

**Effects of Prior Syntactic Information on Thematic Role Processing:
An Event-Related Potentials Study**

Clare Stroud

Final 895 Draft

April 23rd, 2007

1. Introduction

It has been widely assumed in psycholinguistics, and to an even greater degree in linguistics, that semantic composition is parasitic on the computation of syntactic structure. In other words, 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. Even proponents of constraint-based models (e.g., MacDonald et al., 1994; Tanenhaus et al., 1995), in which semantic information typically plays a larger role, assume that semantic information only affects the choice of analysis when the syntax is ambiguous. However, a series of recent ERP studies (starting with: Kuperberg et al. 2003; Kim & Osterhout, 2005a; Kolk et al., 2003) has challenged this assumption. The basic finding is that a certain class of sentences with semantically anomalous interpretations but unambiguous, grammatical syntax, elicit a P600. Given fifteen years of ERP studies (Osterhout & Holcomb, 1992; and many others since) that show P600s in response to the detection of a syntactic anomaly, these were surprising results. The sentences used in the new line of ERP studies are not only unambiguously grammatical, but their structures are very simple and canonical. For example, Kim & Osterhout (2005a) recorded a P600 at the verb in sentences such as *The hearty meal was devouring....* In this sentence, the NP *hearty meal* is unambiguously the subject of the active verb, and because hearty meals do not devour things, we would have expected the processor to consider this sentence to be semantically anomalous. Therefore, we would have expected an N400, the response typically associated with the detection of a semantic anomaly. The elicitation of P600 response instead suggests that perhaps the processor initially analyzed *hearty meal* as the object of *devour*, and subsequently determined that the syntax was not consistent with this interpretation. If

this is true, then it implies that the processor used semantic information of some sort to determine how the words should be combined.

These findings have been taken to mean that both syntactic-first and constraint-based models of language processing should be revised to include a role for semantic information in combinatorial processes, even when the syntax is unambiguous. The proposals, however, have varied in the extent of their departure from a model that computes combinatorial relationships primarily using syntactic information. On the more extreme end, Kim & Osterhout (2005a) argue that these results suggest a semantic composition mechanism that runs parallel to and independent of the syntactic processing system. Similarly, the model proposed by Kolk et al. (Kolk et al., 2003; Van Herten et al., 2005; Kolk & Chwilla 2007) argues for a heavy reliance on heuristics, and believe that the P600 reflects ‘executive control’ which may be called into action when there is some doubt about the correct representation for the sentence (e.g., if the representation proposed by the heuristics and the grammatical structure differ) and instructs the processor to review the sentence in case the inconsistency arose from mishearing or misreading. At the other end of the spectrum, Kuperberg (2007) argues for a perhaps more modest change of moving thematic role assignment into the same processing stream as syntactic structure-building (a ‘combinatorial stream’), while retaining a semantic association stream that is not involved in combining words into hierarchical relations.

A critical feature of all of these experiments, however, makes it difficult to distinguish between more extreme versions of the model, such as completely independent combinatorial mechanisms for semantics and syntax, and models that require less drastic changes, such as Kuperberg’s (2007) proposal of thematic role information playing a larger role in combinatory processing, or a model in which a representation is built using multiple cues from different sources. Concretely, most of the experiments published to date focus on the processing of a critical verb, with the new semantic and syntactic information arriving more or less simultaneously. In the example above, *was* is compatible with both the passive and the progressive forms, so it is the suffix *-ing* that determines the syntactic structure of the sentence, and the verb to which the suffix is attached (e.g., *devour*) which provides the semantic information that could be used to decide that *hearty meal* is a plausible theme. Thus, the two pieces of information arrive simultaneously, or depending on the lexical access model posited, the semantic information may even arrive before the syntactic information. Although these experiments rule out a model in which the combinatorial relationships relevant to the incoming word are completely determined using only phrase structure rules, ignoring all semantic information, they do not distinguish

between models in which semantic information computes combinatorial relationships in an independent processing stream, and one in which semantic information (or more specifically, thematic role information) plays a more minor role in combinatorial processing.

The current study investigates whether prior information about the upcoming syntactic structure can restrict interpretations based on plausibility that are not consistent with the predicted syntactic structure. In a model with a high degree of parallelism and independence between semantic and syntactic processing streams, information at the previous word about the upcoming syntactic structure will not necessarily restrict the actions of the independent semantic processor at the current word. However, in a model in which use of the two types of information is more tightly linked, the processor should only consider interpretations which are consistent with all the information that has been gleaned from the previous words.

In summary, various research groups have recently been investigating how two different kinds of information – syntactic and semantic/thematic – interact, as well as their relative influence in determining combinatorial structure. The experimental results suggest a larger role for semantic/thematic information than had previously been thought, however it is not clear how just how large. The current experiment attempts to find an upper limit on semantic/thematic influence in combinatorial processing, by asking the question: if syntactic information about the upcoming structure is relatively strong *and* presented in advance, can semantic/thematic information about plausible combinations nevertheless override the way the syntax predicts the words will be combined?

1.1 Standard Views of the N400 and P600

From the earliest discovery of ERP components sensitive to linguistic manipulations, it seemed that there was a clean division between semantic anomalies, which elicited an N400 response, and syntactic anomalies, which elicited a P600 response. In this section, I review some of the findings that supported this division, prior to the discovery of the class of semantic anomalies that elicit a P600 and “brutally disturbed” (Kolk & Chwilla, 2007, p. 257) this clean picture.

Kutas and Hillyard (1980) were the first to report an ERP component that is sensitive to manipulations of semantic information. They compared incongruous sentence-final words with congruous controls (*He spread the warm bread with socks/butter*) and found a negative-going

component with a higher amplitude for the incongruous word relative to the congruous word. This negative-going wave peaked at about 400 ms after the critical word, and the effect became known as the N400. In subsequent studies, it became clear that the N400 effect was not confined to sentence-final position but rather also occurred to sentence-medial anomalies as well (Kutas & Hillyard, 1983). Furthermore, it was determined that the N400 was not just found to anomalous words but rather is a natural response to every open-class word (Kutas, Van Petten & Besson, 1988; Van Petten & Kutas 1990). It has since become clear that the N400 is not just sensitive to semantic anomaly, but rather to a variety of manipulations of semantic information within a sentential or discourse context. The N400 effect is found when the word is congruous but unpredictable (Kutas & Hillyard, 1984), when the word is congruous but not consistent with expectations (Hagoort et al., 2004), and when the word is locally congruous but incongruent with the larger discourse (van Berkum et al., 1999).

In addition to being sensitive to semantic congruity with context, the N400 also appears to be sensitive to the organization of lexical items within semantic memory (for a review, see Kutas & Federmeier, 2000). Lower frequency words elicit higher-amplitude N400s than high frequency words (Smith & Halgren, 1987; Van Petten & Kutas, 1990; Allen, et al., 2003; Barber et al., 2004), semantically-related word pairs elicit lower N400 amplitudes than word pairs without a semantic relationship (Bentin et al., 1985; Holcomb 1988, 1993; Rugg, 1985), and concrete nouns and more imageable nouns also elicit N400s with smaller amplitudes (Holcomb et al., 1999; West & Holcomb, 2000; Swaab et al., 2002).

Although the N400 appears to be sensitive to semantic congruity with context and to the organization of items within semantic memory, it is not clear whether the N400 reflects semantic compositional processes. We know that the N400 is not elicited in response to grammatical or morphological violations (Kutas & Hillyard, 1983, and many others). However, it is also not obvious whether the N400 is modulated by the felicity of a semantic interpretation that depends on the syntax structure of the sentence. For example, no N400 effect has found in sentences such as *A robin is not a bird* (Fischler, Bloom, Childers, Roucos & Perry, 1983). In this sentence, the word *bird* fits well within the semantic context of robins, and *robin* and *bird* are associated within semantic memory. However, the overall meaning of the sentence, once the negation is computed, is anomalous. The lack of N400 effect in this case suggests that the N400 does not reflect semantic information that is derived from compositional processes. Similar results have been found in experiments involving quantifiers (Kounios & Holcomb, 1992; Kounios, 1996; Noveck & Posada 2003; Drenhaus, Graben & Frisch

2006; Beltrán, Carreiras, Alvarez & Santamaría 2006), suggesting that the effect is not an artifact of sentences with negation.

The P600, a large positive waveform peaking at about 600ms after the presentation of the critical word, was first reported by Osterhout and Holcomb (1992) in response to critical words that were incongruous with the expected syntactic structure in garden-path sentences such as *The broker persuaded to sell the stock was sent to jail*. Hagoort, Brown and Groothusen (2003) also reported this response to syntactic agreement violations and word order violations, and named the component the ‘Syntactic Positive Shift’, which is sometimes used as an alternate name for this component. Since then, this component has been elicited by a wide range of syntactic anomalies, including category and agreement violations (Friederici et al., 1993; Hagoort, Brown & Groothusen, 1993; Hahne & Friederici, 1999; Kaan, 2002), syntactic garden-paths (Friederici et al., 1996; Kaan & Swaab, 2003; Osterhout, Holcomb, & Swinney, 1994), and subcategorization violations (Friederici & Frisch, 2000; Osterhout & Holcomb, 1992). In recent years the P600 has also been elicited in well-formed sentences that require more processing than simple, canonical sentences. For example, the P600 has been elicited due to the presence of ambiguity (Frisch, Schlesewsky, Saddy, & Alpermann, 2002), as well as during the construction of long-distance dependencies (Fiebach, Schlesewsky, & Friederici, 2002; Kaan, Harris, Gibson, & Holcomb, 2000; Phillips, Kazanina, & Abada, 2005).

In summary, the N400 has been seen to reflect semantic congruity within context and the organization of lexical items within semantic memory, but not necessarily the outcome of semantic compositional processes licensed by the syntax, while the P600 has been seen to reflect syntactic anomalies and the necessity for additional syntactic processing.

1.2 New findings

In light of the relatively consistent view that the N400 reflects semantic processing and the P600 reflects syntactic processing, the following findings seem to contradict the standard assumption of composition primarily based on syntactic information. In this section I present the findings and generalizations, as well as a discussion of some of the accounts that have been proposed to account for the presence of the P600. I save most of the discussion of the large-scale models of language comprehension currently supported by the different groups until the following section, because it is not clear that the current data can be used to definitively support one over another.

The earliest studies in English to report this effect are Kuperberg et al. (2003) and Kim and Osterhout (2005a), shown in (1) and (2), respectively.

1. a. For breakfast the *eggs* would only **eat** toast and jam.
b. For breakfast the *boys* would only eat toast and jam.
2. a. The *hearty meal* was **devouring** the kids.
b. The *hearty meal* was devoured by the kids
c. The *hungry boy* was devouring the cookies.

In these sentences, the verb shown in bold elicited a P600 relative to the verb(s) in the corresponding control sentences. Based on the standard view of the N400 as a response to semantic anomaly, one might have expected that they would elicit an N400 since the syntax of both (1.a) and (2.a) unambiguously indicates that the NP is the subject of an active sentence. Given this syntactic configuration, it is the semantics that is anomalous in these examples.

The first Dutch study, Hoeks et al. (2004) found this effect by accident – the sentences were intended to elicit an N400 through semantic implausibility, but instead elicited a P600 at the verb in bold. An example set of sentences is:

3. a. De speer werd door de atleten geworpen.
The javelin was by the athletes thrown
'The javelin was thrown by the athletes.'
- b. De speer heft de atleten **geworpen**.
The javelin has the athletes **thrown**
'The athletes were thrown by the javelin.'

At around the same time the effect was also attested in Dutch by Kolk, Chwilla, van Herten and Oor (Kolk et al., 2003, expt 1). For example:

4. a. De stropers die op de vos joegen slopen door het bos.
The poachers who at the fox hunted_{pl} stalked through the woods
'The poachers who hunted_{pl} the fox stalked through the woods.'
- b. De vos die op de stopers **joeg** sloop door het bos.
The fox that at the poachers **hunted**_{sg} stalked through the woods
'The fox who hunted_{sg} the poachers stalked through the woods.'

A P600 was elicited at the relative clause verb in (4.b), relative to (4.a). As in the English cases, the sentences were syntactically well-formed but semantically anomalous.

1.2.1 Animacy and suffix mismatch

The simplest explanation that could be proposed for many of these results is that the P600 is elicited in the above examples because the subject NP violates the animacy subcategorization restrictions of the verb. If this is a syntactic feature, there would be no reason to change the interpretation of the P600 as a response to syntactic anomaly. In support of this account, the following experiment suggests that animacy violations do modulate the appearance of the P600. Kuperberg's group (Paczynski et al., 2006) conducted a study in which the semantically unexpected NP was either animate (5.b. and c.) or inanimate (5.d. and e.), and semantically associated (5.b. and d.), or not (5.c. and e.):

5. a. At long last the man's pain was understood by the doctor ...
- b. At long last the man's pain was understood by the **hypochondriac** ...
- c. At long last the man's pain was understood by the **violinist** ...
- d. At long last the man's pain was understood by the **medicine** ...
- e. At long last the man's pain was understood by the **pens** ...

Only the animacy violation cases (5.d. and e.) elicited a P600; when the NP did not violate animacy restrictions, such as in 5.b. and c. only an N400 was elicited. However, this account cannot cover all the findings, as there are other cases in which the P600 is attested even in the absence of an animacy violation. This is seen, for example, in the original Kolk et al. study (2003, experiment 1), repeated here from (4):

6. a. De stopers die op de vos joegen slopen door het bos.
 The poachers who at the fox hunted_{pl} stalked through the woods
 'The poachers who hunted_{pl} the fox stalked through the woods.'
- b. De vos die op de stopers **joeg** sloop door het bos.
 The fox that at the poachers **hunted**_{sg} stalked through the woods
 'The fox who hunted_{sg} the poachers stalked through the woods.'

This experiment also included stimuli with cats chasing mice, and in the anomalous versions, mice chasing cats, so the relevant distinction can also not be human/non-human.

A second possibility that would require little rethinking of the standard view is that perhaps the P600 arises because the suffix on the verb is easily 'repairable'. In the Kim & Osterhout (2005a) experiment, the suffix on the verb could be 'repaired' from *-ing* to *-ed*, making the sentence

acceptable. This could be categorized as a syntactic process, and therefore would not require any change in the interpretation of the P600 as reflecting syntactic processing. However, Van Herten et al. (2005, experiment 1) used similar materials as the Kolk et al. experiment (2003), except both NPs appeared in third person singular. This meant that the critical verb was in the same form in both the anomalous and the acceptable versions.

