Getting real about Semantic Illusions: Rethinking the functional role of the P600 in language comprehension

Harm Brouwer[a,*], Hartmut Fitz[b], John Hoeks[a]

[a] *Center for Language and Cognition / BCN Neuro-Imaging Center*
University of Groningen
[b] *Max Planck Institute for Psycholinguistics, Nijmegen*

Abstract

In traditional theories of language comprehension, syntactic and semantic processing are inextricably linked. This assumption has been challenged by the 'Semantic Illusion Effect' found in studies using Event Related brain Potentials. Semantically anomalous sentences did not produce the expected increase in N400 amplitude but rather one in P600 amplitude. To explain these findings, complex models have been devised in which an independent semantic processing stream can arrive at a sentence interpretation that may differ from the interpretation prescribed by the syntactic structure of the sentence. We review five such multi-stream models and argue that they do not account for the full range of relevant results because they assume that the amplitude of the N400 indexes some form of semantic integration. Based

---

[*]Corresponding author:
Center for Language and Cognition / BCN Neuro-Imaging Center
University of Groningen, Faculty of Arts
PO Box 716, 9700 AS Groningen, The Netherlands
Phone: +31 (0) 50 363 6683
Fax: +31 (0) 50 363 6855

Email addresses: `harm.brouwer@rug.nl` (Harm Brouwer), `hartmut.fitz@mpi.nl` (Hartmut Fitz), `j.c.j.hoeks@rug.nl` (John Hoeks)

on recent evidence we argue that N400 amplitude might reflect the retrieval of lexical information from memory. On this view, the absence of an N400-effect in Semantic Illusion sentences can be explained in terms of priming. Furthermore, we suggest that semantic integration, which has previously been linked to the N400 component, might be reflected in the P600 instead. When combined, these functional interpretations result in a single-stream account of language processing that can explain all of the Semantic Illusion data.

1. Introduction

As a sentence or story unfolds in time, a language user incrementally constructs an interpretation of the linguistic input. Creating this interpretation draws on various different information sources such as syntactic, semantic, pragmatic, prosodic, and visual information, as well as world knowledge. Exactly when and how these different types of information are combined is a matter of debate. Many theories of language comprehension claim that there is a tight coupling between syntactic and semantic processing (e.g., Frazier, 1987; MacDonald et al., 1994; McRae et al., 1998; van Gompel et al., 2000). According to these theories, it would be impossible to construct an interpretation of language input without immediately taking syntactic information into account. This assumption has been challenged by recent evidence from Event Related brain Potentials (ERPs). Hoeks et al. (2004) for example, studied Dutch sentences in which two plausible verb arguments appeared in

a semantically anomalous order, as in 'De speer heeft de atleten geworpen' (lit: The javelin has the athletes thrown). Relative to a control sentence 'De speer werd door de atleten geworpen' (lit: The javelin was by the athletes thrown), no shift in N400 amplitude (a negative deflection of the ERP signal peaking at about 400ms after the onset of a critical stimulus) was found. This was surprising because the amplitude of the N400 has been associated with difficulty in *semantic integration* (see Kutas and Federmeier, 2011, for an overview). Instead, Hoeks et al. (2004) found that the sentence-final verb *thrown* produced a P600-effect (a positive deflection of the ERP signal that reaches maximum around 600ms post stimulus onset) relative to control. Again, this was unexpected since P600 amplitude has been linked with syntactic revision (see Gouvea et al., 2010, for an overview) but the test items were perfectly grammatical. This phenomenon in which a semantically anomalous, syntactically well-formed sentence elicits a P600-effect, but no N400-effect, has been called a 'Semantic Illusion'[1]. This is because the absence of an N400-effect suggested that participants were temporarily under the illusion that these sentences made sense. The presence of a P600-effect, on the other hand, indicated that participants eventually realized that their interpretations were infelicitous, and that they were trying to resolve this conflict through 'effortful syntactic processing' (Hoeks et al., 2004, p. 71).

To account for Semantic Illusions, so-called *multi-stream* models have been proposed in which a separate semantic analyzer can put forward an

---

[1]The term 'Semantic Illusion' was adapted from a study by Erickson and Mattson (1981). Others have labeled the phenomenon at hand a 'semantic P600-effect', stressing the non-syntactic nature of the evoked late positivities in these materials.

interpretation of a sentence that may not be in line with its surface structure (Kim and Osterhout, 2005; Kolk et al., 2003; van Herten et al., 2005, 2006; Kuperberg, 2007; Bornkessel-Schlesewsky and Schlesewsky, 2008; Hagoort et al., 2009; Kos et al., 2010). In this paper, we critically review five *multi-stream* architectures and ask whether they can account for the available data on the Semantic Illusion. The review suggests that none of them can explain the full range of relevant results. We argue that this is the case because all of them adopt the view that the amplitude of the N400 indexes the relative difficulty of integrating the meaning of an incoming word into a partial interpretation of a sentence. In contrast, we show that most evidence is consistent with an alternative interpretation of N400-effects in which N400 amplitude reflects the *retrieval* of the meaning of a word from long-term memory (Kutas and Federmeier, 2000; Lau et al., 2008, 2009; van Berkum, 2009; Federmeier and Laszlo, 2009; van Berkum, 2010; Kutas and Federmeier, 2011). On this view, the absence of an N400-effect in Semantic Illusion sentences follows directly from the workings of the memory system. Crucially, the processes reflected in N400 amplitude are thus not integrative or compositional in nature. We suggest that the process of integration, in which a mental representation of what is being communicated is updated with new information, is reflected in the amplitude of the P600 instead.

Combining this novel view on P600 amplitude with the retrieval perspective on N400 amplitude leads to a parsimonious *single-stream* account that can explain the full range of Semantic Illusion data. On this account, sentences like 'De speer heeft de atleten geworpen' (lit: The javelin has the athletes thrown) do not give rise to a Semantic Illusion. Language users

do not perceive such sentences as if they made sense, not even temporarily. There is no N400-effect because participants do not have problems activating the meaning of the anomalous word. A P600-effect is observed instead because the situation or event that is described is difficult to interpret.

2. Explaining the Semantic Illusion phenomenon

The results of Hoeks et al. (2004) (as well as similar results of Kuperberg et al., 2003; Kolk et al., 2003) raised two critical questions: 1) Why did the thematic violation brought about by the critical verb fail to give rise to an N400-effect effect? and 2) Why did the critical verb produce a P600-effect although these sentences were syntactically well-formed?

According to Hoeks et al., the absence of an N400-effect suggested that participants had been tricked into some kind of a 'Semantic Illusion', leading them to believe that the sentences made perfect sense. They argued that this illusion could arise because of difficulty in creating a coherent interpretation of the sentence fragment preceding the main verb (lit: 'The javelin has the athletes'). Due to problems with assigning the correct thematic roles, the interpretation of this fragment was not specific enough to immediately affect integration of the critical verb *thrown*. The subsequent P600-effect then indicated that this illusion lasted only for a couple of hundred milliseconds, as the reader quickly realized that something was wrong with the interpretation, and engaged in 'effortful syntactic processing' in an attempt to revise it. On this account, however, it is unclear what exactly triggered these revision processes, as it seems paradoxical for the processor to initiate syntactic processing to revise an 'implausible' interpretation that was just constructed

as if it made perfect sense. Hoeks et al. (2004) did not suggest a solution for this 'paradox'.

2.1. Multi-stream models

2.1.1. Semantic Attraction

Kim and Osterhout (2005) observed a Semantic Illusion Effect (henceforth SIE) in response to an animacy-based thematic role violation. They failed to find an N400-effect but did find a P600-effect in sentences such as 'The hearty meal was devouring...' relative to the non-anomalous 'The hearty meal was devoured...'. In these sentences, the fragments preceding the main verb are compatible with, for instance, a passive analysis, and hence do not seem to cause problems in thematic role assignment. It is therefore doubtful whether the thematic role-based explanation proposed by Hoeks et al. (2004) is key to the Semantic Illusion. Kim and Osterhout suggested that these sentences exhibited strong 'semantic attraction' (high activation of well-established semantic relationships between the argument and the verb). They also looked at sentences in which there was no strong semantic attraction. In contrast to the strong attraction sentences ('The hearty meal was devouring...'), the no-attraction items ('The dusty tabletops were devouring ...') did produce an N400-effect, and no P600-effect, relative to 'The hearty meal was devoured...'. Based on these results, they argued that the semantic attraction between a verb (*devoured*) and its argument (*the hearty meal*) can be so strong that a syntactically well-formed sentence is perceived as ungrammatical. Thus *devouring* in 'The hearty meal was devouring...' may have been interpreted as a wrong inflection of the intended past participle *devoured*, triggering a P600-effect. Kim and Osterhout concluded that in case

6

of strong semantic attraction, semantics can override syntax during on-line comprehension. They argued that this is strong support for a model in which syntax and semantics are being processed autonomously and can give rise to different interpretations of a sentence that are maintained concurrently (see also Martín-Loeches et al., 2006; Kim and Sikos, 2011).

The results of a study by van Herten et al. (2005), however, provide a difficult case for the Semantic Attraction account. They observed an SIE in response to relative clauses containing a semantic anomaly: 'De vos die op de stroper joeg...' (lit: The fox that on the poacher hunted...) relative to 'De stroper die op de vos joeg...' (lit: The poacher that on the fox hunted...). In these sentences, the anomaly only becomes apparent through a violation of world knowledge and not through animacy violations as in Hoeks et al. (2004) and Kim and Osterhout (2005). Both *poachers* and *foxes* can *hunt*, but it is more probable, given what we know about the world, that *poachers hunt foxes* rather than the other way around. Moreover, both argument NPs agree in number with the inflection of the verb. This is problematic for the Semantic Attraction account, because the P600-effect found by van Herten et al. cannot be attributed to a syntactic mismatch between an observed and an expected verb inflection. Even if a semantic processor produces an interpretation that is not in line with the syntactic structure of these sentences, no P600-effect should be produced at the critical verb because the inflection of the verb is also consistent with an analysis in which the *the poacher hunted the fox*. Therefore, the Semantic Attraction account cannot explain the P600-effect for the sentences of van Herten et al. (2005).

van Herten et al. (2005, 2006) offered an explanation for the presence of an SIE in terms of a framework called *Monitoring Theory* (see also Kolk et al., 2003; Kolk and Chwilla, 2007; Vissers et al., 2007; Ye and Zhou, 2008; van de Meerendonk et al., 2009, 2010). They proposed an architecture in which an algorithmic, syntax-driven stream works in parallel to a plausibility heuristic driven by world knowledge. Provided sufficient time, the algorithmic stream always arrives at an interpretation of a sentence that is in line with its syntactic structure. The plausibility heuristic, however, only draws on word meanings and world knowledge to rapidly spell out the most likely interpretation of a sentence. The relative ease with which this heuristic can spell out an analysis is reflected in N400 amplitude. The two processing streams can arrive at conflicting interpretations of a sentence, and the processor is assumed to 'monitor' for such conflicts. In case of a mismatch between the streams, the processor will attempt to resolve the conflict through reanalysis, producing an increase in P600 amplitude. Hence, Monitoring Theory can yield a number of possible processing outcomes. For instance, the streams can agree that the sentence is plausible. In this case no N400- or P600-effect should be observed. A second possibility is that the streams agree that the sentence is implausible. This should lead to an increase in N400 amplitude because the plausibility heuristic encounters difficulty in combining individual words and world knowledge into a plausible interpretation, but there should be no P600-effect because the streams are in unison. On a third possible outcome, the algorithmic processor spells out an implausible analysis, whereas the plausibility heuristic comes up with a plausible interpretation;

according to Monitoring Theory, this is exactly what happens in Semantic Illusion sentences. For example, in the reversal anomaly 'De vos die op de stroper joeg...' (lit: The fox that on the poacher hunted...), the algorithmic processor arrives at an analysis in which *the fox* has *hunted the poacher*, and the plausibility heuristic at a reading in which *the poacher* has *hunted the fox*. Because the analysis from the heuristic stream is semantically plausible there is no N400-effect, but there is a P600-effect reflecting the revision processes arising from the conflict with the algorithmic stream.

