Examining the evidence for an independent semantic analyzer: An ERP study in Spanish

Clare Stroud a, Colin Phillips a,b,⇑

aDepartment of Linguistics, 1401 Marie Mount Hall, University of Maryland, College Park, MD 20742-7505, United States
bNeuroscience and Cognitive Science Program, 2131 Biology-Psychology Building, University of Maryland, College Park, MD 20742, United States

ARTICLE INFO

Article history:
Accepted 2 February 2011
Available online 5 March 2011

Keywords:
ERPs
Syntax
Semantics
Thematic roles
Spanish
Animacy
Incremental parsing

ABSTRACT

Recent ERP findings challenge the widespread assumption that syntactic and semantic processes are tightly coupled. Syntactically well-formed sentences that are semantically anomalous due to thematic mismatches elicit a P600, the component standardly associated with syntactic anomaly. This ‘thematic P600’ effect has been attributed to detection of semantically plausible thematic relations that conflict with the surface syntactic structure of the sentence, implying a processing architecture with an independent semantic analyzer. A key finding is that the P600 is selectively sensitive to the presence of plausible verb-argument relations, and that otherwise an N400 is elicited (‘The hearty meal was devouring … vs. The dusty tabletop was devouring …’). The current study investigates in Spanish whether the evidence for an independent semantic analyzer is better explained by a standard architecture that rapidly integrates multiple sources of lexical, syntactic, and semantic information. The study manipulated the presence of plausible thematic relations, and varied the choice of auxiliary between passive-biased fue and active-progressive biased estaba. Results show a late positivity that appeared as soon as comprehenders detected an improbable combination of subject animacy, auxiliary bias, or verb voice morphology. This effect appeared at the lexical verb in the fue conditions and at the auxiliary in the estaba conditions. The late positivity elicited by surface thematic anomalies was the same, regardless of the presence of a plausible non-surface interpretation, and no N400 effects were elicited. These findings do not implicate an independent semantic analyzer, and are compatible with standard language processing architectures.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

1.1. ERP evidence for a parallel processing architecture

It has been widely assumed in psycholinguistics and neurolinguistics that there is a tight coupling of semantic and syntactic processes (e.g., Ferreira & Clifton, 1986; Friederici, 2002; Hagoort, 2008; MacDonald, Pearlmuter, & Seidenberg, 1994; Trueswell & Tanenhaus, 1994), and more specifically that semantic interpretations are built up by combining the meanings of words and phrases based upon the syntactic structure of a sentence. Furthermore, in linguistics it is a standard assumption that semantic composition of words and phrases is parasitic on the computation of syntactic structure (Heim & Kratzer, 1998; Pollard & Sag, 1994; Steedman, 2000). In terms of on-line comprehension, these assumptions imply that when an incoming word is encountered, syntactic information is used first to fit the word into the current structure, and then the semantic contribution of that word is added to the representation based on the syntactic role it has been assigned. This view has been challenged by recent electrophysiological findings that have been taken to show that semantic composition can proceed independently of syntactic structure, i.e., the language processor considers interpretations that are not licensed by the structure of the sentence (e.g., Hoeks, Stowe, & Doedens, 2004; Kim & Osterhout, 2005; Kolk, Chwilla, Van Herten, & Oor, 2003; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003). These findings have potentially broad implications for the architecture of language processing, and thus merit close scrutiny.

Recent ERP studies in English (e.g., Kim & Osterhout, 2005; Kuperberg, Caplan, Sitnikova, Eddy, & Holcomb, 2006; Kuperberg, Kreher, Sitnikova, Caplan, & Holcomb, 2007; Kuperberg et al., 2003), Dutch (e.g., Kolk et al., 2003; Van Herten, Chwilla, & Kolk, 2006; Van Herten, Kolk, & Chwilla, 2005), and Chinese (e.g., Ye & Zhou, 2008) have elicited a P600 in response to semantic anomalies in simple and unambiguously grammatical sentences. This finding appears to directly contradict the traditional view of the N400 as an ERP component that is elicited by semantic anomalies.
the elegant argument that can be constructed based on ERP responses to sentences like (1) deserves further attention. First, in an increasingly crowded set of findings on the ‘thematic P600’ there are few studies that present the clear contrast between semantically attractive and non-attractive verb-argument combinations seen in (1), and a number of findings that appear to be inconsistent with Kim and Osterhout’s generalization. Second, it is important to consider how well the findings fit with the widespread evidence in psycholinguistics for a language processor that rapidly integrates information from multiple sources (syntax, semantics, discourse, lexical probabilities) to continuously update its interpretation of an incoming sentence (e.g., Altmann & Steedman, 1988; Levy, 2008; MacDonald et al., 1994; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Accounts of the thematic P600 have paid relatively little attention to the role that sources of information other than the semantic relationship among open class words may play in guiding the interpretation pursued by the processor (but see Bornkessel-Schlesewsky & Schlesewsky, 2008). Many of the existing experimental results are compatible with the more standard architecture in psycholinguistics.

The current study uses evidence from Spanish to evaluate whether the thematic P600 effect motivates the addition of an independent, compositional semantic analyzer to the processing architecture, or whether this body of results is consistent with a single analyzer that integrates multiple sources of information as soon as they become available. Specifically, the study manipulates several syntactic and semantic factors in the sentence to provide a better understanding of the conditions under which a P600 is elicited. The results suggest that the thematic P600 effect in sentences like (1) is not driven by consideration of an interpretation that is incompatible with the surface structure, but rather by features of the surface form of sentences.

1.2. Previous findings on the ‘thematic P600’

There are now many reports of P600s elicited by anomalies that do not appear to be syntactic in nature. Among these, studies that show a contrast between semantic anomalies that elicit a P600 and semantic anomalies that elicit an N400 are particularly relevant for motivating the addition of an independent, compositional semantic analyzer to the processing model, because they potentially shed light on the circumstances under which the different processing streams arrive at inconsistent analyses.

A number of different accounts have been proposed for thematic P600 effects (Bornkessel-Schlesewsky & Schlesewsky, 2008; Hoeks et al., 2004; Kim & Osterhout, 2005; Kuperberg, 2007; van Herten et al., 2005, 2006). These accounts reflect a range of fundamental assumptions about the nature of the language processor and they focus on accounting for somewhat different sets of results. Nevertheless, the different accounts have a number of features in common. Most importantly, all accounts assume a processing architecture in which it is possible to analyze the semantic relations among words independent of the syntactic parse of the sentence. In most cases, this is achieved via multiple partially independent analyzers, with separate analyzers for structural and semantic information. The semantic analyzer operates over open class words, and uses them to construct interpretations with little regard to the structural relations among them. Meanwhile, the syntactic analyzer provides an analysis of the sentence that is compatible with the morphosyntactic features of the sentence. Kuperberg (2007) allows for the possibility that an additional analyzer calculates likely thematic relations based on a limited set of features such as animacy. According to these accounts, a P600 is elicited when there is inconsistency between the analyses pursued by the different analyzers, or when one analyzer successfully reaches an analysis but another does not. The accounts differ in the precise definition of the
different analyzers, and in their view of the specific process(es) that elicit the P600 (see Table 1). An interesting alternative account offered by Hoeks and colleagues (Hoeks et al., 2004) assumes a more standard processing architecture in which syntactic and semantic information is not analyzed by independent processors. However, functional independence is temporarily achieved in situations where semantic information becomes available faster than the syntactic parse of the sentence. Hoeks and colleagues assume that it is in such situations where comprehenders experience fleeting semantic illusions, followed by the realization that the interpretation is at odds with the syntactic form of the sentence. A recent proposal by Bornkessel-Schlesewsky and colleagues pursues a different approach, based on cross-language differences in ERP responses to reversal anomalies, and treats the late positivity as a case of a P300b component elicited by a binary categorization of the well-formedness of the sentence (Bornkessel-Schlesewsky et al., 2011). Nevertheless, since the categorization is assumed to make specific reference to whether the thematic roles of the arguments of a clause are reversed or not, this implicitly assumes that the processor considers a non-surface interpretation in its evaluation of the sentence. Our goal in this study is not to choose among existing accounts, but rather to examine the shared assumption that thematic P600s reflect consideration of interpretations that are not supported by the surface syntax of the sentence.

The argument for an architecture with an independent, compositional semantic analyzer, as opposed to one in which semantic and syntactic analyses are tightly coupled, depends on a series of specific empirical findings and assumptions about ERP components. The first is the claim that the thematic P600 is specifically sensitive to semantic attraction, i.e., to the existence of a plausible interpretation that is not supported by the surface syntax of the sentence. If the occurrence of thematic P600 effects can be accounted for in terms of anomalies in the surface interpretation of sentences, then there is no need to appeal to a mechanism that ignores surface syntax. Second, the argument for architecturally independent analyzers relies on the assumption that the syntactically inappropriate interpretation is constructed at a moment in time when the conflicting syntactic information is already, in principle, available. If the semantic interpretation is constructed at a point in the parse when it remains consistent with the surface syntax of the sentence, then there is no need to appeal to an independent semantic analyzer. Third, the absence of an N400 effect in certain cases of thematic anomalies implies a (temporary) semantic illusion only under the assumption that N400 effects are specifically associated with semantic anomaly detection. Each of these points merits further scrutiny.

Turning to the first argument in more detail, a review of existing literature indicates that there is only limited evidence that the thematic P600 is selectively elicited by sentences that exhibit plausible thematic relations among open class words. Kim and Osterhout directly manipulated the plausibility of the semantic fit between the subject NP and the verb, and found that a P600 was elicited in the condition in which there was a plausible relationship between the subject and the verb (semantic attraction: 1c), whereas an N400 was elicited in the condition in which there was no plausible relationship between the noun and the verb (no-attraction: 1d). (Fig. 3 of Kim and Osterhout’s study suggests that there was a late positivity in the no-attraction condition, of similar amplitude to the N400, but this difference was not found to be reliable in the statistical analyses.) The manipulation of semantic attraction between the verb and subject noun is well suited to testing the semantic selectivity of the thematic P600, and Kim and Osterhout’s finding fits their account very well. However, there have been very few direct tests of the effect of semantic attraction, despite the importance of this effect for motivating the existence of an independent semantic analyzer.

