Précis of *The Electrophysiology of Language Comprehension: A Neurocomputational Model*

Harm Brouwer
Saarland University
brouwer@coli.uni-saarland.de

Abstract | One decade ago, researchers using Event-Related brain Potential (ERP) measurements stumbled upon what looked like a *Semantic Illusion* in language comprehension: Semantically anomalous, but otherwise well-formed sentences did not affect the meaning-related N400 component, but instead increased the amplitude of the structure-related P600 component. This finding spawned five new models of language comprehension, all of which claim that instead of a single comprehension process, there are two or even more separate processing streams, one of which is not driven by structure, but by word meaning alone. In my thesis *The Electrophysiology of Language Comprehension: A Neurocomputational Model*, I make a case for rethinking the functional role of the N400 and the P600, thereby providing a much simpler way to account for these data. As a 'proof of concept', I present a neurocomputational model that directly instantiates a functional-anatomic mapping of the proposed reinterpretations of N400 and the P600 onto a minimal cortical network for language processing; simulations show that this model it is able to predict all relevant ERP patterns found in the literature. These results have important implications for our understanding of the human language comprehension system.

1 Introduction

Neurophysiological activity can be measured online from the scalp using electroencephalography (EEG). A single EEG recording reflects thousands of brain processes in parallel. For the investigation of a single stimulus, like an image, a sound, or a word in a sentence, raw EEGs are therefore not very useful. However, when averaging over numerous similar EEG recordings, random activity is filtered out, and what is left reflects the neurophysiological activity as elicited by a stimulus. This residual activity is referred to as an Event-Related brain Potential, or in short an ERP.

In electrophysiological research into language comprehension, there are two central ERP responses. The first is the N400 component, a negative deflection of the ERP signal that peaks around 400 ms after stimulus onset, and that is sensitive to semantic anomalies such as 'He spread his warm bread with socks' (relative to *butter*; Kutas and Hillyard, 1980). The second is the P600 component, a positive deflection that can be maximal around 600 ms, and

that was found in response to syntactic violations such as 'The spoilt child throw [...]' (relative to *throws*; Hagoort et al., 1993). The idea that semantic processing difficulty is reflected in the N400 component, and syntactic processing difficulty in the P600 component, has been central in the psycholinguistic literature since the discovery of these components. Ten years ago, however, findings emerged that presented a challenge to this mapping. Around 2003, more and more research groups discovered that certain types of syntactically sound, but semantically anomalous sentences failed to produce an N400-effect, but produced a P600-effect instead (e.g., Kolk et al., 2003; Kuperberg et al., 2003; Hoeks et al., 2004; Kim and Osterhout, 2005). Hoeks et al. (2004), for instance, found that Dutch sentences such as 'De speer heeft de atleten geworpen' (lit: 'The javelin has the athletes thrown') produced an increase in P600 amplitude (relative to a non-anomalous control), but not in N400 amplitude. This was unexpected, because as javelins do not throw athletes, the word *thrown* should create semantic processing difficulty, and hence an increase in N400 amplitude. Equally unexpected was the finding of an effect on P600 amplitude in the absence of a syntactic anomaly. 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.

To account for these 'Semantic Illusion' effects, five complex language processing models were proposed that incorporate multiple, potentially interacting processing streams (*Monitoring Theory*: Kolk et al., 2003; *Semantic Attraction*: Kim and Osterhout, 2005; *Continued Combinatory Analysis*: Kuperberg, 2007; the *extended Argument Dependency Model*: Bornkessel-Schlesewsky and Schlesewsky, 2008, and the *Processing Competition* account: Hagoort et al., 2009). What these models have in common is that they include a processing stream that is purely semantic, unconstrained by any structural information (e.g., word order, agreement, case marking). This independent semantic analysis stream does not run into semantic processing problems on the word *thrown*, and hence does not produce an N400-effect, because the words *javelin*, *athletes*, and *thrown* fit together well semantically. Eventually, the processor does realize that something is wrong with the interpretation that was constructed, and the effort put into solving this problem is reflected in a P600-effect. Critically, these multi-stream models have very different architectural properties, and seem highly incompatible. However, as they are only specified at the 'box-and-arrow' level, it has proven difficult to decide between them. The aim of the research presented in my thesis *The Electrophysiology of Language Comprehension: A Neurocomputational Model* (Brouwer, 2014) was to determine using computational modeling, which one of these models is most viable, thereby offering a formally precise explanation of the 'Semantic Illusion' in language processing.