7. a. De schilder die op de ladder klom viel plotseling.

The painter that on the ladder climbed fell suddenly.

‘The painter who climbed the ladder suddenly fell.’

b. De ladder die op de schilder **klom** viel plotseling.

The ladder that on the painter **climbed** fell suddenly

‘The ladder that climbed the painter suddenly fell.’

Nevertheless, the anomalous (7.b) elicited a P600 at the verb. Therefore, this phenomenon cannot be due to a simple mismatch between the verbal inflection expected and observed, nor to some trivial process to ‘fix’ the morphological inflection.

1.2.2 Thematic roles, semantic attraction and semantic association

It does not appear that a mismatch in either animacy or morphological suffix is sufficient to explain the P600 effect to these semantically anomalous but syntactically acceptable sentences. A fairly obvious generalization across these experiments is that they always seem to be elicited by verb-argument combinations. In addition, in the earlier studies, the anomaly was always detected at the verb (cf. Paczynski et al., 2006, and Nieuwland and van Berkum, 2005; both discussed below). Due to syntactic differences, the anomaly in the English studies was over a combination of an NP and a verb, while in Dutch both the subject and the object NPs appear prior to the verb and must both be taken in combination with the verb. This makes it look as if we are dealing with the processing of thematic roles, in particular in cases where the anomalous material appears prior to the verb, rather than a more general notion of semantic processing, as, for example, Kim & Osterhout (2005a) seem to suggest.

Given that all the early examples involved verb-argument relations, it looked as if the effect might be accounted for by a more general notion of “reparability”. In all of the above examples, if the processor simply reversed the thematic role assignment, the sentence becomes acceptable. This proposal was supported by Kim & Osterhout’s (2005a) second experiment which included sentences in which the subject was a plausible theme but a bad agent of the verb (8.a.), as well as sentences in

which the subject NP was neither a plausible theme nor a plausible agent of the verb (8.b.), and a passive control (8.c.).

8. a. The hearty meal was **devouring**...
- b. The dusty tabletop was **devouring**...
- c. The hearty meal was *devoured*...

Consistent with past ERP studies of semantic anomaly, an N400 was elicited in 8.b. relative to the control in 8.c., but in the comparison of 8.a and 8.c. a P600 was elicited. This suggests that the P600 only occurs when the subject NP would make a good theme for the verb. A similar result from the Dutch group suggests that the relevant relationship really is the theme-verb relationship (remember that in Dutch the agent NP also appears before the verb). Van Herten et al. (2005, experiment 2) manipulated whether the verb was associated with the object (9.b.) or not (9.c.), but neither formed a plausible sentence when combined with the subject NP.

9. a. Jan zag dat de olifanten de bomen omduwden en hun mars door het oerwoud vervolgden.
John saw that the elephants the trees pushed-over and their march through the jungle continued
'John saw that the elephants pushed-over the trees and continued their march through the jungle.'
- b. Jan zag dat de olifanten de bomen **snoeiden** ...
John saw that the elephants the trees **pruned** ...
'John saw that the elephants pruned the trees...'
- c. Jan zag dat de olifanten de bomen **verwendden** ...
John saw that the elephants the trees **caressed** ...
John saw that the elephants the trees caressed ...

Consistent with past ERP studies of semantic anomaly, an N400 effect was attested in (9.c) relative to the control in (9.a), but a P600 appeared in the comparison of (9.b) with (9.a). In this case, both the subject and the object appear before the critical verb, it was the object-verb thematic fit that controlled whether an N400 or a P600 occurred, since in both cases the subject NP was a poor fit as agent of the verb.

Kim & Osterhout (2005a) use the term "semantic attraction" to refer to the situation in which an NP is a plausible argument for a verb, independent of the syntactic relationship between the two. Kuperberg (2007) refers to the same concept as "semantic-thematic fit". Under Kim & Osterhout's view, the P600 only occurs when there is semantic attraction between the subject NP and the verb,

under the hypothesis that an independent semantic combinatorial processor analyzes the NP as a good theme for the verb, and then, based on this analysis, the syntactic processor concludes that the syntax is anomalous.

However, a follow-up study by Kuperberg and colleagues (2006) suggests the P600 can also be elicited if the subject NP and the verb are semantically associated – that it is not necessary for the subject NP to be a plausible theme for the verb. The study included both sentences that are easily repairable (10.a) and sentences where a similar repair is impossible (10.b). Rating studies showed that the NP and the verb were judged to be equally associated in both cases.

10. a. When John arrived at the restaurant, the food would **order** ...

b. When they greeted the Queen of England, the trumpets would **curtsey** ...

A P600 was elicited at both these words, and in fact, the P600 to the verb in (10.b) had a slightly higher amplitude. Because *trumpets* are neither a good agent nor a good theme for the verb *curtsey*, it seems that the appearance of the P600 cannot depend on the ability to repair the sentence by reversing the thematic role assignment.

Although that study suggested that the presence of semantic association between the NP and the following verb is sufficient to elicit the P600, there are also studies that found the P600 even in the absence of semantic association and semantic attraction between the NP(s) and the verb. Kuperberg et al. (2007) compared their original sentences (11.a.) with sentences such as (11.b) in which there is no semantic association between the NP and the verb. This design is very similar to Kim & Osterhout's (2005a) (8.a) and (8.b).

11. a. Every morning at breakfast the eggs would **eat** ...

b. Every morning at breakfast the eggs would **plant** ...

The result was a P600 at the verb in both conditions, casting doubt on the idea that semantic association is the sole triggering factor for the P600. This pattern has also been seen in a German study by Friederici and Frisch (2000), in which a P600 was elicited at the verb in the following sentence:

12. a. Anna knows that the inspector (NOM) the banker (ACC) **stained** and left.

This pattern also appeared in the Hoeks et al. (2004) study initially presented above:

13. a. The javelin has the athletes **summarized** ...

These studies in which the P600 does occur even in the absence of semantic association, taken together with those above, suggests that the presence of semantic association implies that a P600 will occur, but its absence does not necessarily result in the absence of a P600, specifically in cases involving

thematic role assignment between an NP and the following verb. It is unclear what caused the P600 to be absent from Kim & Osterhout's (2005a) *the dusty tabletops were devouring*, but present in the studies described in this paragraph.

Adding to the complicated picture of results, a further pair of examples by Kolk et al. (2003, experiment 1) shown in (14) show that semantic association between the verb and the preceding NP can also result in both N400 and P600 components simultaneously:

14. a. De bomen die in het park **speelden**_{pl} waren stuk voor stuk bijzonder.

The trees that in the park **played**_{pl} were one by one remarkable.

'The trees that **played**_{pl} in the park were all remarkable.'

b. De bomen die in het park stonden_{pl} waren stuk voor stuk bijzonder.

The trees that in the park stood_{pl} were one by one remarkable.

'The trees that stood_{pl} in the park were all remarkable.'

Thus, it is not yet clear what controls the presence or absence of the N400 in these kinds of studies, but the presence of semantic association does appear to be related to the appearance of the P600 in this genre of sentences.

1.2.3 Other stimulus manipulations that affect the P600

In her comprehensive review article, Kuperberg (2007) describes a few additional factors that appear to influence whether the P600 will appear to semantically anomalous sentences of this type. Although the theoretical implications of these factors is at times unclear, they will impact experimental design and interpretation of results.

Kuperberg's group (Geyer et al., 2006) found that the P600 is more likely to occur when the semantic composition is impossible, rather than merely implausible. The relevant experiment is:

15. a. Tyler cancelled the subscription

b. Tyler cancelled the **birthday**

c. Tyler cancelled the **tongue**

In this case, there was a P600 and an N400 in c. relative to the control a., but an N400 in b. relative to a. This example is also interesting because it shows that the phenomenon is attested to a violation at a noun rather than a verb. Nevertheless, the violation still involves the relationship between a verb and its theme.

Context also appears to modulate this phenomenon – it appears that adding context (or simply more words prior to the violation?) makes the P600 more likely. This is particularly apparent in a study conducted in Dutch by Nieuwland and van Berkum (2005). Participants read short stories, in the middle of which an inanimate NP (e.g., *suitcase*) replaced an animate NP (e.g., *tourist*). The story-final sentences would look like:

16. a. Next, the woman told the tourist ...

b. Next, the woman told the **suitcase** ...

The word **suitcase** elicited, as expected, an N400 if the sentence was shown in isolation, but a P600 when it was embedded in the short story. In her review article, Kuperberg (2007) suggests that the effect of context may also be behind the difference in the nonsemantic-attraction condition in her study (17.a) which elicited a P600, and the nonsemantic-attraction condition in Kim and Osterhout (2005) (17.b), which elicited an N400.

17. a. Every morning at breakfast the eggs would **plant** ...

b. The dusty tabletops were **devouring** ...

It remains unclear whether an extra four words is enough context to change the processing mechanism enough that different components emerge. In addition, in the Nieuwland and van Berkum experiment, the context had the particular property of introducing the NP prior to its appearance as the anomalous word. In the Kuperberg (2007) study, the ‘context’ did not do this, so it is not clear whether that study and the Nieuwland and van Berkum study can be united in this way.

Finally, it seems that the P600 appears more robustly when participants are asked to make explicit acceptability judgments. Several groups have run experiments with the same stimulus materials twice, once asking for acceptability or plausibility judgments, and once where the participants simply read or listened for comprehension. Geyer et al (2006) ran the experiment mentioned above in (15) twice.

15. c. Tyler cancelled the **tongue**

In the experiment in which participants had to make a plausibility judgment, the critical word elicited a P600 and an N400, whereas when the task was simply reading for comprehension, only an N400 was elicited. Likewise, Kolk et al. (2003) ran the following kinds of sentences twice (see (14) above for full example set):

14. a. The trees that in the park **played** ...

When participants made explicit judgments a P600 was elicited, but when they read only for comprehension, an N400 was seen.

In summary, the elicitation of a P600 in response to semantically anomalous sentences with simple, canonical and grammatical syntactic structure cannot be explained away by making reference to mismatches in features that could clearly be considered syntactic, such as animacy or verbal suffix. The common feature of all these experiments is that they involve the relationship between a verb and its internal argument, however, the phenomenon does not solely depend on whether the sentence can be easily repaired by switching thematic roles such that the subject NP becomes the theme. The presence of semantic association between the NP in the relevant configuration and the verb also appears to make the P600 more likely to appear, but is not necessary. Finally, the presence of additional material prior to the violation, and explicit judgment tasks have both been hypothesized to increase the likelihood of seeing a P600.

1.3 Current accounts

There have been three groups that have made proposals with regards to this set of results: Kim & Osterhout (2005a), Kuperberg and colleagues (e.g., Kuperberg, 2007), and Kolk, Chwilla and van Herten and colleagues (e.g., Kolk & Chwilla, 2007). All suggest that we allow for a larger role in language processing for some kind of “semantic” processing. Although there is broad agreement in theories based on ambiguity resolution that interpretations are associated with structures (e.g., MacDonald et al., 1994; Tanenhaus et al., 1995), the authors of these studies all argue that the level of semantic involvement in combinatorial processes proposed in these constraint-based theories is not sufficient to explain the effects found in sentences with unambiguous syntactic structure. Nevertheless, there is a range of how much people think we have to deviate from a model in which syntactic structure is the primary source of combinatorial processes, outside of cases in which ambiguity resolution is necessary - for some it could mean as little as adding some kind of thematic role information into the syntactic processor, while others advocate adding a completely parallel stream of semantic processing. All authors argue for different streams of information processing; they disagree on what exactly each stream computes and what the P600 reflects.

	Processing Streams Proposed	Interpretation of the P600
Kim & Osterhout (2005a)	Semantic; Syntactic	Detection of syntactic anomaly
Kuperberg (2007)	Semantic memory-based association; Combinatorial	Continued processing required due to inconsistency within or across streams
Kolk & Chwilla (2007)	Heuristic-based; Syntactic	Executive control indicating detection of an inconsistency and the need to review the representation

Kim & Osterhout (2005a) argue the most forcefully for a complete revision of the language processing model to include parallel streams for syntactic and semantic composition. For them, a semantic composition system builds up its own representation based on the meanings of the individual words, unmediated by the syntactic structure. To make this explicit, they argue that the semantic composition system takes *hearty meal* and *devouring* and decides that the NP is the theme of the verb. It is unclear whether the semantic composition mechanism would function over selectional properties of the verbs, or using a more “world-knowledge”-based system that has little in common with the types of information used by the syntactic system. Although they would like to frame their finding as evidence for an independent semantic composition mechanism, they recognize that this is hard to distinguish from a “semantic heuristic” based on world-knowledge. They note that in order to account for their results, the world-knowledge heuristic must be stronger than the other heuristic that is commonly proposed that assumes that any sentence of the form NVN is agent-verb-theme (Bever, 1970; Townsend & Bever, 2001). Their idea of how the semantic and syntactic streams interact is that if a stream recognizes a “strongly attractive analysis”, it will pursue this analysis, otherwise it may be more susceptible to the analysis of the other stream. They argue that this allows us to capture both the experimental results, as well as the intuitive need for interaction between the two streams. Kim & Osterhout (2005a) maintain a “traditional” view of the separation of syntax (phrase structure rules) and semantics (everything else) and believe that the P600 reflects the perception of a syntactic error based on the semantic representation that is being maintained. This model does not then require a reanalysis of what the P600 has traditionally been thought to index.

In her recent review paper (2007), Kuperberg makes a serious attempt to spell out in detail the many possible models that are compatible with this set of results. Like Kim & Osterhout (2005a), she argues that semantic information plays a role in determining relationships among words. However, she believes that the solution is to rethink the division we commonly make between syntax and semantics. She proposes two processing streams: one based on semantic memory, and the other a combinatorial mechanism that works mainly over morphosyntactic rules but also assigns thematic roles. She remains agnostic about how much of this “semantic” processing we should push over to the combinatorial mechanism (i.e., syntax) – is it just animacy restrictions, or does it include a more sophisticated notion of selectional restriction? She argues that although the semantic stream plays some role in the interpretation of the sentence by relating incoming words based on semantic memory and lexical organization, this stream does not perform regular combinatorial operations. She adds that perhaps much of the processing of thematic-role information may take place in both the semantic and the combinatorial streams. Under this view, the P600 reflects some kind of conflict between the two streams, or within the combinatorial stream, and indicates continued or extra processing. The conflict could either be between the semantic memory-based stream and the combinatorial stream, or within the combinatorial stream if results of the thematic processing results conflict with the morphosyntactic processing results.