However, findings from other studies offer at best limited support for Kim and Osterhout’s generalization. It is certainly the case that a number of studies have used materials that display semantic attraction between verbs and arguments, and have observed P600 effects (Hoeks et al., 2004; Kolk et al., 2003; Kuperberg et al., 2003; Ye & Zhou, 2008). However, a number of studies have observed similar P600 effects in sentences that contain surface thematic anomalies, but lack semantic attraction. For example, Kuperberg and colleagues observed P600 effects in sentences with and without a plausible thematic relationship between the inanimate subject noun and the verb, e.g., Every morning at breakfast the eggs would {eat/plant} (Kuperberg et al., 2007). Similarly, Kolk and colleagues observed a P600 in response to the Dutch counterparts of sentences like The trees that played in the park … (Kolk et al., 2003). In addition to Kim and Osterhout’s study, one other study has successfully manipulated the presence of a P600 through different types of thematic anomalies. Van Herten and colleagues (2006, Experiment 2) varied the strength of semantic association between the verb and the noun in object position of anomalous Dutch sentences. Anomalous sentences in which the object and the verb were highly associated (e.g., the elephant pruned the tree) elicited a robust P600, but anomalous sentences in which the object and the verb were not highly associated (e.g., the elephant carressed the tree) elicited a robust N400 and a greatly attenuated P600 that was small at left hemisphere electrodes and absent at right hemisphere sites. Van Herten and colleagues propose that it is lexical-semantic association between words, rather than the plausibility of the subject NP as the theme of the verb, that is responsible for the effects seen in Kim and Osterhout’s study. For example, in the examples in (1) meal is more strongly associated with dinner than is tabletop.

In addition to the instances of P600 effects elicited in the absence of semantic attraction, there are a number of recent reports of P600s that are not straightforwardly described in terms of thematic relations. For example, a P600 has also been elicited by orthographic errors in high-cloze situations but not in low-cloze situations (Vissers, Chwilla, & Kolk, 2006), and when the meaning of the sentence fails to match an accompanying picture (Vissers, Kolk, Van de Meerenendonk, & Chwilla, 2008). Second, other studies have elicited both an N400 and a P600 to words that are highly

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary of current accounts of the thematic P600.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Processing analyzers</th>
<th>The P600 reflects…</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoeks et al. (2004)</td>
<td>Integrated processor; some information types may become available earlier</td>
</tr>
<tr>
<td>Kim and Osterhout (2005)</td>
<td>Syntactic and semantic</td>
</tr>
<tr>
<td>van Herten et al. (2005, 2006)</td>
<td>Algorithmic and heuristic</td>
</tr>
<tr>
<td>Kuperberg (2007)</td>
<td>Combinatorial (thematic and morphosyntactic) and semantic-association</td>
</tr>
<tr>
<td>Bornkessel-Schlesewsky and Schlesewsky (2008)</td>
<td>Linking and plausibility</td>
</tr>
</tbody>
</table>

Effortful syntactic processing required to obtain semantic coherence
Detection of a syntactic violation
Executive control indicating detection of an inconsistency and the need to review the representation
Continued processing following detection of a conflict within the combinatorial stream(s) or between any of the streams
Problem with linguistic mapping of the two streams
incongruous with the context, relative to words that are only slightly incongruous, which elicited only an N400 (Geyer, Holcomb, Kuperberg, & Pearlmutter, 2006; Van de Meerdonk, Kolk, & Chwilla, 2008). Relatively, it is important to note that many conventional ERP studies of semantic anomaly have elicited an N400 followed by a late positivity, although the later component has received much less attention than the N400 (e.g., Federmeier, Wlotko, de Ochoa-Dewald, & Kutas, 2007; Matsumoto, Idaka, Haneda, Okada, & Sadato, 2005; Van Petten & Luka, 2006).

In sum, Kim and Osterhout’s finding that the P600 is selectively elicited by sentences involving semantic attraction constitutes a key argument in favor of an independent semantic analyzer, but a broader set of findings challenges this generalization about the thematic P600. One of the goals of the current study was to re-examine the selectivity of the thematic P600.

The second key step in the argument for an independent semantic analyzer is an assumption about timing. If a syntactically inappropriate interpretation is constructed at a point when it is already clear that the interpretation is incompatible with the evidence at the point, then this lends support to the claim of an independent semantic analyzer. But if the inappropriate interpretation is constructed before the contradicting syntactic information is available, then there is no need to invoke an independent semantic analyzer.

Many discussions of the thematic P600 have focused on the syntactic and semantic relations that obtain at the point when all of the critical open class words have been processed (e.g., hearty meal ... devour), but have paid less attention to the sequence of interpretive steps that leads up to that point (for related observations see Bornkessel-Schlesewsky and Schlesewsky (2008)). As the sentence unfolds the processor may rapidly draw on multiple sources of information to build expectations about the interpretation of the sentence. In so doing, the processor may construct plausible interpretive hypotheses that subsequently turn out to conflict with the ultimate syntactic form of the sentence. Such conflicts could arise within a single analyzer that rapidly integrates multiple sources of information, rather than processing different information sources in parallel. For example, in the case of sentences like The hearty meal was devouring ... the inanimate subject, the auxiliary, and the verb stem devour- are all most consistent with a passive construction. In a corpus search of the written text materials in the BYU-BNC database we found that when the auxiliaries was and had been are followed by a verb (these were the auxiliaries used in Kim and Osterhout’s materials), the verb is passive in around 82% of cases. This bias is likely reinforced by the use of an inanimate subject NP, and possibly also by a verb stem that is strongly associated with that inanimate NP (e.g., devour). If comprehenders integrate the verb stem into their ongoing interpretation before they recognize the import of the progressive –ing morpheme, then a conflict may arise due to a mismatch between an initial passive interpretation that must later be retracted. This conflict could account for the appearance of a P600, without appeal to an independent semantic analyzer. This account of the thematic P600 effect is similar to the one offered by Hoeks et al. (2004).

If it is the case that the passive interpretation is constructed prior to the verb, then this predicts that a similar conflict should arise in sentences that lack semantic attraction, e.g., The dusty tabletop was devouring ... a prediction that appears to be at odds with Kim and Osterhout’s contrasting ERP findings. If the passive interpretation is constructed only when the verb stem (but not the suffix) is integrated into the sentence, then it could still be possible to capture Kim and Osterhout’s contrast. However, an alternative possibility suggested by the current findings and other results (Stroud, 2008) is that an account that predicts similar P600 effects in sentences with and without semantic attraction can better account for the available evidence.

The third step in the argument for an independent semantic analyzer involves the interpretation of the presence or absence of N400 effects in sentences with thematic anomalies. Based on the assumption that N400 effects reflect semantic anomaly detection, the absence of N400 effects found in some studies of thematic anomalies (e.g., Hoeks et al., 2004; Kim & Osterhout, 2005; Kuperberg et al., 2007; van Herten et al., 2005; Ye & Zhou, 2008) has sometimes been taken as evidence that participants initially failed to notice the semantic anomaly. This interpretation of the findings lends further support to the existence of an independent semantic analyzer. However, this argument is undermined by an ongoing debate about the functional status of the N400 effect. Under the integration view of the N400 the amplitude of the N400 reflects the ease or difficulty of the semantic integration of an incoming word into the previous context (Baggio, Choma, van Lambalgen, & Hagoort, 2010; Brown & Hagoort, 1993; Hagoort, 2008; Kutas & Hillyard, 1980; Osterhout & Holcomb, 1992; Van Berkum, Hagoort, & Brown, 1999). In contrast, under the lexical view the amplitude of the N400 reflects the ease or difficulty with which the incoming word can be accessed from memory (Deacon, Hewitt, Yang, & Nagata, 2000; Federmeier, 2007; Fischer, Bloom, Childers, Roucos, & Perry, 1983; Kutas & Federmeier, 2000; Lau, Phillips, & Poeppel, 2008; van Berkum, 2009). According to this view the N400 does not directly reflect combinatorial semantic processes, although factors such as semantic association and predictability based on the context influence the ease with which an incoming word is accessed. Thus, the lack of an N400 in the hearty meal was devouring ... does not indicate that semantic integration has proceeded smoothly, but rather reflects strong lexical association between the noun and the verb or the predictability of the semantic features of the verb when it follows the subject noun. The N400 contrast between the conditions in Kim and Osterhout’s study may reflect differences in lexical association, e.g., meal and devour are more strongly associated than tabletop and devour (cf. Bornkessel-Schlesewsky & Schlesewsky, 2008). In support of this interpretation, all studies of thematic anomalies that created the anomalous conditions by reversing the arguments of the control condition, thereby holding lexical associations constant, found a P600 effect and no N400 effect (Hoeks et al., 2004; Kolk et al., 2003; van Herten et al., 2005, 2006; Ye & Zhou, 2008). Meanwhile, studies of thematic anomalies that have elicited biphasic N400 + P600 effects, or N400 effects only, have typically compared sentences that were not lexically matched. The N400 effects for an independent semantic analyzer rely upon a number of generalizations and assumptions that are open to question. The current study was designed to further examine the components of this argument.

1.2.1. The current experiment

The aim of the current study was to investigate in greater detail the circumstances in which the P600 component is elicited by syntactically well-formed sentences with thematic anomalies. The ultimate aim of this research is to clarify whether findings of thematic P600 effects motivate the existence of an independent, compositional semantic analyzer that ignores syntactic cues, or whether these findings are also compatible with a single analyzer that rapidly integrates diverse sources of information (syntactic, semantic, contextual) as soon as they become available. The strongest evidence for an independent, compositional semantic analyzer would consist of experimental results showing that (a) the language processor identifies plausible thematic relations despite contradictory syntactic information that is available prior to the semantic composition process, and that (b) the thematic P600 effect is specifically sensitive to the plausibility of thematic relations rather than to lower-level lexical-semantic association effects. In this study we present thematic anomalies in environments that
attempt to provide both of these circumstances, and thus test the extent to which the thematic P600 reflects syntax-independent semantic attraction between the words of a sentence.

We created these two circumstances by using an experimental design in Spanish that added two key features to the manipulation of thematic fit between a subject noun and a verb used in previous studies (e.g., Kim & Osterhout, 2005, Experiment 2). First, we varied the choice of auxiliary verb in order to manipulate the syntactic bias of the sentence towards an active or passive continuation of the sentence. This allowed us to establish stronger syntactic biases before readers encountered potentially contradictory thematic information. Second, we introduced a context sentence prior to each target sentence, which established all critical nouns as discourse referents in advance of the target sentence, thereby reducing the possible confound between lexical predictability and thematic plausibility found in previous studies and allowing us to more directly assess the contribution of thematic plausibility to the P600.

In English, the auxiliary be is used in both passive and active progressive constructions. As already discussed, this leads to uncertainty in the interpretation of P600 responses elicited by semantically anomalous progressive constructions such as *The hearty meal was devouring*. ... Thus, if comprehenders show sensitivity to the possibility that the subject NP is a good theme for the verb *devour* this could reflect either an independent semantic analysis that ignores the surface syntax, or a strongly incremental and non-independent analysis that pursues the possibility of a passive parse until it recognizes that the progressive –ing suffix blocks this analysis. In Spanish, different auxiliaries typically precede verbs in passive and active progressive constructions, and hence the choice of auxiliary can be used to manipulate syntactic expectations in advance of the thematic information conveyed by the main verb. The auxiliary *fue* (past tense of *ser*) provides evidence that an active progressive verb form is unlikely to follow, and the auxiliary *estaba* (past tense of *estar*) provides evidence that a passive verb form is unlikely to follow.