[1]The term 'Semantic Illusion' was adapted from a study by Erickson and Mattson (1981).

2 Rethinking the N400 and the P600

The first part of my thesis provides a critical review of the five multi-stream models that have been put forward to explain the 'Semantic Illusion'-effect (this part is published as Brouwer et al., 2012). This review shows that while all five models can account for a subset of the relevant data, none of them covers the full range of results found in the literature. The reason for this failure is argued not to be architectural in nature, but rather due to an assumption that is common to all five models, namely that the N400 component indexes some form of semantic integration or semantic combinatorial processing. Based on recent evidence, I propose that the N400 rather reflects a *non-combinatorial* (or *non-compositional*) memory retrieval process (see Federmeier and Laszlo, 2009; Kutas and Federmeier, 2000, 2011; van Berkum, 2009, for overviews). On the *memory retrieval view* of the N400 component, N400 amplitude reflects the ease with which the conceptual information associated with an incoming word can be retrieved from long-term memory. Ease of retrieval is, among other things, determined by the retrieval cues present in a word's prior context. Retrieval is facilitated if the conceptual knowledge associated with an incoming word is consistent with the conceptual knowledge already activated by the preceding context, and, conversely, retrieval is not facilitated when the features of this word are not activated by the context. For Semantic Illusion sentences such as "De speer heeft de atleten geworpen" (lit: The javelin has the athletes thrown), the ease with which the lexical features of the critical verb–e.g., *thrown*–can be retrieved from memory depends on conceptual cues in its prior context–e.g., *javelin* and *athletes*–as well as cues from scenario-based world knowledge–e.g., *javelins are usually thrown by athletes*. These retrieval cues should be very similar for the critical verb in the corresponding control sentences, e.g., "De speer werd door de atleten geworpen" (lit: The javelin was by the athletes thrown). The lexical features of the critical verb–e.g., *thrown*–should thus be equally easy to retrieve in the critical and the control sentences, yielding no difference in N400 amplitude, and hence no N400-effect. This provides a parsimonious explanation for the absence of an N400-effect in Semantic Illusion sentences, but also raises an important question: If the N400 component does not reflect integration–or combinatorial/compositional–processing, then how and when does integration of information from multiple sources (e.g., the meaning of the current word with its prior context) take place? As semantic integration (i.e., the creation of a semantic representation of the language input) is without doubt *the* central task of the language comprehension system, it would be very unlikely that it does not show up in ERPs.

I hypothesize that these integrative processes are reflected in P600 amplitude. Under this hypothesis, the P600 component is assumed to be a family of (late) positivities that reflect the effort involved in the word-by-word construction, reorganization, or updating of a 'mental representation of what is being communicated' (MRC for short). MRC composition requires little effort if the existing representation can be straightforwardly augmented to incorporate the information contributed by the incoming word. It is effortful, on the other hand, when the existing representation needs to be reorganized, supplemented with, for

instance, a novel discourse referent, or when the resulting representation does not make sense in light of our knowledge about the world. This last aspect explains the presence of a P600-effect in response to Semantic Illusion sentences like "De speer heeft de atleten geworpen" (lit: The javelin has the athletes thrown) relative to its control "De speer werd door de atleten geworpen" (lit: The javelin was by the athletes thrown). Integration of the critical word leads to a representation that does not make sense in light of what we know about the world (javelins are inanimate and cannot throw athletes), and raises the question of what the speaker meant to communicate with this sentence. Did we perhaps misunderstand the speaker, and did the athletes throw the javelin after all? Are we dealing with non-literal language use, as is the case in irony (cf. Regel et al., 2011; Spotorno et al., 2013)? Or did the speaker really mean that some animated javelin was throwing athletes? Hence, in order for the resulting interpretation to be meaningful, we need to recover what the speaker meant to communicate. These recovery processes lead to an increase in P600 amplitude, and hence a P600-effect relative to control. Importantly, the MRC hypothesis of the P600 component predicts that P600 amplitude is sensitive to combinatorial semantic processing in general, and not only to semantic anomaly. This is consistent with evidence from recent studies investigating the incremental processing of atypical, but non-anomalous sentences (e.g., Urbach and Kutas, 2010; Molinaro et al., 2012).