A number of studies (Hoeks et al., 2004; van Herten et al., 2006; van de Meerendonk et al., 2010; van Petten and Luka, 2006, among others) (for an overview, see van Petten and Luka, 2011) have shown that some sentences engender biphasic N400/P600-effects and not only an N400-effect (for implausibility), or only a P600-effect (for conflicting streams). Hoeks et al. (2004), for instance, observed such a biphasic pattern in response to 'De speer heeft de atleten opgesomd' (lit: The javelin has the athletes summarized) relative to 'De speer werd door de atleten geworpen' (lit: The javelin was by the athletes thrown). For these sentences, no conflict should arise between the algorithmic processor and the plausibility heuristic. Both streams are expected to agree on the implausibility of the sentence and therefore Monitoring Theory predicts an N400-effect due to difficulty in semantic integration of word meanings and world knowledge but no P600-effect.

Biphasic N400/P600-effects are also problematic for the Semantic Attraction account. When processing the sentence 'De speer heeft de atleten opgesomd' (lit: The javelin has the athletes summarized), the semantic stream should not challenge syntax because there is no strong semantic at-

traction between the arguments *javelin* and *athletes* on the one hand, and the verb *summarized* on the other. For this reason, the attraction account would predict an N400-effect reflecting difficulty in semantic integration. However, since the sentence is syntactically correct, and syntax is not challenged, no modulation of P600 amplitude is predicted. Instead of a biphasic N400/P600-effect, the Semantic Attraction account would thus only predict an N400-effect for this sentence. Biphasic N400/P600-effects thus provide evidence against both the Monitoring Theory and the Semantic Attraction account.

2.1.3. Continued Combinatory Analysis

Kuperberg et al. (2007) observed an SIE in response to sentences containing a thematic violation, as in 'For breakfast the eggs would <u>eat</u>...' relative to 'For breakfast the boys would <u>eat</u>...' (see Kuperberg et al., 2003, 2006, for similar results). These sentences exhibit a strong semantic relation between the argument and the verb (*eggs* can be *eaten*). Interestingly, sentences containing a thematic violation in which there was no such relation, such as 'For breakfast the eggs would <u>watch</u>...', also evoked a P600- and no N400-effect relative to the same control. These results are problematic for the Semantic Attraction account. Since there is no attraction between *watch* and *eggs*, it predicts that the critical verb will only generate an N400-effect reflecting difficulty in semantic integration, but not a P600-effect reflecting syntactic revision. The results are also difficult to explain for Monitoring Theory. This account predicts that both the algorithmic stream and the plausibility heuristic agree on the implausibility of the sentence, yielding an N400-effect due to difficulty in semantic integration, but no P600-effect, since there is no conflict between the streams. Thus, the findings of Kuperberg et al. (2003, 2006,

2007) can be seen to undermine both Monitoring Theory and the Semantic Attraction account.

Kuperberg (2007) proposed an architecture in which three streams operate in parallel. The first stream is based on semantic memory and is similar to the plausibility heuristic proposed by the Monitoring Theory. The second is a syntax-driven stream, which uses morpho-syntactic constraints to algorithmically build a representation of meaning. The third stream is thematic-role based, and uses semantic-thematic cues to build an interpretation. The latter two streams are referred to as 'combinatory' streams. Kuperberg's three streams are assumed to be fully interactive in that they can influence each other at any time during on-line processing. This model can account for the results reported in Kuperberg et al. (2003, 2006, 2007). When processing sentences that contain a thematic violation, such as 'For breakfast the eggs would eat...' or 'For breakfast the eggs would watch...', a conflict arises between the syntax-driven stream and the thematic-role based stream, because syntax assigns the inanimate NP (*the eggs*) the role of Actor, whereas the thematic stream assigns it the role of Undergoer. This conflict results in a "continued combinatory analysis", a revision process that attempts to resolve the conflict between the combinatory streams which is reflected in a P600-effect. On this view, these sentences do not produce an N400-effect because the conflict between the combinatory streams supposedly blocks additional semantic processing in the memory-based stream. Furthermore, sentences such as 'For breakfast the boys would watch...', engender an N400-effect because processing problems arise in the semantic memory-based stream (*watching* something is not a typical activity during *breakfast*).

Kuperberg et al. (2007) argued that sentences like 'For breakfast the eggs would <u>watch</u>...' do not produce an N400-effect because semantic processing is blocked as soon as a conflict between the combinatory streams becomes apparent. However, this makes it unclear how the Continued Combinatory Analysis (CCA) account could explain the biphasic N400/P600-effect found by Hoeks et al. (2004) in sentences such as 'De speer heeft de atleten <u>opgesomd</u>' (lit: The javelin has the athletes <u>summarized</u>) relative to 'De speer werd door de atleten <u>geworpen</u>' (lit: The javelin was by the athletes <u>thrown</u>). The syntax-driven stream should spell out an analysis in which *the javelin* has *summarized the athletes*, and the thematic-role based stream an interpretation in which *the athletes* have *summarized the javelin*. Because these analyses are in conflict, semantic processing should be blocked. In other words, the account proposed by Kuperberg (2007) only predicts a P600-effect for these sentences, and no N400-effect. In a recent study, however, Kuperberg et al. (2010) found a biphasic N400/P600-effect for sentences containing an animacy violation such as 'The journalist astonished the <u>article</u>...' relative to 'The journalist wrote the <u>article</u>'. In the anomalous sentences, the syntax-driven stream should produce an analysis in which *the journalist astonished the article*, whereas the thematic-role based stream should spell out an interpretation in which *the article astonished the journalist*. Again, since the analyses of the combinatory streams are in conflict, semantic processing is assumed to be blocked. And again, a P600-effect should occur, and no N400-effect. It is quite unclear why blocking would occur in the one situation and not in the other, as the materials Kuperberg (2007) and Kuperberg et al. (2010) seem highly similar. Biphasic N400/P600-effects therefore re-

main problematic for the CCA account.

2.1.4. The extended Argument Dependency Model

Bornkessel-Schlesewsky and Schlesewsky (2008) argue that there is ample evidence that the language processor constructs an interpretation of verb-final constructions before the critical verb is reached (see Bornkessel et al., 2003, for instance), and claim that this is problematic for the CCA account. For instance, before reaching the verb *thrown* in 'De speer heeft de atleten geworpen' (lit: The javelin has the athletes thrown), the syntax-driven stream of the CCA account should have established an interpretation of the sentence in which *the javelin* is doing something to *the athletes*. The thematic-role based stream, on the other hand, is argued to spell out an analysis in which *the athletes* are doing something to *the javelin*. Hence, according to Bornkessel-Schlesewsky and Schlesewsky (2008), these streams should already be in conflict after processing the second NP, and the CCA should thus predict a P600-effect at this NP instead of at the verb, which is clearly not what was reported in any of the SIE studies.

Bornkessel-Schlesewsky and Schlesewsky (2008) propose that their *extended Argument Dependency Model* (eADM) (Bornkessel and Schlesewsky, 2006) can overcome these problems. The eADM is a model of core argument interpretation (rather than a fully fledged model of sentence comprehension) and its focus lies on explaining thematic role assignment. The model postulates two processing streams. The first stream assigns thematic roles to incoming NPs based on 'prominence' information (e.g., animacy, case marking, and linear word order) and links these roles to the argument structure of an incoming verb. Difficulties in thematic role assignment or verb-argument

13

linking are assumed to produce an increase in N400 amplitude. The other stream engages in semantic processing similar to the plausibility heuristic proposed by Monitoring Theory. Processing difficulties in this plausibility-driven stream are also assumed to lead to an increase in N400 amplitude. The analyses generated by these streams are integrated in a "generalized mapping step". If the streams conflict, integration is problematic and a P600-effect should be produced. This happens in verb-final constructions such as 'De speer heeft de atleten geworpen' (lit: The javelin has the athletes thrown), where thematic role assignment leads to an interpretation in which *the javelin* has *thrown the athletes*, and the plausibility heuristic to one in which *the athletes* have *thrown the javelin*. Importantly, neither thematic role assignment, verb-argument linking, or plausibility processing is difficult for this sentence, so no N400-effect is evoked according to the eADM. After the generalized mapping step the input is checked for "well-formedness". The authors specify that well-formedness "is not meant to contrast strictly with ill-formedness, but rather refers to a gradient mechanism that evaluates the acceptability of a structure under different environments (e.g., a discourse context)." (Bornkessel and Schlesewsky, 2006, p. 790). Problems in "well-formedness" processing are also assumed to produce an increase in P600 amplitude. Thus, in the eADM there are two ways in which a sentence can evoke a P600-effect; due to problems with integrating the two streams, but also due to a lack of 'well-formedness'. The late positivities that are caused by each of these processing phases are hypothesized to be additive. Bornkessel-Schlesewsky and Schlesewsky (2008) suggest that this is the case in sentences like 'De speer heeft de atleten geworpen' (lit: The javelin has the

athletes <u>thrown</u>) where both integration of streams and 'well-formedness' are problematic. In contrast to Kuperberg's CCA, the eADM predicts processing difficulty to only become apparent once the arguments have been linked into the final verb's argument structure, and not at an earlier point in time.

The eADM goes a long way in explaining the SIE data. However, Kos et al. (2010) recently posed a difficult case for this model. They found an N400-effect in sentences such as 'Fred eet een <u>restaurant</u>' (lit: Fred eats a <u>restaurant</u>) relative to control sentences such as 'Fred eet een <u>boterham</u>' (lit: Fred eats a <u>sandwich</u>). Kos et al. argue that based on prominence information, the eADM should assign *a restaurant* the role of Undergoer, whereas the plausibility processor should spell out an analysis in which this NP has a locative role, as in *eating in a restaurant*. As thematic role assignment proceeds smoothly, no N400-effect is produced in this stream. Similarly, also no N400-effect arises in the plausibility processing stream because it spells out a plausible analysis. The different streams eventually conflict in the generalized mapping step. Moreover, the sentence may lead to a well-formedness problem. The eADM therefore predicts that the sentence evokes a P600-, but no N400-effect, which is the reverse of what was actually found.

