The Electrophysiology of Language Comprehension: A Neurocomputational Model

Harm Brouwer

The work in this thesis has been carried out under the auspices of the Research School of Behavioural and Cognitive Neurosciences (BCN) and the Center for Language and Cognition Groningen (CLCG). Both are affiliated with the University of Groningen.

The work in this thesis has been funded by the Netherlands Organisation for Scientific Research (NWO) through PGW grant 10-26 (322-70-002).

Groningen Dissertations in Linguistics 123
ISSN: 0928-0030
ISBN: 978-90-367-6892-4 (printed version)
ISBN: 978-90-367-6893-1 (electronic version)

Document prepared with LaTeX 2ε and typeset by pdfTeX (TeX Gyre Pagella font)
Cover design: Harm Brouwer and Oebele Brouwer
Printed by: RCG grafimedia, Groningen, The Netherlands — `http://www.rcg.nl/`

rijksuniversiteit
groningen

The Electrophysiology of Language Comprehension
A Neurocomputational Model

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. E. Sterken
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

donderdag 26 juni 2014 om 12.45 uur

door

Harm Brouwer

geboren op 28 november 1984
te Smallingerland

Promotores
Prof. dr. J. C. J. Hoeks
Prof. dr. ir. J. Nerbonne

Beoordelingscommissie
Prof. dr. P. Hendriks
Prof. dr. M. Kutas
Prof. dr. D. C. Plaut

To my wife, Noortje

Foar Heit en Mem

Acknowledgements

First and foremost, I want to express my gratitude to my *promotores* John Hoeks and John Nerbonne, or John H. and John N. as I used to disambiguate you in e-mail conversations that involved the three of us. John H., you were my daily supervisor, and I could not have wished for a better one. It has been more than a delight writing my PhD thesis under your supervision, and there are many highlights that I will remember about the past four years: our Monday till Thursday morning-coffee-routine, our 'Oxford Handbook chapter'-adventure (accepted without revisions, *BAM!*), co-teaching the 2013 LOT Summer School course on 'The Neurobiology of Language', and of course the many celebratory dinners with you, Catharina, Anna, David, and Noortje. I hope that we will co-author and publish many more papers in the future, and celebrate each of these with at least two Duvels! John N., although we did not work together on a daily basis, you have shared your experience and wisdom with me many times at crucial moments in my career. I believe that I should be especially grateful that you convinced me to apply for an NWO-funded PhD position, as the freedom that came with this position has been critical to the work in this thesis. In the context of supervision, I should also thank Hartmut Fitz. Hartmut, although our ways have parted early on during my PhD, you have been essential in shaping the first three chapters of this thesis.

Next in line are Petra Hendriks, Marta Kutas, and David Plaut, who agreed to be in my reading committee. Thank you for finding the time to read this dissertation. You have all been a source of inspiration, and I feel honoured that my work is deemed scientifically sound by you. I should also thank Gina Kuperberg and Colin Phillips for the interesting discussions that we had during the LOT Summer School in 2012. Furthermore, I am very grateful to my paranymphs Leonie Bosveld-de Smet and Lotte Schoot. Leonie, you were one of my first teachers at the university, and given my troubled journey through high school, I cannot stress enough how much our long talks and discussions have positively affected my career in academia. To me it seems only natural that you are there to assist me with the last steps towards obtaining a PhD. Lotte, you joined John H. and me as a research assistant, and during this appointment we have shared an office. I very much enjoyed your company as a roommate, as well as the dinners with you, Jeroen, and Noortje, and of course our trip to Riva del Garda!

During the almost nine years that I have been affiliated with the Alfa Informatica department in one way or another, I have had the honour to meet many great people. First of all, there are my current colleagues. Thank you, Çağrı, Dieke, George (for singing a song when Noortje and I got married), Gertjan (for twelve hours and eighty kilometers of hiking fun), Gosse, Gregory, Johan, Johannes (and Vivian), John N., Kilian (and Ni), Leonie, Martijn, Noortje, Simon, and Valerio (and Sara). I also want express my gratitude to all my former colleagues at the department, but I feel that I should thank two of them in particular. Daniël, thanks for the many discussions we had over coffee, and for convincing me to start writing a book together (which we should finish, some day). Kostadin, thank you for being you. Finally, there are also some colleagues outside of the Alfa Informatica group that deserve a word of thanks: Alice, Dicky (for singing many songs at our wedding), Laurie, Meriel, Ryanne, and Wyke, thank you.

Next, I want to thank my closest friends, who are always there to remind me that there is more to life than work. Thank you, Bas and Emmy, for being my friends since forever! Your company is always magnificent, and especially now that Collin joined your ranks, extremely calming (although he does trigger an interest in developmental neuropsychology from time to time). Thank you, Ben and Lotte, Dirk, Frank and Manja, Frans and Nathalie and the kids, Guus, Hennie, Julia, Mark, Peter, Richard, Rick and Kasper, and Sytze.

I am also grateful to my family, which expanded significantly after I got married last year. Thank you, Elly, Guus, Oebele and Siska, Sjoerdieke and Charles and Christopher, Sytze, Tonnis and Netty, and Wiebout and Richelle. Two of you deserve some extra words of gratitude, my lovely parents Oebele and Siska. Leave heit and mem, jullie weten als geen ander dat de reis van de brugklas naar dit proefschrift niet is verlopen zonder omwegen, valpartijen, blessures, urenlange discussies, en ruzies. Door dit proefschrift aan jullie op te dragen, wil ik jullie bedanken voor jullie onuitputtelijke geduld en onvoorwaardelijke support. Jimme binne de beste!

Finally, I also want to dedicate this dissertation to my brilliant wife Noortje. Thank you for your endless optimism. Thank you for your unconditional love and support. Thank you for everything and more!

Groningen,
April 2014

Contents

List of Figures

List of Tables

CHAPTER 1

General Introduction

1.1 Language Comprehension

An average language user produces between three to four words per second (Marslen-Wilson, 1973). The average listener, therefore, must 'process' these words at least as fast. Among other things, this processing entails establishing syntactic dependencies between the words, activating the meanings of the words and combining them into a (partial) meaning of what is said, taking into account all kinds of pragmatic aspects such as, for instance, the conversational context in which the sentence is uttered. Uncovering the architecture of this fast and highly efficient comprehension system has been a very important and prominent goal of both psycholinguistics and cognitive science. Over the past decades, our understanding of the comprehension system has grown enormously. For instance, we know by now that language comprehension operates incrementally: an incoming utterance is processed on a more or less word-by-word basis (e.g., Altmann and Kamide, 1999). What this means is that during processing, the meaning of each incoming word is assigned immediately, and related to the interpretation of the sentence, and ultimately to the entire preceding discourse. In addition, we know that the comprehension system works integratively, that is, in creating the meaning of an utterance it combines multiple sources of information: syntactic and semantic knowledge is integrated with pragmatic and world knowledge. Furthermore, it has recently been suggested that one of the characteristics of the comprehension system that makes it so fast, is its ability to anticipate what a speaker will say next. In other words, the language comprehension system works predictively (e.g., Otten et al., 2007).

This thesis is about a computational model of language comprehension that incorporates precisely these three characteristics: incrementality, integrativity, and predictivity. While other models have focused on the behavioral correlates of comprehension, such as reading times or eye-tracking data, we model the neural correlates of comprehension as provided by Event-Related brain Potentials (ERPs). Combining techniques from computational linguistics with neuroscientific methods is a novel approach in studying the human language comprehension system.

1.2 Unraveling comprehension using ERPs

Neurophysiological activity can be measured on-line 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. When averaging over numerous similar EEG recordings, however, 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 psycholinguistics, research using ERPs took off with Kutas and Hillyard (1980) who found that semantically anomalous sentence-final words "He spread his warm bread with socks" elicit a negative shift in the ERP after approximately 400ms, which became known as the N400 component. Soon after, it was found that the N400 is elicited for every content word, and that its amplitude is modulated by the extent to which a word is expected given the preceding context. Words that violate syntactic expectations "The spoilt child throw...", on the other hand, were found to elicit a positive shift after approximately 600ms. This positive shift is known as the P600 component or the Syntactic Positive Shift (Hagoort et al., 1993). Recently, however, this convenient one-to-one mapping between semantics and the N400 on the one hand, and syntax and the P600 on the other, has been called into question. An increasing number of ERP studies have found non-semantic N400 effects, as well as non-syntactic P600 effects (e.g., Kutas et al., 2006).

In particular the so-called Semantic Illusion effect has become a prominent topic of debate (see Bornkessel-Schlesewsky and Schlesewsky, 2008, for a comprehensive overview). Semantic Illusion effects may be observed in sentences where a verb requires a certain assignment of thematic roles, but the actual syntactic structure goes against that, as in this example from Hoeks et al. (2004): "De speer heeft de atleten geworpen" (lit: The javelin has the athletes thrown). The sentence-final verb *thrown* makes this sentence semantically anomalous, as *javelins* do not generally throw *athletes*. Despite this fact, there is no sign of an N400 which normally is highly sensitive to semantic processing difficulty. This suggests that readers are fooled, at least for a very short period of time, into believing that the sentence actually makes sense (Hoeks et al., 2004). This illusion of semantic correctness is temporary, as a few hundreds of milliseconds later a P600 occurs, showing that the reader eventually realizes that this is not a correct sentence. The occurrence of a P600 is unexpected too, as there is no syntactic violation of any kind in the critical sentence. Thus, the phenomenon of Semantic Illusion challenges previously held views on how the brain handles language processing.

Until now, there has been no consensus whatsoever on the factor (or factors) that cause the N400 to be absent, and a P600 to appear instead. On the contrary, almost every study that looked into the Semantic Illusion effect offers its own explanation, often at the exclusion of previous explanations. Because of the lack of precision of the proposed explanations and micro-models, it has remained impossible so settle the issue on either theoretical or empirical grounds. Kim and Osterhout (2005), for instance, explain the Semantic Illusion as a mismatch between an observed, and a semantically expected verb. van Herten et al. (2005) explain it as a mismatch between a syntactic analysis and a "plausibility heuristic". Hoeks et al. (2004) attribute it to a message-level underspecification due to ill-defined "thematic processing problems". Kuperberg (2007) accounts for it by means of a "continued combinatory analysis" due to either thematic, semantic-memory-based or other factors, and Bornkessel-Schlesewsky and Schlesewsky (2008) argue that it results from a conflict between thematic role assignment in a syntactic linking step, and a parallel semantic plausibility processing step. These accounts seem highly incompatible, but due to the lack of a specific, implemented model, it has turned out to be impossible to put any of these proposals to a crucial test.

1.3 How computational modeling will help out

Over the years, a great number of different cognitive architectures have been proposed to explain the language comprehension system. Many of these are quite abstract 'box and arrows' models, although some have been worked out in more detail. Nonetheless, most remain conceptual and do not generate specific, quantitative predictions which makes it difficult to falsify or even test them. This seems to be especially the case in the range of explanations of the Semantic Illusion effect, where no model reaches a level of specificity to allow concrete, testable predictions. To make any progress in understanding semantic illusions—and, eventually, language processing in general—it is essential to construct a computational model.

A computational model is a formally precise description of the mechanisms which are hypothesized to underlie language processing. In contrast to conceptual models, implementing such mechanisms computationally does not leave any model component vague or underspecified. Since model parameters can then be manipulated systematically, we can obtain causal explanations of the observed behavior. Secondly, a computational model is procedural, because it operates as a sequence of steps. Every processing step assumed in an architecture has to be defined in terms of logical, transparent sequential operations. As a consequence, every sequence of processing steps in such a model leads to concrete, testable predictions. A formal model is, therefore, more informative than a conceptual one in that it can be used to empirically validate an architecture, as well as to arrive at novel insights and predictions that follow from it. Furthermore, due to their transparency, such models can be analyzed in depth, both quantitatively as well as qualitatively. Hence, we can gain understanding not only by devising a computational model of an architecture, but also by studying it once it is obtained.

A vast amount of work has been done on computational modeling of the language comprehension system, and different kinds of models have been proposed to account for different kinds of psycholinguistic phenomena. One type of model, for instance, concerns incremental models of language processing that focus on the predictive aspect of comprehension. Such models have been devised to account for processing difficulty due to syntactic ambiguities (Vosse and Kempen, 2000; Levy, 2008). Levy (2008), for example, describes an implemented model of the recent, influential surprisal theory. Within this theory, the processing difficulty of each word in a sentence can be modeled. A word's difficulty is related linearly to its expectancy, i.e., the extent to which it was predicted or expected, given the words preceding it. By means of simulating reading time and eye-tracking data, Levy shows that this model accounts for a wide range of psycholinguistic phenomena like, for instance, the effect of inserting additional material between a direct object and a final verb in German. In similar vein, Brouwer et al. (2010a) show that an implementation of surprisal theory for Dutch can account for the noun phrase versus sentence coordination ambiguity.

Another type of model concerns incremental models that focus more on the integrative aspect of language processing. Models like these have often been devised to account for the resolution of syntactic ambiguities in which multiple information sources, or 'constraints', are available simultaneously (e.g., McRae et al., 1998; Padó et al., 2009). McRae et al. (1998), for instance, propose their 'competition/integration' model to account for the resolution of temporary ambiguities in English reduced relative clauses. Within this model, constraints like thematic fit between a verb and a noun, a verb's relative frequency, and general processing biases (such as the preference for interpreting an ambiguous part of a sentence as a main clause instead of as a relative clause), are rapidly integrated to resolve the ambiguity. The two syntactic alternatives (i.e., main clause reading and relative clause reading) receive support from the various constraints, and are seen to be in competition: the alternative that receives the most support will win the competition. The amount of time it takes for this competition to significantly favor one analysis over another reflects the difficulty in processing an ambiguity. McRae et al. (1998) show that this model is able to account for reading time data regarding this specific type of ambiguity.

A third type of model that has been developed very recently incorporates both prediction and integration. Such architectures are rare, and have mainly been devised to model the on-line integration of visual and linguistic information in language comprehension (Farmer et al., 2007; Crocker et al., 2010). In this thesis, we take this kind of model an important step further, as we present a computational model of comprehension that incorporates incrementality, prediction, and integration in explaining ERP effects in language processing, specifically those pertaining to the Semantic Illusion effect. Doing so will enable us to build a model of language processing that is psychologically and neurolinguistically plausible and that is sufficiently precise to allow for novel predictions regarding semantic and syntactic processing and how these are reflected in ERPs.

1.4 About this thesis

This thesis is organized in four parts. Part I provides a critical review of the models that have been proposed to account for the Semantic Illusion effect (chapter 2). It is shown that none of these models is able to explain the full range of relevant data, which we argue is because they are built upon wrong functional interpretations of the N400 and the P600 component. On the basis of this review, and converging evidence in the literature, we propose a functional reinterpretation of the N400 component as reflecting memory retrieval, and the P600 component as reflecting compositional semantic processing (chapter 3). In Part II, a minimal functional-anatomic mapping is derived by aligning these new functional interpretations of the N400 and the P600 with neuroanatomy, and it is argued that this mapping forms the core of the comprehension system (chapter 4). Part III provides a 'proof of concept' of this functional-anatomic mapping in terms of simulations with a neurocomputational model that accounts for the Semantic Illusion effect as found in the study by Hoeks et al. (2004) (chapter 5). Part IV provides conclusions and future perspectives (chapter 6).

1.5 Publications

The present chapter is adapted from an accepted, peer-reviewed NWO grant proposal:

- Brouwer, H., Fitz, H., Hoeks, J. C. J., and Nerbonne, J. (2010b). How teeth can brush a child: A neurocomputational account of the Semantic Illusion in language comprehension. NWO-PGW, 10-26 (322-70-002).

Others are adapted from peer-reviewed journal publications:

- 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 (chapters 2 and 3).

- 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 (chapter 4).

1.6 Software

For the neurcomputational simulations presented in this thesis, two software packages were developed:

- MESH—A general purpose artificial neural network simulator. See appendix A.

- COALS—An implementation of the Correlated Occurrence Analogue to Lexical Semantics (COALS; Rohde et al., 2009). See appendix B.

PART I

The Electrophysiology of Language Comprehension

CHAPTER 2

The "Semantic Illusion" Phenomenon

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. In this chapter, 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.*[1]

[1]This chapter is adapted from the first part of 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.

2.1 Introduction

As a sentence or story unfolds in time, language users incrementally construct 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'[2]. This is because the absence of an N400-effect suggested

[2]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

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 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.

2.2 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 the-

---

the non-syntactic nature of the evoked late positivities in these materials.

matic 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'.

2.2.1 Multi-stream models

2.2.1.1 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 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).

2.2.1.2 Monitoring Theory

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, 2012) 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 attraction 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 Monitoring Theory and the Semantic Attraction account.

2.2.1.3 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 eat…' relative to 'For breakfast the boys would eat…' (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 watch…', 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 watch…' 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 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). 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 article…' relative to 'The journalist wrote the article'. 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 of Kuperberg (2007) and Kuperberg et al. (2010) seem highly similar. Biphasic N400/P600-effects therefore remain problematic for the CCA account.

2.2.1.4 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 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 an interpretation in which *the athletes* have *thrown the javelin*. Importantly, neither thematic role assignment, verb-argument linking, nor 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 thrown) 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 restaurant' (lit: Fred eats a restaurant) relative to control sentences such as 'Fred eet een boterham' (lit: Fred eats a sandwich). 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.

2.2.1.5 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 has the athletes summarized). 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, p. 10).

2.3 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 suitcase…'

They found that if the story continued as in the example ('Next, the woman told the suitcase…'), 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.

2.4 Conclusion

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 2.1 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 2.2 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, 2012, 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 validity of these *multi-stream* architectures. In the next chapter, 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.

Model	Streams	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 2.1 | *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	**Observed**	**SA**	**MT**	**CCA**	**eADM**	**PC**	**RI**
Hoeks et al. (2004)							
De speer werd door de atleten <u>geworpen</u>	—	—	—	—	—	—	—
De speer heeft de atleten <u>geworpen</u>	P6	P6	P6	P6	P6	P6	P6
De speer werd door de atleten <u>opgesomd</u>	N4/P6	N4	N4	P6*	N4/P6	P6	N4/P6
De speer heeft de atleten <u>opgesomd</u>	N4/P6	N4	N4	P6*	N4/P6	P6	N4/P6
Kim and Osterhout (2005)							
The hearty meal was <u>devoured</u>…	—	—	—	—	—	—	—
The hearty meal was <u>devouring</u>…	P6	P6	P6	P6	P6	N4	P6
The dusty tabletops were <u>devouring</u>…	N4/(P6)[a]	N4	N4	P6	N4/P6	N4	N4/P6
van Herten et al. (2005)							
De stroper die op de vos <u>joeg</u>…	—	—	—	—	—	—	—
De vos die op de stroper <u>joeg</u>…	P6	—	P6	P6	P6	P6	P6
Kuperberg et al. (2007)							
For breakfast the boys would <u>eat</u>…	—	—	—	—	—	—	—
For breakfast the boys would <u>watch</u>…	N4	N4	N4	N4	N4	N4	N4
For breakfast the eggs would <u>eat</u>…	P6	P6	P6	P6	P6	P6	P6
For breakfast the eggs would <u>watch</u>…	(N4)/P6	N4	N4	P6	P6	P6	(N4)/P6
Kos et al. (2010)							
Fred eet een <u>boterham</u>…	—	—	—	—	—	—	—
Fred eet een <u>restaurant</u>…	N4	N4	N4	N4	P6	N4	N4/(P6)
Fred eet in een <u>restaurant</u>…	—	—	—	—	—	—	—
Fred eet in een <u>boterham</u>…	N4	N4	P6	N4	??	N4	N4/(P6)
Nieuwland and van Berkum (2005)							
Prior context…							
Next, the woman told the <u>tourist</u>…	—	—	—	—	—	—	—
Next, the woman told the <u>suitcase</u>…	P6	N4	N4	N4	N4/P6	N4	P6

Table 2.2 | *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. Stroud (2009) found a biphasic N400/P600 response for the condition marked with an 'a'. Predictions marked with an '*' are corrections to erroneous predictions reported in Brouwer et al. (2012), Table 2.*

CHAPTER 3

Rethinking the functional role of the N400 and the P600

Abstract | *In the previous chapter, we have argued that five multi-stream models that have been proposed to explain the Semantic Illusion phenomenon fail to account for all relevant data. This is because all five models assume that the amplitude of the N400 indexes some form of semantic integration. In the present chapter, we argue based on recent evidence 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]

[1]This chapter is adapted from the second part of 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.

3.1 Introduction

In the previous chapter, we have argued that 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 in this chapter, 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 geworpen' (lit: The javelin has the athletes thrown) relative to 'De speer werd door de atleten geworpen' (lit: The javelin was by the athletes thrown), 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.

3.2 Rethinking the N400 and the P600

3.2.1 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 geworpen' (lit: The javelin has the athletes thrown), 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 have recently come to the fore in which N400 amplitude 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 (*hav-*

ing 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 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 by-product 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 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 *pres-*

ence 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 bird' produced an

N400 amplitude equal in size to that elicited by the final word in true sentences such as 'A robin is a bird' (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 white...' produced a larger N400 amplitude than the critical word in sentences like 'The Dutch trains are yellow...'. 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 2.2). 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.

3.2.2 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, 2003a; 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, 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 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., 2012), 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 2.2). One study that requires some additional explanation, however, is that of Kos et al. (2010). For a sentence such as 'Fred eet een restaurant tijdens de lunch' (lit: Fred eats a restaurant during the lunch) (as compared to 'Fred eet een boterham tijdens de lunch'; lit: Fred eats a sandwich 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 and hence the meaning of this sentence needs to be composed from what has already been perceived (see below for a more

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

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 was...' (relative to 'The lawyer charged that the defendant was...'). 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, 2012, 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 imitate…'). 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.

3.2.3 The 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 underlie 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. Increased 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.

3.3 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 thrown) 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 victim...' (*'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 N400 amplitude.

3.4 Predictions

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 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 it can mean that both the squares and the triangles are blue or that only the squares are blue. A prosodic break after *squares* 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 or garden-path, 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, 2012, 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.

3.5 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).

PART II

Aligning Electrophysiology and Neuroanatomy

CHAPTER 4

A Time and Place for Language Comprehension

Abstract | *We propose a new functional-anatomical mapping of the N400 and the P600 to a minimal cortical network for language comprehension. Our work is an example of a recent research strategy in cognitive neuroscience, where researchers attempt to align data regarding the nature and time-course of cognitive processing (from ERPs) with data on the cortical organization underlying it (from fMRI). The success of this 'alignment' approach critically depends on the functional interpretation of relevant ERP components. Models of language processing that have been proposed thus far do not agree on these interpretations, and present a variety of complicated functional architectures. We put forward a very basic functional-anatomical mapping based on the recently developed Retrieval-Integration account of language comprehension (Brouwer et al., 2012). In this mapping, the left posterior part of the Middle Temporal Gyrus (BA 21) serves as an epicenter (or hub) in a neurocognitive network for the retrieval of word meaning, the ease of which is reflected in N400 amplitude. The left Inferior Frontal Gyrus (BA 44/45/47), in turn, serves a network epicenter for the integration of this retrieved meaning*

with the word's preceding context, into a mental representation of what is being communicated; these semantic and pragmatic integrative processes are reflected in P600 amplitude. We propose that our mapping describes the core of the language comprehension network, a view that is parsimonious, has broad empirical coverage, and can serve as the starting point for a more focused investigation into the coupling of brain anatomy and electrophysiology.[1]

4.1 Introduction

The aim of the study of language comprehension is to understand how the brain creates meaning from linguistic input. Starting from the lesion-studies of Broca and Wernicke, and subsequent work by Lichtheim and Geschwind, we have learned that the language system is not rooted in a single cortical area, but rather involves a whole network of interconnected regions (a *neurocognitive network*, henceforth). Neuroimaging and lesion studies have since produced a vast collection of data on the cortical organization of language comprehension (for discussions and overviews, see e.g., Dronkers et al., 2004; Turken and Dronkers, 2011; Cabeza and Nyberg, 2000; Bookheimer, 2002; Binder et al., 2009; Price, 2002, 2010; Andrews, 2011). The challenge we now face is twofold: we need to find out what processes are subserved by these areas, and also how these functional processes are ordered temporally.

To arrive at a neurobiological model of language comprehension, a link is needed between time and place of language comprehension in the brain. This means that we will have to find a way to deal with the limitations of the currently available neuroimaging methods; some methods allow for assessing whether cognitive processes are different in kind, and how they evolve over time (e.g., electroencephalography, EEG, and magnetoencephalography, MEG), whereas other methods can be used to pinpoint the location of the areas that are most active during a given cognitive task (e.g.,

[1]This chapter is adapted from 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.

functional magnetic resonance imaging, fMRI, positron emission tomography, PET, and, functional near-infrared spectroscopy, fNIRS). As hemodynamic and electrophysiological measurements are fundamentally different in nature, it is not immediately clear how they should be combined. That is, due to their differences in spatial and temporal resolution, it is often impossible to simply compare them for a given experimental paradigm (cf. Lau et al., 2008).

A more promising strategy is to start with electrophysiology (EEG, MEG), identify and categorize the processes assumed to be reflected in different event-related measurements (Event-Related brain Potentials or ERPs, Event-Related magnetic Fields or ERFs), and then try to find candidate cortical areas or neurocognitive networks that could host them. The success of this 'alignment' approach, however, critically depends on the interpretation of ERP/ERF components, and in the literature there is no broad agreement on these interpretations. Take as an example two recent models that have been proposed on the basis of this 'process-alignment strategy' (Baggio and Hagoort, 2011; Friederici, 2011). These models disagree on their interpretation of language-related ERP components, and also on where in the brain certain processes are carried out. Baggio and Hagoort (2011), for instance, postulate a complex cortical circuit for word-processing, which they argue is responsible for generating the N400 component (a negative deflection in the ERP waveform which is maximal about 400 ms post-onset). In their model, the N400 is taken to reflect both semantic integration, which they argue takes place in the *frontal* lobe, and the retrieval of word meaning from memory (combined retrieval and integration view), which is carried out in the *temporal* lobe. Baggio and Hagoort make no claims about the processes reflected in the P600 component (a positive deflection in the ERP which is maximal about 600 ms post-onset). Friederici (2011), by contrast, proposes an extensive language comprehension model in which the processes underlying both the N400 component and the P600 component are linked to specific cortical areas. Friederici claims that the N400 component reflects the creation of semantic relations between words or phrases (semantic integration—and critically: no retrieval), and argues that these processes take place in middle and posterior parts of the temporal cortex. The P600 component, in turn, is assumed to reflect the integration of syntactic and semantic information in the Temporo-Parietal Junction or TPJ (see Friederici, 2011, Figure 11).

The interpretation of ERP components and effects thus seems to pose a serious problem, as there is as yet no broad agreement in the language processing literature on what these components mean. This could make the process-alignment strategy a hazardous enterprise, leading to a proliferation of incompatible models. However, we will argue for a research strategy in which one starts from the simplest account of ERP/ERF effects and components, while keeping empirical coverage constant. In a recent paper, Brouwer et al. (2012) present a highly parsimonious account of the two most salient ERP components for language comprehension—the N400 and the P600. They show that their *Retrieval-Integration* account is able to explain a wide spectrum of electrophysiological data on language processing. We will give a brief overview of the Retrieval-Integration account and then apply the process-alignment strategy to derive a *minimal* neurocognitive network of language comprehension that can implement this account. We focus on the *epicenters* (Mesulam, 1990, 1998) or *hubs* (Buckner et al., 2009) of this network that serve to integrate or bind together information from various sub-networks (see also the idea of *convergence zones* Damasio, 1989). The resulting electrophysiological-anatomic mapping can serve as the starting point for a more elaborate coupling of brain anatomy and electrophysiology. That is, we believe that our proposed mapping forms the core of the comprehension system, which can be extended to account for other language-related ERP components—such as the Early Left Anterior Negativity (ELAN; see Steinhauer and Drury, 2011, for a discussion), the Left Anterior Negativity (LAN; see Kutas et al., 2006), and for instance, sustained negativities like the Nref-effect (van Berkum et al., 2007, see Hoeks and Brouwer, 2014, for a recent account of the Nref as a component)—once we know what processes underlie these components, as well as where these processes are carried out in the brain. Hence, our mapping provides a first step towards such an elaborate neurocognitive model of language comprehension.