The views on the N400 and the P600 that were described above are combined in the Retrieval-Integration account. Under this account, language comprehension proceeds in biphasic N400/P600 cycles, brought about by the retrieval and subsequent integration of the information associated with each incoming word. Every word thus modulates N400 amplitude, reflecting the ease with which its lexical information can be retrieved, as well as P600 amplitude, reflecting the effort involved in integrating a word's meaning with a representation of its prior context. The result of this N400/P600 cycle is an updated representation of what is being communicated in the unfolding discourse thus far, which will itself provide a context for a next word. The Retrieval-Integration account has been shown to have broad empirical coverage (Brouwer et al., 2012; Hoeks and Brouwer, 2014).

3 Mapping function to anatomy

In the second part of my thesis, I aligned the Retrieval-Integration view on the N400 and the P600 with neuroanatomy, and suggested that the processes of retrieval and semantic integration are mediated by specific brain areas (this part is published as Brouwer and Hoeks, 2013). First of all, I propose that the left posterior Middle Temporal Gyrus (lpMTG; BA 21) serves as a network *hub* that mediates lexical retrieval ($\sim$N400). Such a hub (cf. Buckner et al., 2009), or epicenter (cf. Mesulam, 1990, 1998), is a brain region that serves to integrate or bind together information from various neighbouring areas (see also Damasio, 1989), and to broadcast this information across larger neuroanatomical networks. The lpMTG finds itself in the middle of brain areas that constitute long-term memory. Words (and also other meaningful stimuli) activate parts of that network, and the resulting information is col-

lected through the lpMTG and shared with other brain networks for further (higher-level) processing. On the Retrieval-Integration account, this information is used to create a valid and coherent mental representation of what is communicated (an MRC). I hypothesize that the construction of such an MRC takes place in and around the left Inferior Frontal Gyrus (lIFG; BA 44/45/47). That is, the lIFG is a hub mediating MRC composition, and thus the most prominent source of the P600. Again, more areas may be involved in making sense, but the lIFG is the most central one, binding together the information from surrounding neural territory.

The information sharing between the two hubs (i.e., from the lpMTG to the lIFG and back) occurs via white matter tracts that structurally connect them. There are two major white matter pathways connecting the lpMTG and the lIFG, the dorsal pathway (dp) and the ventral pathway (vp), the precise functional roles of which are still poorly understood (see Brouwer and Hoeks, 2013, for discussion). Nonetheless, an approximate functional-anatomic Retrieval-Integration cycle of an incoming word can be described (see Figure 1, top). First, a word reaches the lpMTG via either the auditory cortex (ac) or the visual cortex (vc), depending on the input modality. The lpMTG then retrieves the conceptual knowledge associated with this word from the association cortices and binds it together, a process that generates the N400 component. Next, the retrieved knowledge is shared with the lIFG, via one of the white matter pathways, where it is integrated with the prior context. This process is assumed to generate the P600 component. Finally, the new representation feeds back to the lpMTG causing pre-activation of conceptual knowledge in memory.

4 A neurocomputational model

The Retrieval-Integration account developed in the first two parts of my thesis is, in comparison to the five multi-stream models, theoretically the most parsimonious, has the broadest empirical coverage, and seems to fit well with what is presently known about the neuroanatomy of language. However, what it has in common with the other models, is that it is still a conceptual *box-and-arrow* model. This means that the predicted outcome of the model in any specific case is at best *qualitative*, and may be affected by implicit biases and other subjective factors; that is, researchers may for instance disagree on the amount of contextual and lexical priming in any specific sentence, and their predictions on the presence or absence of an N400-effect may vary. The only way to overcome this problem is to implement the model computationally, which means giving a formally precise description of the mechanisms that are supposed to underlie it, and then running this computational model to generate *quantitative* predictions. The predictions in terms of N400 amplitude and P600 amplitude can then be compared to the actual results of empirical studies.