2.1.5. Processing Competition

On the basis of the data discussed in the previous paragraph, Kos et al. (2010) proposed the *Processing Competition* model which is a two-stream architecture with a syntactic and a semantic stream (see also Hagoort et al., 2009). Both streams simultaneously attempt to construct an interpretation of an incoming sentence. Kos et al. argue that if the processing of a sentence leads to a conflict between the two streams, the burden of resolving the

conflict is placed on the stream that has the weakest support. For example, if semantic cues are strong the burden is placed on the syntactic stream, leading to a P600-effect. If, on the other hand, syntactic cues are strong, the burden is placed on the semantic system, producing an N400-effect. Consider the sentence 'Fred eet een restaurant' (lit: Fred eats a restaurant). Here, the syntactic stream spells out a semantically implausible analysis in which *Fred* consumes *a restaurant*. The semantic stream produces a plausible interpretation in which *Fred eats in a restaurant*. Kos et al. (2010) argue that in this case syntactic cues are stronger, supposedly because these cues predict the critical word to be an NP (rather than a prepositional phrase), whereas the semantic cues do not create a specific expectation. Hence, the burden of resolving the conflict lies on semantics, evoking an N400-effect (but no P600-effect). For SIE sentences, the model predicts the reverse pattern. When processing a sentence such as 'De speer heeft de atleten geworpen' (lit: The javelin has the athletes thrown), syntax arrives at an analysis in which *the javelin* has *thrown the athletes* while semantics advocates a reading in which *the athletes have thrown the javelin*. Semantic cues are seen as strong because it is plausible for *athletes* to *throw a javelin*. Consequently, the burden to resolve the conflict is placed on the syntactic stream, leading to a P600-effect (but no N400-effect).

The Processing Competition model postulates a mechanism in which a conflict between competing interpretations is always resolved by one of the two streams, producing either an N400-effect or a P600-effect. It is therefore not clear how the model could account for biphasic N400/P600-effects, such as those observed in 'De speer heeft de atleten opgesomd' (lit: The javelin

16

has the athletes <u>summarized</u>). One might speculate that in these sentences syntactic and semantic cues are of approximately equal strength and that both streams are equally involved in resolving the conflict, leading to both an N400-effect and a P600-effect. In this case, however, it is difficult to see how the processor could ever arrive at a final analysis, unless one of the streams eventually drops out or is overruled. Hence, biphasic N400/P600-effects appear to be a stumbling block for the Processing Competition account as it is currently explicated. Moreover, a model relying on cue strength can only be truly predictive if cue strength can be objectively quantified, which may turn out to be quite difficult, as Kos et al. admit themselves (see Kos et al., 2010, page 10).

2.2. Semantic Illusions in wider discourse

The Semantic Illusion data discussed so far were all obtained in studies using isolated sentences. Nieuwland and van Berkum (2005) tested whether the 'Semantic Illusion' phenomenon extended to wider discourse as well. They presented participants with stories like:

> 'A tourist wanted to bring his huge suitcase onto the airplane. However, because the suitcase was so heavy, the woman behind the check-in counter decided to charge the tourist extra. In response, the tourist opened his suitcase and threw some stuff out. So now, the suitcase of the resourceful tourist weighed less than the maximum twenty kilos. Next the woman told the <u>suitcase</u>...'

They found that if the story continued as in the example ('Next, the woman told the <u>suitcase</u>...'), a P600-, but no N400-effect was produced

relative to a more plausible continuation like 'Next, the woman told the tourist...'. The continuation sentence contained a semantic anomaly as inanimate objects like suitcases are not usually an addressee of speech. Models that rely on a plausibility heuristic-like semantic stream will therefore not produce a plausible analysis for the anomalous continuation. Consequently, all five *multi-stream* models predict an N400-effect for these sentences due to difficulties in semantic processing. Thus, they all fall short of explaining the results of Nieuwland and van Berkum, and it is difficult to see how the models could be repaired to accommodate these findings.

2.3. Summary

We have reviewed five multi-stream models of ERP patterns in Semantic Illusion sentences: 1) the Semantic Attraction account, 2) Monitoring Theory, 3) the Continued Combinatory Analysis account, 4) the extended Argument Dependency Model, and 5) the Processing Competition model (see Table 1 for an overview of their characteristics). While all of them can account for a subset of the SIE data, none of them covers the full range of results (see Table 2 for a summary of the predictions made by each model). One reason for this failure, at least for four out of the five models (1-3 and 5), is that they have difficulty explaining biphasic N400/P600-effects that have been found in several studies (e.g., Hoeks et al., 2004; van Herten et al., 2006; van de Meerendonk et al., 2010) (see van Petten and Luka, 2011, for an overview). The eADM can generate biphasic patterns, but it does not explain the results of Kos et al. (2010). Furthermore, none of the five models can account for the Nieuwland and van Berkum (2005) data on discourse processing. Thus, our review casts doubt on the explanatory power and va-

18

lidity of these *multi-stream* architectures. In the next section, we will propose a *single-stream* account that explains all relevant data, not by changing the complexity of the processing architecture, but by reconsidering the functional role of the ERP components involved: the N400 and the P600.

3. Rethinking the functional role of the N400 and the P600

The Semantic Illusion phenomenon suggested that semantic information can to some extent be processed separately and autonomously, causing a shift from *single-stream* to *multi-stream* models. What we would like to suggest, however, is that instead of abandoning *single-stream* models, we should reconsider the functional interpretation of the ERP components that are involved. Following a growing number of studies, we propose to interpret N400 amplitude as reflecting a *memory retrieval* phase, in which all the information linked to an incoming word (i.e., the syntactic, semantic, and pragmatic information associated with that word) is 'retrieved' from long-term memory (cf. Kutas and Federmeier, 2000; Lau et al., 2008, 2009; van Berkum, 2009; Federmeier and Laszlo, 2009; van Berkum, 2010; Kutas and Federmeier, 2011). We also propose that the *integration* of this activated lexical information into the existing current mental representation of an unfolding sentence is reflected in P600 amplitude. We combine this retrieval view on the N400 amplitude and the 'integration' view on the amplitude of the P600 into a *single-stream* account of language processing, which we will refer to as the Retrieval-Integration (RI) account. This account explains the absence of an N400-effect and the presence of a P600-effect in response to SIE sentences. In a sentence such as 'De speer heeft de atleten <u>geworpen</u>'

19

(lit: The javelin has the athletes <u>thrown</u>) relative to 'De speer werd door de atleten <u>geworpen</u>' (lit: The javelin was by the athletes <u>thrown</u>), there is no N400-effect because retrieving the lexical information associated with *thrown* is facilitated approximately equally in both sentences due to word and context priming. There will be difficulty in integrating this information into the existing mental representation, but only in the anomalous 'illusion' sentence, producing a P600-effect relative to the normal control. Integration is predicted to be difficult because it results in a representation that conflicts with what we know about the world: *javelins* do not *throw athletes*. In what follows, we will motivate the Retrieval-Integration account in more detail. First, we will argue that there is converging evidence for the retrieval nature of N400 amplitude, then we will discuss the integration perspective on the amplitude of the P600.

3.1. N400 as Memory Retrieval

In the five *multi-stream* models we have discussed, it was assumed that N400 amplitude is sensitive to difficulty of *semantic integration* (compositional semantic processing) (Osterhout and Holcomb, 1992; Brown and Hagoort, 1993; Chwilla et al., 1995; van Berkum et al., 1999; Hagoort and Van Berkum, 2007; Hagoort et al., 2009; Wang et al., 2009; Baggio and Hagoort, 2011; Lotze et al., 2011). Under such a view, it should be difficult to integrate the critical verb *thrown* in sentences such as 'De speer heeft de atleten <u>geworpen</u>' (lit: The javelin has the athletes <u>thrown</u>), because this yields an interpretation that is not in line with what we know about the world: *athletes throw javelins* and not the other way around. This difficulty should be reflected in an N400-effect, but none was observed. Precisely

this complication has motivated *multi-stream* architectures to postulate a semantic analyzer which can build an interpretation independent of syntactic surface structure. In this way, an autonomous semantic processing stream can produce a semantically correct analysis in which *athletes have thrown the javelin*, accounting for the absence of an N400-effect.

However, alternative views on the functional role of N400 amplitude have recently come to the fore in which it is interpreted in terms of *memory retrieval* rather than integration (see Kutas and Federmeier, 2000, 2011; Federmeier and Laszlo, 2009, for overviews). On this view, the amplitude of the N400 reflects the mental processes that accompany the retrieval of lexical information from long-term memory. For instance, the observed N400-effect in response to 'For breakfast the boys would watch...' (as compared to 'For breakfast the boys would eat...') indicates that the retrieval of the lexical features of *eat* is facilitated by the activation of semantic and syntactic features of the preceding words, as well as by scenario-based world knowledge (*having breakfast* entails *eating*; see Chwilla and Kolk 2005 for a similar view on script priming, and van Berkum 2009 pp. 295–297, for other possible factors of influence), but retrieval of the lexical features of *watch* is not. This is consistent with theoretical proposals on the organization of the mental lexicon (Pulvermüller, 1999, 2001; Rogers and McClelland, 2004; Elman, 2004, 2009), with theories of memory (Ratcliff, 1978; Gillund and Shiffrin, 1984; Hintzman, 1988) and with the general idea that language processing is highly predictive (Otten et al., 2007; van Berkum, 2010). The retrieval view on N400 amplitude clearly differs from the integration view in that retrieval is a bottom-up process that does *not* involve integrative semantic processing

21

or semantic composition. Crucially, we assume that top-down information, for example from the existing mental representation of the preceding sentence fragment, does play a role, but it *adds* to the activation pattern (van Petten, 1993, 1995; van Berkum, 2009); it does not *constrain* the pattern of activation (i.e., make it more specific). Thus, context has important 'excitatory' power. A byproduct of such a mechanism is that the language processing system is able to anticipate or predict upcoming words (see Schwanenflugel and Shoben, 1983, 1985; Schwanenflugel et al., 1988; Federmeier and Kutas, 1999; van Berkum, 2009). This is also consistent with the idea of 'readiness' of information from the memory and text comprehension literature (see e.g., Gerrig and McKoon, 1998; Gerrig, 2005),

With respect to the SIE data, the retrieval hypothesis provides a different perspective on the absence of an N400-effect in sentences such as 'De speer heeft de atleten geworpen' (lit: The javelin has the athletes thrown) (see Stroud, 2009; Stroud and Phillips, 2011, for a similar proposal). For both this sentence and its control 'De speer werd door de atleten geworpen' (lit: The javelin was by the athletes thrown), retrieval of the lexical features of *thrown* is facilitated. Facilitation occurs because of semantic relatedness between *thrown* and the preceding words *javelin* and *athletes* in its prior context, and also through the activation of scenario-based world knowledge (*javelins* are typically *thrown* by *athletes*). The retrieval hypothesis therefore predicts approximately equally sized N400 amplitudes for the anomalous sentence and its control and hence no N400-effect. A similar explanation can be offered for the absence of an N400-effect in the other studies in which an SIE has been observed. All relevant materials contain some kind of semantic

relation between a verb and its argument(s), leading to facilitated retrieval of the lexical features of a critical word by means of priming from the preceding lexical items, and from the existing mental representation of the preceding part of the sentence, in both the test and the control sentence. In the following, we underlined the words that are in such a priming relationship for the studies discussed in this paper:

(1)   The javelin has the   athletes thrown
         The javelin was by the athletes thrown
         (Hoeks et al., 2004)

(2)   The hearty meal was devoured
         The hearty meal was devouring
         (Kim and Osterhout, 2005)

(3)   The poacher that on the fox   hunted
         The fox   that on the poacher hunted
         (van Herten et al., 2005)

(4)   (*prior context:* a story about a tourist and a suitcase):

         Next, the woman told the suitcase

         Next, the woman told the tourist

         (Nieuwland and van Berkum, 2005)

(5)   For breakfast the eggs would eat
         For breakfast the boys would eat
         (Kuperberg et al., 2007):

For all of these item pairs, the retrieval view thus predicts approximately

23

equal N400 amplitudes for the critical words and no N400-effect when the sentences are contrasted.