The Kolk, Van Herten & Chwilla group (Kolk et al., 2003; Van Herten et al., 2005, 2006; Kolk & Chwilla, 2007) also claims that these results are evidence for the existence of parallel processing streams – one semantic (although they prefer the term “heuristic”) and the other syntactic. Their model appears to be a hybrid between those proposed by the other groups. Like Kim & Osterhout (2005a), they continue to posit a purely syntactic stream. However, unlike Kim and Osterhout, they call for a significant revision of what the P600 component reflects. They posit a monitoring analysis, in which the P600 indexes the detection of a mismatch of the analyses proposed by the two streams. The detection of an inconsistency would cause the parser to review the sentence and processing done so far, to check whether there had been some error in the processing. Their analysis of the P600 appears to be similar in many ways to Kuperberg’s proposal, although couched in different terms. Kuperberg proposes that the P600 indexes additional processes needed to reconcile conflicting analyses, while the Dutch group proposes that the P600 reflects the actual recognition that some reconciliation may need to take place and the review of the processing completed so far.

1.4 Unresolved Issues

It remains unclear, however, whether the current experimental data can be used to distinguish a model in which there are two separate independent streams, semantic and syntactic, from a model in which there is more interaction or a model in which composition is based on thematic role processing plus morphosyntactic rules. In this section I outline in further detail three families of models that can be proposed for the relationship between syntactic and semantic processing.

On one end of the continuum, there are models in which there are fully parallel and independent processing streams for syntactic and semantic processing. At the other end of the continuum are the models in which all combinatorial work is based on syntactic structure. Somewhere in the middle of these two kinds of models are a set of models in which both syntactic and semantic information are taken into account in determining a single representation – a multiple cues model. This is perhaps what Kuperberg (2007) had in mind for the ‘combinatorial stream’ - multiple sources of information, including morphosyntax and thematic role assignment, collaborating to determine how the words should be combined. Under this view, there are not separate processing streams for syntax and for thematic role assignment, but depending on the particular circumstances, the influence of each may be more or less prominent.

The first model contains separate and independent streams for syntactic and semantic information. Although the strongest view would be that they never interact, this is clearly too strong since the ultimate interpretation could be completely independent of the syntactic structure of the sentence. Instead, under this strong view we could say that the two streams work separately for large chunks of the incoming material, and therefore do influence each other in the end, but there are large portions of time during which they work independently. Another possibility for a strong model is that the interaction is more frequent, but the influence of one upon the other is fairly weak. This would imply that two contradictory representations could be carried forwards. Under these kinds of models, we would not expect the syntactic prediction made at the previous word to have any effect on the semantic interpretation. This model speaks against a strongly incrementalist view in which all possible processing is done as soon as possible, since the syntax will not influence the semantics until some time downstream, and vice-versa. In the strongly incrementalist view, the syntax and semantic information would have to be reconciled after each word since leaving the reconciliation until after a larger portion

of the sentence had passed would imply that some relevant information was not used as early as it could have been.

A more moderate version of this model is one with separate information-processing streams which interact and are reconciled with one another fairly regularly, perhaps after each word. In this model it seems that there will be a single interpretation carried forward, but as the incoming word arrives, its semantic information is integrated into the semantic information of the carried-forward representation, and the syntactic information is integrated with the syntactic part of the representation carried forward. Under this view, any prior restricting information from either system will influence the possible interpretations of both systems at a later word, since there is only one single representation carried forward after each incoming word. The advantage of this system is that we can have two processors that are each specialized for a particular kind of information and therefore may work more quickly. However, this puts the burden on the “reconciliation” part of this model – if the syntactic and semantic information of an incoming word are treated separately, then after each incoming word there must be some way to check whether the two proposed representations are consistent. This is particularly difficult given that the representations will be stated in different terms, one using information relevant to syntactic structures and the other using information relevant for semantic interpretation. Furthermore, if the two representations contradict each other, there must be some way to reconcile the two.

In the multiple-cues family of models, new material is added to a single representation based on multiple sources of information, some of which might be semantic in nature, some of which might be syntactic. An obvious advantage here is that we remove the issues of how to compare and reconcile multiple streams. However, now we need a processing algorithm that can simultaneously handle qualitatively different types of information. However, if we follow Kuperberg’s argument and distinguish between different kinds of semantic information – thematic role information based on subcategorization and selectional restrictions, and then semantic-memory based information that may include world-knowledge and experiential information – then it is easier to imagine that thematic role information may be relatively easy to process alongside syntactic information because it is more limited and more clearly defined (e.g., specifications for two arguments per verb), in a way that memory-based semantic information may not be. However, under this view it is not so obvious why we get two different ERP components.

Although Kuperberg (2007), Kim & Osterhout (2005a) or Kolk and Chwilla (2007) all argue that their results do not support a model in which morphosyntactic rules are the sole basis for combinatorial processes, there are a couple of ways that the results might be viewed as largely consistent with such a model. The arguments they put forward all depend on the assumption that the syntactic composition that occurs matches the syntactic composition that is licensed by the surface forms observed. However, it is also possible that although morphosyntax is the basis for all combinatorial processes, in some cases comprehenders at least fleetingly map the external stimulus onto the wrong syntactic representation. Under this view, the semantic attraction effect would imply incorrect parsing rather than semantic composition independent from syntactic composition. The fact that semantic attraction and semantic association make the P600 more likely to appear indicates that the likelihood of incorrect parsing is modulated by semantic cues. Therefore, although the combinatorial mechanism is solely based on morphosyntactic rules, it is not isolated from incoming semantic information.

1.5 The Current Experiment

In all of the experiments conducted so far, the results reflect interpretations that are compatible with the syntax up until the critical word. For example, in the Kim and Osterhout study (2005a), the auxiliary *was* occurs with both the passive and the progressive. Therefore, the critical syntactic information differentiating these two syntactic constructions comes at exactly the same time as the semantic information – at the word *devour{-ed/-ing}*. In fact, it is possible that the syntactic information arrives even later than the semantic information because it is carried by the suffix. The English results, therefore, are compatible with models in which there is no independent composition based upon semantic information, but because of the particular timing of information arrival in these kinds of sentences, the effect of the semantic cues appears larger.

The current study aims to reverse this timing effect, such that the syntactic information appears prior to the semantic information carried by the verb. In Spanish, the choice of auxiliary can be manipulated such that the passive structure can be effectively ruled out even before seeing the verb and its suffix. The question is whether this information can be used to modulate the effect of semantic attraction at the verb. If the processor knows that the passive structure is not correct for this sentence, even before the critical verb, will it still try to process the subject NP as theme of the verb in cases of semantic attraction? If there were such a thing as a semantic composition mechanism that worked completely independently of the syntactic processor, changing the syntactic restrictions at the previous

word would not impact the representations proposed by the semantic processor based on the most plausible way to combine the subject NP and the verb. On the other hand, in any theory in which there is no independent semantic composition, or in which the semantic and syntactic processors are more tightly linked, prior information of either kind should restrict downstream interpretations to those that are consistent with the prior information. The experiment, therefore, attempts to put an “upper cap” on how independent semantic compositional processes can be.

In Spanish, the auxiliaries preceding the passive and the progressive verbs are different, forms of *ser* and *estar*, respectively (although, of course, these are not the only structures in which these auxiliaries appear). Therefore, the choice of auxiliary can be used to manipulate the upcoming syntactic expectation *prior* to the semantic information present in the main verb. In particular, if the form of *ser* is used, a verb in progressive form is unlikely to follow, and if a form of *estar* is used, a passive is unlikely to follow. In a similar design to that of Kim & Osterhout’s (2005a) experiment 2, grammatical controls were compared with anomalous sentences in which there was semantic attraction between the verb and the first NP (i.e., that NP would make a good theme for the verb) and also with anomalous sentences in which there was no semantic attraction between the NP and the verb. Two different auxiliaries were used – *fue* which is the third person singular past tense of the auxiliary *ser*, and *estaba* which is the first and third person singular progressive tense of the auxiliary *estar*. *Fue* is used in a passive syntactic structure, and *estaba* is used in a progressive syntactic structure. It should be noted that both can also be followed by adjectives, but crucially it would be highly unlikely for a passive to follow *estaba* and for a progressive to follow *fue*. All six versions of an item were preceded by an identical context sentence that introduced all three NPs that began the target sentences (good theme, bad theme, agent). An example set of materials follows below in table 1.

Table 1. Sample materials for the main experiment.

“Semantic Attraction” target experiment		
Context sentence		Después del fuego, el bombero formuló un aviso para prohibir la entrada al piso quemado. ... <i>After the fire, the fireman prepared a warning to forbid entrance into the burnt apartment.</i>
<i>fue</i> - passive prediction	grammatical control	El aviso fue declarado por el bombero ante todos los vecinos. <i>The warning was declared by the fireman in front of all the neighbors.</i>
	+Sem. Attr.	El aviso fue declarando con un megáfono para que todo el mundo lo oyera. <i>The warning was declaring with a megaphone so everyone would hear it.</i>
	-Sem. Attr.	El piso fue declarando con un megáfono para que todo el mundo lo oyera. <i>The apartment was declaring with a megaphone so everyone would hear it.</i>
<i>estaba</i> - progressive prediction	grammatical control	El bombero estaba declarando el aviso cuando el fuego se formó otra vez. <i>The fireman was declaring the warning when the fire rose up again.</i>
	+Sem. Attr.	El aviso estaba declarando con un megáfono para que todo el mundo lo oyera. <i>The warning was declaring with a megaphone so everyone would hear it.</i>
	-Sem. Attr.	El piso estaba declarando con un megáfono para que todo el mundo lo oyera. <i>The apartment was declaring with a megaphone so everyone would hear it.</i>

Kim & Osterhout had a 2005 Amlap paper (Kim & Osterhout, 2005b) in which they presented results from an auxiliary manipulation in English. They included both *was*, which could be followed by either *devoured* or *devouring*, and also *will*, which is more likely to be followed by *devour*, although still compatible with the passive voice, with an additional *be*. Examples are shown in (18).

18. a. The hearty meal was devouring...
 b. The dusty tabletop was devouring...
 c. The hearty meal will devour...
 d. The dusty tabletop will devour...

They found that the P600 was still present to (18.c), which seems to suggest that an auxiliary that is incompatible with the passive does not block the processor from trying to make *hearty meal* the theme of *devour*. However, as mentioned above, a possible continuation is *will be devoured*, in which case *hearty meal* could still be the theme for the verb. In Spanish, the passive is a much less frequent form than in English. While the passive is felicitous in a situation where the theme NP has already been mentioned (hence the reason for including a context sentence before the target sentences), the

construction of *estaba siendo devorando* (lit. was being eaten) is highly dispreferred. It is therefore unlikely that participants would expect that after an inanimate NP and the auxiliary *estaba*.

1.6 Predictions

If the streams processing semantic and syntactic information are parallel but reconcile after each word, or if the multiple-cues model is correct, then we expect the auxiliary manipulation to restrict the thematic role assignments to only those that are compatible with the upcoming structure. Concretely, after *fue*, the passive structure is still a possibility. Therefore, there are maximum chances of seeing evidence for the processor treating the subject NP as a plausible theme for the verb, and we should see a P600 for the semantic-attraction condition (*hearty meal*) and an N400 for the nonsemantic-attraction condition (*dusty tabletop*), exactly as in Kim & Osterhout's (2005a) experiment 2. However, after *estaba* the processor should not consider the passive structure as an option and therefore should not consider the NP as a possible theme for the passivized sentence. In this case, we should see an N400 in both the semantic-attraction and nonsemantic-attraction conditions.

If, however, the semantic processor operates more independently from the syntactic processor, then the different syntactic structures predicted by the auxiliaries should not affect the interpretations made based on goodness-of-thematic-fit. Concretely, we should see a P600 to both semantic-attraction conditions (one with *fue*, and one with *estaba*), and an N400 to both nonsemantic-attraction conditions (one with *fue*, and one with *estaba*).

Based on prior results, there are a few additional possibilities, although it should be stressed that in all cases, under any view except a model with a very high degree of independence between the semantic and syntactic compositional mechanisms, we expect the results for *fue* to be different than those for *estaba*. The first additional possibility is based on the observation that some studies (Hoeks et al., 2004; Kolk et al., 2003 expt 1; Van Herten et al., 2006 expt 1) found a combination of N400+P600 to violations when there was semantic-attraction between the NP and the verb. If this occurs in this study, we would find a N400+P600 to the *fue* + semantic-attraction condition and an N400 to the *fue* + nonsemantic-attraction condition. By contrast, with the *estaba* conditions we would still expect an N400 to both.

One of the Kuperberg studies (2007) showed that both the semantic-attraction and the non-semantic attraction conditions elicit a P600 (e.g. *the eggs would only eat* and *the eggs would only plant*). Under this scenario, the *fue* conditions should both elicit a P600 and the *estaba* conditions should both elicit an N400. Again, it is important to emphasize that under all of these scenarios, the auxiliary manipulation would cause different components to be elicited at the verb, except if the semantic and syntactic combinatorial streams work independently.

1.7 Control sub-experiments

Because few ERP studies have been conducted in Spanish, we included control conditions to confirm that syntactic and semantic anomalies elicit the standard N400 and P600 effects. This would allow us to compare the components generated in the novel sub-experiment with clear results from an experimental comparison that we understand better. The two control sub-experiments contained each two pairs of conditions that manipulated, respectively, the semantic felicity between noun and a following adjective, and the correctness of number agreement between determiner and noun. The control sub-experiments were run together with the attraction study. A sample set of experimental conditions is presented below in Table 2.

Table 2. Sample materials for the two control experiments.

