Eye Movements and Brain Responses in Natural Reading

Paul-Philipp Metzner

Doctoral Thesis

Submitted to the Faculty of Human Sciences
at the University of Potsdam in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Psychology.

University of Potsdam
2015

Supervisors:
Prof. Dr. Frank Rösler
Prof. Dr. Shravan Vasishth

Submitted March 26, 2015
Accepted April 15, 2015
Defended October 9, 2015

Reviewers:
Prof. Dr. Frank Rösler
Prof. Dr. Pienie Zwitserlood

Published online at the
Institutional Repository of the University of Potsdam:
URN urn:nbn:de:kobv:517-opus4-82806
http://nbn-resolving.de/urn:nbn:de:kobv:517-opus4-82806

To my parents.

Abstract

Intuitively, it is clear that neural processes and eye movements in reading are closely connected, but only few studies have investigated both signals simultaneously. Instead, the usual approach is to record them in separate experiments and to subsequently consolidate the results. However, studies using this approach have shown that it is feasible to coregister eye movements and EEG in natural reading and contributed greatly to the understanding of oculomotor processes in reading. The present thesis builds upon that work, assessing to what extent coregistration can be helpful for sentence processing research.

In the first study, we explore how well coregistration is suited to study subtle effects common to psycholinguistic experiments by investigating the effect of distance on dependency resolution. The results demonstrate that researchers must improve the signal-to-noise ratio to uncover more subdued effects in coregistration. In the second study, we compare oscillatory responses in different presentation modes. Using robust effects from world knowledge violations, we show that the generation and retrieval of memory traces may differ between natural reading and word-by-word presentation. In the third study, we bridge the gap between our knowledge of behavioral and neural responses to integration difficulties in reading by analyzing the EEG in the context of regressive saccades. We find the P600, a neural indicator of recovery processes, when readers make a regressive saccade in response to integration difficulties.

The results in the present thesis demonstrate that coregistration can be a useful tool for the study of sentence processing. However, they also show that it may not be suitable for some questions, especially if they involve subtle effects.

Keywords: natural reading, sentence processing, eye tracking, EEG, coregistration, dependency resolution, neural oscillations, regressive saccades

Acknowledgments

First of all, I would like to thank my supervisor, Frank Rösler. He taught me valuable lessons in patience, attention, and rigor. Thank you, Frank, for sharing your knowledge and experience with me. I would also like to thank Shravan Vasishth, my second supervisor, who invited me to Potsdam to pursue a PhD in his laboratory. He supported me in every endeavor but was never shy to voice his concerns when I got carried away.

Many other colleagues and friends affected me as a person and a researcher, and thereby, this thesis. Titus von der Malsburg was not formally a supervisor, but he had such a large effect on me and the thesis that I would like to think of him as my third supervisor. Furthermore, I would like to name Reinhold Kliegl, Felix Engelmann, Olaf Dimigen, Nicole Gotzner, Pavel Logaçev, and Lena Jäger. Through comments, questions, and criticism, they challenged my perspectives and made me think again.

Throughout the last four years, my wife Jule and my children Oskar and Benno were the safe haven that kept me in balance. Thank you.