At the same time, the retrieval hypothesis can also explain the *presence* of N400-effects in materials that elicit biphasic N400/P600-effects such as 'De speer heeft de atleten opgesomd' (lit: The javelin has the athletes summarized) (compared to 'De speer werd door de atleten geworpen; lit: The javelin was by the athletes thrown). Due to absence of priming, retrieval of the lexical features of *summarized* in the critical sentence is predicted to be more cumbersome, and produce a larger N400 amplitude than retrieval of the lexical features of *thrown* in the control sentence and therefore we observe an N400-effect.

Kuperberg et al. (2007) reported findings that seem to challenge the retrieval view. Sentences such 'For breakfast the eggs would watch...' produced a P600-effect relative to 'For breakfast the boys would eat...' but no N400-effect. On the retrieval view one would predict that the activation of the lexical features of *watch* in the critical sentence is facilitated less than that of the features of *eat* in the control sentence, producing an N400-effect: *eat* is semantically related to *breakfast*, whereas *watch* is not. However, close inspection of the materials used by Kuperberg et al. showed that there were sentence pairs such as 'To explore the area the truck should describe...' relative to 'To explore the area the travelers should rent...'. In neither of these sentences, facilitated retrieval is expected, and hence no difference in N400 amplitude is predicted. The retrieval hypothesis thus predicts an N400-effect for some, but not for other sentences from the Kuperberg et al. materials. This may have been the reason that the N400-effect did not reach significance.

Many researchers, most notably the proponents of the five multi-stream models that we have discussed, still adhere to the integration view of N400 amplitude. Nevertheless, evidence favoring retrieval over integration is accumulating in the ERP literature (see Kutas and Federmeier, 2000; Federmeier and Kutas, 1999; Kutas et al., 2006; Lau et al., 2008, 2009; van Berkum, 2009; Federmeier and Laszlo, 2009; van Berkum, 2010, for discussions). For instance, it has been known for some time that N400 amplitude is insensitive to the truth value of an utterance. Fischler et al. (1983) showed that the final word in sentences such as 'A robin is not a <u>bird</u>' produced an N400 amplitude equal in size to that elicited by the final word in true sentences such as 'A robin is a <u>bird</u>' (see also Kounios and Holcomb, 1992). This is difficult to explain under the integration view because the critical noun *bird* is expected to be more difficult to integrate with the conceptual representation evoked by 'A robin is not a [...]' due to a conflict with world knowledge. The retrieval hypothesis, on the other hand, does not predict an N400-effect. Retrieval of the lexical features of *bird* is facilitated equally in both sentences, because of semantic relatedness to *robin*. Hence, the sentences should generate equal sized N400 amplitudes. In a related study, Hagoort et al. (2004) found that the critical word in sentences such as 'The Dutch trains are <u>white</u>...' produced a larger N400 amplitude than the critical word in sentences like 'The Dutch trains are <u>yellow</u>...'. These findings are in line the integration hypothesis, as Dutch trains are in fact *yellow* and not *white*, so *white* should be harder to integrate because it renders the sentence false. However, the results are also consistent with the retrieval view. Since *yellow* is a semantic feature of Dutch trains, retrieval of *yellow* is facilitated compared to the retrieval of

white leading to an increased N400 amplitude for the latter.

Finally, the retrieval view of N400 amplitude is supported by evidence from neuroimaging studies. Lau et al. (2008) surveyed fifteen functional Magnetic Resonance Imaging (fMRI) studies of semantic priming and found that the posterior middle temporal gyrus (pMTG) was the most likely source of the N400-effect (see Halgren et al., 2002, for converging evidence from magnetoencephalography, MEG). Importantly, the pMTG was active in lexical decision tasks at Stimulus Onset Asynchronies (SOAs) as short as 33ms where it seems implausible that a word is sufficiently processed to be integrated into a larger context. This finding is difficult to explain under the integration view. Lau et al. suggest that processes at these short SOAs must reflect automatic spreading activation processes in lexical-semantic memory. Furthermore, they argue that activity in the pMTG has been consistently linked to tasks requiring semantic categorization or feature judgment (Price et al., 1994; Pugh et al., 1996; Gold et al., 2006), suggesting that this region is engaged in lexical retrieval. This is supported by the fact that aphasics with lesions in the pMTG show difficulties with tasks that require lexical retrieval (Kertesz, 1979; Hart Jr. and Gordon, 1990), and the observation that the pMTG was the only area found consistently active in production tasks involving lexical selection (Indefrey and Levelt, 2004). Thus, there is a strong neuroanatomical link between N400 amplitude and brain areas involved in lexical retrieval.

If we accept the retrieval perspective on the amplitude of the N400, the absence of an N400-effect in Semantic Illusion sentences is explained easily (see Table 2). However, if N400 amplitude does not reflect integration, the

question remains how and when integration takes place. Integrating multiple sources of information is without doubt a core task of the language processor and it would be surprising if the brain activity associated with that process would not be detectable in ERP waveforms. In the following section, we will argue that these integrative processes are reflected in the amplitude of the P600.

3.2. P600 as Mental Representation Composition

The P600 component has long been interpreted as indexing the difficulty of revising the existing syntactic analysis when an incoming word renders the sentence ungrammatical. Hagoort et al. (1993), for instance, found a P600-effect in response to sentences containing a mismatch in number agreement, as in 'Het verwende kind gooien...' (lit: The spoilt child throw...) relative to control sentences such as 'Het verwende kind gooit...' (lit: The spoilt child throws...) (see also Neville et al., 1991; Münte et al., 1998; Allen et al., 2003; Gouvea et al., 2010, among others). P600-effects have also been observed in response to garden-path sentences. These sentences do not contain a syntactic anomaly, but have been found to lead participants to initially adopt an incorrect syntactic analysis of a sentence that needs to be revised later on (Osterhout and Holcomb, 1992, 1993; Osterhout et al., 1994; Frisch et al., 2002; Kaan and Swaab, 2003; Gouvea et al., 2010). Osterhout et al. (1994), for instance, observed a P600-effect in response to 'The lawyer charged the defendant was...' relative to 'The lawyer charged *that* the defendant was...'. Presumably, readers initially pursue a reading in which *the defendant* is the direct object of the verb *charged*. The auxiliary *was* then disambiguates the NP *the defendant* as the subject of a subordinate clause,

27

requiring readers to revise their initial analysis. The P600-effect elicited by syntactic violations and garden-paths thus appears to reflect processes of syntactic revision or syntactic repair. In the Hagoort et al. study, for instance, such revision could involve correcting number agreement by mentally adjusting the critical verb's inflection (for similar proposals see Kim and Osterhout, 2005; Kim and Sikos, 2011; Stroud, 2009; Stroud and Phillips, 2011).

Results by Kaan et al. (2000) challenged the interpretation of P600 amplitude as an index of syntactic revision. They found that long-distance *wh*-dependencies such as 'Emily wonders *who* the performers in the concert imitate…' also produced a P600-effect relative to a sentence lacking such a dependency (as in: 'Emily wonders *whether* the performers in the concert imitate…') (see also Fiebach et al., 2002; Felser et al., 2003; Phillips et al., 2005; Gouvea et al., 2010). This sentence is syntactically well-formed and does not contain a garden-path, so the observed P600-effect is unlikely to reflect syntactic revision processes. Rather, it seems to reflect difficulty in establishing the *wh*-dependency at the verb. The verb *imitate* has to be 'linked' to the *wh*-pronoun *who* while no such dependency needs to be established in the control sentence. Kaan et al. therefore suggested that P600 amplitude must also reflect processes of syntactic integration.

Recently, it has become clear that P600-effects do not only appear in sentences that are structurally complex. Burkhardt (2006), for instance, reported a P600-effect for sentences that require extra processing at the discourse level. She used target sentences containing a referring expression for which 1) the referent was novel, 2) the referent could be inferred from the preceding context, or 3) the referent was given in the context. When a novel

discourse referent had to be established, a P600-effect was observed (relative to the control sentences with a given referent). There was also a P600-effect when the referent could be inferred. None of the conditions contained syntactic violations, garden-paths, or complex syntactic dependencies. The P600-effect observed by Burkhardt can therefore not be attributed to purely syntactic processes. In later studies, Burkhardt (2007); Schumacher (2011) showed that not only the introduction of a new protagonist, but also the inferred presence of an instrument and processes of establishing reference produce a P600-effect. These findings again suggest that the amplitude of the P600 might be more than a reflection of intensive syntactic processing.

Regel et al. (2011) made a similar suggestion when finding a P600-effect in response to *irony*. In their study, participants were presented with short contextual story fragments, for example, about someone hearing many mistakes in the performance of a Bach sonata. The person in question would then look at the orchestra playing the sonata, and ironically say to his or her conversational partner: 'These artists are gifted'. In the control condition the person would be listening in ecstasy and again say 'These artists are gifted', but this time being sincere. Regel et al. interpreted the P600-effect in the irony condition as a reflection of discourse-related interpretative processes involved in computing ironic versus literal meaning of an utterance.

Taken together, the findings of Burkhardt (2006, 2007); Schumacher (2011), and Regel et al. (2011) cast doubt on a purely syntactic interpretation of P600 amplitude. The P600-effects found in these recent studies seem to reflect processes that are related to meaning rather than syntax. What their materials have in common is that they require additional processing (as compared to

the control condition) in order to arrive at a coherent mental representation of what the speaker or writer meant to communicate. In the materials of Burkhardt, for instance, the processor has to first establish a discourse referent before it can arrive at a representation of the described situation. Building such a representation should require more work in the critical conditions compared to the control condition because the referent was not given. In the irony materials of Regel et al., the processor has to compute an interpretation in which an utterance is meant in a 'dishonest' way. Compared to the truthful condition, establishing such a non-literal interpretation may involve additional processing. Hence, the results of Burkhardt and Regel et al. are compatible with the idea that P600 amplitude reflects the additional effort invested in establishing a representation of what the speaker wants to convey.