In a design similar to Kim and Osterhout’s (2005) Experiment 2, felicitous control sentences were compared with semantically anomalous sentences with semantically attractive and non-attractive thematic relations. In the semantic-attraction condition the subject NP was a plausible theme for the verb, whereas in the no-attraction condition the subject NP was not a plausible theme for the verb. The three conditions (control, semantic-attraction and no-attraction) were tested with each of the two auxiliaries. The auxiliaries were used in 3rd person singular past forms, *fue* and *estaba* respectively. *Fue* is typically used in passive constructions, and *estaba* typically is used in active progressive constructions. Both are also used in adjectival constructions. It is important for the current design that these are probabilistic biases: *estaba* may appear in a passive construction, and *fue* may be followed by a verb in progressive form, due to the fact that it is lexically ambiguous with the past tense of the verb *ir* (‘go’). The construction consisting of *fue* + progressive participle is used to convey the equivalent of the English aspectual ‘did something little by little’ meaning, e.g., *fue comiendo la sopa* ‘ate the soup [little by little]’. It is also used when the progressive verb characterizes the way in which someone or something went somewhere: this may be translated as ‘went around doing something’, as illustrated below, or for example, *fue nadando hacia la costa* ‘he went swimming toward the coast.’ In addition, both auxiliaries can also be followed by an adjective. In the terms of Carlson (1977) the auxiliary *ser* is used with individual-level predicates that denote more permanent properties, such as *tall* or *intelligent*, and the auxiliary *estar* is used with stage-level predicates that denote more transient properties, such as *sick* or *hungry*. An example set of materials is shown in Table 2.

A context sentence was presented before each target sentence, in order to increase the naturalness of the passive construction and to reduce the contribution of lexical accessibility and lexical expectations to observed ERP effects. Passive constructions in Spanish have a more limited distribution than their English counterparts, and in particular they are more natural when the passive subject NP is an existing discourse referent. Since three different subject NPs were used across the 6 target conditions in each item set, the target sentences within each item set were preceded by an identical context sentence that introduced all three potential subject NPs. In order to construct a natural context sentence, all three NPs were generally chosen from the same semantic field. Following the context sentence, it was still the case that the thematic fit between the verb and the subject NP was stronger in the control and semantic-attraction conditions than in the no-attraction conditions, but the global semantic association between the verb and the three subject NPs was similar across conditions, because the subject NPs were drawn from the same semantic field and all corresponded to existing discourse referents. This contrasts with the materials in Kim and Osterhout’s study, in which the no-attraction conditions were formed by combining the subject NP from the semantic-attraction condition in one item set with a verb and sentence ending from another item set, resulting in a very low level of semantic association between the subject NP and the verb in the non-attractive condition.

1.2.1.1. Predictions. We can articulate a number of different predictions for the ERP results, based on different possible processing architectures and different functional interpretations of ERP components.

We can first outline predictions for the P600. In a processing architecture that incorporates an independent semantic analyzer, and that elicits a P600 specifically when an independently generated semantic interpretation conflicts with the surface syntax, we predict that the P600 should be affected by our manipulation of the subject-verb relation, but should not be affected by our manipulation of the auxiliary. The P600 should be elicited in the semantic-attraction condition, due to the presence of a plausible noun–verb pairing, but should not be elicited in the no-attraction condition, as there is no motivation to interpret the subject noun as the theme of the verb. If the independent semantic analyzer attends to plausible relations between open class words while ignoring syntactic information, then similar results should obtain in the conditions with *fue* and in the conditions with *estaba*. In other words, we should replicate Kim and Osterhout’s findings, irrespective of the choice of auxiliary.

In contrast, if the thematic P600 reflects the detection of a surface anomaly, with no need to appeal to an independent semantic analyzer, then we should expect to observe a P600 in the semantic attraction and no-attraction conditions alike, contrary to what Kim and Osterhout found. In this scenario the P600 might be associated with lexical associations (van Herten et al., 2006) or selectional violations involving animacy (Kuperberg, 2007). Furthermore, if these effects arise from an architecture in which a single analyzer rapidly integrates multiple sources of information as soon as they are available, then the auxiliary manipulation may impact the results, and we might even expect to see effects at different time points, as a consequence of when the processor detects anomalies. For example, if the auxiliary ‘estaba’ biases comprehenders to expect an active voice clause, which typically has an agentive subject, and if the sentence has an inanimate subject NP, which is generally a poor agent, then this may engender processing difficulty due to this conflict already at the auxiliary.

Similarly, the predictions for the occurrence of N400 effects depend both on the processing architecture that is assumed and on the functional interpretation of the N400. If the presence or
absence of N400 effects in ERP studies of thematic anomalies reflects the occurrence of fleeting semantic illusions arising from an independent semantic analyzer, then we should expect the N400 effects to be absent in the semantic-attraction conditions, but present in the no-attraction conditions, again replicating Kim and Osterhout’s findings. On the other hand, if the N400 effects in such studies reflect lexical associations between the target words and preceding words in the sentence, then a different pattern of results is expected. N400 effects should vary as a function of the association between the target verb and the subject noun and possibly other material in the target items. Due to the design of our study, in which all potential subject nouns were introduced in a lead-in sentence, and where all potential subject nouns in an item set were chosen from the same semantic field, it is possible that N400 effects would be either small or absent. A previous study by Nieuwland and van Berkum may be very relevant to this prediction. These authors embedded sentences with strong semantic anomalies in discourse contexts that made the anomalous words highly salient in the discourse. No N400 effect was elicited, but a P600 effect was observed. The authors interpret the lack of N400 effect as evidence of failure to notice the semantic anomaly, but it may reflect the high discourse accessibility of the anomalous lexical item (Nieuwland & van Berkum, 2005).

1.2.1.2. Control sub-experiments. Because relatively few ERP studies have been conducted in Spanish, and particularly in the Latin American Spanish used in our study, the experiment also included two control sub-experiments that were used to confirm that syntactic and semantic anomalies elicit canonical N400 and P600 effects in our participants. These sub-experiments are potentially valuable in the interpretation of novel or unexpected ERP effects observed in the primary conditions. The semantic sub-experiment manipulated the felicity of the semantic relation between a noun and a following adjective, and the syntactic sub-experiment manipulated the correctness of number agreement between a deteminier and a noun. In order to control for possible effects of number morphology (Wagers, Lau, & Phillips, 2009) half of the syntactically anomalous sentences involved a singular determiner/plural noun mismatch and the other half involved a plural determiner/singular noun mismatch. Likewise, half of the corresponding grammatically correct versions contained a plural determiner and noun, and the other half contained a singular determiner and noun. The items of the control sub-experiments were intermixed with items from the primary study. All items in the control sub-experiments were preceded by a context sentence, so that they would be indistinguishable from the other sentences in the study. A sample set of items for the control sub-experiments is shown in Table 2.

The effect of semantic felicity between a noun and its adjective has been previously investigated in several Spanish ERP studies. Martín-Loeches and colleagues compared sentences containing felicitous noun–adjective combinations such as el sentimiento profundo (‘the profound sentiment’) with infelicitous noun–adjective combinations such as el sentimiento peludo (‘the furry sentiment’) and observed a clear N400 (and also, unexpectedly, a small P600) (Martín-Loeches, Nighur, Casado, Holhfield, & Sommer, 2006). In a series of auditorily and visually presented mini-stories Wicha and colleagues compared Spanish counterparts of Little Red Riding Hood was carrying food to her grandmother in a very pretty basket/crown, which also elicited an N400 at the infelicitous word (Wicha, 2003; Wicha, Moreno, & Kutus, 2003). These experiments indicate that semantic anomalies elicit an N400 in Spanish in a similar fashion to other languages. The syntactic control sub-experiment tested the effects of number (dis)agreement between an article and a noun. Previous studies

| Table 2 |

| Sample materials set. |

| Primary conditions |

| Context sentence | Fue – passive prediction | Grammatical control | El aviso fue declarado por el bombero ante todos los vecinos. |
| | | Semantic attraction | The warning was declared by the fireman in front of all the neighbors. |
| | | No-attraction | El piso fue declarando con un megáfono para que todo el mundo lo oyera. |

| Estaba – progressive prediction | Grammatical control | El bombero estaba declarando el aviso cuando el fuego se formó otra vez. |
| | Semantic attraction | The fireman was declaring the warning when the fire rose up again. |
| | No-attraction | El piso estaba declarando con un megáfono para que todo el mundo lo oyera. |

| Semantic control sub-experiment |

| Context sentence | Semantic Felicity | +Felicitous | El chofier estaba limpiando el carro lujoso, cuando Juan chocó con una moto contra la puerta del garage. |
| | | –Felicitous | El chofier estaba limpiando el carro enfadado, cuando Juan chocó con una moto contra la puerta del garage. |

| Syntactic control sub-experiment |

| Context sentence | Syntactic agreement | +Agree | El ladrón estaba desconectando los hilos de la alarma antirrobo cuando llegó la policía. |
| | | –Agree | El ladrón estaba desconectando el hilo de la alarma antirrobo cuando llegó la policía. |
of grammatical agreement in Spanish have yielded comparable results to similar manipulations in English. Previous studies that manipulated either number or gender agreement between nouns and following adjectives have observed a P600 response and in most instances also a LAN (Barber & Carreiras, 2005; Martín-Löeche et al., 2006; Wicha, Moreno, & Kutas, 2004). In studies using a slightly different paradigm Barber and Carreiras (2003, 2005, Experiment 1) presented word pairs (article plus noun or noun plus adjective) that disagreed in number, and elicited an N400. This finding is surprising in comparison to previous studies of agreement mismatches in English and Spanish, but plausibly reflects differences between the processing of isolated word pairs and words in sentence contexts.

2. Methods

2.1. Participants

There were 31 participants in the ERP study. Data from one participant was excluded due to technical problems; five participants were excluded due to low rates of accuracy in the behavioral task (<75%), and one participant was excluded due to high levels of artifacts in the EEG recordings. All 24 remaining participants (14 females; mean age 27; range 18–41 years) were healthy, native speakers of Latin American Spanish (from Perú (5), Argentina (4), Chile (4), Colombia (4), Puerto Rico (2), Guatemala (2), Venezuela (1), Ecuador (1), Costa Rica (1)). All had been in the United States for 3 years or less (mean: 1.5 years; range: 1 month to 3 years) except for one participant who had been in the country for 5 years but spoke Spanish almost exclusively in everyday life. All were right-handed and had normal or corrected-to-normal vision. All participants gave informed consent and were paid $15/h for their participation, which lasted approximately 4 h, including set-up time. The duration of this study was unusually long, due to the large number of experimental conditions and the need for context sentences to make the target sentences felicitous.