Contents

1 Introduction **1**

1.1 Brain Responses 2

1.1.1 Event-Related Potentials 5

1.1.2 Time-Frequency Representations 8

1.2 Eye Movements 10

1.2.1 Eye Movements during Sentence Comprehension 12

1.3 Coregistration 15

1.3.1 Technical Challenges 17

1.3.1.1 Ocular Artifacts 17

1.3.1.2 Overlap 19

1.3.1.3 Setup 20

1.3.1.4 Synchronizing Timelines 22

1.3.1.5 Analysis 22

1.4 Open Questions 24

1.4.1 Sensitivity 24

1.4.2 Neural Oscillations 26

1.4.3 Regressive Saccades 26

2 Investigating Expectation and Locality with Concurrent Eye Movement and EEG Recordings **29**

2.1 Experiment 1 36

2.1.1 Methods 36

2.1.1.1 Participants 36

2.1.1.2 Materials 36

2.1.1.3 Procedure 38

2.1.1.4 Recording 39
2.1.1.5 Analysis 41
2.1.2 Results 42
2.1.2.1 Comprehension 42
2.1.2.2 Eye Movements 43
2.1.2.3 ERP 45
2.1.3 Discussion 47
2.2 Experiment 2 49
2.2.1 Methods 49
2.2.1.1 Participants 49
2.2.1.2 Materials 49
2.2.1.3 Procedure 50
2.2.1.4 Recording and Analysis 51
2.2.2 Results 52
2.2.2.1 Comprehension 52
2.2.2.2 Eye Movements 53
2.2.2.3 ERP 55
2.2.3 Discussion 55
2.3 Experiment 3 57
2.3.1 Methods 58
2.3.1.1 Participants 58
2.3.1.2 Materials 58
2.3.1.3 Procedure 58
2.3.1.4 Analysis 58
2.3.2 Results 59
2.3.3 Discussion 59
2.4 General Discussion 59

3 Brain Responses to World-Knowledge Violations: A Comparison of Stimulus- and Fixation-Triggered Event-Related Potentials and Neural Oscillations 69

3.1 Experiment 1 73

3.1.1 Methods . . . 73
3.1.1.1 Participants . . . 73
3.1.1.2 Materials . . . 73
3.1.1.3 Procedure . . . 74
3.1.2 Recording and Analysis . . . 74
3.1.2.1 Recording . . . 74
3.1.2.2 Preprocessing . . . 75
3.1.3 Results . . . 77
3.1.3.1 Behavioral Data . . . 77
3.1.3.2 ERP . . . 77
3.1.3.3 TFR . . . 77
3.1.4 Discussion . . . 78
3.2 Experiment 2 . . . 79
3.2.1 Methods . . . 79
3.2.1.1 Participants . . . 79
3.2.1.2 Materials . . . 80
3.2.1.3 Procedure . . . 80
3.2.2 Recording and Analysis . . . 81
3.2.2.1 Eye movements . . . 81
3.2.2.2 EEG . . . 82
3.2.3 Results . . . 83
3.2.3.1 Behavioral Data . . . 83
3.2.3.2 Eye Movements . . . 83
3.2.3.3 ERP . . . 84
3.2.3.4 TFR . . . 85
3.2.3.5 Delta Range (1–3 Hz) . . . 86
3.2.3.6 Upper Alpha Range (11–13 Hz) . . . 86
3.2.3.7 Single-Frequency Analysis . . . 87
3.2.4 Discussion . . . 87
3.3 General Discussion . . . 88

4 The Importance of Reading Naturally: Evidence from Combined Recordings of Eye Movements and Electric Brain Potentials 97

4.1 Materials and Methods 103
4.1.1 Participants 103
4.1.2 Design 104
4.1.3 Apparatus 105
4.1.4 Procedure 105
4.1.5 Data Preprocessing 107
4.1.6 Analysis 108
4.2 Results 110
4.2.1 Judgment Accuracy 110
4.2.2 Eye Tracking Data 111
4.2.3 Event-Related Potentials 115
4.2.3.1 Word-by-Word Presentation 115
4.2.3.2 Natural Reading 116
4.2.4 Regression-Contingent Analysis 116
4.2.4.1 Regression Trials 118
4.2.4.2 No-Regression Trials 118
4.3 Discussion 118

5 Summary and Conclusions 125

5.1 Subtle Effects in Coregistration 125
5.2 Neural Oscillations in Natural Reading 126
5.3 Regressive Saccades and Recovery 128
5.4 Outlook 129