In the third part of my thesis, I present such a neurocomputational model that predicts the amplitude of the N400 and the P600 component at every word of a sentence. This model directly instantiates the functional-anatomic mapping of the Retrieval-Integration account described above. The neurocomputational model consists of five layers of artificial

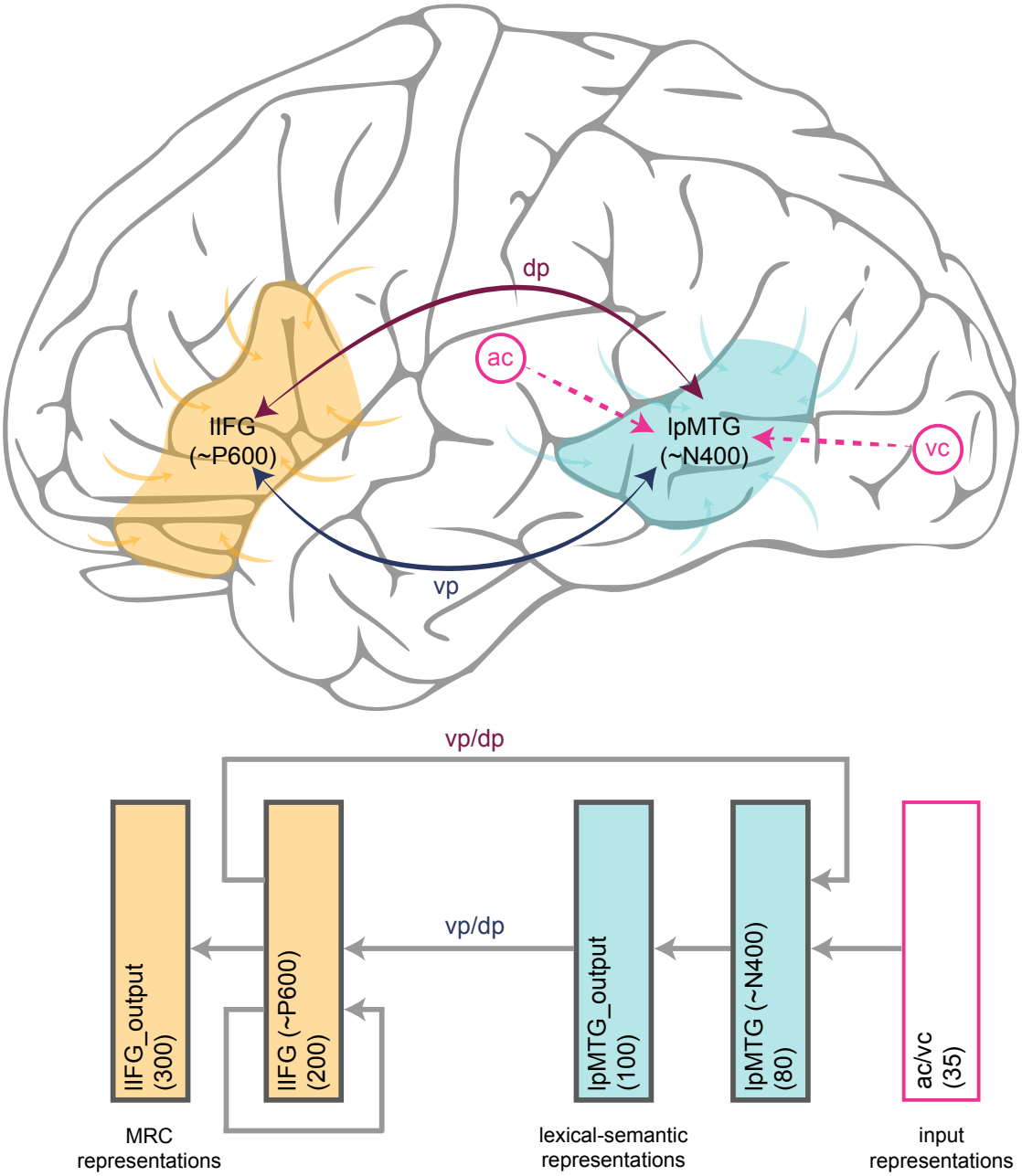

Figure 1: **Top:** A functional-anatomic Retrieval-Integration cycle. See section 3 for details. **Bottom:** A neurocomputational model. See section 4 for details.