4.2 The Retrieval-Integration Account

An ERP is the summation of the post-synaptic potentials of large ensembles (in the order of thousands or millions) of neurons synchronized to an event. When measured from the scalp, continuous ERP signals manifest themselves as voltage fluctuations that can be divided into components. A component is taken to reflect the neural activity underlying a specific computational operation carried out in a given neuroanatomical module (Luck, 2005; Näätänen and Picton, 1987). Components vary in polarity, amplitude, latency, duration, and scalp distribution, suggesting that different components reflect distinct functional processes, carried out in distinct cortical regions. The two most salient ERP components for the study of language comprehension are the N400 and the P600. The N400 component is a negative deflection in the ERP signal that starts around 200-300ms post-word onset, and peaks at about 400ms. This component has been taken to index *semantic integration* processes (Brown and Hagoort, 1993; Chwilla et al., 1995; Hagoort and van Berkum, 2007; Hagoort et al., 2009); words that are semantically incongruent given their preceding context (e.g., *socks* in "He spread his warm bread with socks") produce an increase in N400 amplitude relative to congruent words (e.g., *butter* in "He spread his warm bread with butter"; Kutas and Hillyard, 1980), presumably reflecting that they are more difficult to integrate with their prior context. The P600 component, in turn, is a positive deflection in the signal that starts, on average, around 500ms post-word onset, and reaches its maximum around 600ms. This component was originally considered to be an index of syntactic reanalysis or repair. Its amplitude has, for instance, been found to increase in response to words that induce a syntactic violation (e.g., *throw* in "The spoilt child throw ...") relative to control words (e.g., *throws*; Hagoort et al., 1993). This increased amplitude is taken to reflect the processes involved in repairing the agreement error between the critical verb and its argument. For some time, there appeared to be a clear, one-to-one mapping between the N400 and semantic integration (combinatorial semantic processing), and the P600 and syntactic processing. This mapping forms the core of many neurocognitive models of sentence comprehension (e.g., Friederici, 2002, 2011; Hagoort, 2005; Hagoort et al., 2009, among others). However, in a review of the "Semantic Illusion" phenomenon in sentence processing, Brouwer et al. (2012) have recently shown that an increasing number of experimental findings cannot be explained when adhering to this mapping.

The label "Semantic Illusion" (or "Semantic P600") is used to refer to a finding in the ERP literature, in which a semantically anomalous sentence does not give rise to an expected increase in N400 amplitude, but rather to one in P600 amplitude (Kolk et al., 2003; Kuperberg et al., 2003; Hoeks et al., 2004). Hoeks et al. (2004), for instance, observed a P600-effect, and no N400-effect, in response to 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 "De speer werd door de atleten geworpen" (lit: The javelin was by the athletes thrown). The absence of an N400-effect is puzzling. The critical verb *thrown* should be more difficult to integrate into the prior context "The javelin has the athletes [...]", as the resulting interpretation of the sentence is in conflict with our knowledge about the world (athletes throw javelins, and not the other way around).

Brouwer et al. (2012) review five models that have been proposed to account for this absence of an N400-effect (taken as absence of semantic integration difficulty), and conclude that none of these models is able to account for the relevant data. They attribute this failure to the 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, Brouwer et al. (2012) argue 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 a stimulus can be retrieved from long-term memory. In the case of language comprehension, relevant stimuli are typically words, and in this case we refer to memory retrieval as *lexical retrieval*. When dealing with non-linguistic stimuli, however, like an image or a sound, memory retrieval is referred to as *semantic retrieval*. Memory retrieval, lexical retrieval, and semantic retrieval amount to the same thing, and in the remainder of this chapter we will use these terms interchangeably to refer to the retrieval of the conceptual knowledge associated with a stimulus (in our case a word) 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. Brouwer et al. (2012) hypothesized 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, our 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). These studies report frontally distributed late positive effects for semantically atypical versus typical sentences. Of particular interest are the results of Molinaro et al. (2012), who investigated the processing of different degrees of (a)typicality, and found a significant inverse correlation between P600 amplitude and the "naturality" (the degree to which speakers would produce a given expression) of stimulus sentences, which is clearly consistent with our MRC hypothesis; the less natural an utterance, the more effort it takes to make sense of it, and the higher P600 amplitude.

The views on the N400 and the P600 that were described above are combined in the Retrieval-Integration (RI) account (Brouwer et al., 2012). 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. We expect Retrieval-Integration cycles to be most pronounced for open-class words, as these carry more meaning than closed-class words. However, we also predict closed-class words to modulate N400 and P600 amplitude (see van Petten and Kutas, 1991; King and Kutas, 1995; DeLong et al., 2005; Hoeks and Brouwer, 2014).

4.3 Connecting Electrophysiology and Anatomy

Retrieval-Integration cycles provide a general and parsimonious account of the elicitation patterns of the N400 and the P600. This sheds light on the *how* and the *when* of comprehension, but not on the *where*. In the second part of this chapter, we propose a mapping of the RI account onto a minimal anatomical network. Our approach follows the 'process-alignment strategy'; we seek to identify the most parsimonious cortical network that can host the RI-account, while remaining consistent with the extant collection of neuroimaging and lesion studies on language comprehension.

4.3.1 A List of Requirements

Based on the assumption that language processing involves continuous Retrieval-Integration cycles, we can specify a list of anatomical 'building blocks' that are minimally required to host the RI account in a cortical network. However, before we start with this list, we should define the granularity of the elements to look for. As argued in the introduction, we know that language is not rooted in a single cortical area, but rather in a network of interconnected regions. This is in line with a paradigmatic shift in cognitive neuroscience, in which researchers move from localizationist theories of cognition, towards large-scale, distributed brain networks, called *neurocognitive networks* (see Bressler and Menon, 2010; Meehan and Bressler, 2012, for overviews). The large-scale nature of these neurocognitive networks significantly increases the difficulty of connecting electrophysiology and anatomy. The extent of the language network can, however, still be manageable if we focus on its anatomical and computational *epicenters* (Mesulam, 1990, 1998) or *hubs* (Buckner et al., 2009). These epicenters/hubs are nodes in the network that serve to integrate information from various sub-networks, and are therefore critical gateways for information processing (cf. Buckner et al., 2009). On the basis of such epicenters, we can postulate the following list of requirements in order to create a basic anatomical circuit for language comprehension; we will minimally need two epicenters, and two kinds of pathways connecting them:

1. **An epicenter that mediates *lexical retrieval* (∼N400);** This cortical region (or network node) should host or mediate the mapping of

word forms to conceptual representations. Conceptual representations are most likely stored in a distributed manner across the association cortices (cf. Elman, 2004, 2009; Pulvermüller, 1999, 2001; Rogers and McClelland, 2004). As such, the full range of activity reflected in N400 amplitude will include the activation of conceptual features in the association cortices, which make up a word's conceptual representation. However, the focus of activity underlying the N400 component is presumed to lie in the *retrieval epicenter*, which serves to 'retrieve' and 'tie together' these conceptual representations of incoming lexical items, so they become available for later synthesis.

2. **An epicenter that mediates *mental representation composition* (∼P600);** This cortical region should host or mediate the integration of the lexical information (retrieved via the retrieval epicenter that was described above) with the existing mental representation of previous input, resulting in a mental representation of what is communicated in the discourse thus far. This specific *integration epicenter* thus hosts or mediates the combinatorial/compositional processes involved in meaning construction, and is assumed to initiate the generation of the P600 component. Again, this area serves as an epicenter, and the full range of activity reflected in P600 amplitude may include activity from other areas as well.

3. **A white matter tract containing fibers that connect region (1) to (2);** This (bottom-up) pathway serves to connect a word's conceptual representation, as retrieved in region (1), to region (2) for integration with a representation of its prior context.

4. **A white matter tract containing fibers that connect region (2) to (1).** This (top-down) pathway connects the newly formed representation that is active via region (2) to (1), thereby providing a context for the retrieval of a next word's meaning. Theoretically, the function of this pathway could be subserved by the white matter tract described in

(3) if it were bi-directional, containing fibers that connect (1) to (2), and vice versa.

Provided this list of anatomical requirements, we can try and identify candidate epicenters and white matter tracts. On the basis of several large-scale reviews on the cortical organization of the comprehension system (e.g. Bookheimer, 2002; Dronkers et al., 2004; Turken and Dronkers, 2011; Friederici, 2002, 2011; Hickok and Poeppel, 2004, 2007; Vigneau et al., 2006; Shalom and Poeppel, 2008; Lau et al., 2008), we want to propose two candidate epicenters: the left posterior part of the Middle Temporal Gyrus (lpMTG; BA 21) as *retrieval epicenter* (1), and the left Inferior Frontal Gyrus (lIFG; BA 44/45/47) as *integration epicenter* (2).

4.3.2 A Hub for Lexical Retrieval

Evidence from neuroimaging as well as from lesion studies points towards the lpMTG as an epicenter for lexical retrieval (Cabeza and Nyberg, 2000; Bookheimer, 2002; Dronkers et al., 2004; Lau et al., 2008; Binder et al., 2009; Price, 2010; Turken and Dronkers, 2011). Dronkers et al. (2004), for instance, found that aphasics with lesions in the lpMTG suffered from difficulties in word-level comprehension. They argue that this might be the case because the mapping between word-form and conceptual knowledge (=lexical retrieval) is lost in this patient group. Findings from neuroimaging studies are consistent with this hypothesis. Cabeza and Nyberg (2000), for instance, found in a review of functional PET and fMRI studies that both spoken and written word recognition consistently activates the lpMTG, independently of whether words were presented in isolation or as part of a sentence. In addition, Lau et al. (2008) conducted a large-scale review of neuroimaging research on lexical semantic priming and reported that studies consistently found effects in the MTG. More specifically, they concluded that the MTG was the only region showing effects of semantic priming at both short and long (>600ms) Stimulus Onset Asynchronies (SOAs), which supports the view that the MTG is the generator site of the N400 component. Further evidence for the lpMTG as an epicenter for retrieval comes from MEG studies that investigated the magnetic field equivalent of the N400—the N400m—and used source modeling techniques to localize its generator site. These studies have consistently identified the lpMTG to be involved in the generation of the N400-effect (e.g., Halgren et al., 2002, see Lau et al., 2008, p. 927, for a brief overview). Finally, recent work using connectivity analysis by Turken and Dronkers (2011) provides an important indication for the role of the lpMTG in a network for lexical retrieval. They found that the lpMTG showed a particularly rich connectivity pattern to areas in the frontal, parietal, and temporal cortices of both hemispheres (see Binder et al., 2009; Buckner et al., 2009; Koyama et al., 2010, for similar findings), which is consistent with the idea that the posterior MTG is an epicenter that serves to "retrieve" and "tie together" conceptual knowledge that is represented in a distributed manner across the association cortices.

4.3.3 A Hub for Mental Representation Composition

It has long been known that the lIFG plays a crucial role in sentence comprehension, but it is still disputed what this role precisely entails (see Grodzinsky and Santi, 2008; Rogalsky and Hickok, 2011, for recent overviews). For instance, Rogalsky and Hickok (2011) point out that a substantial part of the lIFG, Broca's Area (BA44/BA45), has been hypothesized to support syntactic movement (Grodzinsky and Santi, 2008), hierarchical processing and phrase structure building (Friederici, 2009), order-related linearization processes (Bornkessel-Schlesewsky et al., 2009), working memory (Buchsbaum et al., 2005; Buchsbaum and D'Esposito, 2008), cognitive control (Novick et al., 2005), semantic unification (Hagoort, 2005), and thematic role checking and reanalysis (Caplan et al., 2008a,b). Others have stressed the role of the lIFG in the control of memory (Badre and Wagner, 2007). These hypotheses are diverse, and some even appear to be outright incompatible. However, we would like to suggest that it is not necessary to choose between these hypotheses, because the lIFG subserves—to a certain degree—all of the hypothesized processes (see Seghier, 2013, for a similar multi-functional approach towards defining the function of the Angular Gyrus, AG). That is, we propose that the lIFG is host to, or mediator of various types of processing involved in creating and maintaining a mental representation of what is communicated. The word-by-word construction, reorganization, or updating of this mental representation involves the accommodation of new discourse entities, establishing a relation between entities and assigning them a thematic role, revising already established relations, revising already established thematic roles, resolving conflicts between information sources (e.g., with respect to world knowledge), and so on (see Brouwer et al., 2012, p. 138). This list subsumes the processes listed by Rogalsky and Hickok (2011), and includes syntactic processes, as well as (working) memory related processes, semantic processes, and control processes. However, our hypothesis is different from the one recently put forward by Bornkessel-Schlesewsky and Schlesewsky (2013), who argue that the lIFG does not sub-serve any linguistic processing at all. On their account, the lIFG only sub-serves cognitive control functions. We find it difficult to reconcile a *cognitive control*-only account of the lIFG with evidence that implies it in combinatorial semantic processing (see the evidence reviewed in Hagoort et al., 2009, for instance).

It is possible that different sub-processes of MRC construction are mediated by or carried out in different, potentially overlapping subparts of the lIFG. Hagoort (2005), for instance, assumes the pars triangularis (BA 45) and the pars orbitalis (BA 47) to be involved in semantic processing, and again BA 45 and the pars opercularis (BA 44) in syntactic processing. Friederici (2009), in turn, suggests that BA 44 supports 'simple' hierarchical structure processing, and that the frontal operculum (located between the anterior insula and lateral opercular part of the IFG, cf. Anwander et al., 2007) is involved in more complex local phrase structure building. Elsewhere, Friederici (2011) also argued for such a functional parcellation of the lIFG, and pointed out that debates on the specific role of BA 44 and BA 45 could be resolved on the basis of further subdivision by the type of neurotransmitter receptors found in these areas; BA 44 can be subdivided into a dorsal (BA 44d) and ventral (BA 44v) part, and BA 45 in a more anterior (BA 45a) and more posterior part (BA 45p). Provided this more granular subdivision, conflicting functions allocated to, for instance, BA 44, could actually coexist when one is allocated to BA 44d and the other to BA 44v. Consistent with this idea, recent studies into the organization of the lIFG have revealed a rather complex neuroarchitectural parcellation of the area and its adjacent regions (e.g., Amunts et al., 2010; Amunts and Zilles, 2012). This anatomical parcellation may underlie a fine-grained functional topology within the lIFG. It has been suggested that such a functional topology may be organized in a systematic manner. Hagoort (2005), for instance, suggests that the lIFG is organized in an anterior-ventral to posterior-dorsal gradient, in which semantic processing is subserved in BA 47 and BA 45, syntactic processing in BA 45 and BA 44, and prosodic processing in BA 44 and partially in BA 6. In their two-process model of memory control, Badre and Wagner (2007) also postulate an anterior-ventral to posterior-dorsal gradient underlying lIFG organization, linking BA 47 to controlled access to stored conceptual representations, and BA 45 to domain-general post-retrieval selection processes. Although it remains to be seen if and how well this gradient unifies with the one proposed by Hagoort (2005), for instance through subdivision of the overlapping Brodmann areas, they both underline the idea of a systematic, functional organization of the lIFG.

To arrive at a more fine-grained functional topology of the lIFG, a more focused and systematic investigation of its subdivision is required. Linking the P600 component to the lIFG opens up a new domain of study, where characteristically different types of P600s (in terms of scalp distribution, amplitude, onset, duration, etc.) can be mapped onto different parcels of the lIFG. Such an endeavor will increase both our understanding of the functional topology of the lIFG, as well as our understanding of the different processes underlying the P600 component. A starting point for such an investigation could be the aforementioned anterior-ventral to posterior-dorsal gradient proposed by Hagoort (2005). If this organizational gradient is correct, we would expect more semantically involved MRC processes to lead to increased activity in BA 47 and BA 45, whereas more structurally involved processes should lead to increased activity in BA 45 and BA 44. Moreover, these different types of processing should be apparent in characteristically distinct P600 modulations.

It has proven difficult to localize the neural generators of the P600 component (Friederici, 2011). Attempts at reconstructing these generators using source localization have identified the middle temporal gyrus and the posterior part of the temporal lobe as generator sites for the P600 (Kwon et al., 2005; Service et al., 2007), which is inconsistent with our hypothesis. However, a number of studies using fMRI have linked the P600 to the lIFG (see van de Meerendonk et al., 2011, for a discussion). Preliminary results from an fMRI study done in our lab are consistent with these findings, and show that the "Semantic Illusion" sentences that produced a P600 effect in the Hoeks et al. (2004) study, also induced increased activity in the pars orbitalis (BA 47) of the lIFG. This correlation between activation in the lIFG and P600 amplitude supports our hypothesis that the P600 is generated in the lIFG. In the discussion, we derive several predictions that can be used to confirm or validate our hypothesis.

4.3.4 Connecting the Two Hubs

Once the lexical knowledge associated with an incoming word is retrieved by the lpMTG, it needs to be connected to the lIFG for integration. This requires a white matter pathway from the temporal to the frontal lobe. The lIFG then integrates the meaning of the incoming word with a representation of its prior context into a representation of what is being communicated. Importantly, the updated mental representation subsequently serves as a context for the retrieval of the next word's meaning, requiring a white matter pathway back, from the frontal to the temporal lobe.

Traditional anatomical models of language assumed that the major white matter pathway connecting the temporal cortex to the IFG was the *arcuate fasciculus*. Recent tractography studies using Diffusion Tensor Imaging (DTI), however, have led to the identification of a more extensive structural connectivity pattern between these two areas (see Catani et al., 2005; Saur et al., 2008; Makris and Pandya, 2009; Turken and Dronkers, 2011, among others). The general view that emerges from these studies is that the inferior frontal cortex and the temporal cortex, are wired together by means of a dorsal and a ventral pathway that can each be subdivided into two sub-pathways. The first dorsal pathway connects the posterior parts of the MTG (and the STG; BA 22) to the pars opercularis (BA 44) via the *arcuate fasciculus* and the *superior longitudinal fasciculus*. The second dorsal pathway connects the posterior STG via the same fiber tracts to the premotor cortex. As for the ventral pathways, one connects the MTG (and the STG) to the pars triangularis (BA 45) via the *extreme fiber capsule system*, and the other connects the anterior part of the STG to the frontal operculum via the *uncinate fasciculus* (the anterior and posterior parts of the temporal lobe are itself connected through the *inferior* and *medial longitudinal fasiculi*). Unfortunately, DTI does not allow us to determine the directionality of white matter pathways (see Friederici, 2011). Nonetheless, three out of the four fiber tracts that we have just described connect the temporal cortex to the inferior frontal cortex (one dorsal pathway connects the STG to the premotor cortex), which provides us with three candidate pathways for our network.

The functional role of these different pathways is, however, still a matter of debate (Hickok and Poeppel, 2004, 2007; Saur et al., 2008; Friederici, 2009, 2011, 2012; Baggio and Hagoort, 2011; Tyler et al., 2011; Weiller et al., 2009; Bornkessel-Schlesewsky and Schlesewsky, 2013). Hickok and Poeppel (2007), for instance, assume in their dual stream model of speech processing that the dorsal pathway maps acoustic speech signals onto articulatory networks in the frontal lobe, whereas the ventral stream subserves speech comprehension. Friederici (2012), on the other hand, attributes different functions to the sub-pathways of the dorsal and ventral fiber tracts, suggesting that whereas the dorsal sub-pathway connecting the temporal lobe to the premotor cortex might indeed be involved in mapping acoustic speech signals to articulatory networks (in line with

Hickok and Poeppel, 2007; Saur et al., 2008), the dorsal sub-pathway connecting the temporal lobe to the pars opercularis (BA 44) is likely to be involved in the processing of syntactically complex sentences and the delivery of top-down predictions to the temporal lobe. As for the ventral pathways, Friederici assumes the sub-pathway connecting the temporal lobe to the pars triangularis (BA 45) to be involved in the transfer of semantic information from temporal to frontal regions, and the sub-pathway connecting the STG to the frontal operculum in syntactic information transfer. Bornkessel-Schlesewsky and Schlesewsky (2013) have recently challenged the accounts by Hickok and Poeppel (2007) and Friederici (2012), and put forward a new ventral-dorsal framework for comprehension, in which the dorsal stream engages in time-dependent combinatory processing, including syntactic structure building, whereas the ventral stream sub-serves "time-independent identification and unification of conceptual [actor-event (AE)] schemata" (Bornkessel-Schlesewsky and Schlesewsky, 2013, p. 67). This framework is attractive as the authors show that it solves several puzzles regarding the requirement for a dorsal stream in the ventral-dorsal stream literature. However, the role attributed to the ventral stream seems to be a direct instantiation of the "plausibility processing"-stream of the extended Argument Dependency Model (Bornkessel and Schlesewsky, 2006; Bornkessel-Schlesewsky and Schlesewsky, 2008), the existence of which has been challenged in the literature (Stroud, 2009; Stroud and Phillips, 2011; Brouwer et al., 2012). Moreover, a core assumption of this framework is that the role of the lIFG is restricted to cognitive control, which is problematic for reasons discussed above. Hence, the validity of this new framework remains open for close scrutiny.

Given the lack of consensus on the functional role of the dorsal and ventral pathways, it is at present difficult to decide which pathway supports the connection of information retrieved in the lpMTG to the lIFG for integration, and which serves to provide a context for retrieval. On a speculative note, there seems to be at least some agreement on the involvement of the ventral pathway in form to meaning mapping (Hickok and Poeppel, 2007; Friederici, 2012; Bornkessel-Schlesewsky and Schlesewsky, 2013), which supports the idea that this pathway may sub-serve the connection of information retrieved in the lpMTG to the lIFG. As for the dorsal route, Friederici's suggestion that one of the dorsal sub-pathways is, among other things, involved in delivering top-down predictions from the frontal to the temporal lobe, is consistent with the idea that the dorsal route is involved in providing a context for retrieval. However, it remains to be seen how this could be reconciled with the proposals put forward by, for instance, Hickok and Poeppel (2007) and Bornkessel-Schlesewsky and Schlesewsky (2013).

In summary, the specific roles of the ventral and dorsal pathways in Retrieval-Integration cycles are as of yet unclear, but their presence does indicate that there is white matter connectivity between the temporal cortex and the inferior frontal cortex that could implement the required circuitry.

4.3.5 A Functional-Anatomical Processing Cycle

Putting the parts together, we can implement the Retrieval-Integration account in a cortical network, and walk through a typical processing cycle. Depending on whether the linguistic input is spoken or written, an incoming word reaches the lpMTG from respectively the auditory or the visual cortex. The lpMTG then acts as an epicenter for the retrieval of the lexical information associated with this word from the association cortices, where it is supposed to be stored in a distributed manner. The onset of this retrieval process corresponds to the onset of the N400 component, and ease with which semantic information can be retrieved corresponds to N400 amplitude. Via one of the candidate white matter tracts, the retrieved lexical information is then connected to the lIFG, where it will be integrated with a representation of the prior context into a representation of what is being communicated. The extent of work required to construct this updated representation corresponds to P600 amplitude. Finally, the representation constructed in the lIFG is fed back to the lpMTG via a white matter pathway resulting in the pre-activation of syntactic and semantic features

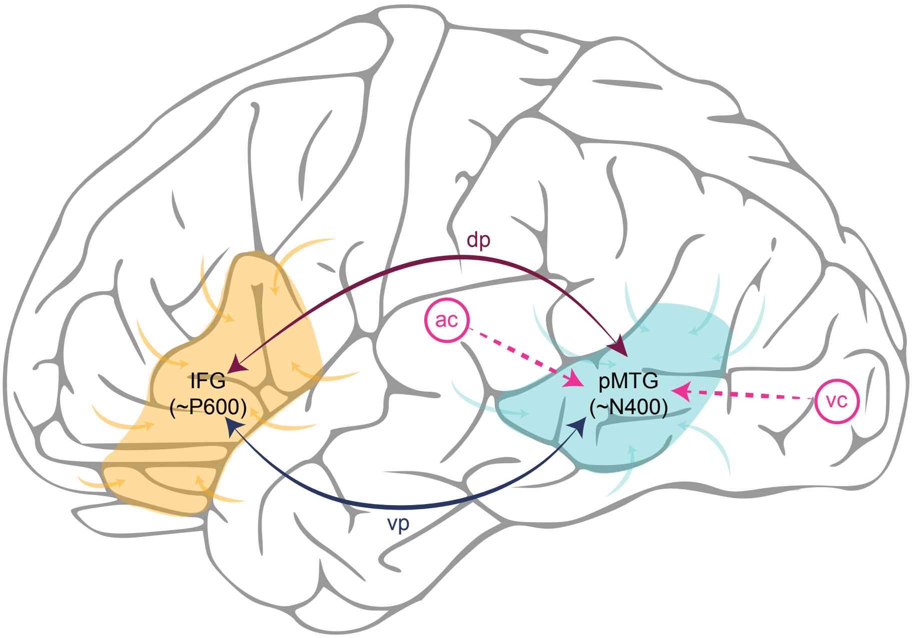

Figure 4.1 | *Schematic illustration of a Retrieval-Integration cycle in the left hemisphere. Words reach the posterior Middle Temporal Gyrus (pMTG) via the auditory cortex (ac) or the visual cortex (vc), depending on whether the linguistic input is spoken or written. The pMTG retrieves the lexical information associated with a word from the association cortices (generating the N400). The retrieved information is then connected to the Inferior Frontal Gyrus (IFG) via one of the white matter tracts in either the dorsal pathway (dp) or the ventral pathway (vp). The IFG integrates this information with a representation of the prior context into an updated representation of what is being communicated (generating the P600). Finally, the representation constructed in the IFG feeds back to the pMTG via white matter tracts in the dorsal or ventral pathway, causing pre-activation of lexical features of possible upcoming words.*

of possible upcoming words. This Retrieval-Integration cycle is repeated as soon as a new word comes in. Figure 4.1 provides a schematic illustration of a typical Retrieval-Integration cycle.

Our functional-anatomical mapping of the Retrieval-Integration account predicts that each incoming word will first evoke activation in the lpMTG (retrieval) and subsequently in the lIFG (integration). A recent study using Event-Related Optical Signal (EROS) supports precisely this prediction (Tse et al., 2007); EROS responses to semantic and syntactic anomalies both showed sequences of increased activity in the lpTMG followed by activity in lIFG.

4.4 Discussion

We have proposed a new functional-anatomical mapping of the N400 component and the P600 component onto a minimal cortical network for language comprehension. In this mapping, the lpMTG (lexical retrieval) and lIFG (semantic integration) play a central role. Our mapping differs from earlier proposals by Friederici (2011) and Baggio and Hagoort (2011), which we have discussed in the introduction. We take the N400 component to index non-combinatory lexical retrieval processes mediated by the posterior temporal cortex, and the amplitude of the P600 to reflect the integration of the retrieved lexical information with a representation of prior context, which is hypothesized to take place in (or to be mediated by) the left inferior frontal regions. The account provided by Baggio and Hagoort (2011) is limited to the N400, which they assume to have generators both in the temporal lobe (reflecting the retrieval of lexical information), and in the frontal cortex (reflecting the construction of multiword units). Our mapping also differs from the one proposed by Friederici (2011). Whereas Friederici (2011) assumes that the processes underlying the N400 are carried out in the temporal cortex (which is consistent with our view), she assumes these processes to involve combinatorial semantic processing (which contrasts with our non-combinatorial, retrieval view).