2.2. Materials

The primary materials consisted of six sentences organized in a 2 x 3 factorial design, illustrated in Table 2. There were two levels of the auxiliary factor (fue and estaba) and three levels of the relatedness factor (control, semantic-attraction and no-attraction). The semantic control sub-experiment consisted of pairs of sentences containing felicitous and infelicitous noun–adjective combinations and the syntactic control sub-experiment consisted of pairs of sentences containing grammatical and ungrammatical article–noun agreement combinations. Examples of sentences from the semantic and syntactic control sub-experiments are shown in Table 2.

The target sentence in all six primary conditions began with a subject NP (determiner + noun), followed by an auxiliary (either fue or estaba), followed by a verb with either passive or progressive verbal morphology. Three conditions used the auxiliary fue. Fue is commonly used in passive constructions and is also used, although much less frequently, in a construction with a verb with progressive morphology. In the former case fue is a form of the auxiliary ser ‘be’, while in the less frequent latter case fue is a form of the verb ir ‘go’, which is lexically ambiguous with ser in the past tense. Three conditions used the auxiliary estaba, which is primarily used in progressive constructions in Spanish. In the grammatical control condition with the auxiliary fue the sentence-initial NP was followed by a verb with passive morphology. In the grammatical control condition with the auxiliary estaba the sentence-initial NP was followed by a verb with progressive morphology. The anomalous semantic-attraction and no-attraction conditions contained verbs with progressive morphology, and differed in the goodness-of-fit of the inanimate subject NP as a theme for the verb.

In all conditions the target sentence was preceded by a context sentence, which was identical for all six conditions in each item set, and mentioned all three possible subject NPs of the target sentence. The three subject NPs of the target sentences were distributed as follows: an inanimate NP that was a good theme for the verb was used in the grammatical control condition with fue and in the semantic-attraction condition with both fue and estaba; an inanimate NP that was a poor theme for the verb was used in the no-attraction conditions with both fue and estaba; and an animate NP was used in the grammatical control condition with estaba. Although five of the six target sentences in each item set turned out to be active progressive constructions, the context sentences were designed to be compatible with both passive and active progressive continuations.

An important feature of the experimental design is that the auxiliary forms fue and estaba are biased to either active or passive voice constructions, but both are also compatible with the other voice. The form fue is biased toward the passive voice, but due to its lexical ambiguity can also be used in a progressive active construction (e.g., fue buscando which translates roughly as ‘went around looking for’ or fue comiendo ‘ate something little by little’). A search of the BYU Corpus del Español suggests that around three-quarters of fue + verb combinations are passive constructions. (This is similar to the passive bias of English was and had been.) Meanwhile, the auxiliary estaba is biased toward the progressive active construction, but can also be used in a passive construction (e.g., estaba cerrado por ‘was in the state of being closed by’). It is more difficult to quantify the relative frequencies for estaba, due to the ambiguity between passive participles and adjectives in Spanish, but under any classification of participles/adjectives it remains the case that most instances of estaba + verb are active progressive forms. If the anomalous conditions were impossible rather than improbable, then those sentences would contain straightforward syntactic violations, which could undermine the investigation of the factors that modulate the thematic P600. We address below the question of whether participants might nevertheless have perceived certain anomalies as simple syntactic violations.

In order to increase naturalness, the grammatical control conditions with fue and estaba used the voice that is most commonly associated with each auxiliary form, with the result that the two grammatical control conditions differed in voice. The grammatical control condition with fue was a passive construction, the Spanish counterpart of sentences like the hearty meal was devoured . . . , and the grammatical control condition with estaba was an active progressive construction, the Spanish counterpart of the hungry boy was devouring . . . These different constructions were chosen for the control conditions in order to ensure that each auxiliary was used in its preferred construction. Previous research suggests that using the two grammatical control conditions was appropriate. Kim and Osterhout (2005, Experiment 1) compared the anomalous the hearty meal was devouring to both an active control (the hungry boy was devouring . . . ) and a passive control (the hearty meal was devoured . . . ), and found that the choice of control did not impact the results.

For each set of six sentences, a verb and three subject NPs were chosen. The verb was chosen to be felicitous in both passive and progressive constructions, as judged by native Spanish speakers who assisted in the development of the materials. The verbs chosen for this experiment tended to impose less rigid selectional restrictions than those used in many previous studies of the thematic P600 (e.g., Kim & Osterhout, 2005; Kolik et al., 2003; Kuperberg et al., 2003, 2006, 2007; van Herten et al., 2005, 2006), due to the fact that few concrete verbs are felicitous in the passive voice in Spanish.
The most felicitous verbs are typically those from business or newspaper language, such as *investigar* ‘investigate’, *firmar* ‘sign’, *entregar* ‘deliver’. A different verb was used in each item set, with the exception of 12 verbs that were repeated once each in order to increase the felicity of certain items. Three NPs were chosen as possible subjects for each verb: an animate noun that was a good agent for the verb, an inanimate noun that was a good theme for the verb, and another inanimate noun that was a poor theme for the verb. In order to ensure that the initial determiner in the target sentence did not provide a clue to which of the three previously-mentioned nouns was the subject of the target sentence, all three nouns in an item set shared the same grammatical gender. In order to guard against possible confounds in the ERP responses due to lexical differences based on factors such as word length or frequency almost all of the inanimate nouns (96%) were used in one item set as the semantically attractive noun and in another item set as the non-attractive noun. This was not the case for 4% of the item sets, in which the inanimate nouns were replaced in order to increase the felicity of the target sentences.

In order to guard against the possibility that participants might consistently assign the same interpretation to the anomalous sentences, the material that followed the critical verb in the anomalous conditions of each item set was varied. This was important, because one account of previous findings about the thematic P600 claims that a P600 provides evidence that the processor entertained a passive interpretation of the sentence, which led to the perception that the progressive form was a syntactic error, and further claims that an N400 provides evidence that the processor considered an active analysis, which led to the perception of a semantic error, since the inanimate subject NPs was a poor agent for the verb (Kim & Osterhout, 2005). The material following the verb in the anomalous conditions might influence the way that participants perceived the violation, and if this perception was reinforced across items, it could potentially change the ERP response that was evoked by the target verb. For example, if participants always saw a by-phrase after the anomalous inanimate NP + progressive form, it might lead them to favor a passive interpretation for the sentence as a whole, and therefore perceive the violation as a syntactic error, due to verbal morphology that mismatches the preferred interpretation. Therefore, the completions in the anomalous conditions were distributed equally among three options: a by-phrase (e.g., ‘by the fireman’), an inanimate direct object NP (e.g., ‘the apartment’), or a preposition or other adverb (e.g., ‘with a megaphone’). In many item sets the word that followed the critical verb was short (i.e., prepositions and definite articles), with some slightly longer adverbs.

Materials for the two control sub-experiments began in a very similar form to the target experimental items. A context sentence did not provide a clue to which of the three previously-mentioned nouns was the subject of the target sentence, all three nouns in an item set shared the same grammatical gender. In order to guard against possible confounds in the ERP responses due to lexical differences based on factors such as word length or frequency almost all of the inanimate nouns (96%) were used in one item set as the semantically attractive noun and in another item set as the non-attractive noun. This was not the case for 4% of the item sets, in which the inanimate nouns were replaced in order to increase the felicity of the target sentences.

Participants were comfortably seated in a dimly lit testing room about 100 cm in front of a computer monitor. Each two-sentence pair (context sentence plus target sentence) was preceded by a fixation cross. All items appeared in black font on a white screen. Participants pressed a button to initiate presentation of the trial. The context sentence was presented in two self-paced sections in 25 pt. font; after reading the first half of the sentence, participants pressed a button to see the second half, which began 180 ms after the button press. After reading the second half of the context sentence, participants pressed a button and the target sentence began 180 ms later. Target sentences were presented one word at a time in 30 pt. font. Each word appeared in the center of the screen for 300 ms, followed by 200 ms of blank screen. The final word of each sentence was marked with a period, and 1000 ms later a question mark prompt appeared on the screen. Participants were instructed to read the sentences carefully without blinking and to indicate with a button press whether the sentence was an acceptable sentence of Spanish. Feedback was provided for incorrect responses. Each experimental session was preceded by a three trial practice session that included both acceptable and unacceptable sentences. Participants received feedback and were invited to ask clarification questions about the task. The experimental session was divided into 7–8 blocks lasting 15 min each, but participants could request additional breaks at their discretion.

**2.4. EEG recording**

EEG was recorded from 28 Ag/AgCl electrodes, mounted in an electrode cap (Electrocap International): midline: Fz, FCz, Cz, CPz, Pz, Oz; lateral: F3/4, F7/8, FC3/4, FT7/8, C3/4, T7/8, CP3/4, TP7/8, P4/5, P7/8, O1/2. Recordings were referenced to the left mastoid. Additional electrodes were placed on the left and right outer canthus, and above and below the left eye to monitor eye movements. The EEG and EOG recordings were amplified by a SynAmps™ Model 5083 EEG amplifier, and sampled at 1 kHz using an analog band-pass filter of 0.1–70 Hz. Impedances were kept below 5 kΩ.

**2.5. EEG analysis**

All comparisons were based upon single word epochs, consisting of the 100 ms preceding and the 1000 ms following the start of the presentation of the critical words. Epochs with ocular and
other large artifacts were rejected from analysis based on visual screening prior to any further analyses. In addition, epochs from trials in which the participant responded inaccurately were excluded. Five participants’ data were excluded because of accuracy below 75%. Epochs that showed a divergence of more than 75 µV from zero were excluded automatically. Data from one participant was excluded because only 22% of critical epochs remained after all these measures were taken. For the remaining 24 participants, after all the exclusion criteria were applied, 65% of the trials were included in the final analysis. This rate is relatively low compared to a typical inclusion rate of about 80% in ERP sentence processing studies. There are two likely reasons for this. First, most ERP studies do not exclude trials to which the participants respond inaccurately on the behavioral task. Second, the fact that participants read context sentences prior to each target sentence lengthened the study and likely resulted in a higher rate of ocular artifacts during the target sentences. Once these features of the current study are taken into consideration, the data inclusion rates are consistent with those from other ERP studies of sentence processing. The waveforms were normalized using a 100 ms pre-stimulus baseline. Averaged waveforms were filtered offline using a 10 Hz low-pass filter for presentation purposes, but all statistical analyses are based on unfiltered data. The following latency intervals were chosen for analysis, based on the intervals used in previous literature and on visual inspection: 300–500 ms (N400), 600–1000 ms (P600). The 0–200 ms interval was also analyzed to test for possible early differences.