It seems uncontroversial that in comprehending a sentence or story, a listener (or reader) constructs some kind of mental representation of what is communicated. What such representations actually look like is still a matter of debate. Representations in comprehension have variously been called *mental models* (Johnson-Laird, 1983), *situation models* (van Dijk and Kintsch, 1983; Kintsch, 1988, 1998; Zwaan et al., 1995; Zwaan and Radvansky, 1998; Zwaan, 1999, 2003; Kerkhofs and Haselager, 2006), *message-level representations* (Morris, 1994), or *discourse representations* (Kamp and Reyle, 1993). For the present discussion, we will adopt the theory-neutral term *mental representation of what is being communicated* (MRC for short). These MRCs probably comprise the propositions that can be derived directly from the linguistic input, but also knowledge from all kinds of inferences (e.g., logical,

causal, or pragmatic) that can be made on the basis of world knowledge, including pragmatic knowledge about communication. Consider, for instance, the following sentence: 'John let loose of the cup, and it covered the floor with splinters'. Comprehension here requires more than extracting its propositions (e.g., something like $let\_loose(john, cup)$ and $cover(splinters, floor)$). It also invites the causal inference that the cup broke when it fell out of John's hands, and that the splinters covering the floor used to make up the cup. Thus, the construction of a mental representation requires knowledge about the causal fabric of the world in which the described situation or event is set. To generate the inference that the cup broke after it fell out of John's hand, for instance, one needs to know that gravity makes the cup fall down, that cups are typically made of glass or porcelain, and that objects made out of such materials are likely to break if they fall from a certain height. On a different level, language use also gives rise to pragmatic inferences, because utterances ultimately have a communicative function. Possibly, a speaker wants to warn for splinters on the floor because John broke his cup, or perhaps he/she wants to emphasize John's clumsiness. Hence, we agree with Gernsbacher (1990); Givón (1992); Kintsch (1992) and Zwaan and Radvansky (1998) that words and sentences "...can be regarded as a set of processing instructions on how to construct a mental representation of the described situation." (Zwaan and Radvansky, 1998, p. 177). It is important to note that MRCs are often incomplete and incorrect (cf. Kintsch and Mangalath, 2011, p. 357), it depends in part on the pragmatic and world knowledge of a specific language user which inferences are possible. What is more, not all inferences that are in principle possible will actually be drawn because our language system seems

31

heavily influenced by task demands and may take the 'good enough' approach to language understanding (Ferreira and Patson, 2007). This means that if the current state of the MRC suffices to successfully take part in a shallow conversation, or to answer a simple question, then no more effort needs to be invested in making the MRC more complete or coherent. On the other hand, if it really matters that the MRC is correct, for instance when asked to determine the plausibility of a sentence in a language task, language users will try harder. It is conceivable that MRCs also play a role in non-linguistic domains such as music (Patel, 2003), sequence learning (Christiansen et al., 2011), and arithmetic (Núñez-Peña and Honrubia-Serrano, 2004). Representations in these domains will of course differ from representations arising in language comprehension, but there may also be important parallels, for instance in how mental representations are reconstructed or evaluated.

Although it is entirely possible that the P600-effects in the studies that were discussed above comprise a rather heterogeneous family, where some family members reflect discourse processing, others semantic processing, and yet others syntactic processing, we would like to start from the hypothesis that all P600-effects can be described in terms of the construction, revision, or updating of a mental representation of what is being communicated. Not only does this explain the findings of Burkhardt (2006, 2007); Schumacher (2011) and Regel et al. (2011), it also casts a different light on the SIE data. For the sentence 'De vos die op de stroper joeg...' (lit: The fox that on the poacher hunted...), for instance, it is more difficult to arrive at a coherent representation than for its control 'De stroper die op de vos joeg...' (lit: The poacher that on the fox hunted...), because our world knowledge tells

us that it is unusual for *foxes* to *hunt poachers*. Similarly, *eggs* do not *eat*, *javelins* do not *throw athletes*, *meals* do not *devour*, and *suitcases* are usually not the addressee of an utterance. In all these materials, MRC composition is thus predicted to be more difficult for the illusion sentences as compared to control because the depicted situation conflicts with world knowledge (see Table 2). One study that requires some additional explanation, however, is that of Kos et al. (2010). For a sentence such as 'Fred eet een <u>restaurant</u> tijdens de lunch' (lit: Fred eats a <u>restaurant</u> during the lunch) (as compared to 'Fred eet een <u>boterham</u> tijdens de lunch'; lit: Fred eats a <u>sandwich</u> during the lunch), the MRC hypothesis would predict a P600-effect reflecting difficulty in constructing a coherent representation of *Fred eating a restaurant*. However, no such P600-effect was found. On closer inspection of the materials used in this study, however, there appear to be quite some items in which the semantic anomaly only becomes apparent at the sentence-final prepositional phrase (PP). As a consequence, in some items, a P600-effect may occur at the critical NP, but in other items only at the sentence-final PP. If ERP components are then averaged at each of the two positions (i.e., critical NP and the sentence-final PP), they might become undetectable, explaining why no significant P600-effect was found at the critical NP[2]. An easy way to test this idea would be to remove the sentence-final PPs in the Kos et al. (2010) materials and to replace them by a *period* (e.g., 'Fred eet een restaurant.';'Fred eats a restaurant.'). In that case, the critical NPs should definitely elicit a P600-effect. This is because the end-of-sentence marker makes it clear that there will be no further input for an unfolding sentence

---

[2]Kos et al. (2010) did not report results for the sentence final PP.

and hence the meaning of this sentence needs to be composed from what has already been perceived (see below for a more elaborate discussion of this issue).

We have argued that understanding 'Semantic Illusion' sentences and sentences in discourse context requires work on the MRC, which is then reflected in an increased P600 amplitude. But what about the P600-effects in response to phenomena such as garden-paths, long-distance *wh*-dependencies and syntactic violations? We would like to suggest that in these sentences P600 amplitude reflects MRC composition as well. Recall that Osterhout et al. (1994) observed a P600-effect in response to garden-path sentences like 'The lawyer charged the defendant <u>was</u>. . . ' (relative to 'The lawyer charged that the defendant <u>was</u>. . . '). One of the hypotheses that they formulated was that P600 amplitude reflects difficulty of syntactic revision. However, the P600-effect they found can also be explained in terms of the effort involved in creating a coherent mental representation. The processor might initially construct a representation in which *the defendant* is a referent that is being *charged* by *the lawyer*. The disambiguating auxiliary *was* signals that this is incorrect, and requires the processor to construct a representation in which which *the lawyer* is *charging* that *the defendant* was doing something. The processor thus has to revise its initial representation of the communicated situation, such that it is consistent with the input again. This revision may involve something like retracting (if that is at all possible) the event-representation depicting *the lawyer charging the defendant*, and postulating a novel event in which the *the lawyer charges that the defendant was doing something*. On this account, the observed P600-effect reflects the effort

in reworking an initial mental representation, rather than the revision of a syntactic analysis.

A similar argument can be applied to the results of Hagoort et al. (1993). They found a P600-effect in sentences containing a number agreement violation 'Het verwende kind gooien...' (lit: The spoilt child throw...) (relative to grammatically correct controls 'Het verwende kind gooit...'; lit: The spoilt child throws...). In terms of mental representation construction, the morphosyntactic error makes it more difficult to arrive at a coherent representation of the state of affairs. One reason why MRC construction is difficult in this case is the fact that number agreement is an essential cue for thematic role assignment. As it is not clear which element has an incorrect inflection (i.e., *child* or *throw*), it is possible that the processor first has to determine whether the depicted situation involves a single child or perhaps more. In the control condition, the construction of a mental representation is unproblematic. Hence, the increased amplitude of the P600 in the critical condition may reflect the additional processing incurred by the attempt to recover the intended meaning from the language input. Such recovery processes might be similar to those that underlie P600-effects for misspelled relative to correctly spelled words (see Vissers et al., 2006; Kim and Lai, 2011, for instance); the increased difficulty in representation construction stems from the effort of trying to recover what the writer meant to communicate.

The P600-effect observed by Kaan et al. (2000) can be explained along the same lines. Kaan et al. found a late positivity in response to sentences containing a long-distance *wh*-dependency 'Emily wonders *who* the performers in the concert imitate...' (relative to sentences lacking such a dependency

'Emily wonders *whether* the performers in the concert <u>imitate</u>...'). Processing of the verb *imitate* in the critical sentence entails establishing a thematic relation between the verb, the entity referred to by the *wh*-pronoun *who*, and the NP *the performers*. In the control sentence no such dependencies need to be established. For this reason, the construction of a coherent representation involves more work in the critical condition than in the control condition. The increase in P600 amplitude, we hypothesize, reflects this extra processing step.

Summarizing, we want to formulate the following hypothesis about the functional interpretation of P600 amplitude in language comprehension:

MRC hypothesis: The P600 component is a family of late positivities that reflect the word-by-word construction, reorganization, or updating of a mental representation of what is being communicated. We hypothesize that the P600 component is evoked by every word in a sentence as the lexical information activated by a word is integrated into the current mental representation. This results in an updated representation of the input given thus far. In other words, the processes of integration and interpretation that were assumed to underlie N400 amplitude on the *integration view* are now assumed to be reflected in the amplitude of the P600 instead. Integration difficulty, we suggest, is determined by how much the current mental representation needs to be adapted to incorporate the current input. This integration difficulty is not simply a function of the plausibility of a word in a given sentence. Although this kind of plausibility may be an important determinant of integration difficulty, there are also other, non-lexical factors at work, such as when a word or phrase introduces a novel entity into the discourse, or when

non-literal (e.g., ironic) meaning is computed at a given point in the sentence.

Differences in amplitude, latency, duration, and scalp distribution of the P600 component suggest that not every P600-effect is created alike (cf. Coulson et al., 1998; Gouvea et al., 2010). We would like to speculate that these electrophysiological properties of the P600 component correlate with the specific subprocesses that may underly the construction of a mental representation. Examples of such subprocesses are accommodating new discourse entities, establishing a relation between the entities and assigning them a thematic role, adding information to entities, revising already established relations, revising already assigned thematic roles, resolving conflicts between information sources (e.g., with respect to world knowledge), and so on. It is important to note that despite being different in kind, these processes have in common that they directly affect the representation of the current linguistic input, and hence they are all involved in the composition of an MRC.

The hypothesis we put forward here differs from the purely syntactic interpretation of P600 amplitude which is still dominant in the field of psycholinguistics. For instance, a recent review of the linguistic processes underlying the P600 component (Gouvea et al., 2010) concludes that the P600 amplitude is an index of the construction and deconstruction of syntactic relations (see also Kim and Osterhout, 2005; Kim and Sikos, 2011; Stroud, 2009; Stroud and Phillips, 2011). Clearly, syntactic complexities and anomalies do elicit a P600-effect, but we believe they do so because they create difficulties in constructing a mental representation. The MRC hypothesis on the amplitude of the P600 also differs from the way *multi-stream* models such as Monitoring Theory, CCA, and eADM interpret P600 amplitude. In-

creased P600 amplitude, in our view, does not result from a conflict between two or more processing streams. As sketched above, we suggest that the P600 component is the brain's natural electrophysiological reflection of updating a mental representation with new information. Finally, the MRC hypothesis is also different from the proposal put forward by Burkhardt (2006, 2007) and Schumacher (2011). Burkhardt assumes that semantic integration and reference computation is reflected in the amplitude of the N400, but that the organization and maintenance of the resulting representation takes place in a system she calls 'discourse memory'. On her account, an increase in P600 amplitude occurs whenever this system is taxed, for instance when a new discourse entity is introduced. Thus, she assumes two distinct levels of representation, one for the meaning of an utterance (reflected in N400 amplitude), and one for its discourse representation (reflected in the P600 amplitude). Under the MRC hypothesis, however, there is a single representation of what is communicated, and the ease of managing this representation with regard to the current input is reflected in P600 amplitude.