Friederici (2011) is less specific about the P600, as she argues that "[a]t the linguistic level, the difficulty of integrating syntactic and semantic information and the need for reanalysis is reflected in a P600", and that "[t]he difficulty of mapping linguistic information onto world knowledge also appears to elicit a P600 effect" (Friederici, 2011, p. 1383). Nonetheless, Friederici assumes these processes to be subserved by one particular area called the Temporo-Parietal Junction or TPJ (see Friederici, 2011, Figure 11), which clearly contrasts with our hypothesis that the P600 is generated in the lIFG.

4.4.1 The Role of Other (Sub-)cortical Areas

The mapping that we propose constitutes what we believe to be the core of the language comprehension system. This is not to say, however, that we believe the comprehension system to be limited to the lpMTG and the lIFG. We merely think that these areas are 'epicenters' that form the *absolute core* of the comprehension network. Other cortical as well as sub-cortical areas are likely to also be important, but for most of them the role they play in language processing is still rather unclear. Moreover, even if we do have a clear idea of the processes sub-served by these areas, the coupling between electrophysiology and anatomy that we seek to arrive at requires us to also identify in which ERP component(s) these processes are reflected.

4.4.1.1 Right hemisphere

One cortical area under discussion, for instance, is the Right Hemisphere. Friederici (2011) assumes that the Right Hemisphere is mainly host to prosodic processes, whereas Vigneau et al. (2011) argue against this, and conclude that at least the frontal part of the right hemisphere appears to be invoked whenever additional processing resources are required (e.g., when material in working memory needs to be manipulated). If Friederici's account is correct, activity in the right hemisphere might for instance contribute to late positivities that resemble the Closure Positive

Shift (CPS; Steinhauer et al., 1999), which we argued is a likely member of the P600-family (Brouwer et al., 2012). On the other hand, if Vigneau et al. are right, processing in the right hemisphere might contribute to positivities that reflect substantial MRC revision. Further research into the role of the Right Hemisphere, especially the lpMTG and the lIFG homologue areas, is required to help us to decide between these hypotheses, and extend our mapping accordingly.

4.4.1.2 Anterior Temporal Lobe (ATL)

Another cortical region that has recently received a vast amount of interest is the Anterior Temporal Lobe (ATL), the role of which is also much debated (see Bi et al., 2011; Tsapkini et al., 2011, for recent overviews). Bi et al. (2011) point out that there are at least two classes of hypotheses about the role of the ATL. On the one hand, the ATL is assumed to be a binding site for semantic properties (e.g., Rogers et al., 2004, 2006), supporting the view that it sub-serves the role of a "semantic hub" (Patterson et al., 2007)—however, see Binder and Desai (2011) for recent arguments against this interpretation. On the other hand, the ATL is assumed to be important for lexical retrieval (e.g, Damasio et al., 1996, 2004). On both of these accounts, activity in the ATL might contribute to the N400 component. Lau et al. (2008), however, argue that the ATL is more likely to support syntactic or thematic combinatorial processing. On this view, activity in the ATL rather contributes to the P600 component. Again, further research is required to help us to decide between these hypotheses and extend our present mapping as necessary.

4.4.1.3 Angular Gyrus (AG)

Similarly, the Angular Gyrus (AG; BA39), a posterior part of the inferior parietal lobule, is also consistently implicated in language processing. In a recent review on AG function, for instance, Seghier (2013) discusses evidence for the AG's involvement in semantic processing and in word reading and comprehension, but also in number processing, attention and spatial recognition, memory retrieval, conflict resolution, and theory-of-mind/social cognition. In an attempt to unify these findings, Seghier argues that the AG is best defined as a cross-modal integrative hub that engages in event categorization, semantic access, retrieval of facts, and guidance of attention toward relevant information (see also Lau et al., 2008). If this view is correct, the question raises *if* and *how* the role of this hub in the parietal lobe is functionally distinct and/or complementary to the retrieval hub in the lpMTG. Interestingly, a potential answer to this question might be found in an hypothesis put forward by Binder and Desai (2011), who suggest that whereas the lpMTG hub might be more involved in the retrieval of conceptual representations of concrete entities, the AG hub may be geared more towards the retrieval of conceptual representations of *events*, which involve spatial and temporal interactions between entities. They exemplify this by means of the concept "birthday party", the understanding of which requires the retrieval of a configuration of concrete entities, such as friends, cake, candles, and gifts, as well as the retrieval of a series of events that unfold in time and space, such as the lighting of the candles on the cake, and the opening of the gifts. If Binder and Desai are correct, we would predict activity in the AG to be reflected in the N400 component (on top of the activity in the lpMTG), reflecting retrieval of the conceptual event representations associated with (part of) the unfolding MRC in the lIFG.

4.4.1.4 Sub-cortical areas

As for sub-cortical areas that have been implicated in language function, such as for instance the *basal ganglia* (e.g., Kotz et al., 2003), the *cerebellum* (e.g., Stowe et al., 2005), and the *thalamus* (e.g., Wahl et al., 2008), the same argument applies. We need to come up with clear hypotheses on the functional role of these areas in the comprehension system, before we can extend our functional-anatomic mapping to incorporate them. The core mapping that we have proposed is intended as a starting point for such further investigation into the functional role of these cortical and sub-cortical areas.

4.4.2 Parcellation of the lIFG and different P600s

Brouwer et al. (2012) hypothesized that the P600 component is a family of positivities, reflecting the word-by-word construction, reorganization, or updating of a mental representation of what is being communicated. Moreover, they argued that differences in amplitude, latency, duration, and scalp distribution of this component suggests that not every P600 is created alike (Kaan and Swaab, 2003b, cf.), and speculated that these electrophysiological properties may correlate with specific sub-processes that underlie the construction of a mental representation. In our functional-anatomic mapping, we put forward the hypothesis that the generation of every P600 is initiated in the lIFG, which raises the question of how differences in the electrophysiological properties of the P600 can arise. In our discussion of the functional role of the lIFG, we argued that it can be divided up into parcels, which may each have a distinct contribution to MRC construction. If different parcels of the lIFG are recruited for different sub-processes, this might explain (part of the) differences in electrophysiological properties of the P600 component. Another factor that might give rise to differences in electrophysiological properties, in particular scalp distribution, is when other nodes (cortical regions) in the neurocognitive network underlying language comprehension are invoked during integration. If, for instance, the right hemisphere homologue to the lIFG is indeed invoked in case of high processing demands (cf. Vigneau et al., 2011) or by the processing of prosodic information (cf. Friederici, 2011), the P600 component would then also reflect activity in the right hemisphere, which would affect its scalp distribution, but possibly also other properties like amplitude and duration. Interestingly, researchers have recently started categorizing differences in P600 properties (e.g. van Petten and Luka, 2012). If our mapping is correct, this means that such a categorization of members of the P600 family might eventually turn out to have an anatomical basis: the different parcels of the lIFG. Moreover, this would mean that a finer-grained mapping of the different functions reflected in the P600 component to these different parcels might provide us with a detailed functional topology of the lIFG.

4.4.3 Communication between the lIFG and lpMTG

The Retrieval-Integration account predicts that every incoming word modulates the amplitude of the N400 component (retrieval of lexical information), followed by a modulation of the P600 component (integration of the retrieved information). In terms of our mapping, this boils down to activity in the lpMTG (retrieval) followed by activity in the lIFG (integration). But the spatial separation of the epicenters for retrieval and integration predicts an additional source of activity reflecting the exchange of information between the lpMTG and the lIFG through the white matter tracts that connect them. However, as EEG measures post-synaptic potentials, rather than action potentials, we do not expect the activity in these white matter tracts to show up in the ERP signal. In addition, as hemodynamic responses may be restricted to gray matter, we also do not expect the information transfer between the lpMTG and lIFG to be visible in blood-flow based signals like the BOLD-response in fMRI (but see Mazerolle et al., 2010). This raises the question of how to study the dynamics of the information flow between the epicenters. One potential approach is to look at the pattern of oscillatory activity between the regions. Oscillatory activity reflects the synchronization of neuronal firing rates, and converging evidence suggests that such synchronization is related to the functional coupling of neural networks (engagement in cooperative processing), while desynchronization is related to their uncoupling (disengagement in cooperative processing; see Fries, 2005; Siegel et al., 2012). Studying the pattern of synchronization between the lpMTG and lIFG, using analysis techniques such as spectral coherence analysis (see Weiss and Mueller, 2003, for a review) or phase-locking statistics (Lachaux et al., 1999), may provide a window into the dynamics of the communication between these epicenters that can help to answer outstanding questions on the time-course of Retrieval-Integration cycles. For example, one could wonder whether the communication between the lpMTG and the lIFG is discrete or continuous. That is, it could be the case that information retrieved in lpMTG is immediately sent to the lIFG, independently of whether additional information is still being retrieved. The lIFG may then immediately attempt to integrate this, potentially incomplete information with a representation of what is being communicated. Under the Retrieval-Integration view, it is very likely that the lIFG is continuously active in creating a coherent representation of the situation that is currently attended. This activity will show an increase every time lexical information from a new incoming word becomes available. If this is correct, N400 and P600 components will overlap in time. As these components are generated independently, in different cortical regions or neurocognitive subnetworks, this means that they will then be summated in the resultant EEG signal. Hence, P600 amplitude may depend on preceding N400 amplitude, which may complicate the interpretation of late positivities following the N400 (for a discussion on this issue, see Hagoort, 2003). In future studies, we need to take this issue of potential overlap into account, and possibly change the way in which we analyze EEG data in order to properly disentangle retrieval and integration processes.

1. **The P600 component is generated (in part) in the lIFG, and reflects MRC composition.** That is, we hypothesize that the generation of the P600 is initiated in the lIFG, but that the full-scale of activity reflected in P600 amplitude might include activity in other, recruited cortical regions as well. This is the most important prediction that follows from our mapping. Despite that there is as of yet no consensus on the neural generators of the P600 component, a number of fMRI studies have linked P600 amplitude to activity in the lIFG (see van de Meerendonk et al., 2011, for a discussion). We did also find evidence for this prediction in our lab, as an unpublished fMRI version of the Hoeks et al. study on "Semantic Illusion" sentences

Hoeks et al. (2004) revealed a correlation between P600 amplitude and activity in the pars orbitalis (BA 47) of the lIFG.

2. **Different types of P600s (e.g., in terms of scalp distribution, onset, and/or amplitude) are generated (in part) in different parcels of the lIFG, and reflect different sub-processes of MRC composition.** That is, differences between P600s may arise because they are initiated in different parcels of the lIFG. However, whereas two distinct P600s may come about by activity in different lIFG parcels, they can also share activity in others, meaning that there may be overlap in the underlying generators of different P600s.

   Researchers have recently started to tease apart and categorize different types of P600s (see van Petten and Luka, 2012, for instance). If our mapping holds, a categorization of functionally different P600s may have an anatomical basis in the functional topology of the lIFG. Moreover, it should be noted that if, for instance, Vigneau et al. (2011) are right in that the right hemisphere is invoked whenever processing demands are high, activity in the right hemisphere may also contribute to differences in P600s.

3. **Every incoming word produces activity in the lpMTG (lexical retrieval ~N400), followed by activity in the lIFG (integration ~P600).** Preliminary EROS evidence for this prediction is provided by Tse et al. (2007), who found that both semantic and syntactic anomalies produced increased activity in the lpTMG followed by activity in lIFG. It might be that this pattern of activity occurs once for every incoming word (integration commences when retrieval is finished), or that activity alternates between the lpMTG and lIFG (integration commences as soon as a processable bit of meaning is retrieved). Albeit speculative, the aforementioned EROS study seems to suggest that the latter may be the case (see Tse et al., 2007, Table 1).

The first two predictions may be investigated using carefully aligned EEG and fMRI studies on stimuli that are known to *only* produce a P600-effect, such as syntactic violations, garden-path sentences, long-distance *wh*-dependencies, and thematic-role reversed "Semantic Illusion" sentences. If the first prediction is correct, we would expect all of these stimuli to produce increased activity in the lIFG, relative to an appropriate control. Moreover, if the second prediction holds, then these different stimuli may evoke activity in different sub-parts of the lIFG, producing characteristically different P600-effects in terms of onset, duration, and scalp-distribution, etc. The third prediction seems more challenging to test, but imaging techniques such as MEG or EROS might shed light on this issue.

4.5 Conclusion

We have proposed a minimal and parsimonious functional-anatomical mapping for language comprehension based on the Retrieval-Integration account (Brouwer et al., 2012). Our mapping entails that the processes of lexical retrieval, reflected in N400 amplitude, are mediated by the lpMTG. In the lIFG, lexical information retrieved in lpMTG is integrated with a representation of its prior context into a representation of what is being communicated. These integrative processes are reflected in P600 amplitude. The representation constructed in the lIFG subsequently provides the lpMTG with retrieval cues for the next word, upon which the Retrieval-Integration cycle repeats itself.

We argued that our mapping forms the core of the comprehension system, and we want to stress that we do not believe the comprehension system to be limited to this minimal, core network. In future research, our mapping—which is based on the 'epicenters' of the language network—may be extended to incorporate other cortical areas, such as regions in the right hemisphere (e.g., the homologues to the lIFG and lpMTG), the anterior temporal lobe (ATL), and the Angular Gyrus (AG), which have also been implicated in language comprehension, but the precise function of which is still unclear. Critically, such an extension will require us to identify the function of these regions, and ascertain what kinds of ERP components (e.g., ELAN/LAN/Sustained Negativities) result from the activation of these areas. The present mapping may serve as a starting point for such an endeavour.

The proposed correlation between activity in the lIFG and P600 amplitude also paves way for further, more fine-grained investigations into the relation between electrophysiology and brain anatomy. Earlier (Brouwer et al., 2012), it has been hypothesized that the P600 is a family of late positivities, of which different members may reflect different processes involved MRC construction. In the present chapter, we have suggested that these differences in electrophysiological properties of the P600, might arise because different instances of the P600 are generated in the different parcels of lIFG. This means that a categorization of different P600-effects (cf. van Petten and Luka, 2012), and a mapping of this categorization to the different parcels of the lIFG, might provide a means of uncovering a fine-grained functional topology of this area.

PART III

A Neurocomputational Model

CHAPTER 5

A Neurocomputational Model of the N400 and the P600 in Language Comprehension

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. Here, we will argue that there is a much simpler way to account for these data, and we will present evidence from neurocomputional simulations showing that our alternative explanation is able to predict all relevant ERP patterns found in the literature.*

5.1 Introduction

In electrophysiological research into language comprehension, there are two central brain responses. The first is the N400 component, a negative deflection of the ERP signal (Event-Related brain Potential) 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 survived until this day. 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, among others). Hoeks et al. (2004), for instance, found that 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. To account for these effects, increasingly complex 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 (often structurally) is reflected in a P600-effect. Despite the attractiveness of these models, however, none of them seems capable of explaining the full range of relevant findings in the literature (see Brouwer et al., 2012, for a review).

5.1.1 A simpler perspective

In contrast to seeking an architectural explanation for the "Semantic Illusion" phenomenon (absence of an N400-effect in response to a semantic anomaly), Brouwer et al. (2012) argued for a functional reinterpretation of the ERP components involved. First of all, in line with others (e.g., Kutas and Federmeier, 2000; Lau et al., 2008; van Berkum, 2009), they suggested that the N400 component reflects *retrieval* of lexical semantic information, rather than compositional semantic processing or semantic integration. Retrieval of the information associated with a word is facilitated if that information is already (partly) activated by its prior context. This explains why the word *socks* engenders a much larger N400 in the context of 'He spread his warm bread with [...]' than the word *butter*: the lexical knowledge associated with *socks* is inconsistent with that already activated by its prior context (*socks* do not fit well with *spread* and *warm bread*), whereas the conceptual knowledge associated with *butter* is. It is important to note that under this account pre-activation can stem from the message representation that has been constructed so far (e.g., a breakfast scene), as well as from the preceding lexical items themselves (*spread* and *bread*).

In addition to this effect on N400 amplitude, presenting the word *socks* also leads to an increase in P600 amplitude, even though the sentence in which it appears is grammatically correct (see Kutas and Hillyard, 1980, Figure C). The same is true for the "Semantic Illusion" sentences, where a P600-effect is elicited in a syntactically well-formed sentence. Brouwer et al. (2012) argued that these results indicate that the P600 is not a reflection of *syntactic* processing, but must represent some form of effortful *semantic* processing or integration. They put forward the hypothesis that the P600 component is actually a family of related components, all of which are late positivities that reflect the word-by-word construction, reorganization, or updating of a *mental representation of what is being communicated* (an MRC for short). Given this view on the P600 component, and the retrieval view on the N400 component, Brouwer et al. (2012) suggest that language is processed in biphasic N400/P600—Retrieval-Integration—cycles; every incoming word modulates the N400 component, reflecting the effort involved in activating its associated conceptual knowledge, as well as the P600 component, reflecting the effort needed to integrate this knowledge into an updated representation of what is being communicated. This parsimonious single stream account was shown to have the broadest empirical coverage of all the extant models (Brouwer et al., 2012), and indicated that the so-called "Semantic Illusion" in sentence processing, the effect that had led to so many new models, was just a simple instance of priming: if a preceding context (e.g., 'The javelin has the athletes [...]') pre-activates the lexical features of an incoming word (e.g, *thrown*), there is no N400-effect. On the other hand, when an incoming word is not consistent with the pre-activated lexical features (e.g., *summarized*), there is an N400-effect. Futhermore, P600 amplitude is expected to increase in every case where there is difficulty in making sense.

5.1.2 Mapping function to anatomy

Brouwer and Hoeks (2013) aligned these functional interpretations of the N400 and the P600 with neuroanatomy, and suggested that the processes of retrieval and semantic integration are mediated by specific brain areas. First of all, they 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 collected 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). Brouwer and Hoeks (2013) suggested 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, we can describe an approximate functional-anatomic Retrieval-Integration cycle of an incoming word (see Figure 5.1, 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.

In summary, compared to the five competing models, the Retrieval-Integration account is 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 it 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.

5.1.3 A neurocomputional model

We present 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 that we described above. Following our hypothesis on the neuroanatomical organization of the language processor, the computational model consists of two connected but relatively independent subsystems: A system for retrieval (∼lpMTG) and a system for integration (∼lIFG). 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).

In what follows, we will first present the model conceptually, and show that it can produce the same patterns of ERPs as found in an actual empirical study. In the 'Methods' section, we will provide the full technical details of the model.

5.2 Results

5.2.1 The neurocomputational model

Our neurocomputational model consists of five layers of artificial neurons, one corresponding to the auditory/visual cortex (ac/vc), two for the left posterior MTG (lpMTG and lpMTG_output), and two for the left IFG (lIFG and lIFG_output) (see Figure 5.1, bottom). Activation flows forward from the ac/vc layer all the way to the lIFG_output layer, and from the lIFG layer both back into itself (in order to update the MRC at the previous time-step) and to the lpMTG layer (providing a context for retrieval). 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).

5.2.2 Linking hypotheses

In our view, 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, we compute the correlates of N400 amplitude at the lpMTG layer,

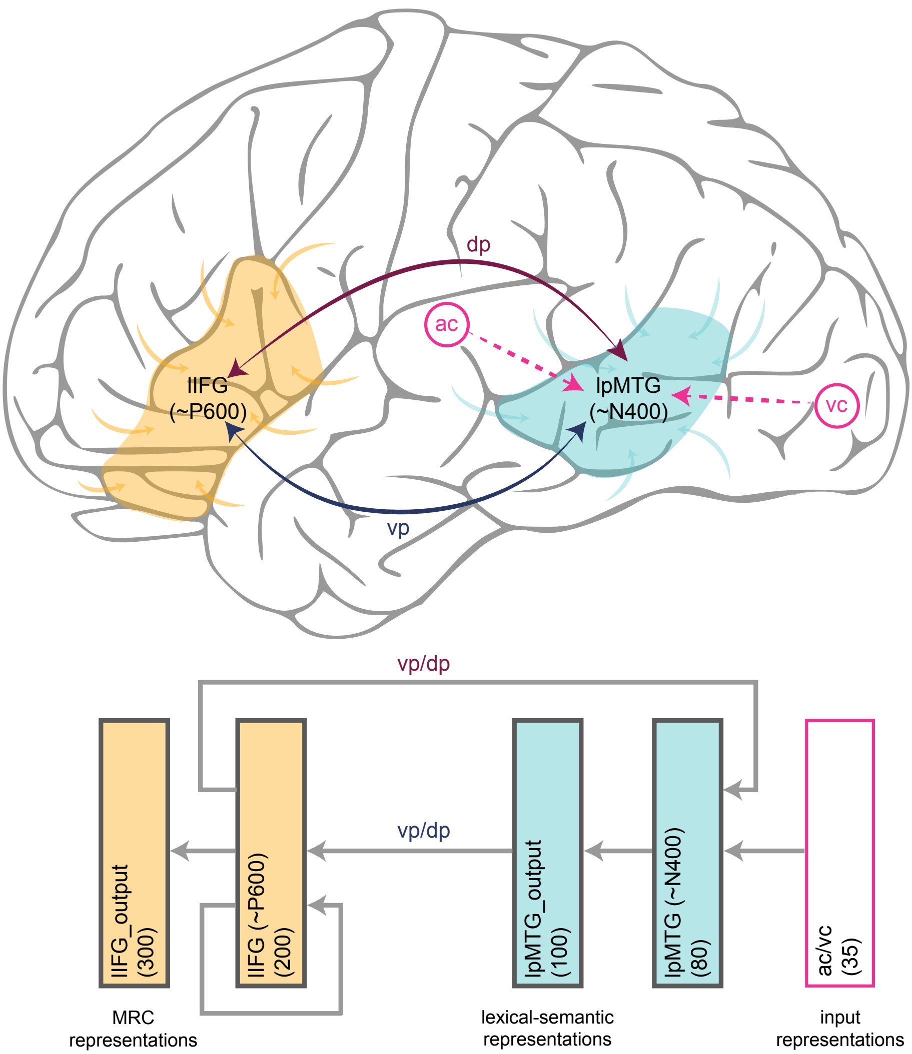

Figure 5.1 | *Schematic illustration of the neurocomputational model (bottom), and its relation to the functional-anatomic mapping of the core language network (top). An incoming word reaches the lpMTG via either the auditory cortex or visual cortex (ac/vc). The lpMTG then retrieves the conceptual knowledge associated with this word from the association cortices (lpMTG → lpMTG_output), thereby generating the N400. Next, this retrieved meaning is sent to the lIFG (lpMTG_output → lIFG), where it is integrated with its prior context (lIFG → lIFG), into an updated MRC (lIFG → lIFG_output). The updated MRC in the lIFG subsequently provides a context for the retrieval of the conceptual knowledge assocatied with the next word (lIFG → lpMTG).*

Item	Condition	Effect
De speer werd door de atleten geworpen *The javelin was by the athletes thrown*	Control (Passive)	—
De speer heeft de atleten geworpen *The javelin has the athletes thrown*	Reversal (Active)	P600
De speer werd door de atleten opgesomd *The javelin was by the athletes summarized*	Mismatch (Passive)	N400/P600
De speer heeft de atleten opgesomd *The javelin has the athletes summarized*	Mismatch (Active)	N400/P600

Table 5.1 | *Materials used in the original ERP experiment, as well as the effects that were observed for each condition.*

where the activation of lexical-semantic features takes place (∼memory retrieval), as the degree to which the pattern of activity induced by the current word, and that induced by the previous word are *different* (see 'Methods' section 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, we compute the correlates of P600 amplitude 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 thematic-role assignment—takes place (see also Crocker et al., 2010).

5.2.3 Modeling Event-Related Potentials

The patterns of N400 and P600 elicitation reported in the literature (see Kutas et al., 2006, for an overview) suggest that there are two main processing outcomes. The most common is the biphasic pattern where an N400-effect co-occurs with a P600-effect. When context does not prepare for a specific upcoming word, retrieval is more difficult (hence an N400-effect), and often the retrieved information is more or less unexpected, leading to more effortful MRC construction (P600-effect). Less frequent are cases that elicit only a P600-effect, such as in "Semantic Illusion" sentences. There, the retrieval of a word is facilitated by context or preceding lexical items (hence no N400-effect), but integration is problematic (P600-effect). Theoretically, there is also a third option: retrieval is more involved, but integration is as easy as in a control sentence (hence only an N400-effect, and no P600-effect). However, isolated N400-effects are scarce, and their presence may be a baseline artefact (Brouwer and Hoeks, 2013; Hoeks and Brouwer, 2014).

Our computational model should thus be able to simulate the two most common findings: biphasic N400/P600-effects as well as isolated P600-effects. Both types of outcomes are present in the study by Hoeks et al. (2004), which we will take as reference point. This study compared semantically anomalous Dutch sentences like 'De speer *heeft* de atleten geworpen' (lit: 'The javelin *has* the athletes thrown') to normal controls like 'De speer *werd door* de atleten geworpen' (lit: 'The javelin *was by* the athletes thrown'). This comparison produced a P600-effect on the final verb *thrown*, but no N400-effect. Two other semantically anomalous conditions were also compared to the same control, which both gave rise to a biphasic N400/P600-effect on the final word: 'De speer *heeft* de atleten opgesomd' (lit: 'The javelin *has* the athletes summarized') and 'De speer *werd door* de atleten opgesomd' (lit: 'The javelin *was by* the athletes summarized'). Table 5.1 provides an overview of these materials and findings, and Figure 5.2 shows the results at the Pz electrode.

5.2.4 Simulation experiments

We modeled the results of the Hoeks et al. (2004) study in two simulation experiments. To assure that any effects that we find are not an artifact of the materials that we used, we used different sets of lexical items, and hence different sets of lexical-semantic representations, in the two simulation experiments.