For statistical analyses, six regions of interest (ROIs) were used in the ANOVAs, consisting of groups of three electrodes at each ROI: left anterior (F3, FC3, C3), anterior midline (FZ, FCZ, CZ), right anterior (F4, FC4, C4), left posterior (CP3, P3, O1), posterior midline (CPZ, PZ, O2), right posterior (CP4, P4, O2). These ROIs were organized into the two topographic factors laterality (left, midline, right) and posteriority (anterior, posterior). ANOVAs were performed separately for the conditions with fue and for those with estaba, due to the fact that the length difference between the two auxiliaries could introduce irrelevant differences into the waveforms. This separation of conditions in the analyses is consistent with the experimental hypotheses, which focused on the relation between the semantic-attraction and no-attraction conditions and the control condition within each level of the auxiliary factor. An omnibus ANOVA was also conducted that included both auxiliaries; these results are included below but were not the focus of the analyses. ANOVAs were performed hierarchically using the within-subjects factor condition (control, semantic-attraction, no-attraction). All p-values reported below reflect the application of the Greenhouse–Geisser correction where appropriate to control for violations of the sphericity assumption (Greenhouse & Geisser, 1959), together with the original degrees of freedom. Due to the large number of possible interactions in this design we discuss only those interactions for which follow-up analyses yielded significant contrasts within the levels of the interacting factors.

3. Results

3.1. Accuracy

Overall accuracy on the behavioral acceptability judgment task for the six primary conditions was 84%. The accuracy on the two acceptable control conditions was higher (fue-control: 92%; estaba-control: 90%) than for the four anomalous conditions (fue + semantic-attraction: 75%; fue + no-attraction: 84%; estaba + semantic-attraction: 82%; estaba + no-attraction: 84%). This difference was likely due to the fact that the anomaly resulted from the addition of a single character in the verb suffix (e.g., passive -ado became progressive -ando) and therefore may have been missed on some trials, especially in trials when the verb was long and the suffix far from the visual fixation point. As noted above, only trials with correct responses were included in the analysis. The semantic control sub-experiment showed somewhat lower overall accuracy (77%), with average accuracy of 84% in the felicitous condition and only 70% in the infelicitous condition, suggesting that some noun–adjective combinations may have been perceived as more felicitous than other anomalous items in the study. The syntactic control experiment showed an overall accuracy of 87% (grammatical: 89%; ungrammatical: 85%).

3.2. Control sub-experiments

The results for the semantic and syntactic control experiments conformed to expectations. There was an N400 effect at the adjective in the infelicitous noun–adjective combinations of the syntactic control sub-experiment, and a P600 at the nouns with mismatching number agreement in the syntactic control sub-experiment.

3.2.1. Semantic control sub-experiment

In the 0–200 ms interval following the critical adjective in the semantic control sub-experiment there were no differences between conditions. In the N400 interval (300–500 ms) voltages across the entire scalp were more negative in the semantic anomaly condition. This negativity led to a main effect of condition in the overall ANOVA, in addition to significant effects in all ROIs except for the right anterior region, where the difference was only marginally significant. Fig. 1 shows an electrode array and F-values are shown in Table 3. Visual inspection suggested a tendency for a broadly distributed positivity following the N400, similar to the late positivity that has been observed in a number of previous studies of semantic anomaly (e.g., Federmeier et al., 2007; Matsumoto et al., 2005; Van Petten & Luka, 2006), but analyses at the 600–1000 ms interval indicated that this difference was not reliable. Although there was a significant 3-way interaction involving the condition factor at this interval, the difference between the conditions was not reliable at any individual region of interest.

3.2.2. Syntactic control sub-experiment

Visual inspection suggested that the ERP waveforms diverged already in the 0–200 ms interval, particularly at central and posterior electrodes. This was confirmed by the ANOVA, which indicated a significant effect of condition (F(1, 23) = 5.24, p < .05) and a marginally significant interaction of condition and posteriority (F(1, 23) = 3.52, p < .1), due to a positivity in the ungrammatical condition. This difference was marginally significant at the anterior midline and anterior right ROIs (anterior midline: F(1, 23) = 3.26, p < .1; anterior right: F(1, 23) = 3.98, p < .1) and was significant at all posterior ROIs (posterior left: F(1, 23) = 6.46, p < .05; posterior right: F(1, 23) = 5.99, p < .05). However, this early difference did not persist to the 300–500 ms interval, where no differences were observed at any ROI. Fig. 2 shows an electrode array and F-values are shown in Table 4.

In the P600 interval (600–1000 ms) ERPs in the ungrammatical condition were more positive than in the control condition. The effect was present across the entire scalp, but was largest at posterior and midline electrode sites, resulting in interactions between condition and posteriority and between condition and laterality, and a marginally significant three-way interaction between condition, posteriority and laterality. The effect was significant in all regions of interest.
Although the difference in the early interval was unexpected, it does not undermine the interpretation of the P600 effect, since there was no significant difference in the interval directly preceding the P600 interval, and the mean amplitude of the early difference (≈1 μV) was much smaller than the mean amplitude of the P600 effect (≈5 μV), averaged across the entire scalp. A further analysis that used a 100 ms post-stimulus baseline interval in order to exclude the effect of the early positivity showed that the strong P600 effect remained in evidence. Also, analyses revealed that the most important effects occurred at different word regions for the two auxiliaries, which made it more informative to focus on individual auxiliaries at each word region. All analyses reported here are based on individual word regions, but additional figures and statistical analyses based on the three-word region starting at the subject noun are available as supplementary materials.

### 3.3. Attraction conditions with the auxiliary *fue*

An omnibus ANOVA that included the verb region in all six critical conditions showed a main effect of the auxiliary type in the 300–500 ms interval ($F(1, 23) = 7.68$, $p < .05$), and in the 600–1000 ms interval it showed main effects of auxiliary type ($F(1, 23) = 11.38$, $p < .01$) and condition ($F(2, 46) = 3.31$, $p < .05$). No other main effects or interactions reached significance. In all further analyses we focus on ERP effects within each level of the auxiliary-type factor, since those effects are more closely related to the experimental predictions. Additionally, length differences between the auxiliaries *fue* and *estaba* could introduce differences that are unrelated to the primary manipulation of the study. Also, analyses revealed that the most important effects occurred at different word regions for the two auxiliaries, which made it more informative to focus on individual auxiliaries at each word region. All analyses reported here are based on individual word regions, but additional figures and statistical analyses based on the three-word region starting at the subject noun are available as supplementary materials.

#### 3.3.1. Subject noun and auxiliary

In the analysis of the three *fue* conditions (grammatical control, semantic-attraction and no-attraction) there were no significant effects of condition, or interactions with condition, in either the 0–200 ms or 300–500 ms intervals following presentation of the subject noun or in the corresponding intervals following presentation of the auxiliary.

#### 3.3.2. Critical verb

The critical verb following the auxiliary *fue* elicited a P600 in both anomalous conditions relative to the acceptable control condition, with no differences in the N400 interval. An ANOVA that compared all three conditions showed no differences in the 0–200 ms interval or in the 300–500 ms interval. In the 600–1000 ms interval both anomalous conditions (semantic-attraction and no-attraction) showed a positivity relative to the control condition. The effect was distributed across the whole scalp and was largest in midline and posterior regions. Fig. 3 shows the grand-average waveforms. This pattern of results was reflected in a main effect of condition ($F(2, 46) = 3.84$, $p < .05$). To determine the source of this effect, planned pairwise comparisons were made between all possible pairs of conditions (Table 5). This confirmed the visual observation that the semantic-attraction condition and the no-attraction condition were both more positive than the control condition, but did not differ from each other.

### Table 3

ANOVA F-values for comparisons of the semantic sub-experiment conditions at the critical adjective.

<table>
<thead>
<tr>
<th>Semantic felicity</th>
<th>0–200 ms</th>
<th>300–500 ms</th>
<th>600–1000 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cond (1, 23)</td>
<td>–</td>
<td>8.00*</td>
<td>–</td>
</tr>
<tr>
<td>Cond × lat (2, 46)</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Cond × post (1, 23)</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Cond × lat × post (2, 46)</td>
<td>–</td>
<td>4.09*</td>
<td>–</td>
</tr>
</tbody>
</table>

Effect of condition at individual ROIs

- Anterior left (1, 23) – 9.35*
- Anterior midline (1, 23) – 7.26*
- Anterior right (1, 23) – 3.59*
- Posterior left (1, 23) – 9.26*
- Posterior midline (1, 23) – 8.26*
- Posterior right (1, 23) – 6.25*

Numbers in parentheses indicate degrees of freedom.

* $0.01 < p < 0.05$.
** $p < 0.01$.
.05 < $p < 1$.  

![Fig. 1. Grand average responses at the critical adjective in the semantic control conditions, showing the felicitous condition (blue) and the infelicitous condition (red).](image)
3.4. Attraction conditions with the auxiliary estaba

Visual inspection suggested that the ERPs in the estába conditions may have diverged prior to the critical verb, and therefore ANOVAs were performed on the 0–200 ms and 300–500 ms intervals after presentation of the subject noun, and on the 0–200 ms, 300–500 ms, and 600–1000 ms intervals after presentation of the auxiliary. The 300–500 ms interval following presentation of the auxiliary corresponds to the interval immediately preceding the presentation of the critical verb.

3.4.1. Subject noun

At the 0–200 ms interval following presentation of the subject noun the three estába conditions showed no significant main effects of condition or interactions with condition. In the 300–500 ms interval following the subject noun the no-attraction condition showed a positivity relative to the control and semantic-attraction conditions at posterior ROIs. The positivity was reflected in a marginally significant interaction between condition and posteriority ($F(1, 23) = 3.88$, $p < .1$; posterior left $F(1, 23) = 3.88$, $p < .1$; posterior midline $F(1, 23) = 4.12$, $p < .1$; posterior right $F(1, 23) = 4.53$, $p < .05$).

The contrast between the no-attraction condition and the two other conditions at the subject noun was not predicted. Although the subject nouns differed across conditions within each item set, the nouns were counterbalanced across item sets such that 96% of the subject nouns used in the semantic-attraction condition also appeared as the subject noun in a different item set in the no-attraction condition. All subject nouns corresponded to existing discourse referents, as they had already been introduced in the context sentence. It is possible that the subject nouns differed across conditions in terms of how likely they were to appear as subject following the context sentence, and this could be responsible for the difference in the estába no-attraction condition at this word, but that should have led to a similar effect in the fue conditions, where no such differences were observed.

3.4.2. Auxiliary

ERPs in the estába conditions showed differences at the auxiliary. An early posterior positivity in the no-attraction condition matched the polarity and scalp distribution of the effect observed at the subject noun immediately prior to the auxiliary, and so it is unlikely to reflect an independent response to the auxiliary. In contrast, a broad positivity in both anomalous conditions that started around 300 ms was likely a response to the auxiliary itself, although any conclusions about the effects at this region must be treated with caution, due to the possibility of overlap with effects at the preceding word. This effect is probably a reflection of the compatibility between the subject noun and the auxiliary: when
**estaba** followed an inanimate subject (semantic attraction and no-attraction conditions) it elicited a positivity relative to the condition where it followed an animate subject (control condition). **estaba** is most frequently used in active progressive constructions with animate subjects, and hence the positivity may reflect the low expectancy for **estaba** following an inanimate subject noun. A number of existing neurocognitive models of sentence processing predict early sensitivity to such factors (e.g., Bornkessel-Schlesewsky & Schlesewsky, 2006; Kuperberg, 2007). Fig. 4 shows grand-average waveforms in all three conditions. Setting aside the early posterior positivity in the no-attraction (red) condition, which was a continuation of a response to the subject noun, the figure suggests that the positivity affected the two anomalous conditions similarly. This is most apparent at anterior electrodes where the ERPs are not contaminated by effects at the previous word, and in the later intervals at posterior electrodes, where the responses to the two anomalous conditions converge.