References 133

List of Tables

2.1 Sample sentences from Santi and Grodzinsky (2007). 34
2.2 Sample sentences from Matchin, Sprouse, and Hickok (2014). 35
2.3 Sample sentences from Experiment 1 with object-verb dependency. . . 38
2.4 Sample sentences from Experiment 1 with antecedent-pronoun dependency. 39
2.5 Summary statistics for first fixation duration in Exp. 1 44
2.6 Summary statistics for gaze duration in Exp. 1 44
2.7 Summary statistics for regression probability in Exp. 1 45
2.8 Sample sentences from Experiment 2 with object-verb dependency. . . 50
2.9 Sample sentences from Experiment 2 with antecedent-pronoun dependency. 51
2.10 Summary statistics for first fixation duration in Exp. 2 53
2.11 Summary statistics for gaze duration in Exp. 2 54
2.12 Summary statistics for regression probability in Exp. 2 54

4.1 Sample set of sentences with English translation 105
4.2 Summary statistics for response accuracy (presentation mode) 111
4.3 Summary statistics for response accuracy (regressions) 113
4.4 Summary statistics for eye movement measures 114
4.5 Summary of ERP results 115

List of Figures

1.1 Sample trial sequence in RSVP 4
1.2 Syntax trees with minimal and non-minimal attachment 13
1.3 Sample trial sequence in natural reading 16
1.4 Schematic diagram of experimental setup 21

2.1 ERPs from Exp. 1 before and after artifact correction 41
2.2 Eye movement measures in Exp. 1 46
2.3 ERPs for object-verb dependencies in Exp. 1 47
2.4 ERPs for antecedent-pronoun dependencies in Exp. 1 48
2.5 ERPs from Exp. 2 before and after artifact correction 52
2.6 Eye movement measures in Exp. 2 55
2.7 ERPs for object-verb dependencies in Exp. 2 56
2.8 ERPs for antecedent-pronoun dependencies in Exp. 2 57
2.9 Reading times in Exp. 3 60
2.10 Eye movement and reading measures (object-verb dependencies) . . . 61
2.11 Eye movement and reading measures (antecedent-pronoun dependencies) 63

3.1 ERPs from Exp. 1 78
3.2 Time-frequency plot and topographies for Exp. 1 79
3.3 ERPs from Exp. 2 before and after artifact correction 83
3.4 Eye movement measures from Exp. 2 84
3.5 ERPs from Exp. 2 85
3.6 Time-frequency plot and topographies for Exp. 2 (Delta) 86
3.7 Time-frequency plot and topographies for Exp. 2 (Alpha) 87

4.1 ERPs before and after artifact correction 108
4.2 Accuracy in the judgment task 112
4.3 First fixation duration, gaze duration, and regression probability 113
4.4 ERPs and topographies from natural reading sessions 117
4.5 ERPs and topographies from word-by-word sessions 119

Abbreviations

µV	Microvolts
dB	Decibel
Hz	Hertz
NP	Noun phrase
PP	Prepositional phrase
M	Mean
SD	Standard deviation
SE	Standard error
CI	Confidence interval
ICA	Independent Component Analysis
PCA	Principal Component Analysis
EEG	Electroencephalogram
EOG	Electrooculogram
ERP	Event-related potential
RSVP	Rapid serial visual presentation
ISI	Inter-stimulus interval
TFR	Time-frequency representation
FFT	Fast Fourier Transform
STFT	Short-term Fourier Transform

1

Introduction

Arguably, the ultimate goal of psycholinguistics is to arrive at a comprehensive description of how the brain acquires, processes, and produces language. Unfortunately, and much to our frustration, this cannot be observed directly in the brain. In other words, dissecting the brain will tell us something about its anatomy but not how it manages to transform unstructured auditory or visual input into a mental representation. As a consequence, we have to treat the brain as a black box and infer its architecture and mechanisms from input and output. In psycholinguistics, the input is typically a linguistic unit like a sound, a word, or a sentence. What constitutes the output is not entirely clear and depends on the research question. It may be how participants respond to the input, how long it takes them to respond, or what they look at while they process the input. Crucially, when we confront participants with two variants of an input that differ in exactly one aspect, we can compare the respective responses and draw inferences about what happened between input and output.