4. Semantic Illusions revisited

The 'Semantic Illusion' phenomenon has led to a paradigmatic shift from *single-stream* towards *multi-stream* models, as it seemed to suggest the existence of an autonomous semantic processing stream. However, from the perspective of the *single-stream* Retrieval-Integration account outlined above, there is no such thing as a 'Semantic Illusion' in any of the SIE studies reported in this paper. Observing a P600-effect instead of an N400-effect in response to sentences such as 'De speer heeft de atleten geworpen' (lit:

The javelin has the athletes <u>thrown</u>) does not entail that participants were temporarily entertaining an illusory interpretation. Rather, they immediately attempted to construct a representation of the anomalous event that is communicated in the sentence. We do not want to suggest, however, that people never perceive semantically anomalous input as meaningful. For instance, Erickson and Mattson (1981) presented participants with questions like 'How many animals of each kind did Moses take on the Ark?'. Most people answered 'two', missing the point that it was Noah, and not Moses, who sailed the Ark. Similarly, Barton and Sanford (1993) found that in response to questions like 'When an airplane crashes on a border with debris on both sides, where should the survivors be buried?', the answers often contained *locations*, as participants failed to realize that survivors should not be buried. These findings suggest that some participants processed these questions as if they made perfect sense, and failed to identify the anomaly. For these materials, the Retrieval-Integration account would predict that detecting the anomaly will have little or no effect on N400 amplitude, because N400 amplitude only depends on how much the retrieval of lexical features of the critical word is facilitated by its context. There should, however, be a modulation of P600 amplitude. If participants detect the anomaly, there should be a P600-effect relative to a non-anomalous control, indicating difficulty in constructing an interpretation of what is communicated (possibly this involves replacing *Moses* with *Noah*). If participants do not detect the anomaly, no such P600-effect should be observed, because the representations for the critical and control sentences should be equally difficult to build. Sanford et al. (2011) recently conducted an ERP investigation of the 'Moses Illusion'. They

presented participants with fragments like:

> '*Child abuse cases are being reported much more frequently these days. In a recent trial, a 10-year {sentence/care order} was given to the <u>victim</u>...*' (*'sentence'* = anomalous; *'care order'* = non-anomalous).

The authors compared the waveforms of participants that detected the anomaly, to those of participants that did not detect it. A P600-effect was reported for detected anomalies relative to undetected ones; there were no significant differences in N400 amplitude.

These results are consistent with the Retrieval-Integration account. First, no N400-effect is predicted for the critical words because retrieval of the critical word's meaning is predicted to be equally facilitated by the preceding words across the two conditions. Secondly, only if participants do identify the anomaly, a P600-effect should appear that reflects the increased processing involved in composing a coherent mental representation. No P600-effect should obtain if participants fail to see the anomaly. This pattern of results is precisely what was found in this study. Thus, genuine Semantic Illusions such as the 'Moses Illusion' affect P600 amplitude but not that of the N400.

5. Predictions and Future Work

On the basis of Retrieval-Integration account, one can make the following predictions:

N400 amplitude does not correlate with plausibility. According to the retrieval view, N400 amplitude reflects bottom-up memory-based activation

processes. Top-down information from the current mental representation can add to the activation patterns in memory, but does not constrain them (van Petten, 1993, 1995; van Berkum, 2009). Consequently, N400 amplitude for a critical word should be relatively insensitive to the plausibility of a sentence within which it is contained. This means that if one of two words makes a given sentence implausible while the other does not, there will not be an N400-effect if both are approximately equally primed by the preceding context. In other words, we predict that N400 amplitude does not correlate with the overall plausibility of a sentence if priming by preceding words and context is taken into account. A fine example consistent with this prediction was presented in the previous paragraph (i.e., the example taken from Sanford et al., 2011). This contrasts with the integration view which attributes the N400 to compositional and integrative processes, and therefore predicts N400 amplitude to be highly sensitive to plausibility.

P600 amplitude correlates with integration difficulty. If integration difficulty is not reflected in the amplitude of the N400, but in that of the P600 instead, every word should produce an increase in P600 amplitude, these increases being largest for words that make integration most difficult, that is, where the current mental representation requires the most substantial revision in order to incorporate the current linguistic input. For instance, larger P600 amplitudes should ensue when words or phrases introduce new referents or reference needs to be inferred, as compared to given entities (cf. Burkhardt, 2006). Also, an increase in P600 amplitude is expected whenever recovering what the speaker or reader means requires intensive pragmatic processing, as for example in the case of irony (Regel et al., 2011). A third of

source of increased integration difficulty are words that render their containing sentence implausible by violating world knowledge (as in the SIE data). Finally, a P600-effect is likely to arise when the position or the form of a word violates syntactic rules, making it hard to arrive at a coherent representation. The amount of integration will thus differ from word to word, as it depends crucially on how the syntactic, semantic and pragmatic information associated with this word necessitates a change of the existing MRC.

P600 amplitude increases at phrase and sentence boundaries. As we have discussed above, P600 amplitude will vary with specific lexical characteristics. In addition, we expect that there may also be positions in a sentence where — regardless of specific word characteristics — integration will be most intensive, namely at phrase and sentence boundaries. Thus, we believe that the so-called clause wrap-up and sentence wrap-up effects very likely reflect MRC composition, and future research should look into these processes more carefully, as to date these wrap-up effects have been treated as a nuisance rather than as a window to meaning construction.

Prosodic boundaries are similar to sentence and phrase boundaries in how they affect the P600. In spoken language, phrase and sentence boundaries often coincide with prosodic breaks. We want to speculate that the late positivity that is found at prosodic boundaries, known as the Closure Positive Shift (CPS) (Steinhauer et al., 1999; Mueller et al., 2005; Kerkhofs et al., 2007, 2008; Wolff et al., 2008; Liu et al., 2009; Bögels et al., 2010, 2011c,a,b), is actually part of the P600 component family. Prosodic breaks help language users structure their input. A construction in which two coordinated NPs are preceded by an adjective 'blue squares and triangles', for instance, is

structurally ambiguous because hat it can mean that both the squares and the triangles are blue or that only the squares are blue. A prosodic break after blue indicates the latter analysis, in which the scope of the adjective is restricted to the first NP. In other words, a prosodic break in this place signals that *blue squares* can be integrated as a complete referent in the mental representation that is being constructed. This suggests that the processes underlying the CPS might be the same as those reflected in P600 amplitude. This hypothesis is supported by similarity of the CPS in onset (around 400ms post stimulus onset), duration (approximately 600ms), and scalp distribution (centro-parietal) to the P600 component (see, e.g., Kerkhofs et al., 2007, figure 2).

P600 amplitude correlates with behavioral measures of processing difficulty. In sentences containing a syntactic anomaly, there is a correlation between behavioral measures of language comprehension, such as reading times or eye-tracking measures, and their electrophysiological counterparts (see Hoeks, 1999; Brown and Hagoort, 1999). Words that introduce a syntactic anomaly require more processing time, and also produce a larger P600 amplitude than correct controls. We hypothesize that this correlation between reading time and P600-effects extends beyond words inducing an ungrammaticality. We predict that P600 amplitude correlates with behavioral measures on a word-to-word basis in every sentence: in garden-path constructions, ungrammatical sentences and semantically anomalous sentences, but also in syntactically well-formed and semantically plausible sentences. Of course, behavioral measures need to be compared to a relevant control, as for instance reading times also include the time needed for lexical retrieval

processes.

The amplitude of the P600 strongly depends on task demands. Language users are very adaptive in finding out how to do a task optimally without investing too much effort. We expect that this will be no different when it comes to MRC composition. For instance, Kolk et al. (2003) found that removing the acceptability judgement task greatly reduced the P600-effect found for their Semantic Illusion stimuli. Thus, P600 amplitude will also vary as a function of the specific demands of the current task. For instance, taking part in conversation may well lead to a higher need for a coherent MRC, and thus to larger P600 amplitudes than reading an isolated sentence from a screen, as is the case in many ERP experiments. This is especially true if participants do not have a task for which proper comprehension of the sentences is necessary, but are merely asked to read for comprehension (see van Petten and Luka, 2011, for an overview).

The predictions given above suggest a host of factors that critically affect P600 amplitude. It is therefore imperative for future studies on the determinants of the P600 component to vary those systematically, or to control for them rigorously.

6. Conclusion

We have proposed a simple and parsimonious *single-stream* account of language processing, the Retrieval-Integration account. On this account, the N400 and the P600 component reflect two successive processing stages, in which the output of the retrieval phase (N400 amplitude) serves as input for integration (P600 amplitude). The N400 component thus reflects a retrieval

stage in which the syntactic properties and semantic features of a current word are 'retrieved' from long-term memory; the N400 amplitude does *not* reflect any integrative or compositional semantic processing. Rather, P600 amplitude reflects the integration of lexical information with the current semantic representation into an updated representation. We have argued that the integrative processes underlying the amplitude of the P600 can best be understood in terms of the construction, reorganization, or updating of a mental representation of what is being communicated in a sentence or story — a proposal we have labeled the MRC hypothesis. Language processing thus seems to be basically characterized by biphasic N400/P600 sequences occurring for every word in a sentence (see Kotchoubey, 2006, for a proposal in which all cognitive processing entails negative-positive cycles). Future research should address how this account can be extended to include other ERP-effects that have been found in language processing studies, such as the Early Left Anterior Negativity (ELAN), the Left Anterior Negativity (LAN), as well as Sustained Negativities (see Friederici, 2011; Steinhauer and Drury, 2011, for critical reviews).

7. Acknowledgements

References

Allen, M., Badecker, W., Osterhout, L., 2003. Morphological analysis in sentence processing: An ERP study. Lang. Cognitive Proc. 18, 405–430.

Baggio, G., Hagoort, P., 2011. The balance between memory and unification in semantics: A dynamic account of the N400. Lang. Cognitive Proc. 26, 1338–1367.

Barton, S.B., Sanford, A.J., 1993. A case study of anomaly detection: Shallow semantic processing and cohesion establishment. Mem. Cognition 21, 477–487.

Bögels, S., Schriefers, H., Vonk, W., Chwilla, D.J., 2011a. Pitch accents in context: How listeners process accentuation in referential communication. Neuropsychologia 49, 2022–2036.

Bögels, S., Schriefers, H., Vonk, W., Chwilla, D.J., 2011b. Prosodic breaks in sentence processing investigated by event-related potentials. Lang. Ling. Compass 5, 424–440.

Bögels, S., Schriefers, H., Vonk, W., Chwilla, D.J., 2011c. The role of prosodic breaks and pitch accents in grouping words during on-line sentence processing. J. Cognitive Neurosci. 23, 2447–2467.

Bögels, S., Schriefers, H., Vonk, W., Chwilla, D.J., Kerkhofs, R., 2010. The interplay between prosody and syntax in sentence processing: The case of subject- and object-control verbs. J. Cognitive Neurosci. 22, 1036–1053.

Bornkessel, I., Schlesewsky, M., 2006. The extended Argument Dependency Model: A neurocognitive approach to sentence comprehension across languages. Psychol. Rev. 113, 787–821.

Bornkessel, I., Schlesewsky, M., Friederici, A.D., 2003. Eliciting thematic reanalysis effects: The role of syntax-independent information during parsing. Lang. Cognitive Proc. 18, 269–298.

Bornkessel-Schlesewsky, I., Schlesewsky, M., 2008. An alternative perspective on semantic P600 effects in language comprehension. Brain Res. Rev. 59, 55–73.

Brown, C.M., Hagoort, P., 1993. The processing nature of the N400: Evidence from masked priming. J. Cognitive Neurosci. 5, 34–44.

Brown, C.M., Hagoort, P., 1999. On the electrophysiology of language comprehension: Implications for the human language system, in: Crocker, M.W., Pickering, M.J., Clifton Jr., C. (Eds.), Architectures and mechanisms for language processing. Cambridge University Press, pp. 213–237.

Burkhardt, P., 2006. Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. Brain Lang. 98, 159–168.

Burkhardt, P., 2007. The P600 reflects cost of new information in discourse memory. NeuroReport 18, 1851–1854.

Christiansen, M.H., Conway, C.M., Onnis, L., 2011. Similar neural correlates for language and sequential learning: Evidence from event-related brain potentials. Lang. Cognitive Proc. To appear.