Statistical analyses showed that in the 0–200 ms interval there was a marginally significant effect of condition and a significant interaction between condition and posteriority. Follow-up pairwise comparisons showed that these effects reflected continuations of the differences already seen following the subject noun: the no-attraction condition was more positive than the other two conditions at posterior ROIs, and the control and semantic-attraction conditions did not differ from each other.

In the 300–500 ms interval following the auxiliary there was a three-way difference between the conditions. ERPs in the no-attraction condition continued to be more positive than in the semantic-attraction condition, which in turn were more positive than those in the control condition. These differences were reflected in a significant main effect of condition. This difference between the semantic-attraction and no-attraction conditions was present at posterior sites and not at anterior sites, as reflected in a significant interaction between condition and posteriority. At anterior sites both anomalous conditions elicited a very similar positivity relative to the control condition.

In the 600–1000 ms interval following the auxiliary there was a main effect of condition and no interaction of condition with either topographic factor. This reflected a broad positivity that was present in the semantic-attraction and no-attraction conditions, relative to the control condition. Planned pairwise comparisons showed that the difference between the anomalous conditions and the control condition was reliable at all regions of interest, and that the semantic-attraction condition and the no-attraction condition did not differ from one another at any region. Table 6 shows F-values for the pairwise comparisons of the three conditions.

The positivity at the auxiliary makes it more difficult to interpret any effects that might be observed at the immediately following verb region. However, the effect suggests that participants were immediately sensitive to the relation between the auxiliary **estaba** and the animacy of the subject noun, and thus it is directly relevant to our questions about the incremental integration of information from open and closed-class words during sentence comprehension. The positivity suggests that the comprehension system does not wait until the verb to evaluate the most likely semantic relations among words in the sentence. If the comprehension system makes such commitments before it encounters grammatical voice information at the main verb, then conflict effects that are observed at the main verb could be accounted for without the need to appeal to an independent semantic analyzer.

### 3.4.3. Verb

Visual inspection suggested that the semantic-attraction condition may have exhibited a positivity relative to the other two conditions from about 300 ms after the critical verb, but the ANOVAs revealed no significant differences at any interval following the verb (0–200 ms, 300–500 ms, 600–1000 ms). This lack of reliable effects may appear puzzling in light of the fact that the semantic anomalies were not subtle and hence are unlikely to have been overlooked by participants. Additionally, it should be noted that the ERPs for the anomalous conditions are based on the 84% of trials in which participants correctly judged the sentences to be unacceptable. The lack of differences at this region likely is due to a combination of the difference at the preceding auxiliary and the standard baselining procedure that sets all averaged waveforms to be matched in a 100 ms pre-stimulus interval (the use of a post-stimulus baseline interval would have the same effect in this case). Fig. 5 shows that at the auxiliary there was already a
difference between the control condition (animate subject noun) and the two anomalous conditions (inanimate subject nouns), and that this difference was long-lasting and affected both anomalous conditions in the same fashion (once the additional early positivity in the no-attraction condition elicited by the subject noun is taken into consideration). Thus, the grand-average waveforms shown in Fig. 5 do not indicate that there was no processing disruption during the presentation of the verb, but rather that there was no additional disruption beyond that caused by the combination of the subject noun and the auxiliary.

4. Discussion

The aim of the current study was to use Spanish to examine the ERP effects of different types of thematic anomalies, with the goal of better understanding the way in which syntactic and semantic information interact in sentence comprehension. Previous research (e.g., Kim & Osterhout, 2005; Kolk et al., 2003; Kuperberg et al., 2003, 2006, 2007; van Herten et al., 2005, 2006; Ye & Zhou, 2008) has been shown to tell the processor that the process may be able to use word meanings to propose interpretations that are incompatible with the surface structure of a sentence. This in turn suggests a sentence processing architecture with a semantic analyzer that is able to operate independently, without relying on syntactic parses to dictate how the meanings of words should be combined. This contrasts with the more common view in the psycholinguistic literature of a single analyzer that rapidly integrates diverse sources of information (syntactic, semantic, contextual) as soon as they become available. The ERP evidence for an independent semantic analyzer comes from two types of findings. First, many studies in English, Dutch, and other languages have observed P600 responses to sentences like The hearty meal was devouring . . . (Kim & Osterhout, 2005), in which there is a plausible thematic relation between the nouns and verbs in the sentence, but where that thematic relation is incompatible with the surface syntactic form of the sentence. Such findings have been taken to suggest that comprehenders consider a passive interpretation despite contradictory syntactic information. Importantly, Kim and Osterhout motivated their conclusion by demonstrating that the P600 response was not elicited in sentences like The dusty tabletop was devouring . . . in which the surface interpretation is similarly anomalous, but where there is also no plausible non-surface interpretation of the subject and verb. This semantic selectivity of the thematic P600 effect is a key piece of evidence for the independent semantic analyzer. Second, a number of ERP studies of thematic anomalies have found that no N400 effect is elicited, despite the presence of an obvious semantic anomaly. Under the assumption that the N400 is an index of semantic processing difficulty, the absence of an N400 effect could indicate that comprehenders temporarily fail to notice the semantic anomaly, again suggesting that they consider interpretations that do not fit the surface syntax of the sentence.

However, as discussed in the Introduction, the argument for an independent semantic analyzer is based upon a series of generalizations and assumptions that merit further scrutiny. First, the claim that the thematic P600 is specifically sensitive to semantic attraction is based on limited empirical evidence. Aside from Kim and Osterhout's study, there have been few other studies that have manipulated semantic attraction between nouns and verbs, and a number of studies indirectly challenge Kim and Osterhout's generalization by showing that P600 effects are elicited in sentences that lack semantic attraction (Kolk et al., 2003; Kuperberg et al., 2007; van Herten et al., 2006). Such studies cast doubt upon the importance of semantic attraction, but it is only by directly comparing semantically attractive and non-attractive noun–verb combinations that one can tell whether this feature impacts ERP responses. The current study included two manipulations of semantic attraction in an effort to address this issue.

Second, the argument for an independent semantic analyzer depends on the assumption that any unlicensed interpretation is constructed at a point when it already conflicts with the surface syntax of the sentence. If the ultimately unsupported interpretation is instead constructed at an earlier point in the sentence, when the syntax is still compatible with that interpretation, then there is no need to invoke an independent semantic analyzer. The current study explored this issue further by manipulating the choice of auxiliary in Spanish, using either the passive-biased auxiliary fue or the active-biased auxiliary estaba. An independent semantic analyzer could potentially combine noun and verb meanings while disregarding cues from closed-class words such as auxiliaries. On the other hand, a processor that rapidly integrates information from multiple sources might quickly take advantage of the cues provided by the auxiliary in forming expectations about the grammatical voice of the upcoming clause.

Third, the finding that thematic anomalies often do not elicit an N400 effect provides evidence of a semantic illusion, and hence an argument for an independent semantic analyzer, only under the assumption that N400 effects directly reflect semantic anomaly detection. In contrast, under the lexical view of the N400 effect, the presence or absence of N400 effects in sentences with thematic

<table>
<thead>
<tr>
<th>Table 5</th>
<th>ANOVA F-values for comparisons of the fue conditions at the critical verb.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fue</td>
<td>0–200 ms</td>
</tr>
<tr>
<td>Control vs. semantic-attraction</td>
<td></td>
</tr>
<tr>
<td>Cond (1, 23)</td>
<td>–</td>
</tr>
<tr>
<td>Cond × lat (2, 46)</td>
<td>–</td>
</tr>
<tr>
<td>Cond × post (1, 23)</td>
<td>–</td>
</tr>
<tr>
<td>Cond × lat × post (2, 46)</td>
<td>–</td>
</tr>
<tr>
<td>Effect of condition at individual ROIs</td>
<td></td>
</tr>
<tr>
<td>Anterior left (1, 23)</td>
<td>–</td>
</tr>
<tr>
<td>Anterior midline (1, 23)</td>
<td>–</td>
</tr>
<tr>
<td>Anterior right (1, 23)</td>
<td>–</td>
</tr>
<tr>
<td>Posterior left (1, 23)</td>
<td>–</td>
</tr>
<tr>
<td>Posterior midline (1, 23)</td>
<td>–</td>
</tr>
<tr>
<td>Posterior right (1, 23)</td>
<td>–</td>
</tr>
<tr>
<td>Control vs. no-attraction</td>
<td></td>
</tr>
<tr>
<td>Cond (1, 23)</td>
<td>–</td>
</tr>
<tr>
<td>Cond × lat (2, 46)</td>
<td>2.71 7</td>
</tr>
<tr>
<td>Cond × post (1, 23)</td>
<td>–</td>
</tr>
<tr>
<td>Cond × lat × post (2, 46)</td>
<td>–</td>
</tr>
<tr>
<td>Effect of condition at individual ROIs</td>
<td></td>
</tr>
<tr>
<td>Anterior left (1, 23)</td>
<td>–</td>
</tr>
<tr>
<td>Anterior midline (1, 23)</td>
<td>–</td>
</tr>
<tr>
<td>Anterior right (1, 23)</td>
<td>–</td>
</tr>
<tr>
<td>Posterior left (1, 23)</td>
<td>–</td>
</tr>
<tr>
<td>Posterior midline (1, 23)</td>
<td>–</td>
</tr>
<tr>
<td>Posterior right (1, 23)</td>
<td>–</td>
</tr>
<tr>
<td>Semantic-attraction vs. no-attraction</td>
<td></td>
</tr>
<tr>
<td>Cond (1, 23)</td>
<td>–</td>
</tr>
<tr>
<td>Cond × lat (2, 46)</td>
<td>–</td>
</tr>
<tr>
<td>Cond × post (1, 23)</td>
<td>–</td>
</tr>
<tr>
<td>Cond × lat × post (2, 46)</td>
<td>–</td>
</tr>
<tr>
<td>Effect of condition at individual ROIs</td>
<td></td>
</tr>
<tr>
<td>Anterior left (1, 23)</td>
<td>–</td>
</tr>
<tr>
<td>Anterior midline (1, 23)</td>
<td>–</td>
</tr>
<tr>
<td>Anterior right (1, 23)</td>
<td>–</td>
</tr>
<tr>
<td>Posterior left (1, 23)</td>
<td>–</td>
</tr>
<tr>
<td>Posterior midline (1, 23)</td>
<td>–</td>
</tr>
<tr>
<td>Posterior right (1, 23)</td>
<td>–</td>
</tr>
</tbody>
</table>

Numbers in parentheses indicate degrees of freedom.