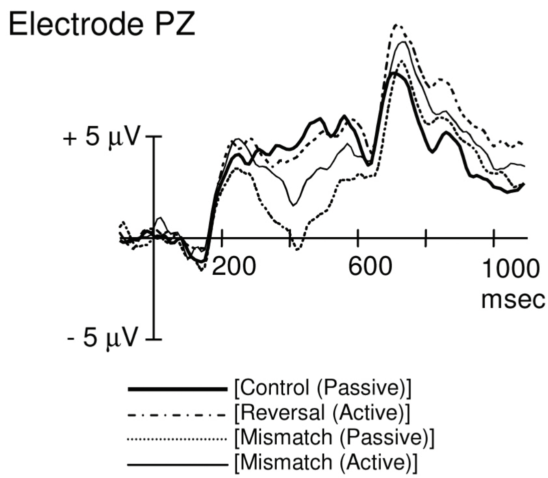

Figure 5.2 | *Results of the original ERP experiment on the Pz electrode. Positive is plotted upwards. Note that the reported ERP effects are based on multiple electrodes, and that this single electrode only serves as a representative example.*

5.2.4.1 N400 results

The computational model should be able to produce two very different findings in the N400 domain: an absence of an N400-effect in the comparison of "Semantic Illusion" sentences ('The javelin *has* the athletes <u>thrown</u>') with normal controls, and presence of an N400-effect in the mismatch anomaly conditions ('The javelin *has/was by* the athletes <u>summarized</u>'). Figure 5.3 shows the N400 results of the two simulation experiments (bottom graphs) in comparison to those of the original ERP experiment (top). Statistical evaluation using Repeated Measures ANOVA (with Condition as four-level within-items factor and Huynh-Feldt correction where necessary) showed a close replication of the Hoeks et al. (2004) findings. The main effect of Condition was significant in each of the simulation experiments (Exp 1: $F(3,27)=35.2$; $p<.001$; Exp 2: $F(3,27)=15.3$; $p<.001$), and pairwise comparisons (Bonferroni corrected) showed that 1) N400 amplitude

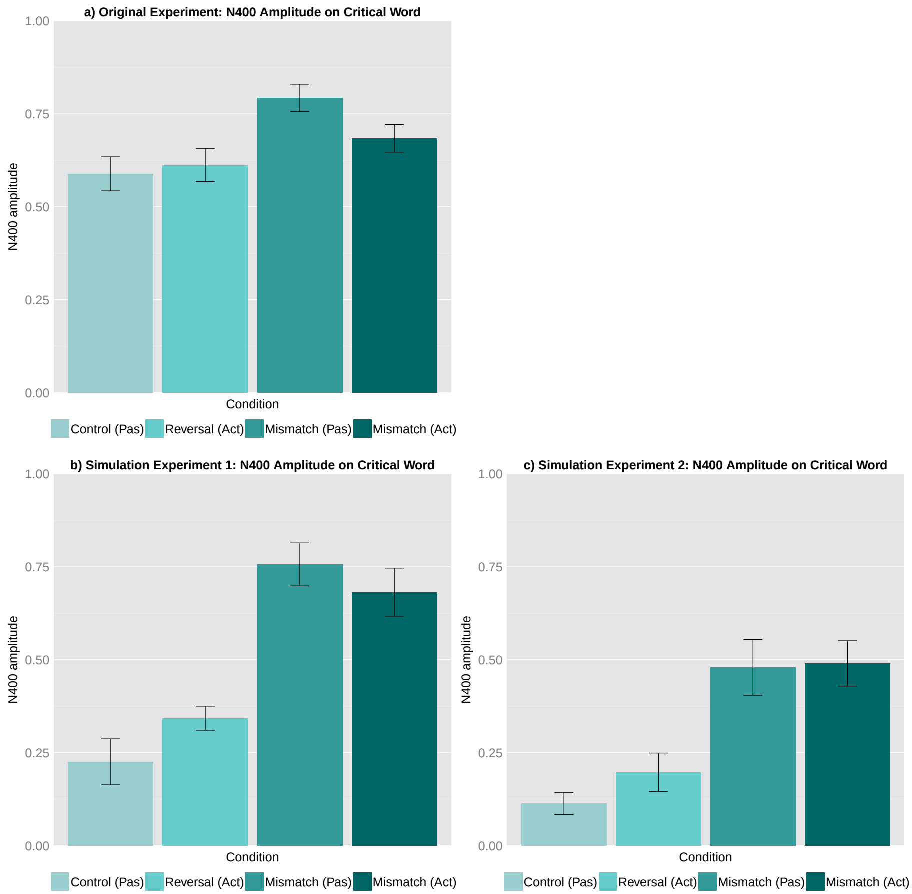

Figure 5.3 | *N400 results of the simulations in comparison to the results of the original experiment. Panel (a) shows the N400 amplitudes as measured in the original experiment (at the Pz electrode). Panel (b) shows the N400 correlates measured in simulation experiment 1, and panel (c) those measured in simulation experiment 2. All scores are transformed to a common scale. Error bars show standard errors of these transformed scores.*

did not differ between "Semantic Illusion" sentences and controls (Exp 1: p=.47; Exp 2: p=.91), and 2) there was a significant N400-effect for the two other anomalous conditions (Exp 1: p-values<.001; Exp 2: p-values<.005).

5.2.4.2 P600 results

For the P600 component, the main result that we wanted to replicate is an increase in amplitude for all anomalous conditions relative to control. Figure 5.4 shows the P600 results of the two simulation experiments (bottom graphs) in comparison to those of the original ERP experiment (top left). Again, we clearly replicated the main result; in both simulations, P600 amplitude was significantly higher for all anomalous sentences (including the "Semantic Illusion" sentences) compared to controls (Main effect of Condition: Exp 1: F(3,27)=118.9; p<.001; Exp 2: F(3,27)=57.1; p<.001); pairwise comparison showed that there was a significant P600-effect for all three anomalous conditions compared to control (Exp 1: p-values<.001; Exp 2: p-values<.001).

One slight difference between the actual and simulated results lies in the ordering of the three implausible conditions. The computational model predicts P600 amplitude to be largest for both *mismatch* conditions, whereas in the original ERP experiment, it is largest for the "Semantic Illusion" condition. A possible reason for this difference is that in the ERP experiment the pattern of P600-effects is affected by an artefact caused by *component overlap* (see Hagoort, 2003; Brouwer and Hoeks, 2013, for discussion), where the size of the preceding N400 affects the size of the P600. Figure 5.4 (top right) shows that if we correct for component overlap (by subtracting N400 amplitude from P600 amplitude in each condition), the pattern of results within the three anomalous conditions comes in line with the results of the simulations (mismatch conditions larger than "Semantic Illusion" sentences).

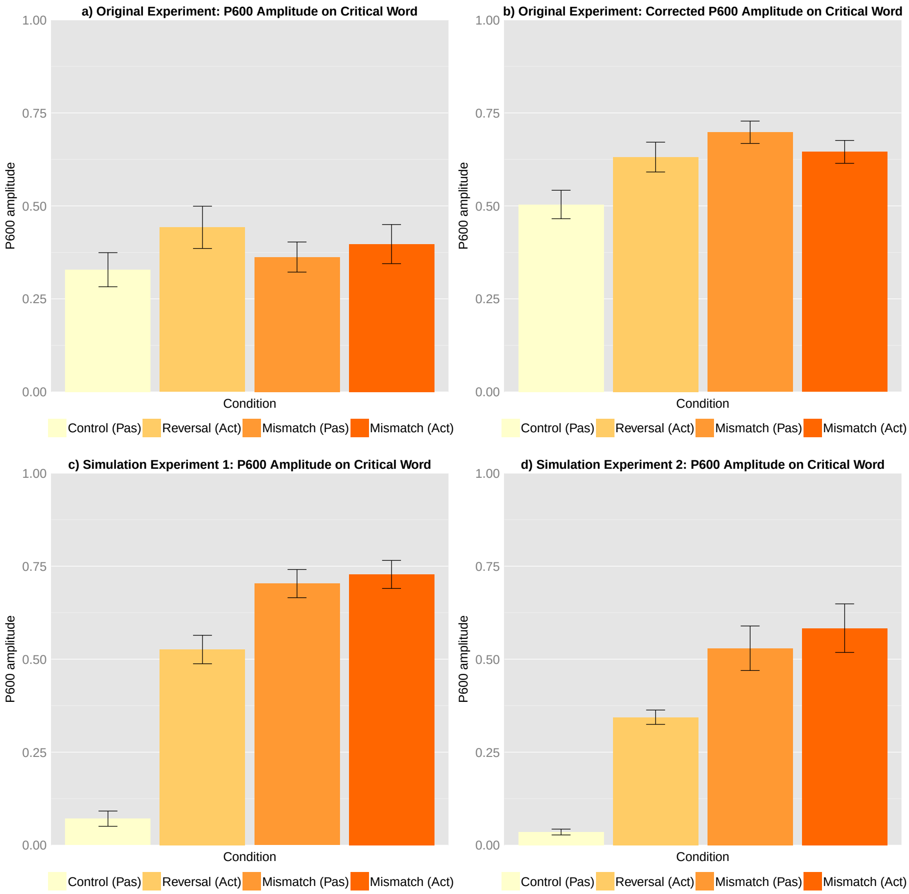

Figure 5.4 | *P600 results of the simulations in comparison to the results of the original experiment. Panel (a) shows the P600 amplitudes as measured in the original experiment (at the Pz electrode). Panel (b) shows the P600 amplitudes corrected for overlap with the N400 component (also at Pz). Panel (c) shows the P600 correlates measured in simulation experiment 1, and panel (d) those measured in simulation experiment 2. All scores are transformed to a common scale. Error bars show standard errors of these transformed scores.*

5.3 Discussion

We presented a neurocomputational model of language processing that directly instantiates a simple functional-anatomic mapping of the core language network. In our model, the N400 component and the P600 component reflect distinct processes that take place in distinct cortical regions; the N400 component reflects lexical retrieval processes mediated by the lpMTG, and the P600 component reflects integrative processes mediated by the lIFG. The computational model was trained to honor this division of labor by means of a two-stage training procedure (see 'Methods' section). The model was able to simulate the two most important patterns of ERPs as reported in the literature (P600-effects and biphasic N400/P600-effects), thereby providing a 'proof of concept' of our Retrieval-Integration view (cf. Brouwer and Hoeks, 2013).

5.3.1 Implications for other models

Our simulations have several implications for previously proposed neurocognitive processing models (Kolk et al., 2003; Kim and Osterhout, 2005; Kuperberg, 2007; Bornkessel-Schlesewsky and Schlesewsky, 2008; Hagoort et al., 2009). First of all, our simulations confirm that there is no need for an independent semantic analysis stream to explain "Semantic Illusions" in sentence processing, or "Semantic P600"-effects (cf. Brouwer et al., 2012). Our model simply 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 an MRC. Each time a word (or 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 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. Our model is both architecturally more parsimonious than previously proposed models, while having broader empirical coverage (see Brouwer et al., 2012, Table 2). Secondly, previously proposed models have been developed as 'box-and-arrow' models, limiting them to *qualitative* predictions about the presence or absence of ERP effects. Instantiating our functional-anatomic mapping as a neurocomputational model, by contrast, adds a *quantitative* dimension to the predictions that it makes. A clear example here, is that whereas all other models simply predict absence of an N400-effect for the reversal condition (Brouwer et al., 2012, Table 2), we predict a slight, albeit non-significant increase in N400 amplitude relative to control. Critically, this slight increase is also present in the original ERP data. Hence, in addition to being more parsimonious, our model also makes more fine-grained predictions.

5.3.2 Future directions

The present computational model represents just a first step towards a full neurocomputational model of language processing. It will be necessary to expand our simulations in several respects. For one, we want to model other types of anomalous sentences. In addition, we want to model P600-effects arising from other processing phenomena, such as syntactic violations (see Gouvea et al., 2010, for an overview), and pragmatic violations (see Hoeks et al., 2013; Hoeks and Brouwer, 2014). Also, we want to include other ERP components into our model, such the Early Left Anterior Negativity (ELAN), which is elicited by word category violations, and the Left Anterior Negativitiy (LAN), which appears to be related to the processing of morphosyntactic marking (see Friederici, 2011). A completely different direction for future exploration is to expand the number of ERP component features covered by our model. For now we have focused on component amplitude (and implicitly polarity), but a component is also defined by its *latency*, *duration*, and *scalp distribution*. It remains to be seen how these features can be incorporated into our model.

5.4 Methods

5.4.1 Model Architecture and Activation Flow

The model is essentially an extended Simple Recurrent Network (SRN) (Elman, 1990) consisting of five layers: ac/vc (35 units), lpMTG (80), lMTG_output (100), lIFG (200), and lIFG_output (300) (see Figure 5.1). Input patterns are clamped to the ac/vc layer, and activation propagates feed forward to the lIFG_output layer following the trajectory: ac/vc → lpMTG → lpMTG_output → lIFG → lIFG_output. In addition, at a given processing timestep t, the lIFG and the lpMTG also receive input from the activation pattern of the lIFG at timestep $t-1$ through an additional 200-unit context layer that receives a copy of the activation pattern of the lIFG prior to feed forward propagation. At timestep $t = 0$, the activation value of each unit in this context layer was set to 0.5. All layers except this context layer and the ac/vc layer also receive input from a bias unit, the value of which is always 1. Unit activation y_j of a unit j is determined by a sigmoid activation function:

$$y_j = \frac{1}{1 + e^{-x_j}} \tag{5.1}$$

where x_j is the net input to unit j:

$$x_j = \sum_i y_i w_{ij} \tag{5.2}$$

which is determined by the activation level y_i of each unit i that propagates to unit j, and the weight w_{ij} on the connection between these units.

The model was trained in two stages (see below for detailed information on the training procedure). In the first stage, the ac/vc layer and the lpMTG layer were excluded from the network, and the model was trained to map sequences of *lexical-semantic representations* (see below for details on the representations used), clamped to the lpMTG_output layer, to *thematic-role assignment representations* in the lIFG_output layer. In the second stage, the ac/vc layer and the lpMTG layer were included, and the model was trained to map sequences of *acoustic/orthographic representations* clamped to the ac/vc layer to thematic-role assignment represenions in the lIFG_output layer. Critically, all the weights that were present in the network during the first training stage were frozen, forcing the network to first map the acoustic/orthographic representations to lexical-semantic representations in the lpMTG_output layer, before mapping these to thematic-role assignment representations in the lIFG_output layer.

5.4.2 Training and Testing Patterns

For both experiments, we created two sets of training items (one for both training stages), and a set of test items. The training items consist of Dutch active and passive sentences with the following template structure:

Active sentences:

De	[Agent]	heeft	het/de	[Patient]	[Action]
The	[Agent]	has	the$_{(+/-\text{NEUTER})}$	[Patient]	[Action]

Passive sentences:

De	[Patient]	werd	door	het/de	[Agent]	[Action]
The	[Patient]	was	by	the$_{(+/-\text{NEUTER})}$	[Agent]	[Action]

From these templates, training sentences were generated by filling in the *agent*, *patient*, and *action* slots using the noun phrases (agent and patient) and verbs (action) listed in Table 5.2. Half of the training sentences were constructed by permuting each of the twenty noun phrases (agents plus patients) with each verb ($2 \times 20 \times 20 \times 10 = 8000$ items). The other half consisted only of sentences with stereotypical agent-patient-action combinations, which are listed in the rows of Table 5.2. Hence, there are a total of 16000 (2×8000) training items, in which each verb appears 1600 times, 802 times ($\approx 50\%$) of which in a stereotypical agent-patient-action construction. As the model processes sentences word-by-word, each training item consists of a sequence of either 6 (active sentences) or 7 (passive sentences) pairs of input and target patterns. The input patterns consist of either lexical-semantic representations (stage one), or acoustic/orthographic representations (stage two). The target patterns, in turn, are always the desired thematic-role assignment representation for a sentence.

The test sets consist of 40 sentences that are evenly divided over four conditions. There are 10 passive stereotypical agent-patient-action sentences (controls), 10 active role-reversed sentences (reversals), which were constructed by swapping the stereotypical agents and patients, and 10 active as well as 10 passive mismatch sentences (active and passive mismatches), in which the stereotypical action verb is replaced by the mismatch verb listed in Table 5.2.

5.4.3 Representations

5.4.3.1 Acoustic/orthographic vectors

The acoustic/orthographic input representations are 35-unit localist vectors, in which each unit corresponds to a single word (20 nouns + 10 verbs + 2 auxiliary verbs + 2 determiners + 1 preposition = 35 words).

5.4.3.2 Lexical-semantic vectors

The lexical-semantic representations are 100-unit binary representations, which were derived from a large corpus of Dutch newspaper texts using the Correlated Occurrence Analogue to Lexical Semantics (COALS; Rohde et al., 2009). In all of the steps described below, we precisely follow the procedure laid out in Rohde et al. to derive these representations (or COALS vectors).

We first derived a co-occurrence matrix using a 4-word ramped window, meaning that a word a co-occurs with b if a occurs within 4 words to the left or right of b, and that this co-occurrence is weighted by the proximity of a to b on a scale of 4 (direct neighbor) to 1 (separated by three words). We then pruned all but the 14.000 columns representing the most frequent words, so that the rows of the matrix then represented 14K-dimensional word feature vectors. Next, the weighted frequency of each co-occurrence

$w_{a,b}$ of words a and b was normalized by converting it to a pairwise correlation:

$$w'_{a,b} = \frac{T \cdot w_{a,b} - \sum_j w_{a,j} \cdot \sum_i w_{i,b}}{(\sum_j w_{a,j} \cdot (T - \sum_j w_{a,j}) \cdot \sum_i w_{i,b} \cdot (T - \sum_i w_{i,b}))^{\frac{1}{2}}} \tag{5.3}$$

where i is a row index, j is a column index, and:

$$T = \sum_i \sum_j w_{i,j} \tag{5.4}$$

In the resulting matrix, we replaced each negative correlation with 0, and each positive correlation with its square root:

$$norm(w'_{a,b}) = \begin{cases} 0 & \text{if } w'_{a,b} < 0 \\ \sqrt{w'_{a,b}} & \text{otherwise} \end{cases} \tag{5.5}$$

To obtain the 100-dimensional feature vectors that we used in our simulations, we pruned all but the 15.000 rows representing the most frequent words, and then reduced the dimensionality of the normalized feature vectors for these words by computing the Singular Value Decomposition of the co-occurrence matrix $X_{15000 \times 14000}$. Here we considered only the first 100 singular values and vectors, such that we obtain matrix $\hat{X}$ that is the best rank-100 approximation to X in terms of sum squared error:

$$\hat{X}_{15000 \times 14000} = \hat{U}_{15000 \times 100} \hat{S}_{100 \times 100} \hat{V}^T_{100 \times 14000} \tag{5.6}$$

A 100-unit feature vector V_c for a word c is then defined as:

$$V_c = X_c \hat{V} \hat{S}^{-1} \tag{5.7}$$

which can be converted to a binary vector by setting its negative components to 0, and its positive components to 1.

5.4.3.3 Thematic-role assignment vectors

The thematic-role assignment representations are 300-unit vectors, which are divided into three 100-unit slots. These three slots respectively represent the lexical-semantic representations of the elements that will be *agent*, *action*, and *patient* (cf. Mayberry et al., 2009).

5.4.4 Training and Testing

In both stages, the model was trained using bounded gradient descent (Rohde, 2002), a modification of the standard backpropagation algorithm (Rumelhart et al., 1986a).

For each input-target pair c, we minimized the sum squared error E_c between the desired activity d_j and the observed activity y_j for each unit j in the lIFG_output layer:

$$E_c = \frac{1}{2}\sum_j (y_j - d_j)^2 \tag{5.8}$$

Error was reduced by adjusting each weight w_{ij} in the model on the basis of a delta that is proportional to the gradient of that weight, and depends on its previous delta:

$$\Delta w_{ij}(t) = -\varepsilon\rho\frac{\partial E}{\partial w_{ij}} + \alpha\Delta w_{ij}(t-1) \tag{5.9}$$

where ε is the network's *learning rate*, ρ a *scaling factor* that depends on the length of the entire gradient:

$$\rho = \begin{cases} \frac{1}{||\partial E/\partial w||} & \text{if } ||\partial E/\partial w|| > 1 \\ 1 & \text{otherwise} \end{cases} \tag{5.10}$$

and α a momentum coefficient, controlling the fraction of the previous weight delta to be added.

The gradient $\frac{\partial E}{\partial w_{ij}}$ of a weight w_{ij}, in turn, is estimated as the product of the *error signal* δ_j of a unit j, and the activation value y_i of a unit i that signals to unit j:

$$\frac{\partial E}{\partial w_{ij}} = \delta_j y_i \tag{5.11}$$

The error signal δ_j for an output unit j is defined as:

$$\delta_j = (y_j - d_j)(y_j(1 - y_j) + 0.1) \tag{5.12}$$

where the constant 0.1 is a flat spot correction constant, preventing the derivative $y_j(1 - y_j)$ of the sigmoid activation function to approach zero when y_j is near 0 or 1 (cf. Fahlman, 1988). The error signal δ_j for a hidden unit j, in turn, is defined as:

$$\delta_j = (y_j(1 - y_j) + 0.1) \sum_k \delta_k w_{jk} \tag{5.13}$$

where all units k are units that receive signals from unit j.

We trained the model for 7000 epochs, in each of which we accumulated gradients over 100 input-output pairs before updating the weights. Training items were presented in a permuted order, such that by the end of training, the model has seen each item at least 43 times ($7000/(16000/100) = 43.75$). After all of the 16000 items were presented once, the training order was permuted again. Weights were initially randomized within a range of $(-0.25, +0.25)$, and were updated using a learning rate ε of 0.2, which was scaled down to 0.11 with a factor of 0.95 after each 700 epochs (that is, after each 10% interval of the total epochs; $0.2 \times 0.95^{10} \approx 0.11$). The momentum coefficient α was set to a constant of 0.9. Finally, we used a *zero error radius* of 0.1, such that no error was backpropagated if $(y_j - d_j) < 0.1$. The training procedure was identical for stage one and two.

After training, we evaluated the comprehension performance of the model using a output-target similarity matrix. For each item, we computed the cosine similarity between the output vector for that item, and each of the 16000 different target vectors. The cosine similarity between two vectors is defined as:

$$cos(x, y) = \frac{\sum_i x_i \times y_i}{\sqrt(\sum_i x_i^2) \times \sqrt(\sum_i y_i^2)} \tag{5.14}$$

The output vector for an item was considered correct if it was more similar to its corresponding target vector than to the target vector of any other item. We computed comprehension performance after stage one on the training set (at the lpMTG_output layer) and after stage two on both the training and the test set (at the lIFG_output layer). In all cases, comprehension performance was perfect (100% correct) for both experiments.

5.4.5 Computing ERP correlates

The ERP correlates for the N400 component were computed as the cosine dissimilarity between the lpMTG vector at time-step t and the lpMTG vector at time-step $t-1$:

$$\text{N400} = 1 - cos(lpMTG_t, lpMTG_{t-1}) \tag{5.15}$$

The P600 correlates were computed in a similar fashion at the lIFG layer:

$$\text{P600} = 1 - cos(lIFG_t, lIFG_{t-1}) \tag{5.16}$$

Exp.	Agent	Patient	NEUTER	Action	Mismatch
1	voetballer *soccer player*	doelpunt *goal*	+	gescoord *scored*	gediend *served*
1	militair *soldier*	land *country*	+	gediend *served*	gescoord *scored*
1	kok *cook*	maaltijd *meal*	-	bereid *prepared*	gezongen *sung*
1	zanger *singer*	lied *song*	+	gezongen *sung*	bereid *prepared*
1	advocaat *lawyer*	bedrijf *company*	+	aangeklaagd *sued*	gelopen *ran*
1	atleet *athlete*	marathon *marathon*	-	gelopen *ran*	aangeklaagd *sued*
1	politicus *politician*	debat *debate*	+	gevoerd *engaged*	uitgegeven *published*
1	uitgever *publisher*	roman *novel*	-	uitgegeven *published*	gevoerd *engaged*
1	arts *doctor*	diagnose *diagnosis*	-	gesteld *made*	geschilderd *painted*
1	schilder *painter*	schilderij *painting*	+	geschilderd *painted*	gesteld *made*
Exp.	**Agent**	**Patient**	NEUTER	**Action**	**Mismatch**
2	rechercheur *detective*	moord *murder case*	-	opgelost *solved*	verhoogd *raised*
2	werkgever *employer*	salaris *salary*	+	verhoogd *raised*	opgelost *solved*
2	dief *thief*	museum *museum*	+	beroofd *robbed*	getrokken *pulled*
2	tandarts *dentist*	tand *tooth*	-	getrokken *pulled*	beroofd *robbed*
2	schipper *sailor*	schip *ship*	+	aangelegd *berthed*	geregisseerd *directed*
2	regisseur *director*	film *movie*	-	geregiseerd *directed*	aangelegd *berthed*
2	piloot *pilot*	vliegtuig *airplane*	+	bestuurd *steered*	afgelegd *taken*
2	student *student*	tentamen *examen*	+	afgelegd *taken*	bestuurd *steered*
2	verzekeraar *insurer*	verzekering *insurance*	-	uitgekeerd *paid*	gereden *rode*
2	wielrenner *cyclist*	etappe *stage*	+	gereden *rode*	uitgekeerd *paid*

Table 5.2 | *Overview of the materials used in the simulation experiments. See text for details.*

PART IV

Conclusions

CHAPTER 6

Conclusions and Future Perspectives

Ever since Kutas and Hillyard (1980), studies using the measurement of Event-Related brain Potentials (ERPs) have been very important to our understanding of the neural basis of language comprehension. Some time ago, however, ERP studies have stumbled upon a puzzle that came to be known as the "Semantic Illusion"-effect (SIE). Hoeks et al. (2004), for instance, found that syntactically well-formed, but semantically anomalous sentences like 'De speer heeft de atleten geworpen' (lit: 'The javelin has the athletes thrown') did not produce an increase in N400 amplitude (relative to a non-anomalous control), which was unexpected given that the N400 component is generally taken to reflect difficulty in *semantic integration*. Instead, the critical verb *thrown* was found to produce an increase in P600 amplitude, which was equally surprising as the P600 component is generally taken to reflect *syntactic processing*. The absence of an N400-effect has been taken to indicate that the processor temporarily assumes (is under the "illusion") that the sentence is semantically felicitous. This illusion is, however, only temporary, as the presence of a P600-effect indicates that the processor quickly realizes that there is something wrong with the sentence. To account for the SIE, five complex neurocognitive processing models have been put forward that allow an independent semantic processing stream to arrive at the interpretation of a sentence independent of its syntactic 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). 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 this thesis, therefore, was to determine using computational modeling, which one of these models is most viable, thereby offering a formally precise explanation of the SIE in language processing.

In part I, we started by critically reviewing the five multi-stream models that have been put forward to explain the SIE (chapter 2). On the basis of this review, we concluded that none of these models succeeds in explaining all of the relevant findings in the literature, and that the reason for this failure is not architectural in itself, but rather due to the interpretation of the N400 component as reflecting semantic integration (compositional semantic processing) and the P600 component as reflecting syntactic processing. In chapter 3, we argued on the basis of converging evidence that the N400 component is better interpreted as reflecting *lexical retrieval* processes (cf. Kutas and Federmeier, 2000; Lau et al., 2008; van Berkum, 2009), and that instead the P600 reflects *semantic integration*. More specifically, we hypothesized that the P600 is a family of late positivities that reflect the word-by-word construction, updating, or reorganization of a mental representation of what is being communicated (an MRC for short). Taken together, these reinterpretations of the N400 and the P600 suggest that language processing proceeds in biphasic N400/P600—Retrieval-Integration—cycles. In these cycles, N400 amplitude reflects the ease with which the conceptual knowledge associated with an incoming word can be activated in memory, and P600 amplitude the effort involved in integrating this activated knowledge with a representation of the prior context, into an updated mental representation of what is being communicated (an updated MRC). Critically, it is shown that these Retrieval-Integration cycles do not only provide a parsimonious, single-stream account of the SIE, but also that they have broad empirical coverage (see also Hoeks and Brouwer, 2014).

In part II, we aligned our reinterpretations of the N400 component and the P600 component with neuroanatomy. We derived a minimal functional-anatomic mapping centered around the left posterior Middle Temporal Gyrus (lpMTG; BA 21) as an *epicenter* or *hub* for retrieval (∼N400), and the left Inferior Frontal Gyrus (lIFG; BA 44/45/47) as an epicenter for MRC composition (∼P600) (chapter 4). In a functional-anatomic Retrieval-Integration cycle, incoming words reach the lpMTG via either the auditory or visual cortex (ac/vc), depending on whether they are spoken or written. The lpMTG then retrieves the conceptual knowledge associated with a given word from the association cortices (generating the N400 component), and connects it to the lIFG via either the dorsal or ventral white matter pathway(s) that connect the lpMTG and the lIFG. In the lIFG, the retrieved knoweldge is integrated with the prior MRC (generating the P600 component), into an updated MRC. Finally, the updated MRC in the lIFG is connected back to lpMTG via the dorsal or ventral pathway in order to pre-activate the conceptual knowledge associated with a possible upcoming word. We concluded that the functional-anatomic mapping that we propose forms the *core* of the language comprehension system that provides a starting point for a more focused investigation into the coupling of electrophysiology and neuroanatomy.