Semantic agreement controlexperiment		
Context sentence		Cuando Juan se hizo millonario, contrató a un chófer y compró un carro costoso. ... <i>When Juan became a millionaire, he hired a chauffeur and bought an expensive car. ...</i>
sem. agree-ment	+felicitous	El chófer estaba limpiando el carro lujoso, cuando Juan chocó con una moto contra la puerta del garaje. <i>The chauffeur was cleaning the luxurious car, when Juan crashed his motorcycle into the garage door.</i>
	-felicitous	El chófer estaba limpiando el carro enfadado, cuando Juan chocó con una moto contra la puerta del garaje. <i>The chauffeur was cleaning the angry car, when Juan crashed his motorcycle into the garage door.</i>
Syntactic agreement controlexperiment		
Context sentence		Cuando el ladrón entró por la ventana de la casa, la alarma antirrobo empezó a sonar muy alto. ... <i>When the thief entered the house through the window, the burglar alarm began ringing loudly. ...</i>
syn. agree-ment	+agree	El ladrón estaba desconectando los hilos del alarma antirrobo cuando llegó la policía. <i>The thief was disconnecting the(pl.) wires of the burglar alarm when the police arrived.</i>
	-agree	El ladrón estaba desconectando el hilos del alarma antirrobo cuando llegó la policía. <i>The thief was disconnecting the(sg.) wires of the burglar alarm when the police arrived.</i>

The semantic felicity between a noun and its adjective has been previously tested in Spanish ERP studies. Martín-Loeches et al (2006) 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 found a clear N400 (and also, surprisingly, a small P600). The N400 has also appeared in Spanish experiments in which the semantic felicity of a word depends on the meaning of the entire sentence, rather than on a combination of adjacent items. In a series of both auditorily- and visually-presented mini-stories, Wicha and colleagues (2003a, 2003b) compared sentences in Spanish that translate to: *Little Red Riding Hood was carrying food to her grandmother in a very pretty {basket/crown}*. In this case they also see a clear N400 to the infelicitous word. These experiments suggest, as would be expected, that the N400 behaves similarly in Spanish as in other languages.

The syntactic control sub-experiment simply looked at number (dis)agreement between an article and a noun. This was tested by Barber and Carreiras (2005), and resulted in a LAN plus a P600. Various similar studies of grammatical agreement have also been tested in Spanish and yielded similar results: Barber and Carreiras (experiment 2, 2005) and Martín-Loeches et al. (2006) also tested number agreement between a noun and its following adjective (LAN and P600 in both cases); Martín-Loeches et al. (2006) tested grammatical gender agreement between a noun and its following adjective (also LAN and P600); finally, Wicha et al. (2004) tested grammatical gender agreement between an article and a noun and found a P600. Barber & Carreiras (2003; 2005 expt 1) presented word pairs (article plus noun or noun plus adjective) that disagreed in number, and elicited an N400, which was surprising when compared to previous number agreement mismatches in English and Spanish. Nevertheless, this result is probably related to the difference between presenting isolated word pairs and embedding the mismatch in a sentential context.

2. Methods

2.1 Participants

22 people participated in the ERP study. Data from 1 participant were excluded due to technical problems, data from 4 participants were excluded due to low rates of correct response (<75%), and data from 1 participant were excluded due to high levels of artifacts in the EEG recordings. All 16 remaining participants (9 female; mean age 27; range 19-41 years) were healthy, native speakers of

Latin-American Spanish (4 from Colombia, 4 from Perú, 3 from Chile, and 1 each from Venezuela, Ecuador, Guatemala, Costa Rica and Puerto Rico). All had been in the United States for three years or less (mean: 1.6 years; range: 1 month to 3 years) except for one participant who had been in the country for four years but spoke Spanish almost exclusively in his/her everyday life. All were right-handed and had normal or corrected-to-normal vision. All participants gave informed consent and were paid \$15/hour for their participation, which lasted approximately 4 hours, including set-up time.

2.2 Materials

The materials for the study consisted of sets of six sentences organized in a 2 x 3 factorial design, illustrated in Table 1. There were two levels of auxiliary (*fue* and *estaba*) and three levels of condition (control, semantic-attraction and nonsemantic-attraction). The semantic control experiment contained pairs of sentences corresponding to the felicitous and infelicitous noun-adjective conditions, and the syntactic control experiment contained pairs of sentences corresponding to the grammatical and ungrammatical article-noun agreement conditions.

In all conditions, the target sentence was preceded by a context sentence. In Spanish, the passive construction is not felicitous if it introduces a new NP, but can be used if the subject NP is referring back to an NP that is already within the discourse. The context sentences were identical for every sentence within each set of items, and mentioned all three possible NP continuations for the following target sentence sextuples (animate NP, inanimate good-theme NP, inanimate bad-theme NP). The context sentences were compatible with both the passive and the progressive grammatical control sentences.

In all six target conditions, the sentences consisted of an NP (always determiner plus noun), followed by an auxiliary (either *fue* or *estaba*), followed by a verb with either passive or progressive verbal morphology. For the anomalous sentences, the two factors varied were goodness-of-fit of the inanimate NP as a theme for the verb, and auxiliary: *fue*, which is used with passives, or *estaba* which is used with progressives.

It is important for this design that *ser* can be actually be followed by the progressive (*fue buscando* which roughly translates as ‘went around looking’) and *estar* can be followed by a passive form (*estaba cerrado por* ‘was closed by’), although both are unlikely and require very specific conditions

in order to make the, felicitous. If these continuations were impossible rather than unlikely, we would have a straight syntactic violation, rather than a semantic anomaly, which would defeat the purpose of the experiment. The baseline control for the two anomalous sentences with *fue* was a passive using the semantic-attraction NP (i.e., *the hearty meal was devoured*), and the baseline control for the two anomalous sentences with *estaba* was an active progressive sentence with an animate NP (i.e., *the hungry boy was devouring*). The different baselines were chosen because, as just described, although the *fue*+progressive and *estaba*+passive form are not impossible, they are highly dispreferred and therefore it is hard to find large numbers of NP+verb combinations that work in these constructions. In addition, in Kim & Osterhout's (2005a) experiment 1, they compared the anomalous *hearty meal was devouring* to both an active (*the hungry boy was devouring...*) and a passive control (*the hearty meal was devoured...*) but the choice of control did not impact the results. Therefore, the most felicitous control sentences were used within each of the auxiliary manipulations.

For each set of six sentences, a verb and three NPs were chosen. The verb was chosen such that it could be felicitous in both passive and progressive constructions. The verbs chosen for this experiment tend to impose less rigid selectional restrictions than those used in many of the experiments discussed above (Kim & Osterhout, Kuperberg et al., van Herten et al. etc) because it is harder in Spanish to find concrete verbs that are felicitous in passive voice – those that are more felicitous tend to be verbs from business or newspaper language, such as *investigar* 'investigate', *firmar* 'sign', *entregar* 'deliver', etc. Different verbs were used for each item set, with the exception of twelve verbs which were repeated once each in order to increase the felicity of the relevant items. Following the selection of the verb, three NPs were chosen: one animate noun that would be a good agent for the verb, one inanimate noun that would be a good theme for the verb, and another inanimate noun that would make a poor theme for the verb. Over the entire set of experimental materials, 96% of the inanimate nouns were balanced such that in one item set they appeared as the good-theme noun and in another set they appeared as the poor-theme noun. This means that any difference between the semantic-attraction and nonsemantic-attraction conditions is not due to basic differences in the processing of the initial NPs. This was not the case for 4% of the items in which the nouns were changed to increase the felicity of specific items.

For the anomalous versions of each item, the word following the verb which was distributed equally among three options: a by-phrase ('by the fireman'), a preposition or other adverb ('with a megaphone' or 'many times'), or the other inanimate NP ('the apartment') – the nonsemantic-attraction NP in the semantic-attraction version, and vice-versa. Within one set of six sentences, all

anomalous versions shared the same ending. The majority of words following the critical verb were short: prepositions of definite articles, with the occasional slightly longer adverb. The variation in endings was chosen to avoid biasing people to consistently attribute a certain interpretation to the violations – for example, if they had always seen a *by*-phrase after the anomalous inanimate NP + progressive form, it might have led them to consistently classify this as a syntactic error (it should have been in passive form) rather than as a semantic error (inanimate NPs do not perform this action).

Materials for the two control experiments began in a very similar form to the target experimental items: context sentence, determiner, noun, auxiliary, verb... They were designed to ensure that, when combined with the target items, the second, target sentences contained equal numbers of the following beginnings: inanimate+*fue*, inanimate+*estaba*, animate+*fue* and animate+*estaba*, in order to minimize predictability.

180 sets of items for the target conditions were distributed across six presentation lists in a Latin Square design such that each list contained 30 items per condition. 30 pairs of grammatical agreement items were distributed across two presentation lists. 30 pairs of items manipulating semantic felicity (these were separate items from those used in the grammatical agreement sub-experiment) were also distributed across two separate presentation lists. These two control sub-experiments were then crossed such that there were four presentation lists for the control experiment items. The six lists of target items were crossed with the four lists of control items, and 60 filler items to create 24 lists with 360 items each. As of this date, only 22 of the lists have been run. The filler items were similar to the experimental items in maintaining a two-sentence format, and were all grammatically correct but had a variety of different syntactic structures. Thus, items from the six target conditions made up half of the items, and the ratio of grammatical to ungrammatical sentences was 1:1. Furthermore, since the violations occurred either earlier in the sentence (target conditions) or later in the sentence (control experiment conditions), participants needed to pay attention to the entire sentence in order to accurately judge the well-formedness of the sentence.

2.3 Procedure

Participants were comfortably seated in a dimly lit testing room around 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, which began 20 ms later. 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 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 on the screen for 300 ms, followed by 200 ms of blank screen. The last 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 3-trial practice session that included both grammatical and ungrammatical sentences. Participants received feedback and were able to ask clarification questions about the task at this time. The experimental session was divided into 7-8 blocks lasting 15 minutes each.

2.4 EEG recording

EEG was recorded from 30 Ag/AgCl electrodes, mounted in an electrode cap (Electrocap International): midline: Fz, FCz, Cz, CPz, Pz, Oz; lateral: FP1/2, 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 bandpass 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. Prior to any further analyses, epochs with ocular and other large artifacts were rejected from analysis based on visual screening. In addition, epochs from trials to which the subject responded inaccurately were excluded. Four participants' data were excluded because of accuracy below 75%. Epochs showing a divergence from 0 of more than 75 μ V were excluded automatically. Data from one participant were excluded because only 22% of critical epochs remained after all these measures were taken. 16 participants' data were included in the subsequent analyses.

The waveforms of the individual trials were normalized using a 100 ms prestimulus baseline. Averaged waveforms were filtered offline using a 10 Hz low-pass filter for presentation purposes, but all statistics are based on unfiltered data. The following latency intervals were chosen for analysis, based on the intervals used in the literature and on visual inspection: 300-500 ms (N400), 600-1000 ms (P600). In the conditions in which it looked like there was both an N400 and a P600 effect, the data was rebaselined using the 300-500ms interval as was done in Hagoort (2003), motivated by the fact that if there is a large N400, it might obscure a following P600. Statistics using both the prestimulus baseline and the new baseline are reported for the P600 window.

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, OZ), right posterior (CP4, P4, O2). These ROIs were organized into the two topographic factors *laterality* (left, midline, right), and *anterior-posterior*.

ANOVAs were performed separately for the conditions with *fue* and for those with *estaba* because the difference in length of the auxiliary would make it difficult to compare waveforms at the critical word directly after the auxiliary. In any case, the hypotheses were that the auxiliary would cause the *semantic-attraction* and *nonsemantic-attraction* anomalous conditions to perform differently relative to the controls; there is no hypothesis about comparing, for example, the *semantic-attraction* condition with *fue* with the *semantic-attraction* condition with *estaba*.

ANOVAs were performed hierarchically, using the within-subjects factor *condition* (*control*, *semantic-attraction*, *nonsemantic-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 report as significant 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 grammaticality judgment task for the main six conditions was 85%. The accuracy on the two grammatical control conditions was higher (*fue*: 91%; *estaba*: 90%) than for the four anomalous conditions (*fue*+semantic-attraction: 77%; *fue*+nonsemantic-attraction: 85%; *estaba*+semantic-attraction: 83%, *estaba*+nonsemantic attraction: 85%). This is likely because the anomaly resulted from the addition of a single letter at the end of the verb (e.g., passive *-ado* became progressive *+ando*) and therefore may be easily missed, especially if the verb is long.

The semantic control experiment showed an overall accuracy that was somewhat lower, 74%. Average accuracy for the felicitous condition was 83%, while for the infelicitous condition it was only 64%, suggesting that perhaps some noun-adjective combinations were not as infelicitous as had been intended, or perhaps that these anomalies were not perceived as being as bad as the other anomalies in the experiment. The syntactic control experiment showed an overall accuracy of 88% with no difference between the accuracies in the grammatical and ungrammatical conditions.

3.2 Control sub-experiments

The results for the semantic and syntactic control experiments were as expected: an N400 effect for the adjective in the infelicitous noun-adjective combinations, and a P600 for the nouns that were preceded by a determiner that mismatched in number.

3.2.1 Semantic violation

In the 0-200 ms window, there were no differences between the conditions. In the N400 window (300-500 ms), the semantic violation condition was more negative than the control condition over the whole head, but particularly in the posterior half. There was a general effect of condition and of an interaction between condition and posteriority, in addition to significant effects in all six regions of interest. Figure 1 shows the electrode map, and the *f-values* can be seen in Table 3 below.