* .01 < p < .05.
** p < .01.
* .05 < p < .1.

To assume that N400 effects directly reflect semantic anomaly detection. In contrast, under the lexical view of the N400 effect, the presence or absence of N400 effects in sentences with thematic
anomalies reflects the relative strength of lexical association among the words in the different experimental conditions, rather than the ability or inability to detect semantic anomalies. The current study did not directly test the functional status of the N400, but this issue is highly relevant to the interpretation of the results.

Because relatively few ERP studies have been conducted in Spanish, the experiment also included two control sub-experiments to confirm that syntactic and semantic anomalies elicit canonical N400 and P600 effects. The semantic sub-experiment manipulated the felicity of the semantic relation between a noun and a following adjective, and the syntactic sub-experiment manipulated the correctness of number agreement between a determiner and a noun. The results of the control sub-experiments showed that, as expected, the canonical N400 and P600 effects are also found in Latin American Spanish. An N400 was elicited at the adjective when the semantic relationship between the noun and the following adjective was infelicitous. Likewise, a P600 was elicited at the noun when the number agreement on the determiner and the noun mismatched. In addition to the P600 effect, there was also an early difference in the 0–200 ms interval following the noun. Although this difference was unexpected, it does not undermine the interpretation of the P600 effect because the early difference was much smaller than the P600 effect and the P600 effect remained clear even when a post-stimulus baseline was used to remove the early effect. Furthermore, the early effect did not overlap in time with the P600 effect. It is therefore unlikely that the difference in the P600 interval could be the result of pre-existing differences. These canonical N400 and P600 effects enable us to interpret the presence or absence of similar effects in our primary comparisons in terms of characteristics of the processing architecture, rather than in terms of fundamental differences in electrophysiological responses in Spanish relative to languages that have received greater attention in the ERP literature, such as English, Dutch, and German.

In the conditions of primary interest we manipulated the thematic fit between an inanimate subject and a verb. This manipulation follows the logic used by Kim and Osterhout (2005), but it has been tested in relatively few studies. We also introduced a novel manipulation of auxiliary type, using the Spanish past tense fue (past tense of both ser and ir), which is primarily used in passive constructions, and estaba (infinitive: estar), which is more frequently used in active progressive constructions. In addition, our study presented a context sentence prior to each target sentence that introduced the three potential subject nouns for the target sentences. This was motivated by the felicity conditions on the use of passive constructions in Spanish, but it had the additional benefit of reducing the effects of lexical availability and global semantic association that may have affected previous studies in this area, and could have contributed to varying findings on the N400.

Results from the conditions with the auxiliary fue showed a P600 in both the semantic-attraction and the no-attraction conditions, demonstrating that the thematic P600 is also found in Spanish, but challenging the claim that this effect is selectively elicited in cases of semantic attraction. The P600 observed in the semantic-attraction condition is consistent with results from other studies (e.g., Hoeks et al., 2004; Kim & Osterhout, 2005; Kolk et al., 2003; Kuperberg et al., 2006, 2007). This condition was closely comparable to the semantic-attraction condition in Kim and Osterhout’s study. In Kim and Osterhout’s experiment the auxiliary was highly compatible with a passive structure, and in the current experiment the auxiliary fue was biased toward a passive structure. Corpus analyses suggest that the probabilistic biases of the auxiliaries used in the English and Spanish studies are similarly strong. In contrast, the P600 observed in the no-attraction condition with fue contrasts with the results of Kim and Osterhout’s study, in which the no-attraction condition did not elicit a statistically reliable P600. However, our finding is consistent with other studies in which a P600 was elicited even in the absence of a plausible thematic relationship among the open class words in the sentence (Kolk et al., 2003; Kuperberg et al., 2006, 2007; van Herten et al., 2006).

Thus, the results from the fue conditions showed a lack of selectivity of the P600 to the thematic fit between the inanimate subject noun and the verb. This finding is surprising under any account that associates the thematic P600 with the consequences of an

**Fig. 4.** Grand average ERP responses at the auxiliary in the estabu conditions, showing the grammatical control (blue), semantic-attraction (black) and no-attraction (red) conditions.
One possibility is that the processor begins to build a passive sentence and does not presume a special status for reversal processes. There is no need to invoke an independent semantic analyzer. A possible objection to this proposal is that previous studies of selectional restriction violations have elicited a biphasic N400/P600 pattern (e.g., Friederici & Frisch, 2000; Kolk et al., 2003), rather than the monophasic P600 that we observed in our ffe conditions. However, we think that this discrepancy is only apparent. The term ‘selectional restriction’ is used to describe a number of different relations in the linguistics and psycholinguistics literature, and so the possibility of false parallels arises. In the case of the studies by Friederici and Frisch and by Kolk and colleagues the materials described as involving selectional restriction violations appeared to involve general violations of real-world plausibility. We consider this to be different from the more specific violation of agentivity requirements that we manipulated in our study. Additionally, the presence or absence of N400 effects in these studies is likely a reflection of the varying degrees of lexical association in the experimental conditions used. Our proposal that the late positivity in the ffe conditions reflects detection of an anomaly in the surface form of the sentence bears some similarity to a recent proposal by Bornkessel-Schlesewsky and colleagues that the thematic P600 is an instance of a P300b associated with binary categorization judgments (Bornkessel-Schlesewsky et al., 2011). But that account presumes that comprehenders specifically attend to the sentence as a ‘reversal anomaly’, which is a consideration of a non-surface interpretation, and also suggests a special status for reversible thematic anomalies. In contrast, our account makes no reference to alternative interpretations of the sentence and does not presume a special status for reversal anomalies.

It is also important to consider why our ffe conditions showed that the P600 was not affected by semantic attraction, whereas Kim and Osterhout concluded that semantic attraction is an important factor. The difference is unlikely to reflect contrasting probabilistic biases of the auxiliaries used in the English and Spanish studies, since our corpus analyses suggest that the passive biases of Spanish ffe and English was/had been are similarly strong. An alternative possibility is that the nature of the disambiguation at the main verb is different in the two languages. When Spanish speakers first encounter ffe they are likely biased to interpret it as the past tense of ser, but encountering the active progressive main verb may force them to consider the possibility that ffe is the past tense of ir. In contrast, the change from a passive to active voice in English requires no lexical reanalysis of the auxiliaries was or had been. We cannot exclude this possibility, but we should note that this account assumes that an initial passive analysis must later be revised to an active analysis in both languages. This assumption

Table 6

<table>
<thead>
<tr>
<th>Condition</th>
<th>N100 (1, 23)</th>
<th>N150 (1, 23)</th>
<th>N250 (1, 23)</th>
<th>N400 (2, 46)</th>
<th>N500 (4, 92)</th>
<th>P100 (1, 23)</th>
<th>P200 (1, 23)</th>
<th>P300 (1, 23)</th>
<th>P400 (2, 46)</th>
<th>P500 (4, 92)</th>
<th>P600 (1, 23)</th>
<th>P600 (2, 46)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control vs. semantic-attraction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cond (1, 23)</td>
<td>6.29</td>
<td>15.72</td>
<td>20.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cond × post (1, 23)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cond × lex × post (2, 46)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Numbers in parentheses indicate degrees of freedom.

* \( p < 0.05 \)

** \( p < 0.01 \)

*** \( p < 0.001 \)

We contend that there is no conflict. Van Herten and colleagues showed that a P600 effect is elicited in role-reversed Dutch sentences, irrespective of whether the grammatical number of the two nouns is matched or mismatched. This shows that the thematic P600 is unlikely to be a simple consequence of encountering unexpected number agreement on a clause-final verb. Van Herten and colleagues draw the stronger inference from these findings that the P600 effect is not due to a syntactic mismatch of any kind. However, it is not necessary that their conclusions about number morphology should generalize to all other syntactic features.

Another possible account of the P600 observed in the ffe conditions treats the effect not as the result of a garden path, but rather as a reflection of a violation in the surface syntactic and semantic form of the sentence, once the main verb is encountered. Specifically, the P600 may reflect the violation of a selectional restriction that occurs when an active verb that requires an agentive subject appears with an inanimate (non-agentive) subject. The possible importance of animacy and agentivity to the P600 has been suggested by other authors (e.g., Kuperberg, 2007). Since it focuses on properties of the surface form of sentences, this account also has no need to invoke an independent semantic analyzer. A possible objection to this proposal is that previous studies of selectional restriction violations have elicited a biphasic N400/P600 pattern (e.g., Friederici & Frisch, 2000; Kolk et al., 2003), rather than the monophasic P600 that we observed in our ffe conditions. However, we think that this discrepancy is only apparent. The term ‘selectional restriction’ is used to describe a number of different relations in the linguistics and psycholinguistics literature, and so the possibility of false parallels arises. In the case of the studies by Friederici and Frisch and by Kolk and colleagues the materials described as involving selectional restriction violations appeared to involve general violations of real-world plausibility. We consider this to be different from the more specific violation of agentivity requirements that we manipulated in our study. Additionally, the presence or absence of N400 effects in these studies is likely a reflection of the varying degrees of lexical association in the experimental conditions used. Our proposal that the late positivity in the ffe conditions reflects detection of an anomaly in the surface form of the sentence bears some similarity to a recent proposal by Bornkessel-Schlesewsky and colleagues that the thematic P600 is an instance of a P300b associated with binary categorization judgments (Bornkessel-Schlesewsky et al., 2011). But that account presumes that comprehenders specifically attend to the sentence as a ‘reversal anomaly’, which is a consideration of a non-surface interpretation, and also suggests a special status for reversible thematic anomalies. In contrast, our account makes no reference to alternative interpretations of the sentence and does not presume a special status for reversal anomalies.

It is also important to consider why our ffe conditions showed that the P600 was not affected by semantic attraction, whereas Kim and Osterhout concluded that semantic attraction is an important factor. The difference is unlikely to reflect contrasting probabilistic biases of the auxiliaries used in the English and Spanish studies, since our corpus analyses suggest that the passive biases of Spanish ffe and English was/had been are similarly strong. An alternative possibility is that the nature of the disambiguation at the main verb is different in the two languages. When Spanish speakers first encounter ffe they are likely biased to interpret it as the past tense of ser, but encountering the active progressive main verb may force them to consider the possibility that ffe is the past tense of ir. In contrast, the change from a passive to active voice in English requires no lexical reanalysis of the auxiliaries was or had been. We cannot exclude this possibility, but we should note that this account assumes that an initial passive analysis must later be revised to an active analysis in both languages. This assumption
concedes the point that passive interpretations may be constructed prior to the main verb, and in so doing gives up a key motivation for an independent semantic analyzer. Yet another possibility is that the non-significant tendency for a late positivity in Kim and Osterhout's non-attraction condition reflected a real effect that was obscured by overlap with the N400 in that condition. We should note that other studies that have shown a P600 in the absence of a plausible thematic fit among the open class words of a sentence in other structural configurations (Kolk et al., 2003; Kuperberg et al., 2007; van Herten et al., 2006). Furthermore, our lab has also found that the P600 is not modulated by thematic fit in two other English studies using the same constructions as the Kim and Osterhout studies, including one study that used identical target sentences (Stroud, 2008).