In part III, we provided a proof of concept of our functional-anatomic mapping in terms of a neurocomputional model. This model, which is an anatomically-constrained Simple Recurrent neural Network (SRN; Elman, 1990), constructs a thematic role assignment representation for an input sentence on an incremental, word-by-word basis. During processing, correlates of N400 amplitude are measured in a layer representing the lpMTG, and those of P600 amplitude in a layer representing the lIFG. It is shown that our neurocomputational model produces ERP elicitation behavior similar to that of humans (cf. Hoeks et al., 2004). We concluded that our model—which synthesizes the ideas on electrophysiology developed in part I and those on neuroanatomy developed in part II—provides strong support for the Retrieval-Integration framework, and that neurocomputational modeling provides a promising methodology for arriving at a neurobiological model of cognition.

In the remainder of this fourth and final part, we will briefly lay out what we believe to be the most important future directions for the neurocomputational modeling of language comprehension.

6.1 Future perspectives

6.1.1 Modeling other processing phenomena

The simulations presented in chapter 5 focused on the N400 component and the P600 component in response to semantic anomalies; we modeled isolated P600-effects, as well as biphasic N400/P600-effects. In future work, we aim to broaden the spectrum of language processing phenomena covered by our simulations. An obvious first step is to include *syntactically*-induced P600-effects (see Gouvea et al., 2010, for an overview). We might model P600-effects in response to agreement violations, for instance, by incorporating plural noun phrases into our training data, such that it is possible for the auxiliary verbs to disagree with the subject of a sentence in number, as in 'De koks <u>heeft</u> de maaltijd bereid' (lit: 'The cooks <u>has</u> the meal prepared'). Besides syntax-related P600 modulations, we also aim to simulate *pragmatically*-induced P600-effects (see Hoeks et al., 2013; Hoeks and Brouwer, 2014). These pragmatically-induced late positivities often involve inferences that require knowledge about the world, but due to the amount of world knowledge that is required for the generation of even the simplest inferences, it is often left out in computational simulations (cf. Frank et al., 2008). One possible strategy to enable processing of world knowledge, however, is to limit the scope of the world rather than the amount of world knowledge made available to a model (cf. Frank et al., 2008, 2009). An interesting direction for future exploration is to see if this *microworld*-strategy can be used to capture P600 modulations that are due to, for instance, pragmatic enrichment (e.g., Regel et al., 2011; Hoeks et al., 2013; Spotorno et al., 2013).

6.1.2 Modeling other features of ERP components

An ERP component is multi-dimensional in that it is defined by a spectrum of features, including *polarity* (it is either more positive or negative than reference electrodes), *amplitude*, *latency*, *duration*, and *scalp distribution*. Our current simulations only modeled amplitude (and implicitly polarity). As such, an important direction for future exploration is to see if and how other features of ERP components can be incorporated into our simulations. A possible approach towards modeling the temporal dynamics (latency and duration) of ERP components, for instance, is to allow activation to propagate through the model for a number of time-steps (where one time-step constitutes propagation of activation from the input to the output layer), such that component amplitude can be measured as a function of time (see Laszlo and Plaut, 2012). Importantly, modeling the temporal dynamics of the N400 and the P600 component, for instance, might also shed light on the communication between the lpMTG and lIFG, as well as on the issue of component overlap (see chapter 4). Modeling scalp distribution is also very interesting, but seems less straightforward, as the distribution of an ERP component across the scalp depends inherently on the folding of the cortex and the shape of the skull.

6.1.3 Beyond the N400 and the P600

The work presented in this thesis has focused on the N400 and the P600 component. Whereas these components are indisputably the most salient for language comprehension, they are not the full story. EEG experiments have identified other ERP components sensitive to aspects of language processing. Two such components are the Early Left Anterior Negativity (ELAN), a negative deflection in the ERP signal that typically occurs between 120–200ms after the onset of a word that induces a category violation, and the Left Anterior Negativity (LAN), a negative deflection between 300-500ms post word onset that seems sensitive to the assignment of grammatical relations when these depend on morphosyntactic marking (see Friederici, 2011, for more dicussion on these components; also see Steinhauer and Drury, 2011 for a critical disussion on the ELAN). Another component is the Nref, a frontally focused, sustained negativity that starts at about 200ms post word onset, and that been taken to reflect increased difficulty in establishing reference when there are multiple possible antecedents (van Berkum et al., 2007). How and where these components fit in our model is an important question for future investigation (but see Hoeks and Brouwer, 2014, for a recent proposal on how to incorporate the Nref component in the Retrieval-Integration framework).

Critically, incorporating these additional ERP components not only requires us to identify the precise processes they reflect, but also where in the brain these processes take place. In chapter 4, we have already speculated about the role of the right hemisphere, the anterior temporal lobe (ATL), and the angular gyrus (AG; BA 39), as well as about the relation of activity in these areas to electrophysiology. At this point in time, this relation is not at all clear. The ATL, for instance, has been implied in semantic memory processes (Rogers et al., 2004, 2006; Damasio et al., 1996, 2004) meaning that activity in this area may contribute to the N400 component, but also in syntactic or thematic combinatorial processing, suggesting that it rather contributes to the P600 component or the LAN. Hence, an important aim for the future is to unequivocally isolate the type of processes that are subserved in cortical regions that are implicated in language processing, and to align these processes with their corresponding electrophysiology.

A related aim is to arrive at a more detailed coupling of electrophysiology and neuroanatomy. We proposed, for instance, that the role of the lIFG is multi-functional, and suggested that its different sub-parts may sub-serve different sub-processes of MRC composition. Moreover, we argued that these different sub-processes may manifest themselves as different types of P600s in terms of scalp distribution, latency, duration, and amplitude. Hence, the different positivies that we hypothesized to be members of the family of P600s (Brouwer et al., 2012), may have their origin in distinct sub-parts of the lIFG. Unravelling this functional topology of the lIFG, and the corresponding categorization of members of the P600 family, is also an important goal for future study.

Appendices

APPENDIX A

Artificial Neural Networks: A Few Nuts and Bolts

Abstract | *This appendix provides a brief introduction to the mathematical nuts and bolts of Artificial Neural Networks. It first introduces the theory behind neural networks, as well as different network topologies. Next, the mathematics of the widely used backpropagation and backpropagation through time training algorithms are laid out. Finally, several customizations in terms of activation functions (sigmoid, softmax, hyperbolic tangent, linear, and step), error functions (sum of squares, cross-entropy, and divergence), weight updating algorithms (steepest and bounded descent, four flavours of Resilient propagation, Quickprop, and Delta-Bar-Delta), and weight randomization methods (Range, Gaussian, and Nguyen-Widrow randomization, as well as the Fan-In method) are presented. All material is described with the aim that it should be possible to implement it using this appendix as a technical reference.*

A.1 Introduction

Artificial Neural Networks are powerful mathematical models that are used in a variety of fields, ranging from artificial intelligence to statistics, psychology, and neuroscience. The modeling work presented in this thesis also makes use of artificial neural networks, and was conducted using MESH, a neural network simulator that was developed along with this thesis (see section A.4). The present appendix was born out of the process of developing this simulator, as I missed a brief overview that presented the mathematics of neural networks, as well as the theory behind different network topologies, activation and error functions, and different variations on the backpropagation algorithm, in a way that would have allowed me to implement these concepts straightaway. This appendix is intended to fill that gap.

First, a brief introduction to artificial neural networks is provided. This introduction is by no means exhaustive, and is meant to provide a context for the remainder of this appendix only. The interested reader is referred to Bishop (1995) and Haykin (1999) for a comprehensive foundation on neural networks. To readers interested in the application of neural network modeling in the study of psychology, I recommend the two volumes on parallel distributed processing by Rumelhart, McClelland, and the PDP research group (Rumelhart et al., 1986b; McClelland et al., 1987), the more recent book by Houghton (2005), and a handbook chapter by Thomas and McClelland (2008). Readers with a more specific interest in neural network modeling in psycholinguistics are referred to Christiansen and Chater (2001) and the second chapter of Rohde (2002).

Next, the *backpropagation* and *backpropagation through time* training algorithms will be introduced. The remaining sections then describe various activation and error functions that can be used in backpropagation networks, as well as different weight updating algorithms and weight randomization schemes.

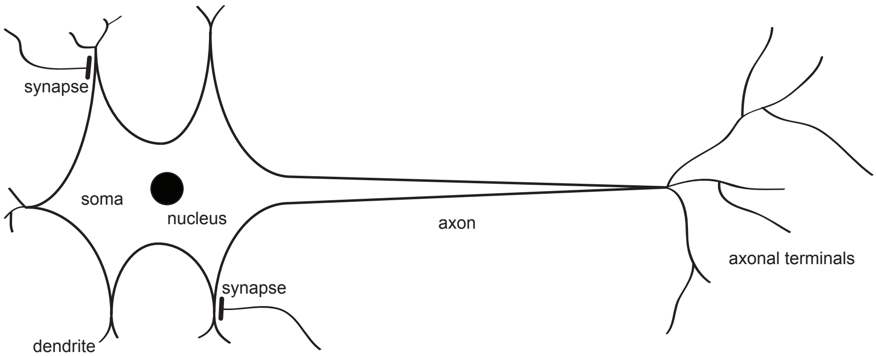

Figure A.1 | *Schematic overview of a biological neuron (or nerve cell).*

A.2 Artificial Neural Networks

A.2.1 A brief introduction

An Artificial Neural Network (ANN)—also known as a connectionist model or parallel distributed processing architecture—is a mathematical model that mimics the processing properties of a biological neural network in the brain. Biological neural networks consist of vast amounts of interconnected information processors called neurons (or *nerve cells*) that give rise to cognitive function by interacting. Each neuron consists of three core components: a *soma* or cell body (containing the cell's core or *nucleus*), *dendrites* at which it receives signals from other neurons, and an *axon* branching into *axonal terminals* with which the neuron itself can send out signals to the dendrites of other neurons. Figure A.1 provides a schematic overview of a neuron. The axonal terminals of a neuron are "connected" to the dendrites of another neuron by means of a *synapse*, which can modulate a signal sent between two neurons to be excitatory or inhibitory. Excitatory signals increase the likelihood that a receiving—*post-synaptic*—neuron will send out a signal to other neurons itself. That is, if the excitatory net input to a neuron is large enough for a brief interval of time, the neuron will fire spikes of electrical activity—*action potentials*—down its axon, and signal other neurons. Inhibitory signals, on the other hand, decrease the likelihood that a post-synaptic neuron will generate action potentials. The collective presence or absence of neuronal firing is what gives rise to cognitive function, and the adjustment of synaptic modulations, which leads to a change in the impact of neurons upon one another, is what underlies learning.

ANNs model the structure of biological neural networks by means of interconnected "model neurons" called *units*. A unit is represented by a single number, an *activation value*, which denotes that unit's firing rate (or activity). The activation value y_j of a unit j is typically a non-linear combination of the *net input* x_j to that unit:

$$y_j = f(x_j) \tag{A.1}$$

where $f(x)$ is a non-linear *activation* (or *transfer*) *function*, such as the commonly used sigmoid function (see section A.3.1 for more activation functions):

$$f(x) = \frac{1}{1 + e^{-x}} \tag{A.2}$$

The net input to a unit, in turn, is determined by the activation values of connected, signaling units, and by the synaptic modulations that these activation values undergo. In ANNs, synaptic modulations are modeled by assigning connections between two units a *weight*; positive weights represent excitatory synaptic modulations, and negative weights inhibitory synaptic modulations. The net input x_j to a unit j is then defined as the weighted sum of the activation values y_i of all units i that signal to unit j:

$$x_j = \sum_i y_i w_{ij} \tag{A.3}$$

where w_{ij} represents the weight on the connection between unit i and unit j. Figure A.2 provides a schematic overview of a model neuron (or unit).

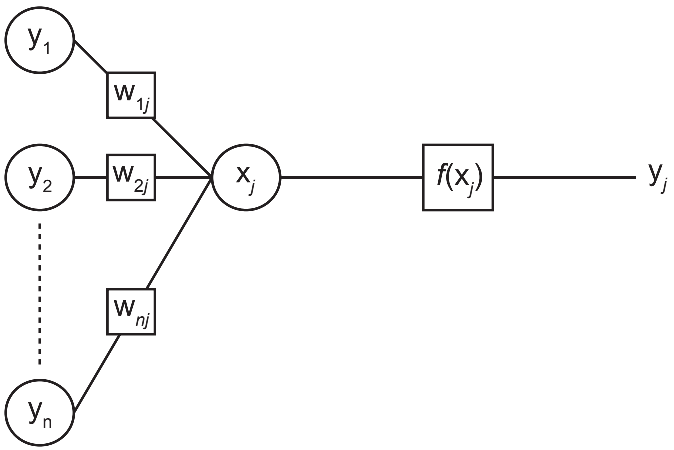

Figure A.2 | *Schematic overview of a unit (or model neuron). The activation level of the unit is a non-linear combination of its net input. The unit's net input, in turn, is the weighted sum of the activation levels of all units that signal to this unit.*

The units in an ANN are generally organized in layers that represent specialized cortical neuronal groups. A typical ANN consists of an *input layer*, the units of which can be taken to correspond to sensory neurons, one or more *hidden layer(s)*, containing processing units such as feature detectors, and an *output layer* that represents the "response" to a stimulus (e.g., a word in a word recognition task). Units in the layers of a neural network can be wired together in different ways. In a Feed Forward neural Network (FFN), units between successive layers are typically fully connected; each unit in the network's input layer is connected to each unit in the hidden layer, and each unit in the hidden layer is connected to each unit in the output layer. Figure A.3 provides a schematic overview of an FFN. In a FFN, activation typically propagates from the input layer, through the hidden layer, to the output layer. As such, the activation levels of units in the hidden layer depend on those of the units in the input layer, and those of the units in the output layer on the activity of the units in the hidden layer. Propagation of activation from the input to the output layer constitutes a single processing time-step. As such, the activity pattern at the output layer of a FFN is fully determined by the network's input pattern at a given time-step. In a Recurrent Neural Network (RNN), by contrast, units within layers can also be connected. In a typical RNN, for instance, each hidden unit is connected to itself and all other hidden units in the same layer. Figure A.4 provides an illustration of an RNN in which each hidden unit has a recurrent connection to itself only (to avoid clutter). Recurrent connections provide the network with a form of memory or feedback; in an RNN with recurrent connections on its hidden units, the activation values of these units depend not only on the current activation values of the units in the input layer (as in a FFN), but also on their own previous activation values. In other words, the activation values of the units in the hidden layer at the current time-step can influence the activation values of the hidden units at a later time-step, allowing the network to process temporally extended input sequences (such as the words of a sentence).

A third type of network, the Simple Recurrent Network (SRN) (Elman, 1990), is essentially an FFN that approximates the recurrent properties of an RNN through the addition of so-called "context" layers that maintain copies of the previous activation values of recurrent units (see also Jordan, 1986). To approximate fully recurrent hidden units, for instance, an SRN employs a context layer that has as many units as the hidden layer, and of which the units are fully connected to the hidden units. Figure A.5 provides a schematic overview of an SRN with such a configuration. Critically, before the activation values of the hidden units are updated, the previous activation values of these units are copied to their corresponding units in this context group[1]. As a result, the updated activation values depend, as in an RNN, on both the activation values of the units in the input layer, and their own previous activation values as stored in the context group.

[1]This process is generally bootstrapped by setting the activation values of the units in the context group to 0.5.

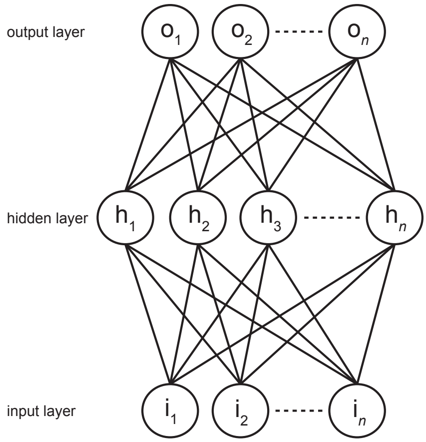

Figure A.3 | *A Feed Forward neural Network (FFN). Units in successive layers are fully connected, whereas units within layers are not.*

Despite architectural differences between FFNs, RNNs, and SRNs, the general idea behind the workings of each of these networks is the same. An input pattern is presented—"clamped"—to a network's input units by setting the activation values of these units to specific values. Activation is then propagated forward through the hidden layer(s) to the output layer (using equations A.3 and A.1), the activation pattern of which determines the network's response to the input pattern. Hence, an ANN essentially maps an input pattern to an output pattern. For a given task, for instance a digit recognition task, the performance of a network is therefore determined by how well it maps each input pattern—a representation of a handwritten digit—to its relevant output pattern—a representation of the numerical value of that digit. Mappings between input-output pairs depend on the weight configuration of a network. This means that how well the network performs on a given task is determined by how well the weight configuration of the network supports the mappings of the relevant set of input-output pairs. This underlines the main challenge of modeling a

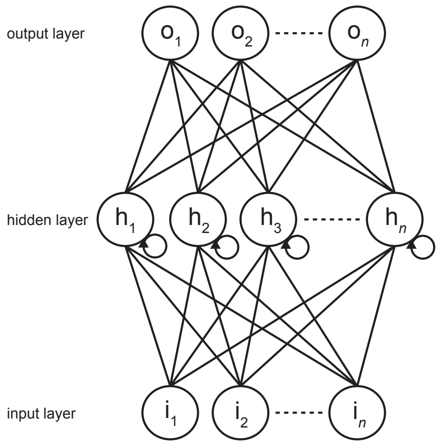

Figure A.4 | *A Recurrent Neural Network (RNN). Units in successive layers are fully connected. In addition, units in the hidden layer have recurrent connections, meaning that each unit in the hidden layer is also connected to itself and all other units in the hidden layer (to avoid clutter, the figure only depicts recurrent connections between each unit and itself). Recurrent connections provide the network with a form of memory.*

task with an ANN: finding a weight configuration that maximizes the network's performance on a relevant set of input-output pairs. An optimal weight configuration for a set of input-output pairs is typically determined by applying an algorithm that iteratively presents the network with items from this set, determines its response error, and adjusts the weights of the network such that its error will be reduced upon the next presentation of these items. FFNs are typically trained with the widely used *backpropagation algorithm* (Rumelhart et al., 1986a) (see section A.2.3 below). RNNs, by contrast, require more complex training algorithms, such as *backpropagation through time* (Rumelhart et al., 1986b; Werbos, 1990; Williams and Peng, 1990) (see section A.2.4). This underlines the important difference between RNNs and SRNs: an SRN is effectively an FFN, and can therefore

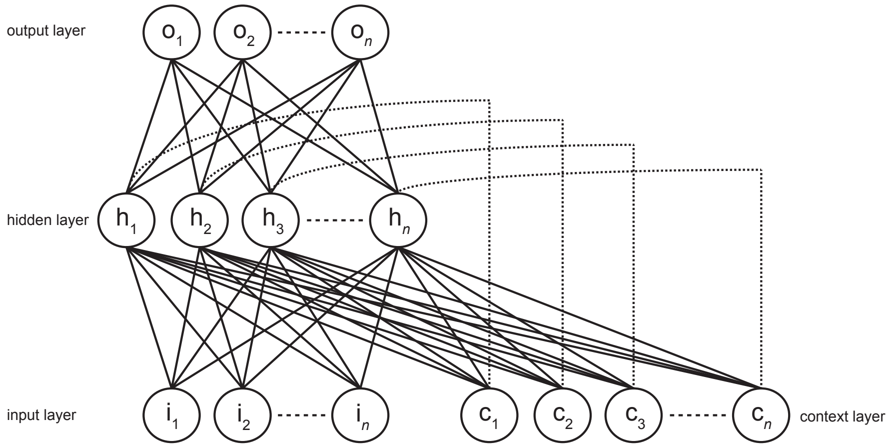

Figure A.5 | *A Simple Recurrent Network (SRN). Before being updated, the activation values of the units in the hidden layer are copied to their corresponding unit in the context layer (the fine dotted lines represent copy connections).*

be trained using the simpler backpropagation algorithm.

In a network with one or more hidden layer(s), applying an iterative training procedure leads the network to develop higher-level internal representations of abstract properties of the training items in its hidden layer(s). Elman (1990), for example, trained an SRN on a simple word prediction task. Each unit in the input layer, as well as each unit in the output layer corresponded to a single word in a localist manner. Hence, in the training data there was no notion about the category or meaning of a word. After training the SRN on a vast amount of sentences, however, analysis of the network's hidden layer revealed that the network developed internal representations reflecting word categories, but also finer-grained distinctions, such as animate versus inanimate nouns, humans versus animals, and distinctions like transitive versus intransitive verbs. Importantly, the development of such higher-level internal representations provides networks with the capability to generalize. Elman found, for instance, that if

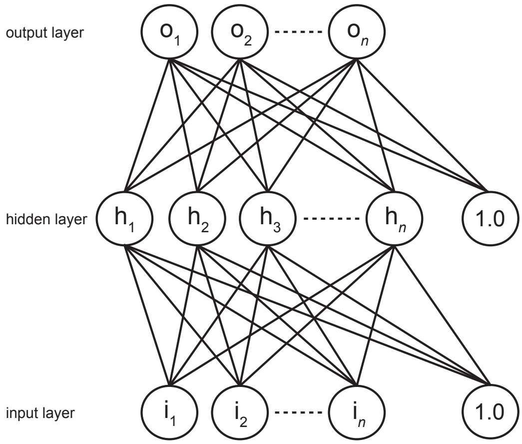

Figure A.6 | *A Feed Forward neural Network (FFN) with bias units connected to each non-input unit. Bias units always have a constant activation value of* 1. *The connection between a non-input unit and a bias unit, the weight of which is trainable like any other weight in the network, therefore effectively poses a threshold that should be overcome by the rest of the net input to a unit, in order for this input to become positive, and the unit to become active.*

the SRN was presented with a novel word that was not seen in the training set, but that is used like the noun "man", this is reflected in the activation pattern of the network's hidden layer. The SRN was thus found capable of generalizing the word prediction task to novel, unseen words.

A.2.2 Regulating unit excitability—the role of biases

In the networks discussed thus far, a unit's activation value—its firing rate—is simply a non-linear function of its net input. Biological neurons, however, only fire action potentials if their excitatory net input is large enough for a brief time interval, presupposing some kind of excitation threshold that needs to be reached before a neuron becomes active. A similar principle is in fact also employed in ANNs by means of *bias* units. Every non-input unit in an ANN typically receives input from a bias unit, which always has a constant activation value of 1 (although other values could in theory also be used). The connection between a non-input unit and a bias unit, the weight of which is trainable like any other weight in the network, therefore effectively poses a *threshold* that should be overcome by the rest of the net input to a unit, in order for this input to become positive, and the unit to become active. Figure A.6 illustrates a FFN with bias units connected to the units in the hidden and output layer.

For means of simplicity, bias units are often omitted in schematic depictions of neural networks. However, in the remainder of this appendix, we assume all non-input units of an ANN to always have a bias unit connected to them.

A.2.3 The Backpropagation algorithm

The most widely used algorithm for finding a weight configuration that minimizes a network's error is the Backpropagation (BP) algorithm (Rumelhart et al., 1986a)[2]. BP minimizes an error function over a finite set of input-output pairs by means of *steepest* or *gradient descent*. Typically, the error function that is minimized is the *sum squared error*:

$$E = \sum_c E_c \tag{A.4}$$

where

$$E_c = \frac{1}{2} \sum_j (y_j - d_j)^2 \tag{A.5}$$

and, in turn, y_j is the observed activity of output unit j for input-output pair c, and d_j the desired activity for this unit.

The BP algorithm requires the weights of the network to be non-zero, and initial weights are therefore randomized, typically with values from

[2]For an elaborate derivation of BP, the reader is referred to (Rumelhart et al., 1986b, chapter 8).

within a small range such as $[-.1, +.1]$ (see section A.3.4 for different randomization schemes). The algorithm then operates in two passes for each input-output pair c. In the *forward pass*, the network's response to the input pattern is computed by means of equations A.3 and A.1. In the subsequent *backward pass*, BP employs the *generalized delta rule* to reduce the network's error E by adjusting each weight proportional to its *gradient*:

$$\Delta w_{ij} = -\varepsilon \frac{\partial E}{\partial w_{ij}} \tag{A.6}$$

where ε is the network's *learning rate*. The gradient of a weight w_{ij}, which quantifies how the error changes as a function of the weight, is estimated as:

$$\frac{\partial E}{\partial w_{ij}} = \delta_j y_i \tag{A.7}$$

where δ_j is the *error signal* of unit j, and y_i is the activation value of unit i that signals to unit j. The error signal of a unit tells how fast the error changes as a function of its total net input. This error signal is defined as the product of a unit's *error derivative*, which tells how the error changes as a function of a unit's activation value, and its *activation derivative*, which tells how the unit's activation value changes as a function of its net input:

$$\delta_j = \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial x_j} \tag{A.8}$$

The error derivative $\frac{\partial E}{\partial y_j}$ depends on whether unit j is an output unit or a hidden unit. If unit j is an output unit, its error derivative is defined as:

$$\frac{\partial E}{\partial y_j} = y_j - d_j \tag{A.9}$$

which is the derivative of the sum squared error function, as defined in A.5, for a given input-output pair c. If unit j is a hidden unit, by contrast, its error derivative is recursively defined as:

$$\frac{\partial E}{\partial y_j} = \sum_k \delta_k w_{jk} \tag{A.10}$$

where all units k are units that receive signals from unit j. The activation derivative $\frac{\partial y_j}{\partial x_j}$, in turn, is simply the derivative $f'(x)$ of the activation function $f(x)$ for the net input x_j to a specific unit j; for the sigmoid function defined in equation A.2, this derivative is defined as:

$$\frac{\partial y_j}{\partial x_j} = f'(x_j) = y_j(1 - y_j) \tag{A.11}$$

Hence, the error signal δ_j for an output unit j is defined as:

$$\delta_j = (y_j - d_j)f'(x_j) \tag{A.12}$$

and the error signal δ_j for a hidden unit j as:

$$\delta_j = f'(x_j) \sum_k \delta_k w_{jk} \tag{A.13}$$

Once the gradient for each weight is gathered, there are two options. The first is to adjust the network's weights after each input-output pair. This method is called *online learning*. Alternatively, one can accumulate gradients over a batch of input-output pairs, and adjust the network's weight based on these accumulated gradients. This is referred to as *batch learning*. During training, the ANN usually processes all input-output pairs in a training set multiple times. One run through all of these input-output pairs is called an *epoch*. Training can be stopped after a fixed number of epochs, or when the error E drops below some predefined threshold value.

A.2.4 Backpropagation Through Time

In an ANN with recurrent hidden units, the processing at time-step t depends on the activation pattern of these hidden units at time-step $t - 1$, which in turn depends on the activation pattern at time-step $t - 2$, and so on. This means that in order to train such a network to map sequences of input patterns to an output pattern (e.g., the individual words of a sentence to a grammatical representation of that sentence), the optimal weights on these recurrent connections need to be determined. The BP algorithm, however, cannot handle cyclic connections. In an SRN, this is overcome by copying the activation pattern of hidden units at a previous time-step to a context layer, and then treating these context units as additional inputs to the hidden units at the next time-step (see Figure A.5). To train the weights on the connections between these context units and hidden units, the forward pass at each time-step is directly followed by a backward pass, meaning that the gradients of these weights are determined only by the error signals of the network's output at that specific time-step. As a consequence, SRNs are effectively trained to use the information in the hidden layer to produce the correct output pattern at the next time-step, without taking into account whether any of this information might be useful later on[3].