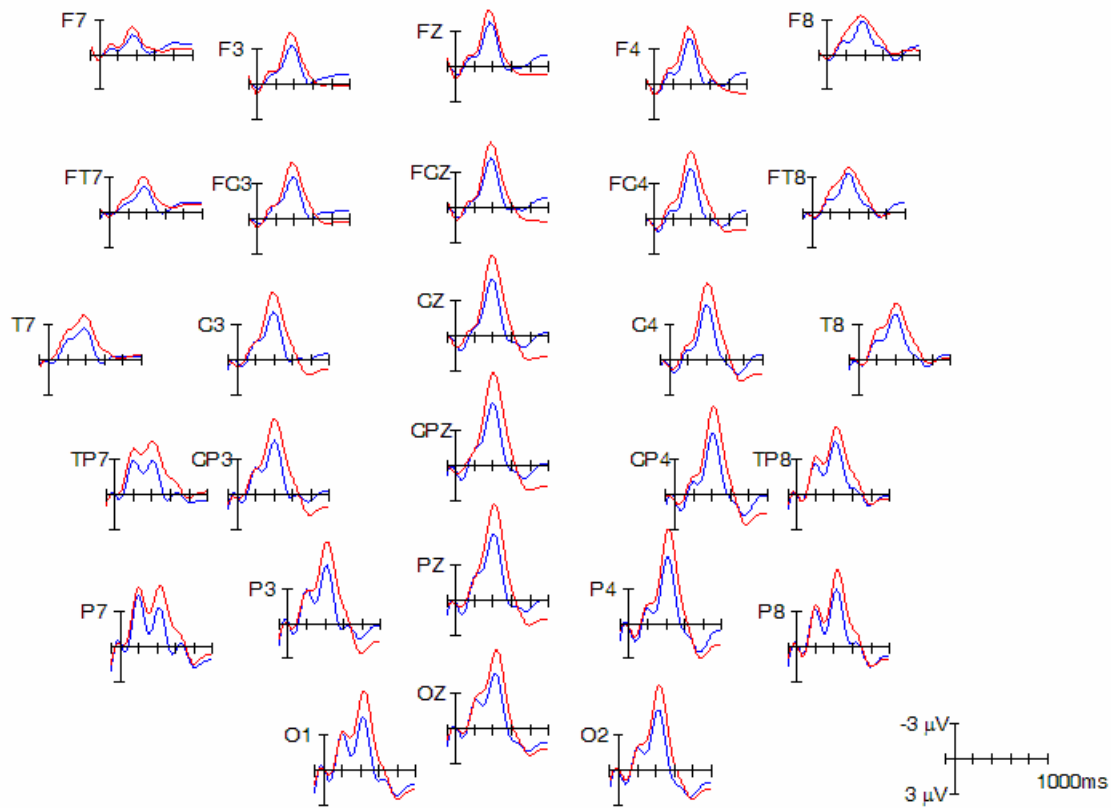


Figure 1. Grand average responses in the felicitous (blue) and infelicitous (red) conditions.

Table 3. Semantic +felicity vs. -felicity: ANOVA *f*-values at the critical adjective

	0-200ms	300-500ms
<i>Overall ANOVA</i>		
<i>cond</i> (1,29)	-	7.11*
<i>cond x lat</i> (1,29)	-	-
<i>cond x post</i> (1,29)	-	4.75*
<i>cond x lat x post</i> (1,29)	-	-
<i>Anterior regions</i>		
<i>left cond</i> (1,29)	-	7.62*
<i>mid cond</i> (1,29)	-	5.34*
<i>right cond</i> (1,29)	-	4.56*
<i>Posterior regions</i>		
<i>left cond</i> (1,29)	-	7.70*
<i>mid cond</i> (1,29)	-	8.39*
<i>right cond</i> (1,29)	-	5.45*

**= <0.01 , *= <0.05 , †= <0.1

3.2.2 Syntactic violation

In the 0-200 ms window, although visual inspection suggests there might be a difference at the central electrodes, there were no significant differences between the conditions. In the 300-500 ms window, there was an interaction of laterality and condition, but this difference was not significant in any individual region. In the P600 window (600-1000 ms), the violation condition was more positive than the control condition. The effect was seen over the entire head, but was largest in the posterior half of the head and along the midline, resulting in interactions between condition and posteriority and between condition and laterality. The effect was significant in all regions of interest.

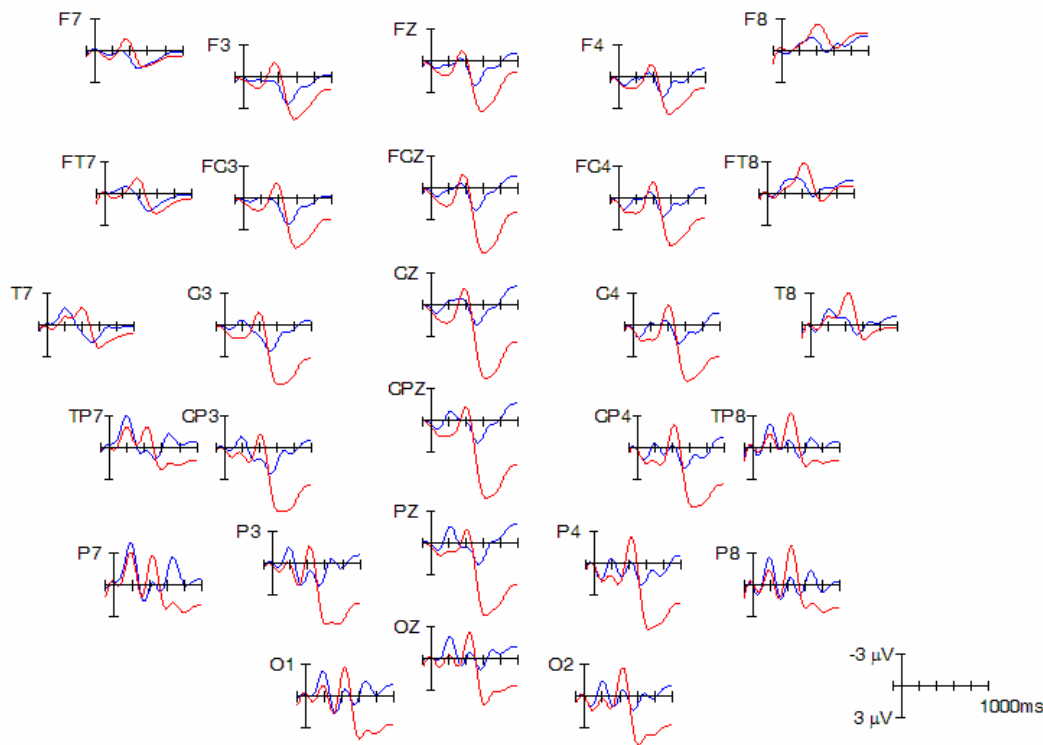


Figure 2. Grand average responses in the grammatical agreement (blue) and mismatch agreement (red) conditions.

Table 4. Syntactic +agreement vs. -agreement: ANOVA *f*-values at the critical noun

	0-200ms	300-500ms	600-1000ms
<i>Overall ANOVA</i>			
<i>cond</i> (1,29)	-	-	23.47**
<i>cond x lat</i> (1,29)	-	4.71*	7.49**
<i>cond x post</i> (1,29)	-	-	9.46**
<i>cond x lat x post</i> (1,29)	-	-	-
<i>Anterior regions</i>			
left <i>cond</i> (1,29)	-	-	15.56**
mid <i>cond</i> (1,29)	-	-	20.19**
right <i>cond</i> (1,29)	-	-	11.56**
<i>Posterior regions</i>			
left <i>cond</i> (1,29)	-	-	29.57**
mid <i>cond</i> (1,29)	-	-	31.99**
right <i>cond</i> (1,29)	-	-	20.60**

**= <0.01 , * = <0.05 , † = <0.1

3.3 Semantic-Attraction Conditions: Pre-critical word

In order to statistically test for pre-existing differences, an ANOVA was performed on the 300-500 ms interval following presentation of the auxiliary, which is the time window immediately preceding the presentation of the critical word.

Fue: Within the three *fue* conditions (grammatical control, semantic-attraction and nonsemantic-attraction), there were no significant effects of condition and no interactions with condition in the 200 ms window immediately preceding the presentation of the critical word.

Estaba: Within the three *estaba* conditions (grammatical control, semantic-attraction and nonsemantic-attraction), there was already an effect of condition ($F(1,29)=5.22$, $p<.05$) in the 200 ms window immediately preceding the presentation of the critical word, as well as an interaction with posteriority ($F(1,29)=5.17$, $p<.05$). This arose because the nonsemantic-attraction condition was more positive than both the control and the semantic-attraction conditions, which were not different from one another. In the ANOVA comparing the nonsemantic-attraction condition to the control condition, there was a large effect of condition ($F(1,29)=9.12$, $p<.01$) which was significant or marginally significant in all the individual regions of interest. This is not, however, likely to be a P600 effect at the auxiliary, since the

divergence begins right after the prestimulus baseline interval and lasts through the entire epoch. It is more likely that this is actually due to a difference at the noun preceding the auxiliary. All three subject NPs (animate for the *estaba* control, semantic-attraction inanimate noun for the *fue* control and both semantic-attraction anomalous conditions, nonsemantic-attraction inanimate noun for both nonsemantic-attraction anomalous conditions) were presented in the context sentence. Given that two of these nouns were chosen to go with a single verb (the animate and semantic-attraction inanimate nouns), and one was chosen because it did *not* go with the verb, it is likely that the context situation made the first two nouns more expected or more salient than the nonsemantic-attraction nouns. This is not a necessity, but simply a subjective observation based on looking at the materials. Therefore, participants may have been more surprised in the cases where the nonsemantic-attraction noun appeared first in the target sentence. If the observation is correct, however, we would also predict a similar effect in the *fue* + nonsemantic-attraction condition, and indeed we find that at the auxiliary (the pre-critical word) the nonsemantic-attraction condition is more positive than the control, although this is not a significant effect. Given that this difference is likely to arise two words before the critical word and arises from surprise rather than anomaly, the effect is likely to be attenuated by the time the critical word is presented. As we will see in the next section, this is supported because the first 0-200 ms of the critical words are very tightly matched and the N400 effect observed is in the opposite direction from the pre-critical word difference. Therefore, we keep this difference at the pre-critical word in mind but it does not exclude analyzing the critical word in the *estaba* + nonsemantic-attraction condition.

3.4 Semantic Attraction Conditions: Critical word

3.4.1 *Fue*

Visual inspection of figure 3 (prestimulus baseline) suggest that there is an N400 effect for the semantic-attraction condition relative to the control, and a smaller N400 effect for the nonsemantic-attraction condition. Figure 4 (300-500 ms baseline) suggests that there is a P600 effect for both anomalous conditions. The P600 for the nonsemantic-attraction condition is also visible with the prestimulus baseline, but the large N400 for the semantic-attraction condition means that the P600, although emerging toward the end of the window, is not obvious without rebaselining.

In the ANOVAs comparing all three *fue* conditions (grammatical control, semantic-attraction, nonsemantic-attraction) in the 300-500 ms window, there was an interaction between condition and posteriority ($F(1,29)=4.45$, $p<.05$). In the window 600-1000 ms with the prestimulus baseline, there was a marginal three-way interaction with condition, laterality and posteriority ($F(1,29)=2.60$, $p=.57$), while with the 300-500 ms baseline, there was a main effect of condition ($F(1,29)=9.17$, $p<.01$), a marginal interaction between condition and posteriority ($F(1,29)=3.03$, $p=.064$), and a three-way interaction between condition, laterality and posteriority ($F(1,29)=4.05$, $p<.01$). Planned comparisons between the control and the semantic-attraction conditions and between the control and the nonsemantic-attraction conditions are presented below the following figures.

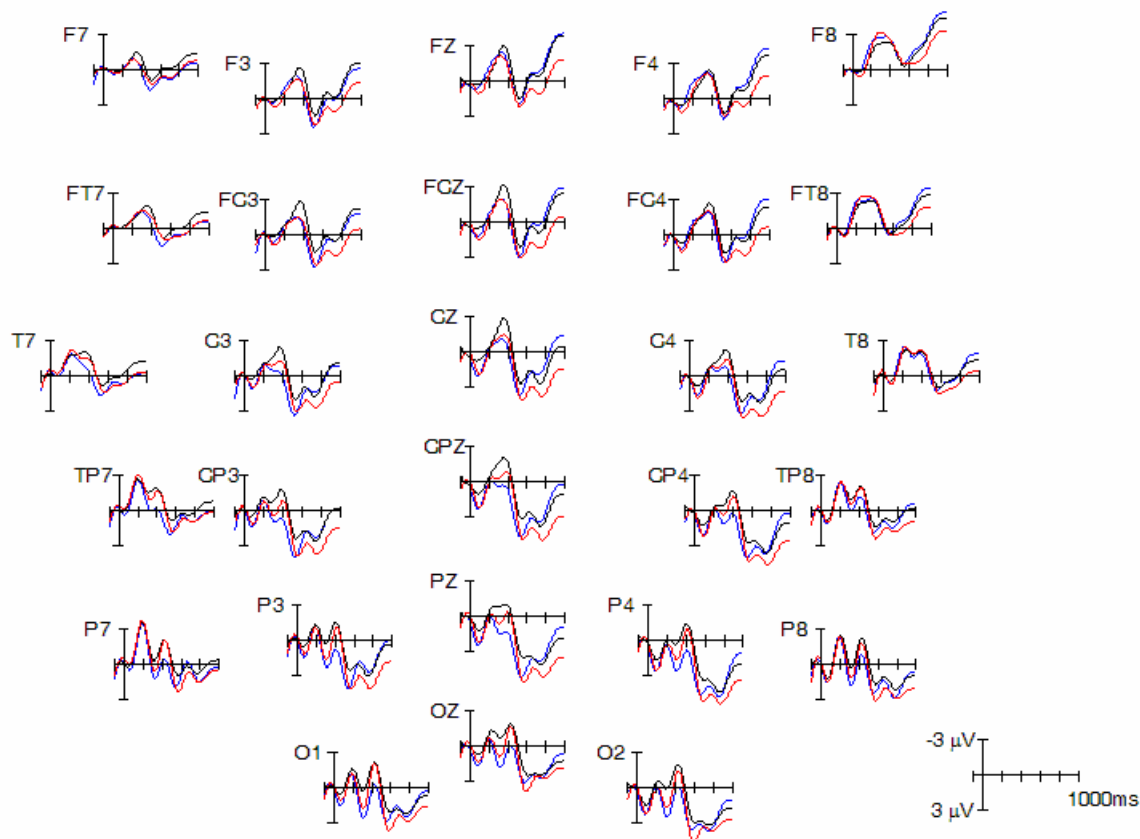


Figure 3. Grand average responses to the *fue* condition using a 100 ms prestimulus baseline, showing the grammatical control (blue), the semantic-attraction (black) and the nonsemantic-attraction (red) conditions.