neurons, implementing two connected but relatively independent sub-systems: A system for retrieval (∼lpMTG) and a system for integration (∼lIFG) (see Figure 1, bottom). The retrieval system is trained to map words onto their lexical-semantic representations (extracted from a corpus using the Correlated Occurrence Analogue to Lexical Semantics, COALS; Rohde et al., 2009). The integration system, in turn, is trained to map these lexical-semantic representations onto an approximation of the 'meaning' of a sentence: a thematic role assignment in terms of the *agent*, *action*, and *patient* (i.e., *who*-did-*what*-to-*whom/what*). Importantly, the mapping of words onto their lexical-semantic representations in the retrieval system (∼lpMTG), can be facilitated by lexical and higher level cues that are present in the unfolding representation in the integration system (∼lIFG). Moreover, the model is taught that any noun phrase can theoretically be an *agent* or a *patient*, but that there are certain stereotypical combinations of *agents*, *patients*, and *actions* (∼minimal world knowledge, see also Mayberry et al., 2009).

On the Retrieval-Integration account, N400 amplitude is a measure of 'unpreparedness'. If no features relevant to an incoming word are pre-activated, N400 amplitude will be maximal; if the lexical-semantic features of an incoming word are consistent with those pre-activated in memory, N400 amplitude will be reduced. Hence, N400 amplitude is a measure of how much the activation pattern in memory changes due to the processing of an incoming word. As such, the correlates of N400 amplitude are computed at the lpMTG layer, where the activation of lexical-semantic features takes place (∼memory retrieval), as the degree to which the pattern of activated features induced by the current word, given the unfolding context, and that induced by the previous word are *different* (see Brouwer, 2014, for mathematical details). P600 amplitude, in turn, reflects the difficulty of establishing coherence. The more the current interpretation (the current MRC) needs to be reorganized or augmented in order to become coherent, the higher P600 amplitude. Hence, P600 amplitude is effectively a measure of how much the representation of the unfolding state of affairs changes due to the integration of an incoming word. As such, the correlates of P600 amplitude are computed as the difference between the previous and the current state of affairs at the lIFG layer, where the (re)construction of an MRC—in terms of a thematic-role assignment—takes place (see also Crocker et al., 2010).

Critically, I show on the basis of simulations of an ERP experiment by Hoeks et al. (2004) that the model is able to produce the two most important patterns of ERP-effects as reported in the literature, namely isolated P600-effects and biphasic N400/P600-effects (see Kutas et al., 2006, for an overview), thereby providing a 'proof of concept' of the Retrieval-Integration account.

5 Conclusions

The work presented in my thesis has important implications for neurocognitive models of language processing. First of all, my simulations confirm that there is no need for an independent semantic analysis stream to explain "Semantic Illusions" in sentence processing.

That is, the model proposes that there is a continuous process of making sense that takes place in the lIFG and that generates the P600, the amplitude of which is proportional to the amount of effort needed to (re-)construct a mental representation of what is being communicated. Each time a word (or any other meaningful stimulus) comes in, this triggers a memory search for information associated with that stimulus, a search mediated by the lpMTG, which generates the N400, the amplitude of which reflects the amount of effort needed to retrieve this meaningful information. If the meaning of an incoming stimulus is primed, retrieval will be easy, and N400 amplitude will be small (less negative). This retrieved information is then used by the lIFG to update the mental representation of what is communicated. The Retrieval-Integration model is both architecturally more parsimonious than previously proposed models, and has broader empirical coverage (see Brouwer et al., 2012; Hoeks and Brouwer, 2014). Critically, neurocomputational simulations show that the Retrieval-Integration account also makes the right *quantitative* predictions, thereby providing a significant improvement on previously proposed 'box-and-arrow' models that are limited to *qualitative* predictions about the presence or absence of ERP effects.

The Retrieval-Integration model thus provides a comprehensive framework for further investigation of language processing, and will also serve as a starting point for a more elaborate model of human language comprehension, incorporating other language-related ERP components as well as other language-sensitive cortical and sub-cortical brain regions.

Acknowledgements

My thesis was supervised by John Hoeks and John Nerbonne. All research was done in collaboration with John Hoeks. The work described in chapters 1-3 (summarized in section 2 of this précis) also involved collaboration with Hartmut Fitz. The research was funded by the Netherlands Organisation for Scientific Research (NWO) PGW grant 10-26.

References

Bornkessel-Schlesewsky, I. and Schlesewsky, M. (2008). An alternative perspective on semantic P600 effects in language comprehension. *Brain Research Reviews*, 59(1):55–73.