The Backpropagation Through Time algorithm (BPTT; Rumelhart et al., 1986b; Werbos, 1990; Williams and Peng, 1990), by contrast, does take the delayed use of information in recurrent units into account. BPTT essentially "unfolds" a RNN in time, such that each time-step is represented by a duplicate of the original network, and that each duplicate is connected to the one representing the previous time-step through the network's recurrent units[4]. For a RNN with fully recurrent hidden units, for instance, this means that the hidden units of the network representing time-step t are connected to hidden units of the one representing time-step $t-1$, which are in turn connected to the ones at time-step $t-2$, and so on[5]. Figure A.7 shows a RNN with fully recurrent hidden units unfolded over three

[3]Despite not explicitly taking this into account, SRNs have nonetheless proven capable of learning mappings that require information to be retained in memory for several time-steps before it is used (e.g. Christiansen and Chater, 1999; Rohde and Plaut, 1999).

[4]Each duplicate has the same weights as the original network.

[5]At time-step $t = 0$, this process is bootstrapped by connecting a layer of hidden units of which the activation values are set to 0.5.

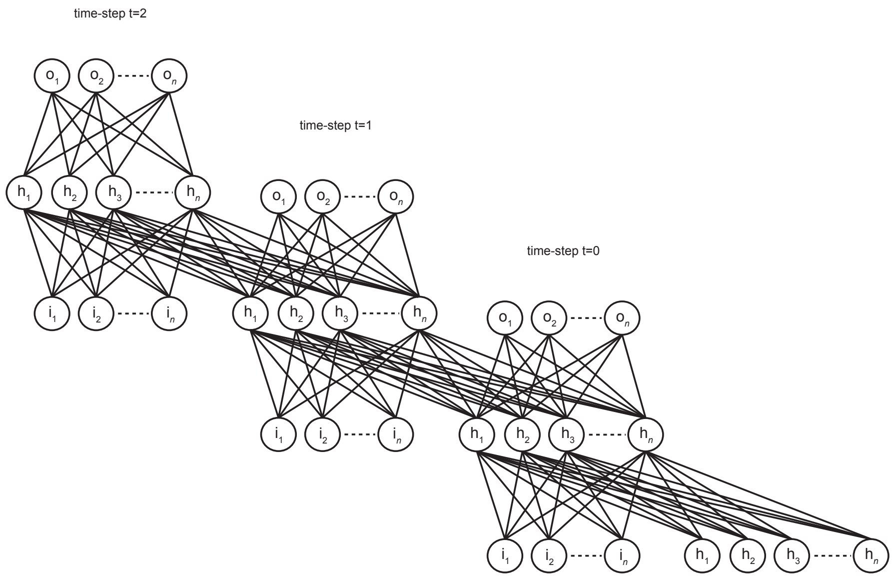

Figure A.7 | *Unfolding of a Recurrent Neural Network (RNN). The depicted network shows a RNN with fully recurrent hidden units unfolded over three time steps. The resulting unfolded network is effectively a FFN, meaning that it can be trained with BP.*

time-steps. Unfolding thus effectively turns the RNN into a FFN, meaning that it can be trained with standard BP. The unfolding, however, makes it possible to do the forward pass for all the individual time-steps first, and to then do a single backward pass on the basis of the network's final output pattern. During this backward pass, error is essentially propagated back "through time", such that the gradients of the weights of duplicate networks at earlier time-steps in the unfolded network take into account any future error for the current training item. These weight gradients are then summed, such that a single gradient is obtained for each weight in the original network, on the basis of which the weight adjustments are made.

Mapping the individual words of a sentence to a grammatical structure is an example of a problem in which a sequence of inputs is mapped to a single output pattern. There are, however, also problems for which a correct output pattern exists at each time-step, such as predicting the trajectory of a moving object. In this case, a single training item exists of a set of input-output pairs. For a RNN with fully recurrent hidden layers, the error signals for hidden layers at a non-final time-step then depend on both the error signals of their corresponding output units, and the error signals of the hidden units at the next time-step. For the backward pass of BP, this means that the error signal for hidden units should generalize over layers, meaning that equation A.13 becomes:

$$\delta_j = f'(x_j) \sum_l \sum_k \delta_k w_{jk} \tag{A.14}$$

where each unit k of a layer l is a unit that receives a signal from unit j.

A.3 Building custom Artificial Neural Networks

The ANNs that were introduced in the previous section can be seen as somewhat 'standard' architectures, in that they employ the sigmoid activation function, minimize sum of squares error, uniformly initialize the weights within a given range, and use 'vanilla' gradient descent to update weights. Each of these aspects, however, can be customized to tailor an ANN to specific requirements, and a number of such customizations have become quite common. The following sections will respectively introduce different *activation functions*, *error functions*, *weight update algorithms*, and *weight randomization schemes* that can be used to build custom ANN architectures. The material will be presented with focus on *how* to do it. The reader is referred to the original papers introducing the customizations, or the books by Bishop (1995) and Haykin (1999) for discussions on *why* to do it.

Name	Function	Derivative
Binary sigmoid	$f(x) = \frac{1}{1+e^{-x}}$	$f'(x) = y(1-y)$
Bipolar sigmoid	$f(x) = -1 + \frac{2}{1+e^{-x}}$	$f'(x) = \frac{1}{2}(1+y)(1-y)$
Softmax	$f(x_j) = \frac{e^{x_j}}{\sum_i e^{x_i}}$	$f'(x) = 1$
Hyperbolic tangent	$f(x) = \frac{e^{2x}-1}{e^{2x}+1}$	$f'(x) = 1 - y * y$
Linear	$f(x) = x$	$f'(x) = 1$
Step	$f(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$	$f'(x) = 1$

Table A.1 | *Overview of different activation functions and their derivatives. For the derivatives it is assumed that $y = f(x)$. If an activation function has no derivative, $f'(x) = 1$.*

A.3.1 Activation functions

Table A.1 presents an overview of different activation functions $f(x)$ and their derivative $f'(x)$. All definitions assume that $y = f(x)$. In case an activation function has no derivative, its "derivative" is defined as 1. A different activation function can be used in an ANN by replacing function $f(x_j)$ in equation A.1, and its derivative term $f'(x_j)$ in equation A.11. Each non-input layer can in principle have a different activation function.

Name	Function	Derivative
Sum of squares	$E = \frac{1}{2}\sum_j (y_j - d_j)$	$\frac{\partial E}{\partial y_j} = y_j - d_j$
Cross-entropy	$E = \sum_j \log \frac{d_j}{y_j}$ $\times\, d_j \log \frac{1-d_j}{1-y_j}$ $\times\, (1-d_j)$	$\frac{\partial E}{\partial y_j} = \frac{(y_j - d_j)}{y_j(1-y_j)}$
Divergence	$E = \sum_j \log \frac{d_j/y_j}{d_j}$	$\frac{\partial E}{\partial y_j} = -d_j/y_j$

Table A.2 | *Overview of different error functions and their derivatives. Error functions and derivatives are defined for a specific input-output pair c. The index c is suppressed in the formulas.*

A.3.2 Error functions

Table A.2 provides an overview of different error functions and their derivative. Error functions and derivatives are defined for a specific input-output pair. To use a different error function to train an ANN, the error function that is minimized has to be replaced in equation A.5. Furthermore, the error derivative in equation A.9 has to be replaced with the error function's corresponding derivative.

A.3.3 Weight update algorithms

Steepest or gradient descent

One of the most commonly used weight updating rules is steepest or gradient descent. This update rule is defined in equation A.6. Upon each update, a weight is adjusted by an amount proportional to its corresponding gradient. This update rule can be extended with a *momentum* term that adds part of the previous weight delta to the current one, and which serves to stabilize the updating and to accelerate convergence:

$$\Delta w_{ij}(t) = \Delta w_{ij}(t) + \alpha \Delta w_{ij}(t-1) \tag{A.15}$$

where the momentum coefficient α controls the fraction of the previous weight delta to be added.

Another addition is *weight decay*, which keeps weights from becoming very large by subtracting a fraction of the current weight from the current weight delta:

$$\Delta w_{ij}(t) = \Delta w_{ij}(t) - d \cdot w_{ij}(t) \tag{A.16}$$

where d is the weight decay coefficient.

Bounded steepest descent

Bounded steepest descent is a modification of 'standard' steepest descent proposed by Rohde (2002). In bounded steepest descent, the gradient term of the weight delta is multiplied by a scaling factor ρ:

$$\Delta w_{ij} = -\varepsilon \rho \frac{\partial E}{\partial w_{ij}} \tag{A.17}$$

where ρ depends on the length of the entire gradient:

$$\rho = \begin{cases} \frac{1}{||\partial E / \partial w||} & \text{if } ||\partial E / \partial w|| > 1 \\ 1 & \text{otherwise} \end{cases} \tag{A.18}$$

Resilient propagation

In Resilient backpropagation (Rprop; Riedmiller and Braun, 1993; Riedmiller, 1994), each weight adjustment is made on the basis of the sign of the gradient of that weight. An Rprop update iteration is a two-stage procedure. In the first stage, a so-called weight specific "update value" Γ_{ij} is computed:

$$\Gamma_{ij} = \begin{cases} \eta^+ \Gamma_{ij}(t-1) & \text{if } \frac{\partial E}{\partial w_{ij}}(t-1)\frac{\partial E}{\partial w_{ij}}(t) > 0 \\ \eta^- \Gamma_{ij}(t-1) & \text{if } \frac{\partial E}{\partial w_{ij}}(t-1)\frac{\partial E}{\partial w_{ij}}(t) < 0 \\ \Gamma_{ij}(t-1) & \text{otherwise} \end{cases} \tag{A.19}$$

where η^+ and η^- are defined as:

$$0 < \eta^- < 1 < \eta^+ \tag{A.20}$$

and Γ_{ij} is bounded by Γ_{max} and Γ_{min}. Reasonable values for these parameters are $\Gamma_{max} = 50$, $\Gamma_{min} = 0$, $\eta^+ = 1.2$ and $\eta^- = 0.5$, and a reasonable initial value for $\Gamma_{ij}(0) = 0.0125$.

In the second stage, the actual weight is updated. The details of this update step depend on the specific flavour of Rprop that is being used. Igel and Hüsken (2000) distinguish between four Rprop flavours:

Rprop with weight-backtracking (Rprop+). After computing the "update values" Γ_{ij} for each weight, the second stage depends on whether the sign of that weight's gradient has changed from time-step $t-1$ to t. If it has not changed, the weight delta Δw_{ij} is defined as:

$$\Delta w_{ij}(t) = -sign(\frac{\partial E}{\partial w_{ij}}(t))\Gamma_{ij} \tag{A.21}$$

where

$$sign(x) = \begin{cases} 1 & \text{if } x \text{ is positive} \\ \text{-1} & \text{if } x \text{ is negative} \\ 0 & \text{otherwise} \end{cases} \tag{A.22}$$

If, on the other hand, the sign has changed, the previous weight update is reverted (*weight backtracking*):

$$\Delta w_{ij}(t) = -\Delta w_{ij}(t-1) \tag{A.23}$$

Algorithm 1: The Rprop+ algorithm.

foreach w_{ij} **do**
  if $\frac{\partial E}{\partial w_{ij}}(t-1)\frac{\partial E}{\partial w_{ij}}(t) > 0$ **then**
    $\Gamma_{ij}(t) \longleftarrow min(\eta^{+}\Gamma_{ij}(t-1), \Gamma_{max})$
    $\Delta w_{ij}(t) \longleftarrow -sign(\frac{\partial E}{\partial w_{ij}}(t))\Gamma_{ij}(t)$
    $w_{ij}(t+1) \longleftarrow w_{ij}(t) + \Delta w_{ij}(t)$
  else if $\frac{\partial E}{\partial w_{ij}}(t-1)\frac{\partial E}{\partial w_{ij}}(t) < 0$ **then**
    $\Gamma_{ij}(t) \longleftarrow max(\eta^{-}\Gamma_{ij}(t-1), \Gamma_{min})$
    $w_{ij}(t+1) \longleftarrow w_{ij}(t) - \Delta w_{ij}(t-1)$
    $\frac{\partial E}{\partial w_{ij}} \longleftarrow 0$
  end
  else if $\frac{\partial E}{\partial w_{ij}}(t-1)\frac{\partial E}{\partial w_{ij}}(t) = 0$ **then**
    $\Delta w_{ij}(t) \longleftarrow -sign(\frac{\partial E}{\partial w_{ij}}(t))\Gamma_{ij}(t)$
    $w_{ij}(t+1) \longleftarrow w_{ij}(t) + \Delta w_{ij}(t)$
  end
end

and its gradient is reset to 0, such that Γ_{ij} will not be updated on the next time-step:

$$\frac{\partial E}{\partial w_{ij}}(t) = 0 \tag{A.24}$$

Now, if we define $min(x, y)$ as:

$$min(x, y) = \begin{cases} x & \text{if } (x < y) \\ y & \text{otherwise} \end{cases} \tag{A.25}$$

and $max(x, y)$ as:

$$max(x, y) = \begin{cases} x & \text{if } (x > y) \\ y & \text{otherwise} \end{cases} \tag{A.26}$$

the above can be algorithmically defined as in Algorithm 1.

Algorithm 2: The Rprop$-$ algorithm.

1 **foreach** w_{ij} **do**
2   **if** $\frac{\partial E}{\partial w_{ij}}(t-1)\frac{\partial E}{\partial w_{ij}}(t) > 0$ **then**
3         $\Gamma_{ij}(t) \longleftarrow min(\eta^{+}\Gamma_{ij}(t-1), \Gamma_{max})$
4   **else if** $\frac{\partial E}{\partial w_{ij}}(t-1)\frac{\partial E}{\partial w_{ij}}(t) < 0$ **then**
5         $\Gamma_{ij}(t) \longleftarrow max(\eta^{-}\Gamma_{ij}(t-1), \Gamma_{min})$
6   **end**
7     $\Delta w_{ij}(t) \longleftarrow -sign(\frac{\partial E}{\partial w_{ij}}(t))\Gamma_{ij}(t)$
8     $w_{ij}(t+1) \longleftarrow w_{ij}(t) + \Delta w_{ij}(t)$
9 **end**

Rprop without weight-backtracking (Rprop$-$). This is similar to Rprop+ except that weight backtracking is omitted, and $\frac{\partial E}{\partial w_{ij}}$ is not reset to 0 if its sign has changed from time-step $t-1$ to t. Hence, Rprop$-$ is algorithmically defined as in Algorithm 2.

Modified Rprop with weight-backtracking (iRprop$+$). This is a modification of Rprop+ in which weight backtracking is only performed if the overall error goes up from time-step $t-1$ to t. Hence, it is the algorithm defined in Algorithm 1, but with the statement in line 8 conditioned on whether the error at the current time-step is larger than the error at the previous time-step, which amounts to replacing it with the following statement:

$$\textbf{if } E(t) > E(t-1) \textbf{ then } w_{ij}(t+1) \longleftarrow w_{ij}(t) - \Delta w_{ij}(t-1) \textbf{ end}$$

Modified Rprop without weight-backtracking (iRprop$-$). This is a modification of Rprop- in which $\frac{\partial E}{\partial w_{ij}}(t)$ is reset to 0 when its sign has changed from time-step $t-1$ to time-step t. Hence, it is the algorithm defined in Algorithm 2, with the following statement inserted between lines 5 and 6:

$$\frac{\partial E}{\partial w_{ij}}(t) \longleftarrow 0$$

Algorithm 3: The Quickprop algorithm.

```
Φ ⟵ μ/(1−μ)
foreach w_ij do
    if Δw_ij(t−1) > 0 then
        if ∂E/∂w_ij(t) < 0 then
            Δw_ij(t) ⟵ Δw_ij(t) + −ε ∂E/∂w_ij(t)
        end
        if ∂E/∂w_ij(t) < Φ ∂E/∂w_ij(t−1) then
            Δw_ij(t) ⟵ Δw_ij(t) + μΔw_ij(t−1)
        else
            Δw_ij(t) ⟵ Δw_ij(t) + [∂E/∂w_ij(t) / (∂E/∂w_ij(t−1) − ∂E/∂w_ij(t))] Δw_ij(t−1)
        end
    else if Δw_ij(t−1) < 0 then
        if ∂E/∂w_ij(t) > 0 then
            Δw_ij(t) ⟵ Δw_ij(t) + −ε ∂E/∂w_ij(t)
        end
        if ∂E/∂w_ij(t) > Φ ∂E/∂w_ij(t−1) then
            Δw_ij(t) ⟵ Δw_ij(t) + μΔw_ij(t−1)
        else
            Δw_ij(t) ⟵ Δw_ij(t) + [∂E/∂w_ij(t) / (∂E/∂w_ij(t−1) − ∂E/∂w_ij(t))] Δw_ij(t−1)
        end
    else if Δw_ij(t−1) = 0 then
        Δw_ij(t) ⟵ −ε ∂E/∂w_ij(t) + αΔw_ij(t−1)
    end
    Δw_ij(t) ⟵ Δw_ij(t) − d · w_ij(t)
    w_ij(t+1) ⟵ w_ij(t) + Δw_ij(t)
end
```

Quickprop

Quickprop (Fahlman, 1988) is a heuristic, second order algorithm, based on two assumptions:

1. For each weight, the relation between error and weight can be approximated by a downward peaking parabola;

2. A change in the gradient of a weight is independent of changes in other weight gradients that occur simultaneously.

Provided these assumptions, the Quickprop algorithm uses the previous and current gradients, as well as the weight deltas at the time-steps at which these gradients were measured, to determine a parabola. On each update, weights are adjusted to jump to the minimum of this parabola:

$$\Delta w_{ij}(t) = \frac{\frac{\partial E}{\partial w_{ij}}(t)}{\frac{\partial E}{\partial w_{ij}}(t-1) - \frac{\partial E}{\partial w_{ij}}(t)} \Delta w_{ij}(t-1) \tag{A.27}$$

At $t = 0$, this process is bootstrapped by using steepest descent, which is also used in the cases where a previous weight delta equals 0. Weight updates are bounded by a max step size μ, a reasonable value for which is $\mu = 1.75$. If a weight step is larger than μ times the previous step for that weight, μ times the previous weight delta is used instead. Moreover, a steepest descent term is included in the weight delta if the current and previous gradients have the same sign, and weight decay is applied to limit the weights sizes. Taken together, this results in Algorithm 3.

Delta-Bar-Delta

In Delta-Bar-Delta backpropagation (Jacobs, 1988), each weight is assigned its own learning rate. On each update, these learning rates are updated together with their weight. Hence, Delta-Bar-Delta is essentially steepest descent, extended with a learning rate update rule, which is defined as:

$$\Delta \varepsilon_{ij}(t) = \begin{cases} \kappa & \text{if } \overline{\frac{\partial E}{\partial w_{ij}}}(t-1)\frac{\partial E}{\partial w_{ij}}(t) > 0 \\ -\phi\varepsilon(t) & \text{if } \overline{\frac{\partial E}{\partial w_{ij}}}(t-1)\frac{\partial E}{\partial w_{ij}}(t) < 0 \\ 0 & \text{otherwise} \end{cases} \tag{A.28}$$

where $\overline{\frac{\partial E}{\partial w_{ij}}}(t)$ is an exponential average of the current and past gradients:

$$\overline{\frac{\partial E}{\partial w_{ij}}}(t) = (1-\theta)\frac{\partial E}{\partial w_{ij}}(t) + \theta\overline{\frac{\partial E}{\partial w_{ij}}}(t-1) \tag{A.29}$$

with θ as its base, a reasonable value for which is $\theta = 0.7$, and time as its exponent. Hence, if the current gradient and the average of the past gradients have the same sign, the learning rate is incremented by κ, a reasonable value for which is $\kappa = 0.1$. If they have opposite signs, by contrast, the learning rate is decremented by ϕ, a reasonable value for which is $\phi = 0.9$, times its current value.

A.3.4 Weight randomization

Range randomization

With range randomization, weights are sampled from a uniform distribution within a specific range $[min, max]$.

Gaussian randomization

With Gaussian randomization, the initial weights are sampled from a normal distribution $N(\mu, \sigma)$. The Box-Muller transform (Box and Muller, 1958) can be used to generate these normally distributed weights.

Nguyen-Widrow randomization

With Nguyen-Widrow randomization (Nguyen and Widrow, 1990), weights are first randomized within the range $[min, max]$ using range randomization. Next, the Euclidean norm is computed for each weight matrix:

$$||w|| = \sqrt{\sum_i w_{ij}^2} \tag{A.30}$$

as well as a β value:

$$\beta = 0.7h^{(1/i)} \tag{A.31}$$

where i represents the number of units in the signaling layer, and h the number of units in the receiving layer. Provided $||w||$ and β, each weight is then adjusted to:

$$w_{ij} = \frac{\beta w_{ij}}{||w||} \tag{A.32}$$

Fan-In randomization

With Fan-In randomization, each weight is first randomized within the range $[-1, +1]$ using range randomization. Next, the weights in each weight matrix are adjusted as:

$$w_{ij} = \frac{min}{h} + w_{ij}\frac{max - min}{h} \tag{A.33}$$

where h is the number of units in a receiving layer, and the range $[min, max]$ defines the minimum and maximum weight values.

A.4 MESH—A Neural Network Simulator

MESH[6] is an open-source[7], multi-threaded[8], artificial neural network simulator written in C99. It is primarily designed as a fast, general-purpose backpropagation simulator, implementing both regular *backpropagation* (Rumelhart et al., 1986a) and *backpropagation through time* (Rumelhart et al., 1986b), with flexibility and extensibility in mind. MESH supports feed forward networks, simple recurrent networks, as well as fully recurrent networks, and implements various error functions (sum of squares, cross-entropy, and divergence) and weight update algorithms (steepest descent,

[6] Available at `http://github.com/hbrouwer/mesh`.

[7] Under the Apache License, Version 2.0; `http://www.apache.org/licenses/LICENSE-2.0.html`.

[8] Powered by the OPENMP API; `http://www.openmp.org/`.

bounded steepest descent, four flavours of Rprop, Qprop, and Delta-Bar-Delta).

A.5 Conclusion

A handful of nuts and bolts of the theory and mathematics behind artificial neural networks was presented. The material was laid out in a way that should allow for direct implementation, using this appendix as a technical reference. This appendix is, however, by no means exhaustive. The reader is referred to Bishop (1995) and Haykin (1999) for more comprehensive treatments of neural networks.

APPENDIX B

COALS—Correlated Occurrence Analogue to Lexical Semantics

B.1 Introduction

The Correlated Occurrence Analogue to Lexical Semantics (COALS; Rohde et al., 2009) is a vector-space model of semantic similarity. COALS is similar to other vector-space approaches to semantics, such as Latent Semantic Analysis (LSA; Deerwester et al., 1990) and the Hyperspace Analogue to Language (HAL; Lund and Burgess, 1996), but has been shown to more accurately model human similarity judgments. This makes COALS the preferred model for generating semantic representations for use within computational models of cognition.

The aim of this appendix is to provide a brief introduction to the COALS model, as well as to a recent extension to it. Furthermore, it introduces an open-source implementation of the COALS system that was used for the modeling work presented in this thesis.

B.2 The COALS Model

The COALS method starts with the construction of a word co-occurrence table. Provided a text corpus, we count for each word how often it co-occurs with each other word in that corpus. A word y is assumed to co-occur with a word x, if it occurs within an n word window to the left or right of x. Typically, a 4-word ($n = 4$), *ramped* window is used[1]:

1 2 3 4 [x] 4 3 2 1

This means that a co-occurrence of x with a direct neighbour y is assigned a weight of 4 (i.e., their co-occurrence count is increased with 4×1), and that of x and a word z that occurs next to y a weight of 3, and so on.

The rows of the resulting co-occurrence table represent feature vectors for each word. Depending on the size of the corpus, these feature vectors may consist of quite a large number of columns. Rohde et al. (2009) show that, when columns are ordered by word frequency in a descending order, performance of the COALS model is equivalent when anywhere between 14.000 to 100.000 columns are used. By default, they use the minimum number of 14.000 columns[2].

Once all but the n most frequent columns have been discarded, the feature vectors are normalized by converting each co-occurrence count of words a and b to a word pair correlation using equation B.1:

$$w'_{a,b} = \frac{T \cdot w_{a,b} - \sum_j w_{a,j} \cdot \sum_i w_{i,b}}{\left(\sum_j w_{a,j} \cdot (T - \sum_j w_{a,j}) \cdot \sum_i w_{i,b} \cdot (T - \sum_i w_{i,b})\right)^{\frac{1}{2}}} \tag{B.1}$$

where i is a row index, j is a column index, and:

$$T = \sum_i \sum_j w_{i,j} \tag{B.2}$$

[1]Alternatively, a *flat* window can be used. Rohde et al. (2009) report no performance differences between a *ramped* and a *flat* window.

[2]Rohde et al. (2009) limit these n most frequent words to open-class words.

However, these correlations are not used directly. If $w'_{a,b}$ is negative, it is set to 0. If it is positive, on the other hand, its square root is taken:

$$norm(w'_{a,b}) = \begin{cases} 0 & \text{if } w'_{a,b} < 0 \\ \sqrt{w'_{a,b}} & \text{otherwise} \end{cases} \tag{B.3}$$

The resulting normalized vectors can then be used to determine the semantic similarity between two words, which is defined as the correlation between their respective vectors:

$$S(a,b) = \frac{\sum(a_i - a)(b_i - b)}{(\sum(a_i - a)^2 \sum(b_i - b)^2)^{\frac{1}{2}}} \tag{B.4}$$

B.2.1 Reduced dimensionality

The procedure outlined above provides vectors that are rather large for modeling experiments using neural networks. The dimensionality of COALS vectors can, however, be reduced using Singular Value Decomposition (SVD). The SVD of a $m \times n$ matrix X is defined as the product of three matrices:

$$X_{m \times n} = U_{m \times r} S_{r \times r} V^T_{r \times n} \tag{B.5}$$

where $r = min(m,n)$, matrix U contains *left singular vectors* along its columns, the diagonal matrix S contains constants called *singular values*, and matrix V^T contains *right singular vectors* along its rows. For both the left and right singular vectors, it holds that each of their dimensions is linearly independent of others (i.e., is orthogonal to others). The singular values in S are constants that define how much of the variance in the data can be explained by a specific dimension. These constants are sorted in descending order, such that left and right singular vectors that are connected to higher singular values, represent more important dimensions in the data.

Provided the SVD of a co-occurrence matrix X, we can reduce the dimensionality of COALS vectors by considering only the first $k < r$ dimensions. That is, if we discard all but the first k singular values and vectors, we obtain a matrix $\hat{X}$, which is the best rank k approximation to X in terms of sum squared error:

$$\hat{X}_{m \times n} = \hat{U}_{m \times k} \hat{S}_{k \times k} \hat{V}^T_{k \times n} \tag{B.6}$$

Given the resulting matrices $\hat{V}^T$ and $\hat{S}$, a reduced COALS-SVD vector V_c for a word c is then defined as:

$$V_c = X_c \hat{V} \hat{S}^{-1} \tag{B.7}$$

Computing the SVD of a matrix is a computionally intensive procedure, which renders it impractical to compute SVDs for a large matrices. To overcome this impracticality, Rohde et al. (2009) suggest to only compute the SVD for the m most frequent words in the co-occurrence matrix. In their simulations, they therefore only use 15.000 words, each with 14.000 features.

B.2.2 Binary vectors

Binary COALS-SVDB vectors can be obtained by converting negative vector components to 0, and positive components to 1.

B.2.3 An extension to the COALS model

Chang et al. (2012) have recently proposed an extension to the COALS model, in which they introduce three new properties to binary COALS-SVD vectors: *sparse coding*, *a fixed number of active units*, and *negative features*.