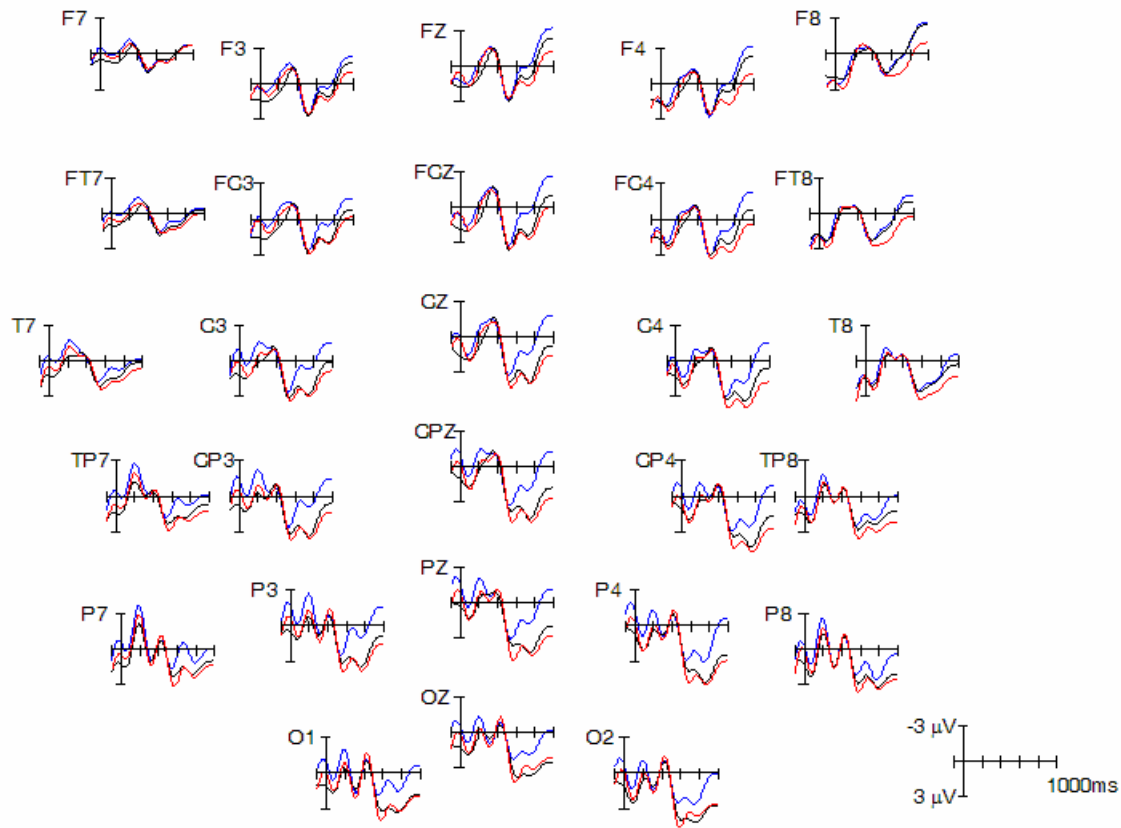


Figure 4. Grand average responses to the *fue* condition using a 300-500 ms baseline, showing the grammatical control (blue), the semantic-attraction (black) and the nonsemantic-attraction (red) conditions.

Semantic-attraction violation: In the 0-200 ms window, there was no difference between the violation condition and the acceptable control. In the N400 window (300-500 ms), the semantic-attraction violation condition was more negative than the acceptable condition particularly in the posterior half of the head. The overall effect of condition was only marginally significant, but there was a significant interaction of condition and posteriority. Looking at individual regions, the effect was significant in the posterior midline region and marginally significant in the left anterior and right posterior regions. In the P600 window (600-1000 ms), prior to rebaselining, the semantic-attraction violation condition was more positive than the acceptable condition, but this was not a significant difference. After rebaselining using the 300-500ms window, the semantic-attraction violation condition was more positive than the acceptable condition and showed a distribution similar to the P600 effect shown to the syntactic violation in the control experiment. The effect of condition was significant, and the

interaction between condition and posteriority was also significant. The effect was significant in all topographical regions, although only marginally significant in the right anterior region.

Table 5. *Fue* control vs. semantic-attraction: ANOVA *f*-values at the critical verb

	0-200ms	300-500ms	600-1000ms (rebaselined)
<i>Overall ANOVA</i>			
<i>cond</i> (1,29)	-	3.11†	10.11**
<i>cond x lat</i> (1,29)	-	-	-
<i>cond x post</i> (1,29)	-	6.75*	6.07*
<i>cond x lat x post</i> (1,29)	-	-	-
<i>Anterior regions</i>			
left <i>cond</i> (1,29)	-	4.03†	7.94*
mid <i>cond</i> (1,29)	-	-	7.56*
right <i>cond</i> (1,29)	-	-	3.18†
<i>Posterior regions</i>			
left <i>cond</i> (1,29)	-	-	17.42**
mid <i>cond</i> (1,29)	-	4.88*	13.42**
right <i>cond</i> (1,29)	-	3.60†	8.38*

**=<0.01, * = <0.05, † = <0.1

Nonsemantic-attraction violation: In the 0-200 ms window, there was no significant difference between the nonsemantic-attraction violation condition and the acceptable control. In the N400 window, the nonsemantic-attraction violation condition was more negative than the acceptable condition. This showed up as an interaction between condition and posteriority, but was only marginally significant in one region (left posterior). In the P600 window, prior to rebaselining, there was a significant interaction of condition x laterality x posteriority ($F(1,30)=4.70$, $p<0.05$) due to the violation condition being more positive than the acceptable condition, but this effect was only marginally significant in one of the regions (anterior midline, $F(1,30)=4.35$, $p=.054$). After rebaselining, the violation condition remained more positive than the acceptable condition. The effect of condition was significant, as was the three-way interaction between condition, laterality and posteriority. The effect was significant in all six regions of interest.

Table 6. *Fue* control vs. nonsemantic-attraction: ANOVA *f*-values at the critical verb

	0-200ms	300-500ms	600-1000ms (rebaselined)
<i>Overall ANOVA</i>			
<i>cond</i> (1,29)	-	-	16.32**
<i>cond x lat</i> (1,29)	-	-	-
<i>cond x post</i> (1,29)	-	5.75*	-
<i>cond x lat x post</i> (1,29)	-	-	6.03**
<i>Anterior regions</i>			
left <i>cond</i> (1,29)	-	-	9.87**
mid <i>cond</i> (1,29)	-	-	10.90**
right <i>cond</i> (1,29)	-	-	14.01**
<i>Posterior regions</i>			
left <i>cond</i> (1,29)	-	3.29†	26.69**
mid <i>cond</i> (1,29)	-	-	14.05**
right <i>cond</i> (1,29)	-	-	14.32**

**=<0.01, * = <0.05, † = <0.1

Discussion: Both the semantic-attraction and the nonsemantic-attraction conditions elicited P600s, in line with Kuperberg et al (2006), although different from Kim & Osterhout (2005) experiment 2, where instead an N400 was elicited to the nonsemantic-attraction condition. In both anomalous conditions + *fue* of the current experiment it also looks like there was an N400, although this is more clearly visible in the semantic-attraction condition. It was not particularly surprising to find an N400 and a P600 since several Dutch studies (Hoeks et al, 2004; Kolk et al, 2003; van Herten et al 2006) also found this combination of components. It was, however, surprising to see a more robust N400 to the semantic-attraction condition than to the nonsemantic-attraction condition, since the Kim & Osterhout (2005a) experiment 2 results led us to expect the reverse effect. Nevertheless, this could possibly be due to component overlap, since the P600 component in the nonsemantic-attraction condition is actually slightly larger than in the semantic-attraction condition. This account is consistent with the Kuperberg et al. (2006) study in which the amplitude of P600 is greater in the nonsemantic-attraction condition (...*the eggs would plant...*) than in the semantic-attraction condition (...*the eggs would eat...*).

3.4.2 *Estaba*

Semantic-attraction violation: In the 0-200 ms window, there was no difference between the violation and control conditions. There was no effect of condition in either the N400 window or the P600 window. Because there was no effect in the N400 window, the data from this condition were not rebaselined using the 300-500 ms window.

Nonsemantic-attraction violation: In the 0-200 ms window, there was no difference between the violation and control conditions. In the 300-500 ms window, visual inspection of the figure suggested that there was an N400 effect, however, this was not significant. There was no significant difference in the P600 window. Because there was no effect in the N400 window, the data from this condition were not rebaselined using the 300-500 ms window.

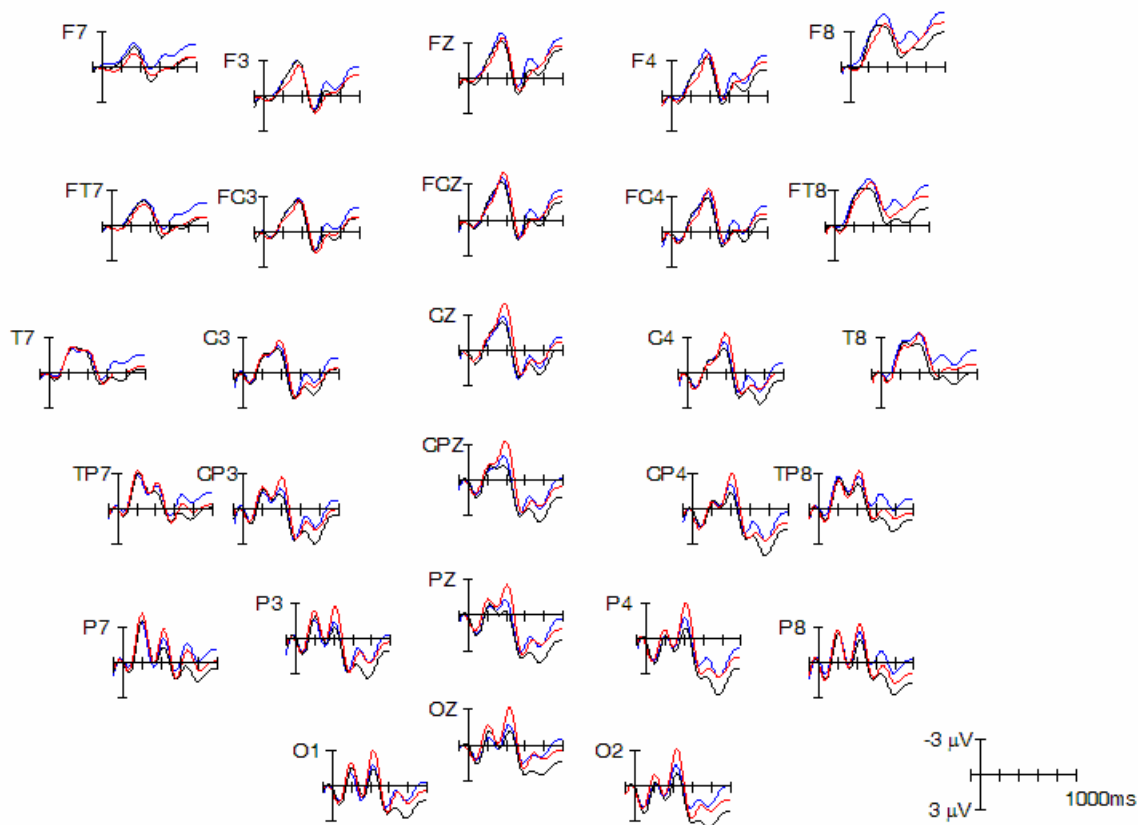


Figure 5. Grand average responses to the *estaba* conditions using a 100 ms prestimulus baseline, showing the grammatical control (blue), the semantic-attraction (black) and the nonsemantic-attraction (red) conditions.

The lack of significant effect in the N400 window, despite the visual pattern, can possibly be explained by the increased variance in this condition relative to the others. The table below shows standard error by condition.

Table 7. Standard error for each condition, including all 18 electrodes used in the ANOVAs.

	<i>Fue</i>	<i>Estaba</i>		<i>Semantic control</i>	<i>Syntactic control</i>
Grammatical control	.72	.58		Grammatical .52	.84
Semantic attraction	.69	.75		Violation .65	.78
Nonsemantic attraction	.80	.94			

Discussion: One thing is very clear from this data: changing the auxiliary changes the results at the following word. With *estaba* there is no evidence of a P600. Therefore, it seems that once participants read *estaba*, which makes a passive structure for the sentence very unlikely, then at the following word their “semantic processor” is no longer considering the first NP to be a possible theme for the verb. Instead, we see only a P600 in the conditions in which the reader has evidence that a passive structure is a possibility. In the case where there is no possibility for a passive structure coming up, we only see some evidence for an N400.

It is quite strange that the semantic-attraction condition with *estaba* showed no effect of the violation; this is the equivalent of *the hearty meal {estaba} devouring*, so we would expect a strong violation response – an N400 according to our hypothesis of more tightly-linked systems, or a P600 if the processing streams really are highly independent. This can possibly be explained by a marginally-acceptable continuation to the sentence:

19. a. El aviso estaba declarando **el bombero**...
 ‘The warning was declaring the fireman...’

This OVS order is possible in Spanish, although in this case it is marginal with a definite article (as was used in all stimuli) and would improve significantly if it began with a demonstrative (e.g., *ese* ‘that’). In this case one might expect to see evidence for detection of a violation at the word following the verb which was distributed equally among three options: a by-phrase (‘by the fireman’), a preposition or other adverb (‘with a megaphone’ or ‘many times’), or the nonsemantic-attraction NP

(‘the apartment’). There was no clear evidence indicating detection of the violation on the following word, but this could be because the following words differed from one another, despite all being short.

3.5 An alternative explanation for the results

Visual inspection of the individual subjects’ data suggested that in the *fue* conditions, there were actually 2 groups of subjects: those with a P600 to both semantic-attraction and nonsemantic-attraction conditions, and those who showed an N400 to both semantic-attraction and nonsemantic-attraction conditions. This would be similar to findings from Osterhout (1997) in which he concluded that although syntactically anomalous open-class words elicited a P600 in most participants, there was a smaller group which responded to these anomalies with N400s. If this is truly the case, it would help account for why the N400 and P600 in the grand average results, although present in the figures, were mostly only marginally significant (using the prestimulus baseline). If true, this would also imply that some participants behaved more like those from Kuperberg et al. (2006) who showed a P600 to both semantic-attraction and nonsemantic-attraction, and other participants show the results that we would have expected prior to this body of work, i.e., they show an N400 to sentences with unambiguous syntax but anomalous interpretation. This is by no means a completed analysis, therefore I will go through and show the evidence for why we might believe that there are actually two groups of participants, but I will also point out reasons why we might be skeptical of this analysis.