Chwilla, D.J., Brown, C.M., Hagoort, P., 1995. The N400 as a function of the level of processing. Psychophysiology 32, 274–285.

Chwilla, D.J., Kolk, H.H.J., 2005. Accessing world knowledge: Evidence from N400 and reaction time priming. Cognitive Brain Res. 25, 589–606.

Coulson, S., King, J.W., Kutas, M., 1998. Expect the unexpected: Event-related brain response to morphosyntactic violations. Lang. Cognitive Proc. 13, 21–58.

Elman, J.L., 2004. An alternative view of the mental lexicon. Trends Cogn. Sci. 8, 301–306.

Elman, J.L., 2009. On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. Cognitive Sci. 33, 547–582.

Erickson, T.D., Mattson, M.E., 1981. From words to meaning: A semantic illusion. J. Verb. Learn. Verb. Be. 20, 540–551.

Federmeier, K.D., Kutas, M., 1999. A rose by any other name: Long-term memory structure and sentence processing. J. Mem. Lang. 41, 469–495.

Federmeier, K.D., Laszlo, S., 2009. Time for meaning: Electrophysiology provides insights into the dynamics of representation and processing in semantic memory. Psychol Learn. Motiv. 51, 1–44.

Felser, C., Clahsen, H., Münte, T.F., 2003. Storage and integration in the processing of filler-gap dependencies: An ERP study of topicalization and wh-movement in german. Brain Lang. 87, 345–354.

Ferreira, F., Patson, N.D., 2007. The 'good enough' approach to language comprehension. Lang. Ling. Compass 1, 71–83.

Fiebach, C.J., Schlesewsky, M., Friederici, A.D., 2002. Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. J. Mem. Lang. 47, 250–272.

Fischler, I., Bloom, P.A., Childers, D.G., Roucos, S.E., Perry Jr., N.W., 1983. Brain potentials related to stages of sentence verification. Psychophysiology 20, 400–409.

Frazier, L., 1987. Syntactic processing: Evidence from Dutch. Nat. Lang. Linguist. Th. 5, 519–559.

Friederici, A.D., 2011. The brain basis of language processing: From structure to function. Physiol. Rev. 91, 1357–1392.

Frisch, S., Schlesewsky, M., Saddy, D., Alpermann, A., 2002. The P600 as an indicator of syntactic ambiguity. Cognition 85, B83–B92.

Gernsbacher, M.A., 1990. Language Comprehension as Structure Building. Lawrence Erlbaum Associates, Hillsdale, NJ.

Gerrig, R.J., 2005. The scope of memory-based processing. Discourse Process. 39, 225–242.

Gerrig, R.J., McKoon, G., 1998. The readiness is all: The functionality of memory-based text processing. Discourse Process. 26, 67–86.

Gillund, G., Shiffrin, R.M., 1984. A retrieval model for both recognition and recall. Psychol. Rev. 91, 1–67.

Givón, T., 1992. The grammar of referential coherence as mental processing instructions. Linguistics 30, 5–55.

Gold, B.T., Balota, D.A., Jones, S.J., Powell, D.K., Smith, C.D., Andersen, A.H., 2006. Dissociation of automatic and strategic lexical-semantics: functional magnetic resonance imaging evidence for differing roles of multiple frontotemporal regions. J. Neurosci. 26, 6523–6532.

Gouvea, A.C., Phillips, C., Kazanina, N., Poeppel, D., 2010. The linguistic processes underlying the P600. Lang. Cognitive Proc. 25, 149–188.

Hagoort, P., Baggio, G., Willems, R.M., 2009. Semantic unification, in: Gazzaniga, M.S. (Ed.), The Cognitive Neurosciences. MIT Press, Cambridge, MA, 4th edition. pp. 819–836.

Hagoort, P., Brown, C.M., Groothusen, J., 1993. The Syntactic Positive Shift (SPS) as an ERP measure of syntactic processing. Lang. Cognitive Proc. 8, 439–483.

Hagoort, P., Hald, L., Bastiaansen, M., Petersson, K.M., 2004. Integration of word meaning and world knowledge in language comprehension. Science 304, 438–441.

Hagoort, P., Van Berkum, J.J.A., 2007. Beyond the sentence given. Philos. T. R. Soc. B 362, 801–811.

Halgren, E., Dhond, R.P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J.D., Dale, A.M., 2002. N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. Neuroimage 17, 1101–1116.

Hart Jr., J., Gordon, B., 1990. Delineation of single-word semantic comprehension deficits in aphasia, with anatomical correlation. Ann. Neurol. 27, 226–231.

Hintzman, D.L., 1988. Judgments of frequency and recognition memory in a multiple-trace memory model. Psychol. Rev. 95, 528–551.

Hoeks, J.C.J., 1999. The processing of coordination: semantic and pragmatic constraints on ambiguity resolution. Ph.D. thesis. Radboud University Nijmegen.

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 Res. 19, 59–73.

Indefrey, P., Levelt, W.J.M., 2004. The spatial and temporal signatures of word production components. Cognition 92, 101–144.

Johnson-Laird, P.N., 1983. Mental Models. Cambridge University Press, Cambridge, UK.

Kaan, E., Harris, A., Gibson, E., Holcomb, P.J., 2000. The P600 as an index of syntactic integration difficulty. Lang. Cognitive Proc. 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 Res. 17, 621–635.

Kamp, H., Reyle, U., 1993. From Discourse to Logic. Kluwer Academic, Dordrecht, NL.

Kerkhofs, R., Haselager, W.F.G., 2006. The embodiment of meaning. Manuscr. Rev. Int. Filos. 29, 753–764.

Kerkhofs, R., Vonk, W., Schriefers, H., Chwilla, D.J., 2007. Discourse, syntax, and prosody: The brain reveals an immediate interaction. J. Cognitive Neurosci. 19, 1421–1434.

Kerkhofs, R., Vonk, W., Schriefers, H., Chwilla, D.J., 2008. Sentence processing in the visual and auditory modality: Do comma and prosodic break have parallel functions? Brain Res. 1224, 102–118.

Kertesz, A., 1979. Aphasia and Associated Disorders: Taxonomy, Localization, and Recovery. Grune & Stratton Inc, New York.

Kim, A., Lai, V., 2011. Rapid interactions between lexical semantic and word form analysis during word recognition in context: Evidence from ERPs. J. Cognitive Neurosci. To appear.

Kim, A., Osterhout, L., 2005. The independence of combinatory semantic processing: Evidence from event-related potentials. J. Mem. Lang. 52, 205–225.

Kim, A., Sikos, L., 2011. Conflict and surrender during sentence processing: An ERP study of syntax-semantics interaction. Brain Lang. 118, 15–22.

Kintsch, W., 1988. The role of knowledge in discourse comprehension: A construction-integration model. Psychol. Rev. 95, 163–182.

Kintsch, W., 1992. How readers construct situation models for stories: The role of syntactic cues and causal inferences, in: Healy, A.F., Kosslyn, S.M.,

Shiffrin, R.M. (Eds.), From Learning Processes to Cognitive Processes: Essays in Honor of William K. Estes. Psychology Press, pp. 261–278.

Kintsch, W., 1998. Comprehension: A Paradigm for Cognition. Cambridge University Press, Cambridge, UK.

Kintsch, W., Mangalath, P., 2011. The construction of meaning. Top. Cognitive Sci. 3, 346–370.

Kolk, H.H.J., Chwilla, D.J., 2007. Late positivities in unusual situations. Brain Lang. 100, 257–261.

Kolk, H.H.J., Chwilla, D.J., van Herten, M., Oor, P.J.W., 2003. Structure and limited capacity in verbal working memory: A study with event-related potentials. Brain Lang. 85, 1–36.

Kos, M., Vosse, T., van den Brink, D., Hagoort, P., 2010. About edible restaurants: Conflicts between syntax and semantics as revealed by ERPs. Front. Psychol. 1, 1–11.

Kotchoubey, B., 2006. Event-related potentials, cognition, and behavior: A biological approach. Neurosci. Biobehav. R. 30, 42–65.

Kounios, J., Holcomb, P.J., 1992. Structure and process in semantic memory: Evidence from event-related brain potentials and reaction times. J. Exp. Psychol. Gen. 121, 459–479.

Kuperberg, G.R., 2007. Neural mechanisms of language comprehension: Challenges to syntax. Brain Res. 1146, 23–49.

Kuperberg, G.R., Caplan, D.N., Sitnikova, T., Eddy, M., Holcomb, P.J., 2006. Neural correlates of processing syntactic, semantic, and thematic relationships in sentences. Lang. Cognitive Proc. 21, 489–530.

Kuperberg, G.R., Choi, A., Cohn, N., Paczynski, M., Jackendoff, R., 2010. Electrophysiological correlates of complement coercion. J. Cognitive Neurosci. 22, 2685–2701.

Kuperberg, G.R., Kreher, D.A., Sitnikova, T., Caplan, D.N., Holcomb, P.J., 2007. The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. Brain Lang. 100, 223–237.

Kuperberg, G.R., Sitnikova, T., Caplan, D.N., Holcomb, P.J., 2003. Electrophysiological distinctions in processing conceptual relationships within simple sentences. Cognitive Brain Res. 17, 117–129.

Kutas, M., Federmeier, K.D., 2000. Electrophysiology reveals semantic memory use in language comprehension. Trends Cogn. Sci. 4, 463–470.

Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). Annu. Rev. Psychol. 62.

Kutas, M., van Petten, C., Kluender, R., 2006. Psycholinguistics electrified II: 1994–2005, in: Traxler, M.J., Gernsbacher, M.A. (Eds.), Handbook of Psycholinguistics. Elsevier, New York, 2nd edition. pp. 659–724.

Lau, E.F., Almeida, D., Hines, P.C., Poeppel, D., 2009. A lexical basis for N400 context effects: Evidence from MEG. Brain Lang. 111, 161–172.

Lau, E.F., Phillips, C., Poeppel, D., 2008. A cortical network for semantics: (de)constructing the N400. Nat. Rev. Neurosci. 9, 920–933.

Liu, B., Jin, Z., Li, W., Li, Y., Wang, Z., 2009. The pragmatic meanings conveyed by function words in chinese sentences: an ERP study. J. Neurolinguist. 22, 548–562.

Lotze, N., Tune, S., Schlesewsky, M., Bornkessel-Schlesewsky, I., 2011. Meaningful physical changes mediate lexical-semantic integration: Top-down and form-based bottom-up information sources interact in the N400. Neuropsychologia 49, 3573–3582.

MacDonald, M.C., Pearlmutter, N.J., Seidenberg, M.S., 1994. The lexical nature of syntactic ambiguity resolution. Psychol. Rev. 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 Res. 1093, 178–189.

McRae, K., Spivey-Knowlton, M.J., Tanenhaus, M.K., 1998. Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. J. Mem. Lang. 38, 283–312.

Morris, R.K., 1994. Lexical and message-level sentence context effects on fixation times in reading. J. Exp. Psychol. Learn. 20, 92.

Mueller, J.L., Hahne, A., Fujii, Y., Friederici, A.D., 2005. Native and non-native speakers' processing of a miniature version of Japanese as revealed by ERPs. J. Cognitive Neurosci. 17, 1229–1244.

Münte, T.F., H., H., Matzke, M., Wieringa, B.M., Johannes, S., 1998. Brain potentials and syntactic violations revisited: no evidence for specificity of the syntactic positive shift. Neuropsychologia 36, 217–226.