This modified COALS method starts with the construction of real-valued, k-dimensional COALS-SVD vectors, using the procedure described above. Next, each k-dimensional vector is extended into a $2k$-dimensional vector by appending a duplicate. In the first k dimensions of this extended vector, only the n most positive components are set to 1, whereas all others are set to 0. In the second k dimensions, the same is done for the m most negative components (negative features). The number of active positive features n and active negative features m are typically small (sparse coding), and are the same for all vectors (fixed number of active units). The result is thus a $2k$-dimensional COALS-SVDB vector with n active positive features in the first k dimensions, and m active negative features in the second k dimensions.

B.3 An implementation of the COALS model

COALS[3] is an open-source[4], implementation of the COALS model, written in C99[5]. It can be used to generate COALS vectors, COALS-SVD, and COALS-SVDB vectors, as well as extended COALS-SVDB vectors (cf. Chang et al., 2012). Moreover, it can use these vectors to generate top-k similar word lists.

[3] Available at http://github.com/hbrouwer/coals.

[4] Under the Apache License, Version 2.0; http://www.apache.org/licenses/LICENSE-2.0.html.

[5] COALS uses UTHASH, which is available at http://troydhanson.github.com/uthash/, and SVDLIBC, available at https://github.com/lucasmaystre/svdlibc.

Bibliography

Allen, M., Badecker, W., and Osterhout, L. (2003). Morphological analysis in sentence processing: An ERP study. *Language and Cognitive Processes*, 18(4):405–430.

Altmann, G. and Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3):247–264.

Amunts, K., Lenzen, M., Friederici, A. D., Schleicher, A., Morosan, P., Palomero-Gallagher, N., and Zilles, K. (2010). Broca's region: Novel organizational principles and multiple receptor mapping. *PLoS Biology*, 8(9).

Amunts, K. and Zilles, K. (2012). Architecture and organizational principles of broca's region. *Trends in Cognitive Sciences*, 16(8):418–426.

Andrews, E. (2011). Language and brain: Recasting meaning in the definition of human language. *Semiotica*, 2011(184):11–32.

Anwander, A., Tittgemeyer, M., von Cramon, D. Y., Friederici, A. D., and Knösche, T. R. (2007). Connectivity-based parcellation of Broca's area. *Cerebral Cortex*, 17(4):816–825.

Badre, D. and Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, 45(13):2883–2901.

Baggio, G. and Hagoort, P. (2011). The balance between memory and unification in semantics: A dynamic account of the N400. *Language and Cognitive Processes*, 26:1338–1367.

Barton, S. B. and Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition*, 21(4):477–487.

Bi, Y., Wei, T., Wu, C., Han, Z., Jiang, T., and Caramazza, A. (2011). The role of the left anterior temporal lobe in language processing revisited: Evidence from an individual with ATL resection. *Cortex*, 47(5):575.

Binder, J. R. and Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, 15(11):527–536.

Binder, J. R., Desai, R. H., Graves, W. W., and Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, 19(12):2767–2796.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford University Press: New York.

Bögels, S., Schriefers, H., Vonk, W., and Chwilla, D. J. (2011a). Pitch accents in context: How listeners process accentuation in referential communication. *Neuropsychologia*, 49(7):2022–2036.

Bögels, S., Schriefers, H., Vonk, W., and Chwilla, D. J. (2011b). Prosodic breaks in sentence processing investigated by event-related potentials. *Language and Linguistics Compass*, 5:424–440.

Bögels, S., Schriefers, H., Vonk, W., and Chwilla, D. J. (2011c). The role of prosodic breaks and pitch accents in grouping words during on-line sentence processing. *Journal of Cognitive Neuroscience*, 23(9):2447–2467.

Bögels, S., Schriefers, H., Vonk, W., Chwilla, D. J., and Kerkhofs, R. (2010). The interplay between prosody and syntax in sentence processing: The case of subject- and object-control verbs. *Journal of Cognitive Neuroscience*, 22(5):1036–1053.

Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, 25(1):151–188.

Bornkessel, I. and Schlesewsky, M. (2006). The extended Argument Dependency Model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, 113(4):787–821.

Bornkessel, I., Schlesewsky, M., and Friederici, A. D. (2003). Eliciting thematic reanalysis effects: The role of syntax-independent information during parsing. *Language and Cognitive Processes*, 18(3):269–298.

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

Bornkessel-Schlesewsky, I. and Schlesewsky, M. (2013). Reconciling time, space and function: A new dorsal–ventral stream model of sentence comprehension. *Brain and language*, 125(1):60–76.

Bornkessel-Schlesewsky, I., Schlesewsky, M., and von Cramon, D. Y. (2009). Word order and broca's region: Evidence for a supra-syntactic perspective. *Brain and Language*, 111(3):125–139.

Box, G. E. P. and Muller, M. E. (1958). A note on the generation of random normal deviates. *The Annals of Mathematical Statistics*, 29(2):610–611.

Bressler, S. L. and Menon, V. (2010). Large-scale brain networks in cognition: Emerging methods and principles. *Trends in Cognitive Sciences*, 14(6):277–290.

Brouwer, H., Fitz, H., and Hoeks, J. C. J. (2010a). Modeling the noun phrase versus sentence coordination ambiguity in Dutch: Evidence from surprisal theory. In *Proceedings of the 2010 Workshop on Cognitive Modeling and*

Computational Linguistics, pages 72–80. Association for Computational Linguistics.

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., Fitz, H., Hoeks, J. C. J., and Nerbonne, J. (2010b). How teeth can brush a child: A neurocomputational account of the Semantic Illusion in language comprehension. NWO-PGW, 10-26 (322-70-002).

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.

Brown, C. and Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, 5(1):34–44.

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

Buchsbaum, B. R. and D'Esposito, M. (2008). The search for the phonological store: From loop to convolution. *Journal of Cognitive Neuroscience*, 20(5):762–778.

Buchsbaum, B. R., Olsen, R. K., Koch, P., and Berman, K. F. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron*, 48(4):687–697.

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.

Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language*, 98(2):159–168.

Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory. *NeuroReport*, 18(17):1851–1854.

Cabeza, R. and Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience*, 12(1):1–47.

Caplan, D., Chen, E., and Waters, G. (2008a). Task-dependent and task-independent neurovascular responses to syntactic processing. *Cortex*, 44(3):257–275.

Caplan, D., Stanczak, L., and Waters, G. (2008b). Syntactic and thematic constraint effects on blood oxygenation level dependent signal correlates of comprehension of relative clauses. *Journal of Cognitive Neuroscience*, 20(4):643–656.

Catani, M., Jones, D. K., and ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Annals of Neurology*, 57(1):8–16.

Chang, Y., Furber, S., and Welbourne, S. (2012). Generating realistic semantic codes for use in neural network models. In Miyake, N., Peebles, D., and Cooper, R. P., editors, *Proceedings of the 34th annual meeting of the Cognitive Science Society (CogSci 2012)*.

Christiansen, M. H. and Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, 23(2):157–205.

Christiansen, M. H. and Chater, N. (2001). *Connectionist psycholinguistics*. Greenwood Publishing Group.

Christiansen, M. H., Conway, C. M., and Onnis, L. (2012). Similar neural correlates for language and sequential learning: Evidence from event-related brain potentials. *Language and Cognitive Processes*, 27(2):231–256.

Chwilla, D. J., Brown, C. M., and Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology*, 32(3):274–285.

Chwilla, D. J. and Kolk, H. H. J. (2005). Accessing world knowledge: Evidence from N400 and reaction time priming. *Cognitive Brain Research*, 25(3):589–606.

Coulson, S., King, J. W., and Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, 13(1):21–58.

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.

Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D., and Damasio, A. R. (1996). A neural basis for lexical retrieval. *Nature*, 380(6574):499–505.

Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., and Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition*, 92(1):179–229.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):391–407.

DeLong, K. A., Urbach, T. P., and Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8):1117–1121.

Dronkers, N. F., Wilkins, D. P., van Valin, R. D., Redfern, B. B., and Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, 92(1):145–177.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14(2):179–211.

Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences*, 8(7):301–306.

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, 33(4):547–582.

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.

Fahlman, S. E. (1988). An empirical study of learning speed in back-propagation networks. Technical report, Carnegie Mellon University.

Farmer, T. A., Anderson, S. E., and Spivey, M. J. (2007). Gradiency and visual context in syntactic garden-paths. *Journal of Memory and Language*, 57(4):570–595.

Federmeier, K. D. and Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, 41(4):469–495.

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.

Felser, C., Clahsen, H., and 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 and Language*, 87(3):345–354.

Ferreira, F. and Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, 1(1-2):71–83.

Fiebach, C. J., Schlesewsky, M., and Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of german wh-questions. *Journal of Memory and Language*, 47(2):250–272.

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., and Perry Jr, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, 20(4):400–409.

Frank, S. L., Haselager, W. F. G., and van Rooij, I. (2009). Connectionist semantic systematicity. *Cognition*, 110(3):358–379.

Frank, S. L., Koppen, M., Noordman, L. G. M., and Vonk, W. (2008). World knowledge in computational models of discourse comprehension. *Discourse Processes*, 45(6):429–463.

Frazier, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language & Linguistic Theory*, 5(4):519–559.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6(2):78–84.

Friederici, A. D. (2009). Pathways to language: Fiber tracts in the human brain. *Trends in Cognitive Sciences*, 13(4):175–181.

Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews*, 91(4):1357–1392.

Friederici, A. D. (2012). The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, 16(5):262–268.

Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, 9(10):474–480.

Frisch, S., Schlesewsky, M., Saddy, D., and Alpermann, A. (2002). The P600 as an indicator of syntactic ambiguity. *Cognition*, 85(3):B83–B92.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Lawrence Erlbaum.

Gerrig, R. J. (2005). The scope of memory-based processing. *Discourse Processes*, 39(2):225–242.

Gerrig, R. J. and McKoon, G. (1998). The readiness is all: The functionality of memory-based text processing. *Discourse Processes*, 26(2):67–86.

Gillund, G. and Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, 91(1):1–67.

Givón, T. (1992). The grammar of referential coherence as mental processing instructions. *Linguistics*, 30(1):5–56.

Gold, B. T., Balota, D. A., Jones, S. J., Powell, D. K., Smith, C. D., and Andersen, A. H. (2006). Dissociation of automatic and strategic lexical-semantics: Functional magnetic resonance imaging evidence for differing roles of multiple frontotemporal regions. *Journal of Neuroscience*, 26(24):6523–6532.

Gouvea, A. C., Phillips, C., Kazanina, N., and Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes*, 25(2):149–188.

Grodzinsky, Y. and Santi, A. (2008). The battle for broca's region. *Trends in Cognitive Sciences*, 12(12):474–480.

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, 15(6):883–899.

Hagoort, P. (2005). On broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, 9(9):416–423.

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.

Hagoort, P., Hald, L., Bastiaansen, M., and Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304(5669):438–441.

Hagoort, P. and van Berkum, J. J. A. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481):801–811.

Halgren, E., Dhond, R. P., Christensen, N., van Petten, C., Marinkovic, K., Lewine, J. D., and Dale, A. M. (2002). N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *NeuroImage*, 17(3):1101–1116.

Hart Jr., J. and Gordon, B. (1990). Delineation of single-word semantic comprehension deficits in aphasia, with anatomical correlation. *Annals of Neurology*, 27(3):226–231.

Haykin, S. (1999). *Neural networks: A comprehensive foundation*. Prentice Hall.

Hickok, G. and Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1):67–99.

Hickok, G. and Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5):393–402.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, 95(4):528–551.

Hoeks, J. C. J. (1999). *The processing of coordination: Semantic and pragmatic constraints on ambiguity resolution*. PhD thesis, University of Nijmegen.

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*. 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.

Hoeks, J. C. J., Stowe, L. A., Hendriks, P., and Brouwer, H. (2013). Questions left unanswered: How the brain responds to missing information. *PLoS ONE*, 8(10).

Houghton, G. (2005). *Connectionist models in cognitive psychology*. Psychology Press.

Igel, C. and Hüsken, M. (2000). Improving the Rprop learning algorithm. In *Proceedings of the second international ICSC symposium on neural computation (NC 2000)*, pages 115–121.

Indefrey, P. and Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92(1):101–144.

Jacobs, R. A. (1988). Increased rates of convergence through learning rate adaptation. *Neural networks*, 1(4):295–307.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge: Cambridge University Press.

Jordan, M. I. (1986). Serial order: A parallel distributed processing approach. *Advances in Psychology*, 121:471–495.

Kaan, E., Harris, A., Gibson, E., and Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, 15(2):159–201.

Kaan, E. and Swaab, T. Y. (2003a). Electrophysiological evidence for serial sentence processing: A comparison between non-preferred and ungrammatical continuations. *Cognitive Brain Research*, 17(3):621–635.

Kaan, E. and Swaab, T. Y. (2003b). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience*, 15(1):98–110.

Kamp, H. and Reyle, U. (1993). From discourse to logic. *Dordrecht, The Netherlands: D. Reidel*.

Kerkhofs, R. and Haselager, W. F. G. (2006). The embodiment of meaning. *Manuscrito–Revista Internacional de Filosofia*, 29:753–764.

Kerkhofs, R., Vonk, W., Schriefers, H., and Chwilla, D. J. (2007). Discourse, syntax, and prosody: The brain reveals an immediate interaction. *Journal of Cognitive Neuroscience*, 19(9):1421–1434.

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

Kertesz, A. (1979). *Aphasia and associated disorders: Taxonomy, localization, and recovery*. New York: Grune & Stratton.

Kim, A. and Lai, V. (2012). Rapid interactions between lexical semantic and word form analysis during word recognition in context: Evidence from ERPs. *Journal of Cognitive Neuroscience*, 24(5):1104–1112.

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.

Kim, A. and Sikos, L. (2011). Conflict and surrender during sentence processing: An ERP study of syntax-semantics interaction. *Brain and Language*, 118(1):15–22.

King, J. W. and Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, 7(3):376–395.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2):163–182.

Kintsch, W. (1992). How readers construct situation models for stories: The role of syntactic cues and causal inferences. *From learning processes to cognitive processes: Essays in honor of William K. Estes*, 2:261–278.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.

Kintsch, W. and Mangalath, P. (2011). The construction of meaning. *Topics in Cognitive Science*, 3(2):346–370.

Kolk, H. and Chwilla, D. J. (2007). Late positivities in unusual situations. *Brain and Language*, 100(3):257–261.

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.

Kos, M., Vosse, T., van den Brink, D., and Hagoort, P. (2010). About edible restaurants: Conflicts between syntax and semantics as revealed by ERPs. *Frontiers in Psychology*, 1:1–11.

Kotchoubey, B. (2006). Event-related potentials, cognition, and behavior: A biological approach. *Neuroscience & Biobehavioral Reviews*, 30(1):42–65.

Kotz, S. A., Frisch, S., von Cramon, D. Y., and Friederici, A. D. (2003). Syntactic language processing: ERP lesion data on the role of the basal ganglia. *Journal of the International Neuropsychological Society*, 9(7):1053–1060.

Kounios, J. and Holcomb, P. J. (1992). Structure and process in semantic memory: Evidence from event-related brain potentials and reaction times. *Journal of Experimental Psychology: General*, 121(4):459–479.

Koyama, M. S., Kelly, C., Shehzad, Z., Penesetti, D., Castellanos, F. X., and Milham, M. P. (2010). Reading networks at rest. *Cerebral Cortex*, 20(11):2549–2559.

Kuperberg, G., Caplan, D., Sitnikova, T., Eddy, M., and Holcomb, P. (2006). Neural correlates of processing syntactic, semantic, and thematic relationships in sentences. *Language and Cognitive Processes*, 21(5):489–530.

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

Kuperberg, G. R., Choi, A., Cohn, N., Paczynski, M., and Jackendoff, R. (2010). Electrophysiological correlates of complement coercion. *Journal of Cognitive Neuroscience*, 22(12):2685–2701.

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., and Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active english sentences: Evidence from event-related potentials. *Brain and Language*, 100(3):223–237.

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.

Kwon, H., Kuriki, S., Kim, J. M., Lee, Y. H., Kim, K., and Nam, K. (2005). MEG study on neural activities associated with syntactic and semantic violations in spoken korean sentences. *Neuroscience Research*, 51(4):349–357.

Lachaux, J., Rodriguez, E., Martinerie, J., and Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping*, 8(4):194–208.

Laszlo, S. and Plaut, D. C. (2012). A neurally plausible parallel distributed processing model of event-related potential word reading data. *Brain and Language*, 120(3):271–281.

Lau, E. F., Almeida, D., Hines, P. C., and Poeppel, D. (2009). A lexical basis for N400 context effects: Evidence from MEG. *Brain and Language*, 111(3):161–172.

Lau, E. F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, 9(12):920–933.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177.

Liu, B., Jin, Z., Li, W., Li, Y., and Wang, Z. (2009). The pragmatic meanings conveyed by function words in chinese sentences: An ERP study. *Journal of Neurolinguistics*, 22(6):548–562.

Lotze, N., Tune, S., Schlesewsky, M., and 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(13):3573–3582.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. MIT Press.

Lund, K. and Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods*, 28(2):203–208.

MacDonald, M. C., Pearlmutter, N. J., and Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4):676–703.

Makris, N. and Pandya, D. N. (2009). The extreme capsule in humans and rethinking of the language circuitry. *Brain Structure and Function*, 213(3):343–358.

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, 244(5417):522–523.

Martín-Loeches, M., Nigbur, R., Casado, P., Hohlfeld, A., and Sommer, W. (2006). Semantics prevalence over syntax during sentence processing:

A brain potential study of noun-adjective agreement in spanish. *Brain Research*, 1093:178–189.

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.

Mazerolle, E. L., Beyea, S. D., Gawryluk, J. R., Brewer, K. D., Bowen, C. V., and D'Arcy, R. C. N. (2010). Confirming white matter fMRI activation in the corpus callosum: Co-localization with DTI tractography. *NeuroImage*, 50(2):616–621.

McClelland, J. L., Rumelhart, D. E., et al. (1987). *Parallel distributed processing, Vol. 2: Psychological and biological models*. MIT press.

McRae, K., Spivey-Knowlton, M. J., and Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38(3):283–312.

Meehan, T. P. and Bressler, S. L. (2012). Neurocognitive networks: Findings, models, and theory. *Neuroscience & Biobehavioral Reviews*, 36(10):2232–2247.

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.

Morris, R. K. (1994). Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(1):92.

Mueller, J. L., Hahne, A., Fujii, Y., and Friederici, A. D. (2005). Native and nonnative speakers' processing of a miniature version of Japanese as revealed by ERPs. *Journal of Cognitive Neuroscience*, 17(8):1229–1244.

Münte, T. F., Heinze, H., Matzke, M., Wieringa, B. M., and Johannes, S. (1998). Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia*, 36(3):217–226.

Näätänen, R. and Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, 24(4):375–425.

Neville, H., Nicol, J. L., Barss, A., Forster, K. I., and Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, 3(2):151–165.

Nguyen, D. and Widrow, B. (1990). Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. In *Proceedings of the International Joint Conference on Neural Networks*, pages 21–26. IEEE.

Nieuwland, M. S. and 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 Research*, 24(3):691–701.

Novick, J. M., Trueswell, J. C., and Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, 5(3):263.

Núñez-Peña, M. I. and Honrubia-Serrano, M. L. (2004). P600 related to rule violation in an arithmetic task. *Cognitive Brain Research*, 18(2):130–141.

Osterhout, L. and Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31(6):785–806.

Osterhout, L. and Holcomb, P. J. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes*, 8(4):413–437.

Osterhout, L., Holcomb, P. J., and Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4):786–803.

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

Padó, U., Crocker, M. W., and Keller, F. (2009). A probabilistic model of semantic plausibility in sentence processing. *Cognitive Science*, 33(5):794–838.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, 6(7):674–681.

Patterson, K., Nestor, P. J., and Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8(12):976–987.

Phillips, C., Kazanina, N., and Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, 22(3):407–428.

Price, C. J. (2002). The anatomy of language: Contributions from functional neuroimaging. *Journal of Anatomy*, 197(3):335–359.

Price, C. J. (2010). The anatomy of language: A review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences*, 1191(1):62–88.

Price, C. J., Wise, R. J. S., Watson, J. D. G., Patterson, K., Howard, D., and Frackowiak, R. S. J. (1994). Brain activity during reading: The effects of exposure duration and task. *Brain*, 117(6):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., and Gore, J. C. (1996). Cerebral organization of component processes in reading. *Brain*, 119(4):1221–1238.

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, 22:253–336.

Pulvermüller, F. (2001). Brain reflections of words and their meaning. *Trends in Cognitive Sciences*, 5(12):517–524.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2):59–108.

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.

Riedmiller, M. (1994). Advanced supervised learning in multi-layer perceptrons: From backpropagation to adaptive learning algorithms. *Computer Standards & Interfaces*, 16(3):265–278.

Riedmiller, M. and Braun, H. (1993). A direct adaptive method for faster backpropagation learning: The RPROP algorithm. In *Proceedings of the IEEE International Conference on Neural Networks*, pages 586–591. IEEE.

Rogalsky, C. and Hickok, G. (2011). The role of Broca's area in sentence comprehension. *Journal of Cognitive Neuroscience*, 23(7):1664–1680.

Rogers, T. T., Hocking, J., Noppeney, U., Mechelli, A., Gorno-Tempini, M. L., Patterson, K., and Price, C. J. (2006). Anterior temporal cortex and semantic memory: Reconciling findings from neuropsychology and functional imaging. *Cognitive, Affective, & Behavioral Neuroscience*, 6(3):201–213.

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., and Patterson, K. (2004). Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review*, 111(1):205.

Rogers, T. T. and McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. The MIT Press.

Rohde, D. L. T. (2002). *A connectionist model of sentence comprehension and production*. PhD thesis, Carnegie Mellon University.

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.

Rohde, D. L. T. and Plaut, D. C. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition*, 72(1):67–109.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986a). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.

Rumelhart, D. E., McClelland, J. L., et al. (1986b). *Parallel distributed processing, Vol. 1: Foundations*. MIT Press.

Sanford, A. J., Leuthold, H., Bohan, J., and Sanford, A. J. S. (2011). Anomalies at the borderline of awareness: An ERP study. *Journal of Cognitive Neuroscience*, 23(3):514–523.

Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M., Umarova, R., Musso, M., Glauche, V., Abel, S., Huber, S., Rijntjes, M., Hennig, J., and Weiller, C. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences*, 105(46):18035–18040.

Schumacher, P. B. (2011). The hepatitis called ...: Electrophysiological evidence for enriched composition. In Meibauer, J. and Steinbach, M., editors, *Experimental Pragmatics/Semantics*, pages 199–219. John Benjamins.

Schwanenflugel, P. J., Harnishfeger, K. K., and Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, 27(5):499–520.

Schwanenflugel, P. J. and Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(1):82–102.

Schwanenflugel, P. J. and Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language*, 24(2):232–252.

Seghier, M. L. (2013). The angular gyrus: Multiple functions and multiple subdivisions. *The Neuroscientist*, 19(1):43–61.

Service, E., Helenius, P., Maury, S., and Salmelin, R. (2007). Localization of syntactic and semantic brain responses using magnetoencephalography. *Journal of Cognitive Neuroscience*, 19(7):1193–1205.

Shalom, D. B. and Poeppel, D. (2008). Functional anatomic models of language: Assembling the pieces. *The Neuroscientist*, 14(1):119–127.

Siegel, M., Donner, T. H., and Engel, A. K. (2012). Spectral fingerprints of large-scale neuronal interactions. *Nature Reviews Neuroscience*, 13(2):121–134.

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.

Steinhauer, K., Alter, K., and Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, 2(2):191–196.

Steinhauer, K. and Drury, J. E. (2011). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language*, 120(2):135–162.

Stowe, L. A., Haverkort, M., and Zwarts, F. (2005). Rethinking the neurological basis of language. *Lingua*, 115(7):997–1042.

Stroud, C. M. A. (2009). *Structural and semantic selectivity in the electrophysiology of sentence comprehension*. PhD thesis, University of Maryland.

Stroud, C. M. A. and Phillips, C. (2011). Examining the evidence for an independent semantic analyzer: An ERP study in Spanish. *Brain and Language*, 120(2):108–126.

Thomas, M. S. C. and McClelland, J. L. (2008). Connectionist models of cognition. In Sun, R., editor, *The Cambridge Handbook of Computational Psychology*, pages 23–58. London: Cambridge University Press.

Tsapkini, K., Frangakis, C. E., and Hillis, A. E. (2011). The function of the left anterior temporal pole: Evidence from acute stroke and infarct volume. *Brain*, 134(10):3094–3105.

Tse, C. Y., Lee, C. L., Sullivan, J., Garnsey, S. M., Dell, G. S., Fabiani, M., and Gratton, G. (2007). Imaging cortical dynamics of language processing with the event-related optical signal. *Proceedings of the National Academy of Sciences*, 104(43):17157–17162.

Turken, A. U. and Dronkers, N. F. (2011). The neural architecture of the language comprehension network: Converging evidence from lesion and connectivity analyses. *Frontiers in Systems Neuroscience*, 5:1.

Tyler, L. K., Marslen-Wilson, W. D., Randall, B., Wright, P., Devereux, B., Zhuang, J., Papoutsi, M., and Stamatakis, E. A. (2011). Left inferior frontal cortex and syntax: Function, structure and behaviour in patients with left hemisphere damage. *Brain*, 134(2):415–431.

Urbach, T. P. and Kutas, M. (2010). Quantifiers more or less quantify online: 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.

van Berkum, J. J. A. (2010). The brain is a prediction machine that cares about good and bad–any implications for neuropragmatics? *Italian Journal of Linguistics*, 22(1):181–208.

van Berkum, J. J. A., Hagoort, P., and Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, 11(6):657–671.

van Berkum, J. J. A., Koornneef, A. W., Otten, M., and Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research*, 1146:158–171.

van de Meerendonk, N., Indefrey, P., Chwilla, D. J., and Kolk, H. H. J. (2011). Monitoring in language perception: Electrophysiological and hemodynamic responses to spelling violations. *NeuroImage*, 54(3):2350–2363.

van de Meerendonk, N., Kolk, H. H. J., Chwilla, D. J., and Vissers, C. T. W. M. (2009). Monitoring in language perception. *Language and Linguistics Compass*, 3(5):1211–1224.

van de Meerendonk, N., Kolk, H. H. J., Vissers, C. T. W. M., and Chwilla, D. J. (2010). Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, 22(1):67–82.

van Dijk, T. A. and Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

van Gompel, R. P. G., Pickering, M. J., and Traxler, M. J. (2000). Unrestricted race: A new model of syntactic ambiguity resolution. In Kennedy, A., Radach, R., Heller, D., and Pynte, J., editors, *Reading as a Perceptual Process*, pages 621–648. Elsevier, Oxford.

van Herten, M., Chwilla, D. J., and Kolk, H. H. J. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience*, 18(7):1181–1197.

van Herten, M., Kolk, H. H. J., and Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22(2):241–255.

van Petten, C. (1993). A comparison of lexical and sentence-level context effects in event-related potentials. *Language and Cognitive Processes*, 8(4):485–531.

van Petten, C. (1995). Words and sentences: Event-related brain potential measures. *Psychophysiology*, 32(6):511–525.

van Petten, C. and Kutas, M. (1991). Influences of semantic and syntactic context on open-and closed-class words. *Memory & Cognition*, 19:95–112.

van Petten, C. and Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language*, 97(3):279–293.

van Petten, C. and Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and erp components. *International Journal of Psychophysiology*, 83(2):176–190.

Venhuizen, N. J., Bos, J., and Brouwer, H. (2013). Parsimonious semantic representations with projection pointers. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)-Long Papers*, pages 252–263.

Venhuizen, N. J., Bos, J., Hendriks, P., and Brouwer, H. (to appear). How and why conventional implicatures project. In *Proceedings of the 24rd Semantics and Linguistics Theory Conference (SALT24)*, New York.

Vigneau, M., Beaucousin, V., Hervé, P., Duffau, H., Crivello, F., Houdé, O., Mazoyer, B., and Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage*, 30(4):1414–1432.