Taking the P600 window (600-1000 ms), I took the difference of the *fue* + semantic-attraction condition and the *fue* control condition. Any participant that showed a positive difference (i.e., the semantic-attraction condition was more positive than the control) was placed in one group, and the rest of the participants were placed in the other group. 3 participants were not easily classified and therefore are excluded from this section of the analysis. This process yielded 4 participants in the “P600 group” and 9 in the “N400 group”. The figures are shown below:

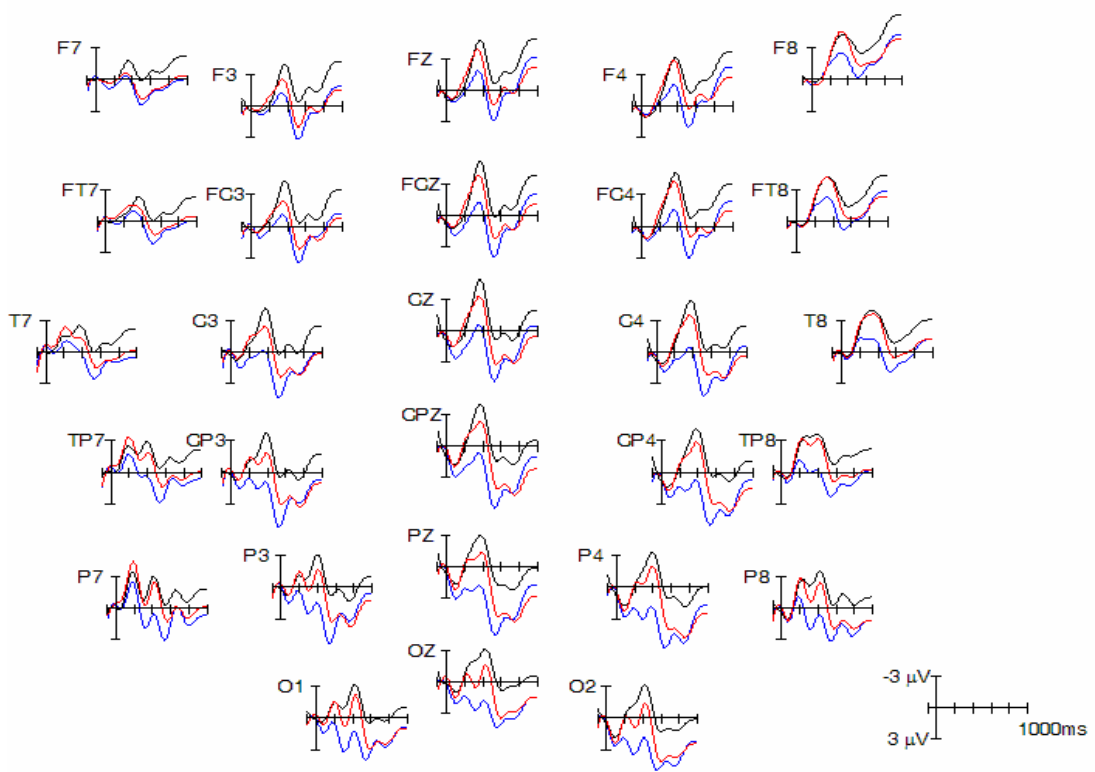
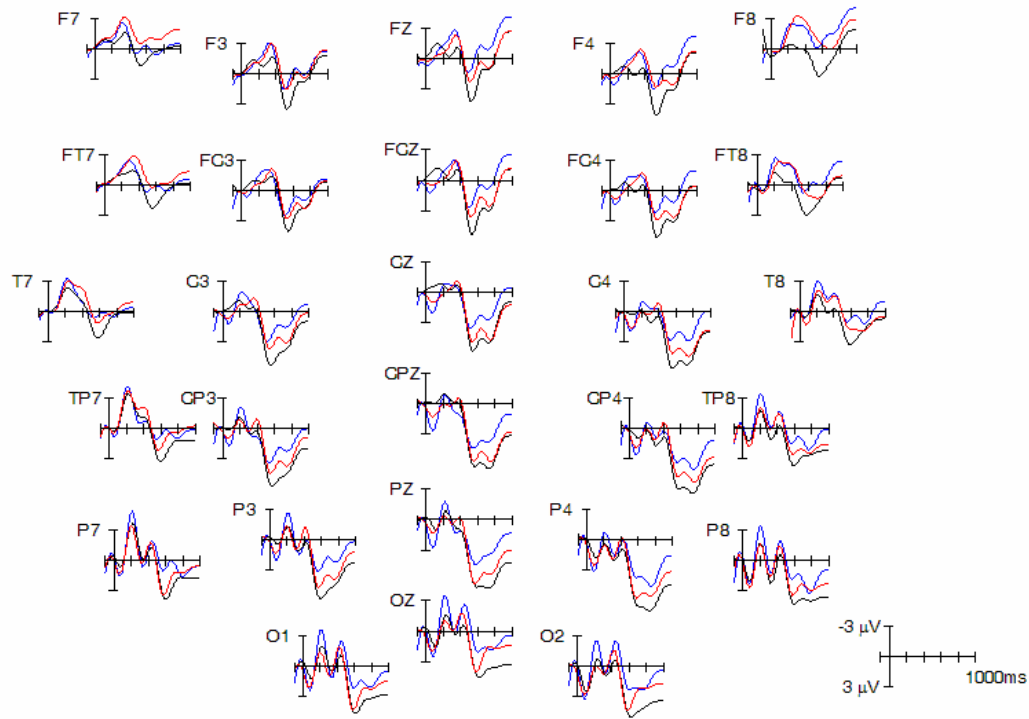


Figure 6. Grand average responses in the *fue* conditions using a 100 ms prestimulus baseline, showing the grammatical control (blue), the semantic-attraction (black) and the nonsemantic-attraction (red) conditions. **Top:** 4 participants showing P600s. **Bottom:** 9 participants showing N400s.

Looking at these figures, it is clear why in the grand average we see N400s, but small ones, and P600s, but also small ones. There are a number of reasons why this pattern does not seem coincidental. First, the division into two groups was made based only on the difference between the semantic-attraction condition and the control condition in the 600-1000 ms window. The figures indicate that this division generalizes to the nonsemantic-attraction condition; those participants who showed a P600 to the semantic-attraction condition also show a P600 to the nonsemantic-attraction condition. The distribution of the effect is largest in the posterior half of the head and along the midline. This is the same distribution as is seen for the P600 effect for the syntactic control experiment. If the division were purely coincidental, we would not necessarily expect the waveform pattern to generalize to the other anomalous condition (but not the control condition), nor would we expect it to have the distribution of a standard P600.

Secondly, the division into two groups took place without even looking at the 300-500 ms window. Nevertheless, those participants that showed a P600 *did not* show an N400 effect. And, the group not showing a P600 *did* show an N400 effect. The N400 effect is distributed over the whole head, although is smallest in the most lateral anterior electrodes, just as in the N400 effect in the semantic control experiment. If the division into two groups were purely coincidental, we would not necessarily expect to see the complementarity between the two components, nor would we expect the effect to have the distribution of a standard N400.

As a final piece of evidence in favor of this division, the four participants in the “P600 group” showed a standard N400 effect in the semantic control experiment, with no difference in magnitude or in distribution relative to the other group. Likewise, the nine participants in the “N400 group” showed a standard P600 effect in the syntactic control experiment, also with no difference in magnitude or distribution relative to the other group. This suggests that we are not simply dealing with participants who, for whatever reason, do not show one of the components.

However, there are a few reasons to think that this may not be a real division. Firstly, it is clear that one group is larger than the other. In fact, it is the smaller group that gave results consistent with the recent body of work (Kim & Osterhout, 2005; Kuperberg, 2007). This does not mean that the division is not real, it just would have been easier to make the case if the division into groups had been more equal.

The next obvious question is: how does the division into groups impact the *estaba* conditions? Recall that we predicted N400s for both anomalous conditions with *estaba* because the auxiliary makes the passive sentence extremely unlikely. We predicted that we would only see a P600 here if the semantic processor (or thematic role processor) is forming its interpretation without making reference to *prior* syntactic information that restricts the semantic interpretations that would be licensed.

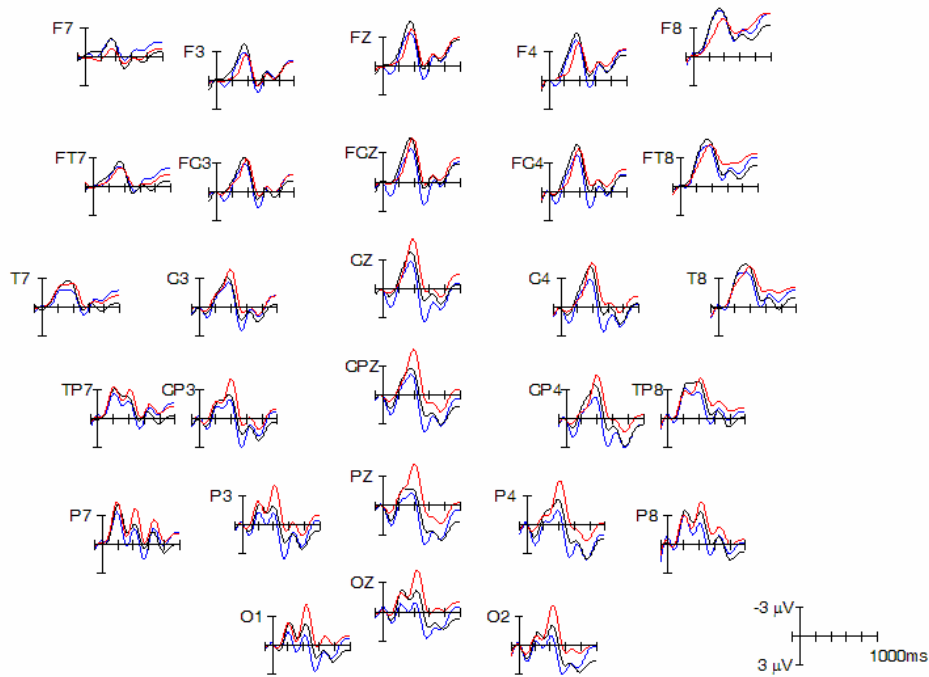
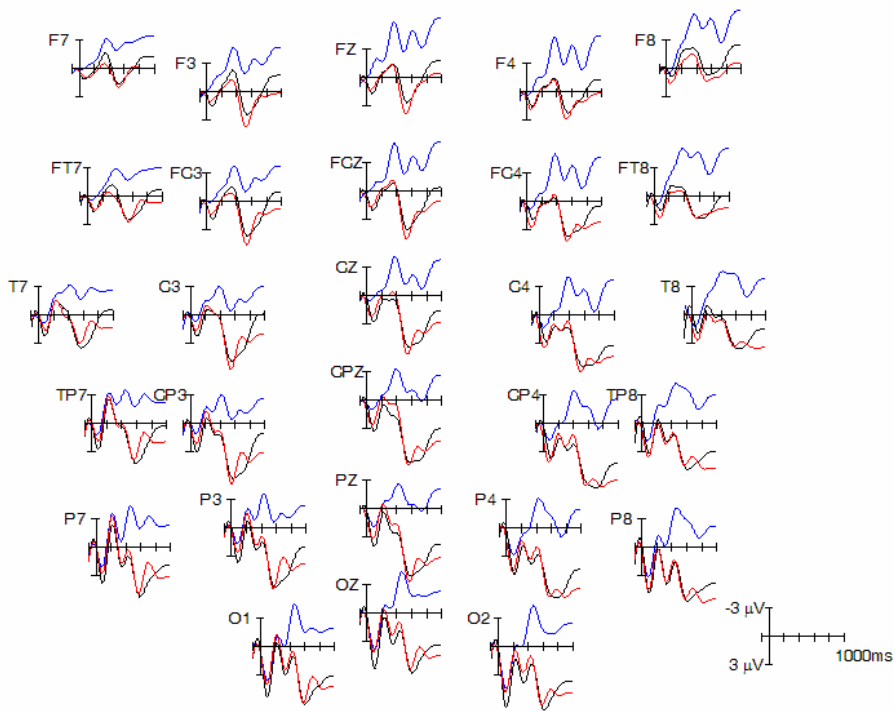


Figure 7. Grand average responses in the *estaba* conditions using a 100 ms prestimulus baseline, showing the grammatical control (blue), the semantic-attraction (black) and the nonsemantic-attraction (red) conditions. **Top:** 4 participants showing P600s in the *fue* conditions. **Bottom:** 9 participants showing N400s in the *fue* conditions.

Starting with the participants who showed the N400 in the *fue* condition, we see they behave as expected here too. We see an N400 to both the semantic-attraction and the nonsemantic-attraction condition. Their relative amplitudes are reversed here: in the *fue* conditions, the semantic-attraction condition had a larger N400, while in the *estaba* conditions the nonsemantic-attraction condition had the larger N400. Comparing these nine participants' results to the grand average results presented in section 3.4.2, it is a relief to actually see evidence for anomaly detection in the *estaba* + semantic-attraction condition.

However, the results for the nine participants who showed a P600 for the *fue* conditions are not so clear. We had expected an N400 to both anomaly conditions with *estaba*. What we find is something that sort of looks like a P600, except it starts very early. Over the whole head, the divergence between anomalous and grammatical begins by 200 ms, and in the anterior portion of the head the divergence actually begins at 0 ms. There are a number of different explanations we could make for this. First, it could be that these four participants for whatever reason showed more positivity overall in the recording. This is unlikely to be the whole story because it is only the control sentence that differs while the two anomalous sentences are closely matched, and these very subjects showed good N400s in the semantic control experiment. Another possibility is that we could conclude that our hypothesis was wrong and in fact these participants are interpreting the inanimate NP as theme of a passive sentence, despite evidence from the auxiliary against this analysis, and therefore we are seeing a real P600. However, this argument makes it hard to explain why the divergence is so early. The last possibility is that the effect seen at the critical verb for these four participants in the *estaba* conditions is a spill-over from the previous word (i.e., the auxiliary). In other words, perhaps there is some problem with the combination of an inanimate NP plus *estaba*, or simply that the processing of these words is different than for an animate NP plus *estaba*. Visual inspection of the pre-critical word suggests that there are problems with the baseline for these four participants in the *estaba* conditions, however it is difficult to make definitive conclusions based on only four participants.

