In contrast, if semantics plays a role in the beginning of the pronoun resolution process, different sizes of the P600 effect depending on the animacy of the antecedent would be expected (i.e., syntax and semantics interaction). Our results showed a P600 effect in two consecutive windows showing a greater positivity for gender disagreement than agreement conditions. The first window (500–700 ms) showed that the effects were marginally larger and more posterior for the animate antecedents than for the inanimate antecedents (bear in mind also that no differences were found in the previous window, 300–500 ms). In the second part of the P600 effect (700–900 ms), a Gender agreement × Animacy interaction was statistically significant. Since the P600 effects in the first window were not statistically different between animate and inanimate antecedents, nor were the interactions with animacy, these findings are more consistent with a scenario where syntactic processes guide the antecedent choice on pronoun resolution. This two-stage P600 is what Garnham et al. (1995) found in a study where a person-name or an object-noun was introduced as an antecedent and later referred to by a pronoun in a second clause. They looked at the effects of gender cue and animacy on reading times for the clauses containing the pronouns and times/accuracy in response to simple questions following each sentence. To answer these questions, participants had to resolve the pronoun in the second clause. Reading times showed strong effects of gender cue in both people and object antecedent sentence types and no interaction between gender cue and type of sentence. As in the present study, there was clear evidence for the use of a gender cue when that cue was grammatical in the sentences about objects. However, the response times in the Garnham et al. (1995) study indicated that the effect of cue was considerably stronger for sentences about people, suggesting a later effect of semantics during a second stage, as reflected in the second window of the P600 component (700–900 ms) in the current study.

General Discussion

The aim of the present study was to investigate the role of conceptual/semantic and syntactic information in the process of establishing coreference between a pronoun and its antecedent. The experimental stimuli used in the present study were carefully chosen to try to avoid word class and grammatical case-marking ambiguities in order to create a cleaner design to study the role of syntactic and semantic information during pronoun resolution. Comparing the same pronoun lo under gender agreement and disagreement conditions, with antecedents with semantic versus non-semantic gender, allowed us to elegantly investigate the nature of the information used for linking pronouns to antecedents. No N400 response was found in the present study, as has consistently been observed in some previous studies (Hammer et al., 2005 (first experiment); Lamers et al., 2006, 2008; Schmitt et al., 2002). One might argue that the present study yielded only P600 effects because of the high self-monitoring demands (see Bornkessel-Schlesewsky & Schlesewsky, 2008) because in the experiments subjects had to judge after each sentence whether it was correct or incorrect. Although correct responses only were included in the analyses, sentence-acceptability tasks could have enhanced the P600 effect while eliminating the N400 effect. In Hammer et al.’s study, as well as in Schmitt et al.’s study, subjects answered simple yes/no questions only from time to time within a block of sentences. Similarly, in Lamers et al.’s studies (2006, 2008), subjects had to answer questions addressing the content of the first part of the sentence for only 10 experimental and 10 filler sentences occurring at random positions within a block. However, some experiments have shown the opposite; that is, an N400/P600 pattern has been observed when participants were asked to perform a judgment task (Kolk, Chwilla, van Herten, & Oor, 2003). Furthermore, given that in Kolk et al.’s Experiment 2 there was a higher percentage of correct responses for animate than inanimate antecedents and because detected errors are likely to elicit larger late positivities than those that are overlooked, this accuracy difference would have influenced ERPs when correct and incorrect trials were averaged together. But, in the present study, the averaged trials included only those with correct responses.

Two contrasting hypotheses were evaluated. If syntactic information is initially used during pronoun resolution, then a P600 larger for gender disagreement than agreement and of similar size for animate and inanimate antecedents during the first stage of the pronoun resolution is to be expected (Garnham et al., 1995). In contrast, if semantic information is kicking in early in the process of pronoun resolution, a P600 with greater effects for gender disagreement than for gender agreement but significantly larger for sentences with animate versus inanimate antecedents should be expected (Hammer et al., 2005; Lamers et al., 2006, 2008; Schmitt et al., 2002).

In support of the two-stage bonding model, our results showed two different P600 effects (larger P600 to gender disagreement than agreement) across time. Previous ERP studies (Barber & Carreiras, 2005; Silva-Pereyra & Carreiras, 2007) have analyzed two time windows for the P600 effects, to distinguish between the early and late P600s, as Hagoort, Brown, and Osterhout (1999) proposed. In the present study, for the earlier time window there was no significant interaction between agreement and animacy. In a later window, such syntax and semantics interaction was statistically significant, as has been shown by others (Hammer et al., 2005, 2008). Accordingly, at least regarding our later P600 effect, the processing system seems to use first syntactic and immediately after biological/conceptual gender information in its attempt to establish a coreferential relationship. It is worth noting that amplitude effects in the 700–900 ms time window could be problematic to interpret, as the word following the pronoun has already been presented. However, in terms of time and space, amplitude effects for the following word can be detected if the onset and baseline are moved to the beginning of this next word, because this effect will ride the incoming electrophysiological signal that was elicited by the pronoun. This is why significant interaction effects continue through the two time windows. During the 700–900 ms time window, the P600 gender agreement effect continues to be significant for animate antecedent conditions, mainly over parietal regions, but it does not remain significant for inanimate conditions. Thus, the differences observed between animate and inanimate conditions seem to reflect the differential use of semantic and syntactic information not only in temporal terms but also in terms of topographical distribution.

A two-stage model for pronoun interpretation (Garrod & Sanford, 1994) would accommodate these two different temporal P600 effects. A first immediate recovery stage, which is a more
syntax-related process, is associated with establishing a link between the pronoun and a previous expression. Pronoun resolution is processed first in a bonding stage, which is fast. Later, during a second integration stage or resolution (testing and resolving the link between the pronoun and the antecedent), there is a rapid resolution and decay of the process in the case of inanimates, while the animate information seems to be more long lasting. Two possible explanations will now be considered that cannot be dissociated by the current set of data, but that both point to the early use of syntax and semantics: (1) syntactic and semantic processing run in parallel during the bonding process with animate antecedents, and (2) only syntax is initially considered for the bonding process. However, neither scenario rules out the possibility that the syntactic aspect guides the bonding process, and the use of semantic information is just to integrate the pronoun. The last is in accordance with the Carreiras et al. (1993), Garnham et al. (1995), and Sturt (2003) studies, where there is no evidence for the influence of semantic aspects in choosing the antecedent during bonding processes in pronoun resolution.

Conclusion

This study shows that word class and grammatical case-marking ambiguities make an important contribution to occurrence of the N400/P600 pattern in pronominal reference studies. By removing this source of confound, it was possible to observe the influence of semantics on grammatical gender agreement in pronoun resolution. Results clearly showed a P600 effect, in support of the idea that both formal and semantic information play a role in pronoun resolution. Two possible interpretations are compatible with the present results: Syntactic and semantic information are initially considered in parallel, or there is an initial use of formal (syntactic) information during pronoun comprehension. At a later stage, semantic information clearly shows a stronger impact on pronoun resolution.

References


(Received November 1, 2011; Accepted June 18, 2012)