Neville, H., Nicol, J.L., Barss, A., Forster, K.I., Garrett, M.F., 1991. Syntactically based sentence processing classes: Evidence from event-related brain potentials. J. Cognitive Neurosci. 3, 151–165.

Nieuwland, M.S., van Berkum, J.J.A., 2005. Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. Cognitive Brain Res. 24, 691–701.

Núñez-Peña, M.I., Honrubia-Serrano, M.L., 2004. P600 related to rule violation in an arithmetic task. Cognitive Brain Res. 18, 130–141.

Osterhout, L., Holcomb, P.J., 1992. Event-related brain potentials elicited by syntactic anomaly. J. Mem. Lang. 31, 785–806.

Osterhout, L., Holcomb, P.J., 1993. Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. Lang. Cognitive Proc. 8, 413–437.

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. J. Exp. Psychol. Learn. 20, 786–803.

Otten, M., Nieuwland, M.S., van Berkum, J.J.A., 2007. Great expectations: Specific lexical anticipation influences the processing of spoken language. BMC Neurosci. 8, 89.

Patel, A.D., 2003. Language, music, syntax and the brain. Nat. Neurosci. 6, 674–681.

Phillips, C., Kazanina, N., Abada, S.H., 2005. ERP effects of the processing of syntactic long-distance dependencies. Cognitive Brain Res. 22, 407–428.

Price, C.J., Wise, R.J.S., Watson, J.D.G., Patterson, K., Howard, D., Frackowiak, R.S.J., 1994. Brain activity during reading: The effects of exposure duration and task. Brain 117, 1255–1269.

Pugh, K.R., Shaywitz, B.A., Shaywitz, S.E., Constable, R.T., Skudlarski, P., Fulbright, R.K., Bronen, R.A., Shankweiler, D.P., Katz, L., Fletcher, J.M., Gore, J.C., 1996. Cerebral organization of component processes in reading. Brain 119, 1221–1238.

Pulvermüller, F., 1999. Words in the brain's language. Behav. Brain Sci. 22, 252–279.

Pulvermüller, F., 2001. Brain reflections of words and their meaning. Trends Cogn. Sci. 5, 517–524.

Ratcliff, R., 1978. A theory of memory retrieval. Psychol. Rev. 85, 59–108.

Regel, S., Gunter, T.C., Friederici, A.D., 2011. Isn't it ironic? an electrophysiological exploration of figurative language processing. J. Cognitive Neurosci. 23, 277–293.

Rogers, T.T., McClelland, J.L., 2004. Semantic cognition: A parallel distributed processing approach. The MIT Press, Cambridge, MA.

Sanford, A.J., Leuthold, H., Bohan, J., Sanford, A.J.S., 2011. Anomalies at the borderline of awareness: An ERP study. J. Cognitive Neurosci. 23, 514–523.

Schumacher, P.B., 2011. The hepatitis called...: Electrophysiological evidence for enriched composition, in: Meibauer, J., Steinbach, M. (Eds.), Experimental Pragmatics/Semantics. John Benjamins, Amsterdam, NL, pp. 199–219.

Schwanenflugel, P.J., Harnishfeger, K.K., Stowe, R.W., 1988. Context availability and lexical decisions for abstract and concrete words. J. Mem. Lang. 27, 499–520.

Schwanenflugel, P.J., Shoben, E.J., 1983. Differential context effects in the comprehension of abstract and concrete verbal materials. J. Exp. Psychol. Learn. 9, 82–102.

Schwanenflugel, P.J., Shoben, E.J., 1985. The influence of sentence constraint on the scope of facilitation for upcoming words. J. Mem. Lang. 24, 232–252.

Steinhauer, K., Alter, K., Friederici, A.D., 1999. Brain potentials indicate immediate use of prosodic cues in natural speech processing. Nat. Neurosci. 2, 191–196.

Steinhauer, K., Drury, J.E., 2011. On the early left-anterior negativity (ELAN) in syntax studies. Brain Lang. To appear.

Stroud, C.M.A., 2009. Structural and semantic selectivity in the electrophysiology of sentence comprehension. Ph.D. thesis. University of Maryland.

Stroud, C.M.A., Phillips, C., 2011. Examining the evidence for an independent semantic analyzer: An ERP study in Spanish. Brain Lang. To appear.

van Berkum, J.J.A., 2009. The 'neuropragmatics' of simple utterance comprehension: An ERP review, in: Sauerland, U., Yatsushiro, K. (Eds.), Semantics and Pragmatics: From Experiment to Theory. Palgrave Macmillan, New York, pp. 276–316.

van Berkum, J.J.A., 2010. The brain is a prediction machine that cares about good and bad–Any implications for neuropragmatics? Ital. J. Ling. 22, 181–208.

van Berkum, J.J.A., Hagoort, P., Brown, C.M., 1999. Semantic integration in sentences and discourse: Evidence from the N400. J. Cognitive Neurosci. 11, 657–671.

van de Meerendonk, N., Kolk, H.H.J., Chwilla, D.J., Vissers, C.T.W.M., 2009. Monitoring in language perception. Lang. Ling. Compass 3, 1211–1224.

van de Meerendonk, N., Kolk, H.H.J., Vissers, C.T.W.M., Chwilla, D.J., 2010. Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns. J. Cognitive Neurosci. 22, 67–82.

van Dijk, T.A., Kintsch, W., 1983. Strategies of Discourse Comprehension. Academic Press, New York.

van Gompel, R.P.G., Pickering, M.J., Traxler, M.J., 2000. Unrestricted race: A new model of syntactic ambiguity resolution, in: Kennedy, A., Radach,

R., Heller, D., Pynte, J. (Eds.), Reading as a Perceptual Process. Elsevier, Oxford, pp. 621–648.

van Herten, M., Chwilla, D.J., Kolk, H.H.J., 2006. When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. J. Cognitive Neurosci. 18, 1181–1197.

van Herten, M., Kolk, H.H.J., Chwilla, D.J., 2005. An ERP study of P600 effects elicited by semantic anomalies. Cognitive Brain Res. 22, 241–255.

van Petten, C., 1993. A comparison of lexical and sentence-level context effects in event-related potentials. Lang. Cognitive Proc. 8, 485–531.

van Petten, C., 1995. Words and sentences: Event-related brain potential measures. Psychophysiology 32, 511–525.

van Petten, C., Luka, B.J., 2006. Neural localization of semantic context effects in electromagnetic and hemodynamic studies. Brain Lang. 97, 279–293.

van Petten, C., Luka, B.J., 2011. Prediction during language comprehension: Benefits, costs, and ERP components. Int. J. Psychophysiol. To appear.

Vissers, C.T.W.M., Chwilla, D.J., Kolk, H.H.J., 2006. Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. Brain Res. 1106, 150–163.

Vissers, C.T.W.M., Chwilla, D.J., Kolk, H.H.J., 2007. The interplay of heuristics and parsing routines in sentence comprehension: Evidence from ERPs and reaction times. Biol. Psychol. 75, 8–18.

Wang, L., Hagoort, P., Yang, Y., 2009. Semantic illusion depends on information structure: ERP evidence. Brain Res. 1282, 50–56.

Wolff, S., Schlesewsky, M., Hirotani, M., Bornkessel-Schlesewsky, I., 2008. The neural mechanisms of word order processing revisited: Electrophysiological evidence from Japanese. Brain Lang. 107, 133–157.

Ye, Z., Zhou, X., 2008. Involvement of cognitive control in sentence comprehension: Evidence from ERPs. Brain Res. 1203, 103–115.

Zwaan, R.A., 1999. Situation models: The mental leap into imagined worlds. Curr. Dir. Psychol. Sci. 8, 15–18.

Zwaan, R.A., 2003. The immersed experiencer: Toward an embodied theory of language comprehension. Psychol. Learn. Motiv. 44, 35–62.

Zwaan, R.A., Langston, M.C., Graesser, A.C., 1995. The construction of situation models in narrative comprehension. Psychol. Sci. 6, 292–297.

Zwaan, R.A., Radvansky, G.A., 1998. Situation models in language comprehension and memory. Psychol. Bull. 123, 162–185.

Model	Stream(s)	Absence of N400-effect	P600-effect reflects
SA*	Syntax-driven and semantics-driven	Plausible combination of arguments and verb	Syntactic revision
MT	Algorithmic stream and plausibility heuristic	Plausible combination of arguments and verb	Conflict resolution
CCA*	Syntax-driven, thematic-role based, and semantic-memory based	Blocking of semantic integration	Continued Combinatory Analysis
eADM	Thematic-role based and plausibility heuristics	Plausible combination of arguments and verb	Problematic integration of streams
PC*	Syntax-driven and semantics-driven	Strong semantic cues	Syntactic processing

Table 1: Summary of the different models/accounts, and their explanations of the absence of an N400-effect and the presence of a P600-effect in SIE data (SA = Semantic Attraction; MT = Monitoring Theory; CCA = Continued Combinatory Analysis; eADM = extended Argument Dependency Model, and PC = Processing Competition. Models marked with a '*' are fully interactive, meaning that their streams can influence each other during online processing.

Item(s)	Observed	SA	MT	CCA	eADM	PC	RI
Hoeks et al. (2004)							
De speer werd door de atleten geworpen	—	—	—	—	—	—	—
De speer heeft de atleten geworpen	P6	P6	P6	P6	P6	P6	P6
De speer werd door de atleten opgesomd	N4/P6	N4	N4	N4	N4/P6	P6	N4/P6
De speer heeft de atleten opgesomd	N4/P6	N4	N4	N4	N4/P6	P6	N4/P6
Kim and Osterhout (2005)							
The hearty meal was devoured...	—	—	—	—	—	—	—
The hearty meal was devouring...	P6	P6	P6	P6	P6	N4	P6
The dusty tabletops were devouring...	N4 (N4/P6)[a]	N4	N4	P6	N4/P6	N4	N4/P6
van Herten et al. (2005)							
De stroper die op de vos joeg...	—	—	—	—	—	—	—
De vos die op de stroper joeg...	P6	—	P6	P6	P6	P6	P6
Kuperberg et al. (2007)							
For breakfast the boys would eat...	—	—	—	—	—	—	—
For breakfast the boys would watch...	N4	N4	N4	N4	N4	N4	N4
For breakfast the eggs would eat...	P6	P6	P6	P6	P6	P6	P6
For breakfast the eggs would watch...	P6 (N4/P6)	N4	N4	P6	P6	P6	(N4)/P6
Kos et al. (2010)							
Fred eet een boterham...	—	—	—	—	—	—	—
Fred eet een restaurant...	N4	N4	N4	N4	P6	N4	N4
Fred eet in een restaurant...	—	—	—	—	—	—	—
Fred eet in een boterham...	N4	N4	P6	N4	??	N4	N4
Nieuwland and van Berkum (2005)							
Prior context...							
Next, the woman told the tourist...	—	—	—	—	—	—	—
Next, the woman told the suitcase...	P6	N4	N4	N4	N4/P6	N4	P6

Table 2: Model predictions (SA = Semantic Attraction; MT = Monitoring Theory; CCA = Continued Combinatory Analysis; eADM = extended Argument Dependency Model; PC = Processing Competition, and RI = Retrieval-Integration account. N4 = N400 and P6 = P600. Gray cells indicate wrong predictions, and the cell containing question marks denotes that no predictions could be made.). See text for literal translation of the items.

---

[a]Stroud (2009) found a biphasic N400-P600 in response to this condition.