In summary, there is reasonable evidence to suggest that the participant group is not homogenous and that it is made up of two groups, one that responded to the *fue* anomalous conditions with a P600 and another that responded to these conditions with an N400. When looking at the *estaba* conditions according to these two groups, the results clearly show an N400 for the "N400 group", but are not so clear for the "P600 group".

Discussion: If the participants really do form two different groups in the response elicited to the *due* conditions, then that is very interesting. For one thing it helps us understand why some groups have found only P600s to this kind of stimulus (Kuperberg et al., 2006), while others have found both an N400 and a P600 (Hoeks et al., 2004; Kolk et al., 2003; van Herten et al., 2006), although it does less to explain why Kim & Osterhout (2005) found a P600 for the semantic-attraction condition but an N400 for the nonsemantic-attraction condition.

The implications, however, are actually far more interesting because it seems to speak in favor of a multiple-cue model. It could be that the “P600 group” is more reliant in these cases on semantic information, while the “N400 group” is more reliant on syntactic information. It is interesting to note that the “P600 group” have a higher accuracy (92%) than the “N400 group” (84%) on the main experimental conditions, although the difference is much lower on the semantic control experiment (6% difference) and even lower on the syntactic control experiment (2% difference).

4. General discussion

Starting with the general question of when we see a P600 to sentences with semantic anomalies but unambiguous syntax, Kuperberg’s (2007) review paper outlines a few factors that seem to increase the likelihood of seeing a P600. This experiment included many of these factors and yet the P600 effect is fairly weak – if we take the grand average analysis we need to rebaseline to see the P600, while if we take the two-groups analysis it is the group with only four members that shows the P600. So, it is surprising that we do not see a more robust P600 effect. The factors that were included in this experiment that seem to increase the likelihood of seeing a P600 were: an animacy mismatch between the likely agent and the NP that actually appears, additional contextual information, and an explicit acceptability judgment task.

A significant difference between this experiment and the others is the concreteness of the nouns and verbs used. In most other experiments, most nouns were concrete objects while many verbs were more concrete actions (e.g., *devour*). Because of restrictions in the kind of passive sentences that are felicitous in Spanish, about half of the nouns and verbs were less concrete (e.g., *govern the country*, *fulfill the agreement*, etc.). It is possible that non-concrete nouns in English would also be less likely to elicit the P600.

These experimental results also serve to back up the generalization pointed out by Kuperberg (2007), that the lack of semantic attraction does not insure that the P600 will not occur, as Kim & Osterhout (2005) had claimed. In these results, there is no indication that the semantic-attraction and nonsemantic-attraction do not pattern alike, at least as far as the type of component elicited. Looking within the “N400 group” we saw that the relative amplitude of the N400 effects reversed from *fue* to *estaba*, but with just nine participants in that group it would be premature to conclude anything from an amplitude change.

The original question was whether the auxiliary manipulation could restrict the interpretation that the semantic processor considers. Looking at the grand average data, it is clear that the auxiliary affected the processing of the verb. With *fue*, there was evidence of a P600, indicating that the analysis of the NP as the theme for the verb was being considered. With *estaba*, there was no evidence for a P600, suggesting that the analysis of the NP as the theme for the verb was not considered. Looking at the data by splitting it into two groups, we do not have evidence against this view, at least for one of the groups. The larger “N400 group” showed an N400 whether the auxiliary was *fue* or *estaba*. Therefore, we cannot see whether *estaba* blocks the interpretation of the NP as the theme, since there is no evidence that this interpretation is being considered in either condition in this group of participants. The “P600 group” participants did seem to be considering this interpretation in the *fue* conditions, but because of the small sample and the noise in the *estaba* anomalous conditions, it is difficult to tell whether the interpretation is being blocked here. If the pre-critical word and baseline problems with these *estaba* + anomaly conditions were not present with the “P600 group” we would be forced to say that the participants show a P600 in both *fue* and *estaba* conditions, and therefore it might seem that prior information about syntactic structure does not restrict interpretations proposed by the semantic (or thematic role) processor.

Based on the grand average analysis, the data support the hypothesis that the information carried by the auxiliary that helps restrict the possible syntactic structures of the sentence does affect the interpretations that the semantic processor makes. Specifically, in the case where the passive is ruled out at the word prior to the verb, the semantic processor no longer seems to propose the NP as possible theme for the verb, and therefore the sentence looks semantically anomalous and we see an N400. This is compatible with a model in which incoming semantic and syntactic information are treated separately but then the representations are reconciled after each word, or with a model in which multiple cues assist with adding the incoming word to a common representation. It is incompatible

with a model in which the semantic and syntactic information are processed in parallel and independently, only reconciling after a larger section of the sentence has passed.

Looking at the two-group analysis of the data, because of the lack of clarity with the *estaba* anomaly conditions, the results are consistent with either one of the models described in the paragraph above, or with a model with a larger degree of parallelism and independence between semantic and syntactic processing. However, the fact that the participants divide into two groups seems more in line with a multiple cues model. If many different cues are competing for influence on the incoming structure, it is reasonable to think that we may see a fair degree of heterogeneity in the results.

5. Conclusion

The results of the experiment support the prediction that the auxiliary can be used to provide advance restrictions on the upcoming syntactic structure, and this knowledge can prevent the semantic processor from proposing interpretations that are incompatible with the syntax. However, the lack of homogeneity among the participants suggests that further investigation is needed as to whether these results actually disguise a more complex situation that may be best explained using a multiple-cues model. The clean results from the two control experiments underline the fact that it is the results from this particular area of research that are complicated, rather than messy data in general.

Acknowledgments

This work was supported in part by grants to Colin Phillips from the National Science Foundation (BCS-0196004), the Human Frontier Science Program (RGY-0134), and the James S. McDonnell Foundation (#99-31T). I was supported by fellowships from the Social Sciences and Humanities Research Council of Canada and Alberta Advanced Education. I would like to thank Antonia Carrasco, Laura Fuentes, Ivan Ortega Santos and Leticia Pablos for their help in creating the materials, and Brian Dillon, Ellen Lau and Ming Xiang for assistance in data acquisition and analysis.

References

- Allen, M., Badecker, B. & Osterhout, L. (2003). Morphological analysis in sentence processing: and ERP study. *Language and Cognitive Processes, 18*, 405-430.
- Barber, H. & Carreiras, M. (2003). Integrating gender and number information in Spanish word pairs: An ERP study. *Cortex, 39*, 465-482.
- Barber, H. & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP Comparison. *Journal of Cognitive Neuroscience, 17*, 137-153.
- Barber, H., Vergara, M. & Carreiras, M. (2004). Syllable-frequency effects in visual word recognition: evidence from ERPs. *Neuroreport, 15*, 545-548.
- Beltrán, D. Carreiras, M., Álvarez, M. & Santamaría, C. (2006). Verification of quantified sentences: An ERP study. *Poster presented at the Cognitive Neuroscience Society annual meeting, San Francisco.*
- Bentin, S., McCarthy, G. & Wood, C.C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology, 60*, 343-355.
- Bever, T.G. (1970). The cognitive basis for linguistic structures. In J.R. Hayes (Ed.), *Cognition and the development of language* (pp. 279-362). New York: Wiley.
- Drenhaus, H., Graben, P. beim & Frisch, S. (2006). Not all but some ERP results on the scalar expressions 'some' and 'all'. *Poster presented at the Cognitive Neuroscience Society annual meeting, San Francisco.*
- Fiebach, C., Schlesewsky, M., & Friederici, A. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language, 47*, 250-272.
- Fischler, I., Bloom, P., Childers, D.G., Roucos, S.E. & Perry, N.W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology, 20*, 400-409.

- Friederici, A. D. & Frisch, S. (2000). Verb argument structure processing: The role of verb-specific and argument-specific information. *Journal of Memory and Language*, 43, 476-507.
- Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). The temporal structure of syntactic parsing: Early vs. late effects elicited by syntactic anomalies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1219-1248.
- Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological, and syntactic violations. *Cognitive Brain Research*, 1, 183-192.
- Frisch, S., Schlesewsky, M., Saddy, D., & Alpermann, A. (2002). The P600 as an indicator of syntactic ambiguity. *Cognition*, 85, B83-B92.
- Geyer, A., Holcomb, P., Kuperberg, G. & Perlmutter, N. (2006). Plausibility and sentence comprehension. An ERP study. *Cognitive Neuroscience Supplement, Abstract*.
- Greenhouse, S. & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95-112.
- Hagoort, P. (2003) Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, 15, 883-899.
- Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The Syntactic Positive Shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8, 439-484.
- Hagoort, P., Hald, L., Bastiaansen, M.C.M. & Petersson, K.M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304, 438-440.

- Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, *11*, 193-204.
- Hoeks, J.C.J., Stowe, L.A. & Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, *19*, 59-73.
- Holcomb, P.J. (1988). Automatic and attentional processing: an event-related brain potential analysis of semantic priming. *Brain and Language*, *35*, 66-85.
- Holcomb, P.J. (1993). Semantic priming and stimulus degradation: implications for the role of the N400 in language processing. *Psychophysiology*, *30*, 47-61.
- Kaan, E. (2002). Investigating the effects of distance and number interference in processing subject-verb dependencies: An ERP study. *Journal of Psycholinguistic Research*, *31*, 165-193.
- Kaan, E., Harris, A., Gibson, E. & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language & Cognitive Processes*, *15*, 159-201.
- Kaan, E., & Swaab, T. Y. (2003). Electrophysiological evidence for serial sentence processing: A comparison between non-preferred and ungrammatical continuations. *Cognitive Brain Research*, *17*, 621-635.
- Kim, A. & Osterhout, L. (2005a). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*, 205-225.
- Kim, A. & Osterhout, L. (2005b). The modulation of “semantic attraction” P600 effects by syntactic factors during sentence comprehension. Poster at Amlap, Sept. 7.
- Kolk, H.H.J. & Chwilla, D.J. (2007). Late positivities in unusual situations: a commentary to (a) Kuperberg, Kreher, Sitnikova, Caplan and Holcomb and (b) Kremmerer, Weber-Fox, Price, Zdanczyk and Way. *Brain and Language*, *100*, 257-262.

Kolk, H.H.J., Chwilla, D.J., Van Herten, M. & Oor, P. (2003). Structure and limited capacity in verbal working memory: a study with event-related potentials. *Brain and Language*, 85, 1-36.

Kounios, J. & Holcomb, P.J. (1992). Structure and process in semantic memory: evidence from event-related brain potentials and reaction times. *Journal of Experimental Psychology: General*, 121, 459-479.

Kuperberg, G.K. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146, 23-49.

Kuperberg, G.R., Sitnikova, T., Caplan, P. & Holcomb, P.J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, 17, 117-129.

Kuperberg, G., Caplan D., Sitnikova, T., Eddy, M. & Holcomb, P. (2006). Neural correlates of processing syntactic, semantic and thematic relationships in sentences. *Language and Cognitive Processes*, 21, 489-530.

Kuperberg, G., Kreher, D.A., Sitnikova, T., Caplan, D. & Holcomb, P. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain and Language*, 100, 223-238.

Kutas, M. & Federmeier, K.D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends on Cognitive Science*, 4, 463-470.

Kutas, M. & Hillyard, S.A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, 207, 203-205.

Kutas, M. & Hillyard, S.A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory and Cognition*, 11, 539-550.

Kutas, M. & Hillyard, S.A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161-163.

Kutas, M., Van Petten, C. & Besson, M. (1988) Event-related potential asymmetries during the reading of sentences. *Electroencephalography and Clinical Neurophysiology*, 69, 218-233.

MacDonald, M.C., Pearlmutter, N.J., Seidenberg, M.S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676-703.

Martín-Loeches, M., Nigbur, R., Casado, P., Hohlfeld, A. & Sommer, W. (2006). Semantics prevalence over syntax during sentence processing: A brain potential study of noun-adjective agreement in Spanish. *Brain Research*, 1093, 178-189.

Nieuwland, M.S. & van Berkum, J.J. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Brain Research: Cognitive Brain Research*, 24, 691-701.

Noveck, I.A. & Posada, A. (2003). Characterizing the time course of an implicature: an evoked potentials study. *Brain and Language*, 85, 203-210.

Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language*, 59, 494-522.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785-806.

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden path sentences: evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 20, 768-803.

Osterhout, L., Kim, A. & Kuperberg, G.R. (2007). The neurobiology of sentence comprehension. To appear in M. Spivey, M. Joanisse & K. McRae (Eds.) *The Cambridge Handbook of Psycholinguistics*, Vol. Cambridge University Press, Cambridge.

Paczynski, M., Kreher, D.A., Ditman, T., Holcomb, P. & Kuperberg, G.R. (2006).

Electrophysiological evidence for the role of animacy and lexico-semantic associations in processing nouns within passive structures. *Cognitive Neuroscience Supplement, Abstract*.

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, 22, 407-428.

Smith & Halgren 1987

Tanenhaus, M.K., Spivey-Knowlton, M.J. & Eberhard, K.M. & Sedivy, J.C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268. 1632-1634.

Townsend, D.J. & Bever, T.G. (2001). *Sentence Comprehension: The Integration of Habits and Rules*. Cambridge, MA: MIT Press.

Van Berkum, J.J.A., Hagoort, P. & Brown, C.M. (1999). Semantic integration in sentences and discourse: evidence from the N400. *Journal of Cognitive Neuroscience*, 11, 657-671.

Van Herten, M., Chwilla, D.J. & Kolk, H.H. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience*, 18, 1181-1197.

Van Herten, M., Kolk, H.H.J. & Chwilla, D. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22, 241-255.

Van Petten, C. & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related potentials. *Memory and Cognition*, 18, 380-393.

West, W.C. & Holcomb, P.J. (2000). Imaginal, semantic and surface-level processing of concrete and abstract words: An electrophysiological investigation. *Journal of Cognitive Neuroscience*, 12, 1024-1037.

Wicha, N.Y.Y. (2003) Potato not pope: human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*, 346: 165-168.

Wicha, N.Y.Y., Moreno, E.M. & Kutas, M. (2003). Expecting gender: An event-related potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in Spanish. *Cortex*, 39, 483-508.