Vigneau, M., Beaucousin, V., Hervé, P., Jobard, G., Petit, L., Crivello, F., Mellet, E., Zago, L., Mazoyer, B., and Tzourio-Mazoyer, N. (2011). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. *NeuroImage*, 54(1):577–593.

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

Vissers, C. T. W. M., Chwilla, D. J., and Kolk, H. H. J. (2007). The interplay of heuristics and parsing routines in sentence comprehension: Evidence from ERPs and reaction times. *Biological Psychology*, 75(1):8–18.

Vosse, T. and Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, 75(2):105–143.

Wahl, M., Marzinzik, F., Friederici, A. D., Hahne, A., Kupsch, A., Schneider, G., Saddy, D., Curio, G., and Klostermann, F. (2008). The human thalamus processes syntactic and semantic language violations. *Neuron*, 59(5):695–707.

Wang, L., Hagoort, P., and Yang, Y. (2009). Semantic illusion depends on information structure: ERP evidence. *Brain Research*, 1282:50–56.

Weiller, C., Musso, M., Rijntjes, M., and Saur, D. (2009). Please don't underestimate the ventral pathway in language. *Trends in Cognitive Sciences*, 13(9):369–1370.

Weiss, S. and Mueller, H. M. (2003). The contribution of EEG coherence to the investigation of language. *Brain and Language*, 85(2):325–343.

Werbos, P. J. (1990). Backpropagation through time: What it does and how to do it. *Proceedings of the IEEE*, 78(10):1550–1560.

Williams, R. J. and Peng, J. (1990). An efficient gradient-based algorithm for on-line training of recurrent network trajectories. *Neural Computation*, 2(4):490–501.

Wolff, S., Schlesewsky, M., Hirotani, M., and Bornkessel-Schlesewsky, I. (2008). The neural mechanisms of word order processing revisited: Electrophysiological evidence from Japanese. *Brain and language*, 107(2):133–157.

Ye, Z. and Zhou, X. (2008). Involvement of cognitive control in sentence comprehension: Evidence from ERPs. *Brain Research*, 1203:103–115.

Zwaan, R. A. (1999). Situation models: The mental leap into imagined worlds. *Current Directions in Psychological Science*, 8(1):15–18.

Zwaan, R. A. (2003). The immersed experiencer: Toward an embodied theory of language comprehension. *Psychology of Learning and Motivation*, 44:35–62.

Zwaan, R. A., Langston, M. C., and Graesser, A. C. (1995). The construction of situation models in narrative comprehension. *Psychological Science*, 6(5):292–297.

Zwaan, R. A. and Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2):162–185.

Lekensamenvatting

Hoe kennen onze hersenen betekenis toe aan woorden die we horen of lezen? En hoe worden de betekenissen van opeenvolgende woorden gecombineerd tot de betekenis van een zin of een verhaal? Om antwoord te krijgen op deze vragen moeten we het brein bestuderen terwijl het aan het werk is, en dus veilig opgeborgen ligt in de menselijke schedel. Hersenmetingen met behulp van elektro-encefalografie (EEG) bieden een uitzicht op het werkende brein, vergelijkbaar met hoe de glazen bodem van een rondvaartboot een uitzicht biedt op het leven onder water. Middels elektroden die op de hoofdhuid worden geplakt, kunnen de zwakke elektrische signalen worden gemeten die hersencellen—de bouwstenen van ons brein—produceren wanneer zij aan het werk zijn. Deze elektrische signalen kunnen ons veel vertellen over de werking van de hersenen en zo ook over de mechanismen van taalverwerking.

Wanneer men bijvoorbeeld de zin "Jan besmeert zijn warme brood met sokken" leest, leidt het woord *sokken* steevast tot een negatieve piek in het EEG signaal met een maximum na ongeveer 400 milliseconden. Deze piek wordt de N400 genoemd (N voor Negatief en 400 voor het tijdstip waarop de piek het grootst is). Belangrijk is dat de grootte van deze N400 piek veel groter is voor *sokken* dan voor bijvoorbeeld *boter* in de zin "Jan besmeert zijn warme brood met boter." In het licht van onze wereldkennis is het inderdaad gek om brood te besmeren met sokken, de N400 laat zien dat onze hersenen dit dus al na 400 milliseconden opmerken. De N400 is echter niet simpelweg gevoelig voor schendingen van wereldkennis. Neem bijvoorbeeld de volgende zin: “De eigenaren wilden het hotel een tropisch uiterlijk geven, dus beplantten zij de oprijlaan met rijen [. . .].” Wanneer deze zin wordt afgemaakt met *palmbomen*, levert dit een kleinere N400 piek op dan wanneer de zin wordt afgemaakt met *dennenbomen*, terwijl *dennenbomen* juist weer een kleinere N400 piek oplevert dan *tulpen*. De N400 lijkt dus een maat te zijn voor hoe goed een woord qua betekenis in de eraan voorafgaande context past. Dit heeft er toe geleid dat onderzoekers de hypothese hebben bedacht dat de grootte van de N400 piek laat zien hoeveel moeite ons brein heeft met het inpassen van de betekenis van een woord in een zin of een verhaal: het *maken* van betekenis. In vakjargon wordt deze hypothese ook wel de *semantische integratie* hypothese genoemd.

Deze hypothese is vandaag de dag nog springlevend, maar heeft de afgelopen tien jaar wel de nodige barsten en deuken opgelopen. Vanaf 2003 publiceerden meer en meer onderzoeksgroepen resultaten die niet goed stroken met de semantische integratie hypothese. Sommige zinnen die qua betekenis duidelijk niet kloppen, leiden niet tot de verwachte toename in de N400 piek. Het woord *geworpen* in “De speer heeft de atleten geworpen” leidt bijvoorbeeld niet tot een grotere N400 piek dan hetzelfde woord in de zin “De speer werd door de atleten geworpen.” Dit is raar omdat de eerste zin onze kennis van de wereld schendt en de tweede niet. Om deze resultaten te verklaren, hebben onderzoeksgroepen van over de hele wereld nieuwe modellen van taalbegrip ontwikkeld, met als resultaat dat er nu vijf complexe modellen centraal staan in het vakgebied. In het eerste deel van dit proefschrift laat ik echter zien dat *geen* van deze modellen alle relevante bevindingen kan verklaren. Ik concludeer dat de reden hiervoor niet schuilt in hoe de verwerkingsmodellen in elkaar steken, maar in een foutieve aanname over de processen die ten grondslag liggen aan de N400 piek: De N400 is niet een maat die laat zien hoeveel moeite ons brein heeft met het *maken* van betekenis, maar een maat die laat zien hoeveel werk onze hersenen ondervinden aan het *ophalen* van betekenis uit het langetermijngeheugen. In “De speer heeft de atleten geworpen” en “De speer werd door de atleten geworpen” is de N400 piek voor *geworpen* even groot, omdat het ophalen van de betekenis van dit woord in beide zinnen even gemakkelijk is. Dit komt omdat in deze zinnen de aanwezigheid van de woorden *speer* en *atleten*, alsmede onze kennis van de wereld (*werpen* is iets wat *atleten* doorgaans met een *speer* doen), ervoor zorgt dat de betekenis van *geworpen* al actief is in ons geheugen voordat we het woord lezen of horen. Deze alternatieve hypothese over de N400 piek—die ook al voorzichtig door andere onderzoekers is voorgesteld—wordt ook wel de *lexicale pre-activatie* hypothese genoemd. Het aanvaarden van deze hypothese roept echter wel een belangrijke vraag op: Als de N400 geen maat is die laat zien hoeveel moeite de taalverwerker heeft met het *maken* van betekenis (het inpassen van de betekenis van een woord in een zin of een verhaal), waar in het EEG signaal zien we dit proces dan terug? Ik stel voor dat we dit terug zien in de P600 piek (P voor positief en 600 voor het tijdstip waarop deze piek het grootst is).

De P600 werd oorspronkelijk ontdekt als zijnde gevoelig voor grammaticale schendingen, zoals bijvoorbeeld in reactie op het woord *gooi* in de zin "Het verwende kind gooi zijn speelgoed op de grond." De onjuiste vervoeging van het woord *gooi* levert een grotere P600 piek op dan de juiste vervoeging *gooit*. Gegeven deze bevindingen, hebben onderzoekers de hypothese opgesteld dat de P600 een maat is voor *grammaticale reparatie*. Onder deze hypothese—die ook centraal staat binnen de eerdergenoemde vijf modellen—treedt bij het woord *gooi* een grotere P600 piek op omdat onze hersenen bij het verwerken van dit woord de onjuiste vervoeging direct corrigeren, om zo tot de juiste interpretatie van de zin te komen. Het woord *geworpen* in de zin "De speer heeft de atleten geworpen" leidt echter *ook* tot een grotere P600 piek dan hetzelfde woord in "De speer werd door de atleten geworpen." Dit trekt de *grammaticale reparatie* hypothese in twijfel, want beide zinnen zijn grammaticaal prima in orde en er valt dus niets te repareren. Ik stel dan ook voor dat deze hypothese niet klopt en formuleer een nieuwe hypothese over de P600 piek: de MRC hypothese.

Onder de MRC hypothese is de P600 piek een maat voor het opbouwen, reorganiseren, of bijwerken van een *mentale representatie van wat er wordt* *gecommuniceerd* ('Mental Representation of what is being Communicated'; MRC). Deze hypothese verklaart de aanwezigheid van een toename in de P600 piek voor *geworpen* in de zin "De speer heeft de atleten geworpen." De verwerker ondervindt moeilijkheden met het inpassen van de betekenis van *geworpen* omdat de resulterende interpretatie (de resulterende MRC) niet strookt met onze wereldkennis.

In combinatie met de *lexicale pre-activatie* hypothese voor de N400 piek, heeft deze nieuwe MRC hypothese over de P600 piek vergaande consequenties voor modellen van taalverwerking. Zo kunnen we de eerder genoemde vijf modellen die centraal staan in de vakliteratuur verwerpen, en een nieuw model voorstellen waarin het taalverwerkingssysteem feitelijk in cycli van 'Retrieval'—het *ophalen* van woordbetekenis (∼N400)—en 'Integration'—het *inpassen* van woordbetekenis in een MRC (∼P600 piek)—werkt. Ik noem dit nieuwe model dan ook het *Retrieval-Integration* model.

In deel twee van dit proefschrift laat ik zien dat deze Retrieval-Integration cycli prima passen binnen de anatomie van het menselijk brein, en in deel drie beschrijf ik een computermodel van het taalvewerkingssysteem op basis van deze anatomie, en laat ik zien dat het Retrieval-Integration model inderdaad de relevante bevindingen in de literatuur kan verklaren. Als ik het model bijvoorbeeld de zin "De speer heeft de atleten geworpen" laat 'lezen', leidt het woord *geworpen* net als bij mensen niet tot een grotere N400, maar wel tot een grotere P600 piek. In het vierde en laatste deel presenteer ik mijn conclusies en schets ik kort een aantal belangrijke onderzoeksvragen voor de toekomst.

Curriculum Vitae (English)

Harm Brouwer was born on November 28th, 1984 in Drachten, The Netherlands. After obtaining a pre-university secondary education (VWO) diploma, he started the bachelor *Information Science* at the University of Groningen in 2005. In the course of this programme, he developed a strong interest in neuro- and psycholinguistics. During his studies, Harm was involved in teaching through various teaching assistantships. He graduated *cum laude* in 2008, and was admitted to the research master *Behavioural and Cognitive Neurosciences* at the same university, where he followed the track *Cognitive Neuroscience and Cognitive Modeling*. In 2010, he graduated *cum laude*, on the basis of his master's theses *Automated Lexical Acquisition for Dutch Verbs* and *Competition in Syntactic Ambiguity Resolution*. In the same year, he was awarded the *Duurzame Geesteswetenschappen* subsidy from the Netherlands Organisation for Scientific Research (NWO) to finance his PhD dissertation research *How teeth can brush a child: A neurocomputational account of the Semantic Illusion in language comprehension*. During his PhD trajectory, Harm wrote software for simulating neural networks, taught the summer school course *The Neurobiology of Language* for PhD and master students together with John Hoeks, and was nominated for the *Faces of Science* project of the Royal Netherlands Academy of Arts and Sciences (KNAW).

Curriculum Vitae (Nederlands)

Harm Brouwer werd op 28 november 1984 geboren in Drachten. Na het behalen van een VWO diploma, begon hij in 2005 met de studie *Informatiekunde* aan de Rijksuniversiteit Groningen. Binnen deze opleiding ontwikkelde hij een sterke interesse in de neuro- en psycholinguïstiek. Tijdens zijn studie was Harm veel betrokken bij onderwijs door middel van verscheidene studentassistentschappen. In 2008 haalde hij *cum laude* zijn bachelordiploma en werd hij toegelaten tot de research master *Behavioural and Cognitive Neurosciences* aan dezelfde universiteit, waarbinnen hij de track *Cognitive Neuroscience and Cognitive Modeling* volgde. Hierin studeerde hij in 2010 *cum laude* af, op basis van de masterscripties *Automated Lexical Acquisition for Dutch Verbs* en *Competition in Syntactic Ambiguity Resolution*. In ditzelfde jaar ontving hij van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) een subsidie binnen het kader *Duurzame Geesteswetenschappen* voor zijn promotieonderzoek *How teeth can brush a child: A neurocomputational account of the Semantic Illusion in language comprehension*. Tijdens zijn promotietraject schreef Harm onder andere software voor het simuleren van neurale netwerken, gaf hij samen met John Hoeks de zomerschoolcursus *The Neurobiology of Language* voor promovendi en masterstudenten, en werd hij genomineerd voor het *Faces of Science* project van de Koninklijke Nederlandse Akademie van Wetenschappen (KNAW).

List of Publications

1. Venhuizen, N. J., Bos, J., Hendriks, P., and Brouwer, H. (to appear). How and why conventional implicatures project. In *Proceedings of the 24rd Semantics and Linguistics Theory Conference (SALT24)*, New York

2. 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*. New York: Oxford University Press

3. 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

4. Hoeks, J. C. J., Stowe, L. A., Hendriks, P., and Brouwer, H. (2013). Questions left unanswered: How the brain responds to missing information. *PLoS ONE*, 8(10)

5. Venhuizen, N. J., Bos, J., and Brouwer, H. (2013). Parsimonious semantic representations with projection pointers. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)-Long Papers*, pages 252–263

6. 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

7. Brouwer, H., Fitz, H., and Hoeks, J. C. J. (2010a). Modeling the noun phrase versus sentence coordination ambiguity in Dutch: Evidence from surprisal theory. In *Proceedings of the 2010 Workshop on Cognitive Modeling and Computational Linguistics*, pages 72–80. Association for Computational Linguistics

Groningen Dissertations in Linguistics (GRODIL)

1. Henriëtte de Swart (1991). *Adverbs of Quantification: A Generalized Quantifier Approach.*
2. Eric Hoekstra (1991). *Licensing Conditions on Phrase Structure.*
3. Dicky Gilbers (1992). *Phonological Networks. A Theory of Segment Representation.*
4. Helen de Hoop (1992). *Case Configuration and Noun Phrase Interpretation.*
5. Gosse Bouma (1993). *Nonmonotonicity and Categorial Unification Grammar.*
6. Peter Blok (1993). *The Interpretation of Focus: an epistemic approach to pragmatics.*
7. Roelien Bastiaanse (1993). *Studies in Aphasia.*
8. Bert Bos (1993). *Rapid User Interface Development with the Script Language Gist.*
9. Wim Kosmeijer (1993). *Barriers and Licensing.*
10. Jan-Wouter Zwart (1993). *Dutch Syntax: A Minimalist Approach.*
11. Mark Kas (1993). *Essays on Boolean Functions and Negative Polarity.*
12. Ton van der Wouden (1994). *Negative Contexts.*
13. Joop Houtman (1994). *Coordination and Constituency: A Study in Categorial Grammar.*
14. Petra Hendriks (1995). *Comparatives and Categorial Grammar.*
15. Maarten de Wind (1995). *Inversion in French.*
16. Jelly Julia de Jong (1996). *The Case of Bound Pronouns in Peripheral Romance.*
17. Sjoukje van der Wal (1996). *Negative Polarity Items and Negation: Tandem Acquisition.*

18. Anastasia Giannakidou (1997). *The Landscape of Polarity Items.*
19. Karen Lattewitz (1997). *Adjacency in Dutch and German.*
20. Edith Kaan (1997). *Processing Subject-Object Ambiguities in Dutch.*
21. Henny Klein (1997). *Adverbs of Degree in Dutch.*
22. Leonie Bosveld-de Smet (1998). *On Mass and Plural Quantification: The Case of French 'des'/'du'-NPs.*
23. Rita Landeweerd (1998). *Discourse Semantics of Perspective and Temporal Structure.*
24. Mettina Veenstra (1998). *Formalizing the Minimalist Program.*
25. Roel Jonkers (1998). *Comprehension and Production of Verbs in Aphasic Speakers.*
26. Erik F. Tjong Kim Sang (1998). *Machine Learning of Phonotactics.*
27. Paulien Rijkhoek (1998). *On Degree Phrases and Result Clauses.*
28. Jan de Jong (1999). *Specific Language Impairment in Dutch: Inflectional Morphology and Argument Structure.*
29. Hae-Kyung Wee (1999). *Definite Focus.*
30. Eun-Hee Lee (2000). *Dynamic and Stative Information in Temporal Reasoning: Korean Tense and Aspect in Discourse.*
31. Ivilin Stoianov (2001). *Connectionist Lexical Processing.*
32. Klarien van der Linde (2001). *Sonority Substitutions.*
33. Monique Lamers (2001). *Sentence Processing: Using Syntactic, Semantic, and Thematic Information.*
34. Shalom Zuckerman (2001). *The Acquisition of "Optional" Movement.*
35. Rob Koeling (2001). *Dialogue-Based Disambiguation: Using Dialogue Status to Improve Speech Understanding.*
36. Esther Ruigendijk (2002). *Case Assignment in Agrammatism: a Cross-linguistic Study.*
37. Tony Mullen (2002). *An Investigation into Compositional Features and Feature Merging for Maximum Entropy-Based Parse Selection.*
38. Nanette Bienfait (2002). *Grammatica-onderwijs aan allochtone jongeren.*
39. Dirk-Bart den Ouden (2002). *Phonology in Aphasia: Syllables and Segments in Level-specific Deficits.*
40. Rienk Withaar (2002). *The Role of the Phonological Loop in Sentence Comprehension.*

41. Kim Sauter (2002). *Transfer and Access to Universal Grammar in Adult Second Language Acquisition.*
42. Laura Sabourin (2003). *Grammatical Gender and Second Language Processing: An ERP Study.*
43. Hein van Schie (2003). *Visual Semantics.*
44. Lilia Schürcks-Grozeva (2003). *Binding and Bulgarian.*
45. Stasinos Konstantopoulos (2003). *Using ILP to Learn Local Linguistic Structures.*
46. Wilbert Heeringa (2004). *Measuring Dialect Pronunciation Differences using Levenshtein Distance.*
47. Wouter Jansen (2004). *Laryngeal Contrast and Phonetic Voicing: A Laboratory Phonology Approach to English, Hungarian and Dutch.*
48. Judith Rispens (2004). *Syntactic and Phonological Processing in Developmental Dyslexia.*
49. Danielle Bougaïré (2004). *L'approche communicative des campagnes de sensibilisation en santé publique au Burkina Faso: les cas de la planification familiale, du sida et de l'excision.*
50. Tanja Gaustad (2004). *Linguistic Knowledge and Word Sense Disambiguation.*
51. Susanne Schoof (2004). *An HPSG Account of Nonfinite Verbal Complements in Latin.*
52. M. Begoña Villada Moirón (2005). *Data-driven identification of fixed expressions and their modifiability.*
53. Robbert Prins (2005). *Finite-State Pre-Processing for Natural Language Analysis.*
54. Leonoor van der Beek (2005). *Topics in Corpus-Based Dutch Syntax.*
55. Keiko Yoshioka (2005). *Linguistic and gestural introduction and tracking of referents in L1 and L2 discourse.*
56. Sible Andringa (2005). *Form-focused instruction and the development of second language proficiency.*
57. Joanneke Prenger (2005). *Taal telt! Een onderzoek naar de rol van taalvaardigheid en tekstbegrip in het realistisch wiskundeonderwijs.*
58. Neslihan Kansu-Yetkiner (2006). *Blood, Shame and Fear: Self-Presentation Strategies of Turkish Women's Talk about their Health and Sexuality.*
59. Mónika Z. Zempléni (2006). *Functional imaging of the hemispheric contribution to language processing.*
60. Maartje Schreuder (2006). *Prosodic Processes in Language and Music.*

61. Hidetoshi Shiraishi (2006). *Topics in Nivkh Phonology.*
62. Tamás Biró (2006). *Finding the Right Words: Implementing Optimality Theory with Simulated Annealing.*
63. Dieuwke de Goede (2006). *Verbs in Spoken Sentence Processing: Unraveling the Activation Pattern of the Matrix Verb.*
64. Eleonora Rossi (2007). *Clitic production in Italian agrammatism.*
65. Holger Hopp (2007). *Ultimate Attainment at the Interfaces in Second Language Acquisition: Grammar and Processing.*
66. Gerlof Bouma (2008). *Starting a Sentence in Dutch: A corpus study of subject- and object-fronting.*
67. Julia Klitsch (2008). *Open your eyes and listen carefully. Auditory and audiovisual speech perception and the McGurk effect in Dutch speakers with and without aphasia.*
68. Janneke ter Beek (2008). *Restructuring and Infinitival Complements in Dutch.*
69. Jori Mur (2008). *Off-line Answer Extraction for Question Answering.*
70. Lonneke van der Plas (2008). *Automatic Lexico-Semantic Acquisition for Question Answering.*
71. Arjen Versloot (2008). *Mechanisms of Language Change: Vowel reduction in 15th century West Frisian.*
72. Ismail Fahmi (2009). *Automatic term and Relation Extraction for Medical Question Answering System.*
73. Tuba Yarbay Duman (2009). *Turkish Agrammatic Aphasia: Word Order, Time Reference and Case.*
74. Maria Trofimova (2009). *Case Assignment by Prepositions in Russian Aphasia.*
75. Rasmus Steinkrauss (2009). *Frequency and Function in WH Question Acquisition. A Usage-Based Case Study of German L1 Acquisition.*
76. Marjolein Deunk (2009). *Discourse Practices in Preschool. Young Children's Participation in Everyday Classroom Activities.*
77. Sake Jager (2009). *Towards ICT-Integrated Language Learning: Developing an Implementation Framework in terms of Pedagogy, Technology and Environment.*
78. Francisco Dellatorre Borges (2010). *Parse Selection with Support Vector Machines.*
79. Geoffrey Andogah (2010). *Geographically Constrained Information Retrieval.*
80. Jacqueline van Kruiningen (2010). *Onderwijsontwerp als conversatie. Probleemoplossing in interprofessioneel overleg.*

81. Robert G. Shackleton (2010). *Quantitative Assessment of English-American Speech Relationships.*
82. Tim Van de Cruys (2010). *Mining for Meaning: The Extraction of Lexico-semantic Knowledge from Text.*
83. Therese Leinonen (2010). *An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects.*
84. Erik-Jan Smits (2010). *Acquiring Quantification. How Children Use Semantics and Pragmatics to Constrain Meaning.*
85. Tal Caspi (2010). *A Dynamic Perspective on Second Language Development.*
86. Teodora Mehotcheva (2010). *After the fiesta is over. Foreign language attrition of Spanish in Dutch and German Erasmus Students.*
87. Xiaoyan Xu (2010). *English language attrition and retention in Chinese and Dutch university students.*
88. Jelena Prokić (2010). *Families and Resemblances.*
89. Radek Šimík (2011). *Modal existential wh-constructions.*
90. Katrien Colman (2011). *Behavioral and neuroimaging studies on language processing in Dutch speakers with Parkinson's disease.*
91. Siti Mina Tamah (2011). *A Study on Student Interaction in the Implementation of the Jigsaw Technique in Language Teaching.*
92. Aletta Kwant (2011). *Geraakt door prentenboeken. Effecten van het gebruik van prentenboeken op de sociaal-emotionele ontwikkeling van kleuters.*
93. Marlies Kluck (2011). *Sentence amalgamation.*
94. Anja Schüppert (2011). *Origin of asymmetry: Mutual intelligibility of spoken Danish and Swedish.*
95. Peter Nabende (2011). *Applying Dynamic Bayesian Networks in Transliteration Detection and Generation.*
96. Barbara Plank (2011). *Domain Adaptation for Parsing.*
97. Çağrı Çöltekin (2011). *Catching Words in a Stream of Speech: Computational simulations of segmenting transcribed child-directed speech.*
98. Dörte Hessler (2011). *Audiovisual Processing in Aphasic and Non-Brain-Damaged Listeners: The Whole is More than the Sum of its Parts.* Herman Heringa (2012). *Appositional constructions.*
99. Herman Heringa (2012). *Appositional constructions.*

100. Diana Dimitrova (2012). *Neural Correlates of Prosody and Information Structure.*
101. Harwintha Anjarningsih (2012). *Time Reference in Standard Indonesian Agrammatic Aphasia.*
102. Myrte Gosen (2012). *Tracing learning in interaction. An analysis of shared reading of picture books at kindergarten.*
103. Martijn Wieling (2012). *A Quantitative Approach to Social and Geographical Dialect Variation.*
104. Gisi Cannizzaro (2012). *Early word order and animacy.*
105. Kostadin Cholakov (2012). *Lexical Acquisition for Computational Grammars. A Unified Model.*
106. Karin Beijering (2012). *Expressions of epistemic modality in Mainland Scandinavian. A study into the lexicalization-grammaticalization-pragmaticalization interface.*
107. Veerle Baaijen (2012). *The development of understanding through writing.*
108. Jacolien van Rij (2012). *Pronoun processing: Computational, behavioral, and psychophysiological studies in children and adults.*
109. Ankelien Schippers (2012). *Variation and change in Germanic long-distance dependencies.*
110. Hanneke Loerts (2012). *Uncommon gender: Eyes and brains, native and second language learners, & grammatical gender.*
111. Marjoleine Sloos (2013). *Frequency and phonological grammar: An integrated approach. Evidence from German, Indonesian, and Japanese.*
112. Aysa Arylova (2013). *Possession in the Russian clause. Towards dynamicity in syntax.*
113. Daniël de Kok (2013). *Reversible Stochastic Attribute-Value Grammars.*
114. Gideon Kotzé (2013). *Complementary approaches to tree alignment: Combining statistical and rule-based methods.*
115. Fridah Katushemererwe (2013). *Computational Morphology and Bantu Language Learning: an Implementation for Runyakitara.*
116. Ryan C. Taylor (2013). *Tracking Referents: Markedness, World Knowledge and Pronoun Resolution.*
117. Hana Smiskova-Gustafsson (2013). *Chunks in L2 Development: A Usage-Based Perspective.*
118. Milada Walková (2013). *The aspectual function of particles in phrasal verbs.*

119. Tom O. Abuom (2013). *Verb and Word Order Deficits in Swahili-English bilingual agrammatic speakers.*
120. Gülsen Yılmaz (2013). *Bilingual Language Development among the First Generation Turkish Immigrants in the Netherlands.*
121. Trevor Benjamin (2013). *Signaling Trouble: On the linguistic design of other-initiation of repair in English Conversation.*
122. Nguyen Hong Thi Phuong (2013). *A Dynamic Usage-based Approach to Second Language Teaching.*
123. Harm Brouwer (2014). *The Electrophysiology of Language Comprehension: A Neurocomputational Model.*

GRODIL
Secretary of the Department of General Linguistics
Postbus 716
9700 AS Groningen
The Netherlands