Brouwer, H. (2014). *The Electrophysiology of Language Comprehension: A Neurocomputational Model*. PhD thesis, University of Groningen.

Brouwer, H., Fitz, H., and Hoeks, J. C. J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, 1446:127–143.

Brouwer, H. and Hoeks, J. C. J. (2013). A time and place for language comprehension: Mapping the N400 and the P600 to a minimal cortical network. *Frontiers in Human Neuroscience*, 7:758.

Buckner, R. L., Sepulcre, J., Talukdar, T., Krienen, F. M., Liu, H., Hedden, T., Andrews-Hanna, J. R., Sperling, R. A., and Johnson, K. A. (2009). Cortical hubs revealed by intrinsic functional connectivity: Mapping, assessment of stability, and relation to Alzheimer's disease. *The Journal of Neuroscience*, 29(6):1860–1873.

Crocker, M. W., Knoeferle, P., and Mayberry, M. R. (2010). Situated sentence processing: The coordinated interplay account and a neurobehavioral model. *Brain and Language*, 112(3):189–201.

Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, 33(1):25–62.

Erickson, T. D. and Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, 20(5):540–551.

Federmeier, K. D. and Laszlo, S. (2009). Time for meaning: Electrophysiology provides insights into the dynamics of representation and processing in semantic memory. *Psychology of Learning and Motivation*, 51:1–44.

Hagoort, P., Baggio, G., and Willems, R. M. (2009). Semantic unification. In Gazzaniga, M. S., editor, *The Cognitive Neurosciences, 4th ed.*, pages 819–836. MIT Press.

Hagoort, P., Brown, C., and Groothusen, J. (1993). The Syntactic Positive Shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8(4):439–483.

Hoeks, J. C. J. and Brouwer, H. (2014). Electrophysiological research on conversation and discourse processing. In Holtgraves, T. M., editor, *The Oxford Handbook of Language and Social Psychology*, pages 365–386. New York: Oxford University Press.

Hoeks, J. C. J., Stowe, L. A., and Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, 19(1):59–73.

Kim, A. and Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52(2):205–225.

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

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

Kuperberg, G. R., Sitnikova, T., Caplan, D., and Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, 17(1):117–129.

Kutas, M. and Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4(12):463–470.

Kutas, M. and Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, 62:621–647.

Kutas, M. and Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427):203–205.

Kutas, M., van Petten, C., and Kluender, R. (2006). Psycholinguistics electrified II: 1994–2005. In Traxler, M. J. and Gernsbacher, M. A., editors, *Handbook of Psycholinguistics, 2nd Edition*, pages 659–724. Elsevier, New York.

Mayberry, M. R., Crocker, M. W., and Knoeferle, P. (2009). Learning to attend: A connectionist model of situated language comprehension. *Cognitive Science*, 33(3):449–496.

Mesulam, M. M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Annals of Neurology*, 28(5):597–613.

Mesulam, M. M. (1998). From sensation to cognition. *Brain*, 121(6):1013.

Molinaro, N., Carreiras, M., and Duñabeitia, J. A. (2012). Semantic combinatorial processing of non-anomalous expressions. *NeuroImage*, 59(4):3488–3501.

Regel, S., Gunter, T. C., and Friederici, A. D. (2011). Isn't it ironic? An electrophysiological exploration of figurative language processing. *Journal of Cognitive Neuroscience*, 23(2):277–293.

Rohde, D. L. T., Gonnerman, L. M., and Plaut, D. C. (2009). An improved model of semantic similarity based on lexical co-occurrence. *Cognitive Science*, pages 1–33.

Spotorno, N., Cheylus, A., Van Der Henst, J., and Noveck, I. A. (2013). What's behind a P600? Integration operations during irony processing. *PloS ONE*, 8(6):e66839.

Urbach, T. P. and Kutas, M. (2010). Quantifiers more or less quantify on-line: ERP evidence for partial incremental interpretation. *Journal of Memory and Language*, 63(2):158–179.

van Berkum, J. J. A. (2009). The 'neuropragmatics' of simple utterance comprehension: An ERP review. In Sauerland, U. and Yatsushiro, K., editors, *Semantics and Pragmatics: From experiment to theory*, pages 276–316. Palgrave Macmillan, Basingstoke.