The conditions with the auxiliary fue also showed no N400 effect in any condition, including the no-attraction condition in which the noun was a poor thematic fit for any of the verb's arguments. As with the presence of the P600 in this condition, this result is inconsistent with the results from Kim and Osterhout's study, but the absence of an N400 effect has also been attested in other studies in which there was no plausible thematic fit among open class words (Kuperberg et al., 2007; van Herten et al., 2006), including most studies in which thematic anomalies have been created via argument role reversal (e.g., Kolk et al., 2003; Ye & Zhou, 2008). In the current study, the design of the context sentence may have mitigated differences in semantic association between the subject noun and the verb and overall lexical predictability across conditions. In contrast, in Kim and Osterhout's materials the semantic-attraction condition generally involved a high level of semantic association between the noun and the verb and the no-attraction condition generally involved a low level of semantic association between the two. This suggests that the N400 effect in that study may have reflected differences in lexical association across conditions rather than the plausibility of the thematic relationship among the open class words. This interpretation is supported by the results of van Herten and colleagues, which showed an N400 only when the object noun and the verb had a low level of lexical association (van Herten et al., 2006). It is similarly supported by other findings from our lab that show a reduced N400 amplitude when any nearby noun is semantically associated with the verb, regardless of whether it occupies one of the verb's argument positions (Stroud, 2008).

Another set of conditions tested for effects of thematic fit in sentences with the auxiliary estaba, which is most frequently used in active progressive constructions. In this set of conditions the ERPs at the lexical verb showed no reliable differences between conditions. This lack of differences is initially puzzling, as the semantic-attraction and no-attraction conditions are clearly anomalous and the results from the judgment task showed that participants detected the anomalies with a high degree of reliability (84%). Furthermore, our analyses included only those trials on which the judgment task was answered correctly. Results from the fue conditions and from the syntactic and semantic control sub-experiments show that our participants were clearly able to generate familiar N400 and P600 effects within the same experimental session.

However, examination of ERPs at earlier word positions suggests a reason why no effects were observed at the lexical verb. The impact of the inanimate subject nouns and the active-biased auxiliary was seen already at the auxiliary itself, where the anomalous conditions elicited a positivity relative to the control condition, which included an animate subject noun. This positivity had an onset of around 300 ms and was broadly distributed across the scalp.

The exact time-course and distribution of the ERPs elicited by the auxiliary estaba is compromised by an additional difference in the no-attraction condition that began 300–500 ms after the subject noun. This posterior positivity overlapped with responses to the auxiliary, making it more difficult to interpret early effects at posterior electrodes following the auxiliary. The additional response to the subject noun in the no-attraction condition was unexpected, since the nouns used in the semantic-attraction and no-attraction conditions were almost completely counterbalanced across item sets, and since all potential subject nouns were
introduced in the context sentence and thus corresponded to existing discourse referents. It is possible that the additional response reflected differences in discourse status of the possible subject nouns in the context sentence, but this must remain a speculation at present.

Notwithstanding the concern raised by the overlapping effect from the subject noun in the no-attraction condition, the responses at the auxiliary _estar_ suggest that the semantic-attraction condition and the no-attraction condition affected the ERPs in exactly the same fashion. At anterior electrodes, where there was no confound due to effects from the previous word, the positivity in the two anomalous conditions began around 300 ms after the auxiliary and was closely matched in the two conditions. At posterior electrodes the responses in the two anomalous conditions diverged in the 0–200 ms and 300–500 ms intervals, probably due to the effect at the previous word, but at the 600–1000 ms interval the anomalous conditions showed a similar positivity relative to the control condition. Overall, then, this suggests that the ERPs showed effects of animacy, but no effect of the thematic fit between the subject noun and the verb, just as in the conditions with the auxiliary _fue_. This conclusion must, of course, be treated with some caution, due to the effects of overlapping responses to adjacent words.

The effect of subject animacy observed at the auxiliary _estar_ suggests that different sources of syntactic and semantic information are rapidly integrated on-line. The fact that this effect was broad and long-lasting may account for why we observed no additional effects of the anomaly at the lexical verb. It is unclear whether this positivity should be treated as functionally equivalent to the positivity elicited at the lexical verb in the _fue_ conditions. The early onset (around 300 ms) and the broad scalp distribution may suggest that this effect is different from standard P600 effects, which tend to exhibit a posterior focus. However, comparison with the ERPs in the _fue_ conditions and the syntactic control sub-experiment reveals that the positivities were broadly distributed across the scalp in each case. The positivity at _estar_ certainly has an earlier onset than the positivity in the _fue_ conditions, but it is already known that lexical and syntactic factors can affect the timing of late positivities within studies (Gouvea, Philips, Kazanina, & PoeppeL, 2010; Osterhout et al., 1994; Phillips, Kazanina, & Abada, 2005), and the earlier effect in the _estar_ condition may be due to the high lexical frequency of the auxiliary, which could allow for more rapid integration of semantic cues.

The effect of animacy that we observed at the auxiliary in the _estar_ conditions contrasts with results from its closest English counterpart. Kim and Osterhout’s (2005) first experiment included two control conditions: a passive control (the hearty meal was devoured . . .) and an active control (the hungry boy was devouring . . .). In that study the choice of control did not affect the status of the P600 elicited at the lexical verb in the anomalous condition. Visual inspection suggests that responses were slightly more positive in the passive control condition (inanimate subject) than the active control condition (animate subject), but this difference was not statistically reliable. Consequently, Kim and Osterhout’s subsequent study included only the passive control condition. The large number of conditions in our study made it impractical to include both active and passive control conditions, and we therefore chose to use an active control for the _estar_ conditions because it was easier to make these sentences felicitous following a context sentence that also needed to be compatible with the other five versions of each item. The effect of animacy at _estar_ in our study may reflect the stronger biases associated with Spanish _estar_ than with English _be_. _Estar_ is commonly used in active progressive constructions, which typically include an agitative subject. It is also used with stage-level predicates (Carlson, 1977), which indicate temporary states—e.g., _I am happy right now_, or _I am sick_, whereas the auxiliary _ser_ is used with individual-level predicates that indicate more permanent states, as in _I am a happy person in general_, or _I am tall_. The fact that _estar_ is used with stage-level predicates may also contribute to the expectation for its subject to be animative.

Taken together, the results from both the _fue_ and _estar_ conditions in the current study show effects of mismatches between subject animacy, auxiliary bias, and verb voice, and these effects occur as soon as these effects are detectable in the surface form of the sentence. Meanwhile, we find no clear evidence of ERP effects that are selectively elicited by semantic attraction between the subject and the verb. These results can be explained in terms of a language processing architecture that rapidly integrates information from different sources, including the surface syntax of a sentence, with no need to invoke an independent semantic analyzer that ignores surface syntactic cues. Such architectures are entirely standard in psycholinguistics and neurolinguistics (e.g., Ferreira & Clifton, 1986; Friederici, 2002; Hagoort, 2008; MacDon-ald et al., 1994; Trueswell & Tanenhaus, 1994). Although none of the psycholinguistic or neurocognitive models listed here offer specific accounts of thematic P600 effects, we think that their fundamental architectural assumptions are compatible with the findings reported here.

We should emphasize that the current study does not account for all P600 effects that have been attributed to an independent semantic processing stream. Our discussion here has focused on evidence from sentences with verb-argument animacy mismatches, which have made up a substantial proportion of previous studies on the thematic P600 in English and Dutch. Our current findings do not extend to cases of subject-object-verb (SOV) sentences with reversed animate arguments, such as the Dutch counterpart of _The fox that at the poachers_ hunted (Van Herten et al., 2005). Such reversals elicited a P600 in Dutch, despite their syntactic well-formedness, and we have found similar effects in SOV sentences in Mandarin Chinese (Chow & Phillips, 2010). Such cases await further investigation.

5. Conclusion

The current study tested whether the language processor identifies plausible thematic relations despite contradictory syntactic information that is available prior to the semantic composition process, and whether the ‘thematic P600’ effect is specifically associated with cases of ‘semantic attraction’ between nouns and verbs. ERP recordings from Spanish showed that a broadly distributed late positivity is elicited as soon as comprehenders detect an unlikely combination of the animacy of a subject noun, the bias of an auxiliary, or the voice morphology (passive vs. active) of a lexical verb. In conditions with the passive-biased auxiliary _fue_ the positivity was elicited at the lexical verb, and in conditions with the active-biased auxiliary _estar_ the positivity was elicited at the auxiliary itself. In an important previous study by Kim and Osterhout, manipulation of thematic attraction between a subject and a verb elicited contrasting ERP responses, and the contrast has been interpreted as a key piece of evidence for an independent semantic analyzer that considers meanings that are not licensed by the syntactic structure of a sentence. In the current study we found no effects of the manipulation of thematic attraction, and thus no evidence that the language processor considers interpretations that are inconsistent with the surface structure of the sentence. We suggested that the effect of thematic attraction found in previous studies may be due to lower-level lexical association effects. Further we suggested that the P600 effect is sensitive to features of surface syntactic and semantic analyses, such as selectional restrictions involving animacy. In sum, this study did not find
evidence that the language processor considers interpretations that are not licensed by the syntactic structure of the sentence. Also, we found that once differences in lower-level lexical-semantic association were minimized, the thematic anomalies did not elicit an N400 response, regardless of whether there was a plausible thematic relationship between the noun and the verb. We propose that rather than motivating an independent, semantic compositional analyzer, these results reflect a language processor that rapidly integrates information from multiple sources (e.g., syntax, semantics, discourse, lexical probabilities) to continuously update its interpretation of an incoming sentence.

Acknowledgments

Preparation of this paper was supported in part by fellowships to CS from the Social Sciences and Humanities Research Council of Canada and the Alberta Government, and by Grants to CP from the National Science Foundation (BCS-0948554 and DGE-0801465). We would like to thank Iván Ortega-Santos, Leticia Pablos, Angel Gallego, Laura Fuentes, Antonia Carrasco, and Juan Uriagereka for help in creation of the Spanish experimental materials, and Gino Kuperberg, Norbert Hornstein, Albert Kim, Herman Kolk, David Poeppel, and Wing Yee Chow for valuable discussion of earlier versions of this work.

Appendix A. Supplementary material


References