To collect such data, researchers have developed a diverse experimental toolbox. Naturally, all methods have benefits and drawbacks and they vary in terms of complexity and ecological validity. For instance, a large body of experimental studies is based on a method that requires participants to read something on a computer display and then give a response (e.g., with a button press). Such an experimental configuration is relatively straightforward and easy to implement. However, the response collection is also quite detached from natural reading. Additionally, the dependent variable in such studies is not derived from the activity of interest (i.e., reading) but from a more or less closely related task. Other methods may have higher ecological validity, but that often comes at the expense of higher complexity or the need for expensive equipment. A consequence of this methodological diversity is that it is becoming increasingly difficult to consolidate the results from different studies.

Eye tracking and EEG are among the most frequently used methods in psycholinguistic research. Both tap cognitive processes in reading, and derived measures have been found to be influenced by a number of linguist manipulations. The two methods have historically been used separately, complementing each other in an indirect fashion. Researchers have recently started to combine eye tracking and EEG, overcoming their respective limitations (e.g., Dimigen, Kliegl, & Sommer, 2012; Dimigen, Sommer, Hohlfeld, Jacobs, & Kliegl, 2011; Simola, Holmqvist, & Lindgren, 2009), but it is not clear how useful their combination (henceforth, *co-registration*) is for the study of human sentence processing. To answer this question, the present thesis investigates the pitfalls and potential of coregistration for sentence processing research.

The remainder of this introduction is supposed to give a brief overview of how brain responses and eye movements are recorded, how they are used in psycholinguistics, and what their respective strengths and weaknesses are. A dedicated section covers the coregistration of both methods with a focus on technical challenges. Finally, the studies in Chapters 2 to 4 will be motivated and summarized.

1.1 Brain Responses

Measuring the EEG allows researchers to non-invasively investigate neuronal activity, which is possible because neurons generate small electrical charges: action potentials and post-synaptic potentials. When large assemblies of neurons fire simultaneously, their post-synaptic potentials add up and propagate via volume conduction to the scalp surface where they can be measured with electrodes (Luck, 2005). A problematic property of the EEG is that the potential from a cortical generator does not flow directly to the surface but spreads out, particularly as the current hits the large resistance of the skull. As a result, EEG has relatively poor spatial resolution.

Crucially, cognitive processes are reflected in the recorded EEG, such that it can be used as a dependent variable in psychological experiments. The most important aspect for this approach is the temporal relation between an experimental manipulation and the voltage changes in the recorded EEG that are investigated and interpreted. By contrast, the location of an effect is secondary. In part, this is owing to the fact that the EEG has poor spatial resolution. Either way, a presupposition for interpreting the recorded EEG is that the experimental manipulation, the observed effect, and the hypothesized underlying process are closely and functionally related.

There are two major problems with this approach. First, the recorded EEG has a low signal-to-noise ratio, that is, the signal induced by the experimental manipulation is superimposed by signals from other generators. One solution for this problem is to compute and analyze *event-related potentials* (ERPs, see Section 1.1.1).

Second, the volume conduction in the skull and on the scalp has another consequence besides poor spatial resolution: The large deflections in the recorded EEG that are caused by eye movements are captured not only by electrodes right next to the eyes but by electrodes on the entire scalp (see Section 1.3.1). Most researchers have therefore resorted to a presentation format that does not require participants to move their eyes. In *rapid serial visual presentation* (RSVP), sentences are presented word by word at a fixed pace in the center of the display (see Figure 1.1). Blank displays after each slide reduce the overlap of successive potentials and allow readers to prepare for the next word.

Of course, such a procedure is different from natural reading in several respects. First, reading speed is not under the reader's control but fixed. Additionally, RSVP is typically much slower than in natural reading. At around 600 ms, the time between the presentation of two successive words is often more than twice as long as in normal reading. Second, readers are prevented from rereading earlier parts of the sentence. This has several implications. In the event of processing difficulties, readers may adopt a different strategy than in normal reading, where they can move their eyes back to the preceding words. Moreover, even in the absence of processing difficulties, readers may build and maintain a more elaborate mental representation of the sentence. Third, readers cannot skip words which they frequently do in natural reading (see Section 1.2). Finally, readers have no preview of the next word when words are presented in isolation.

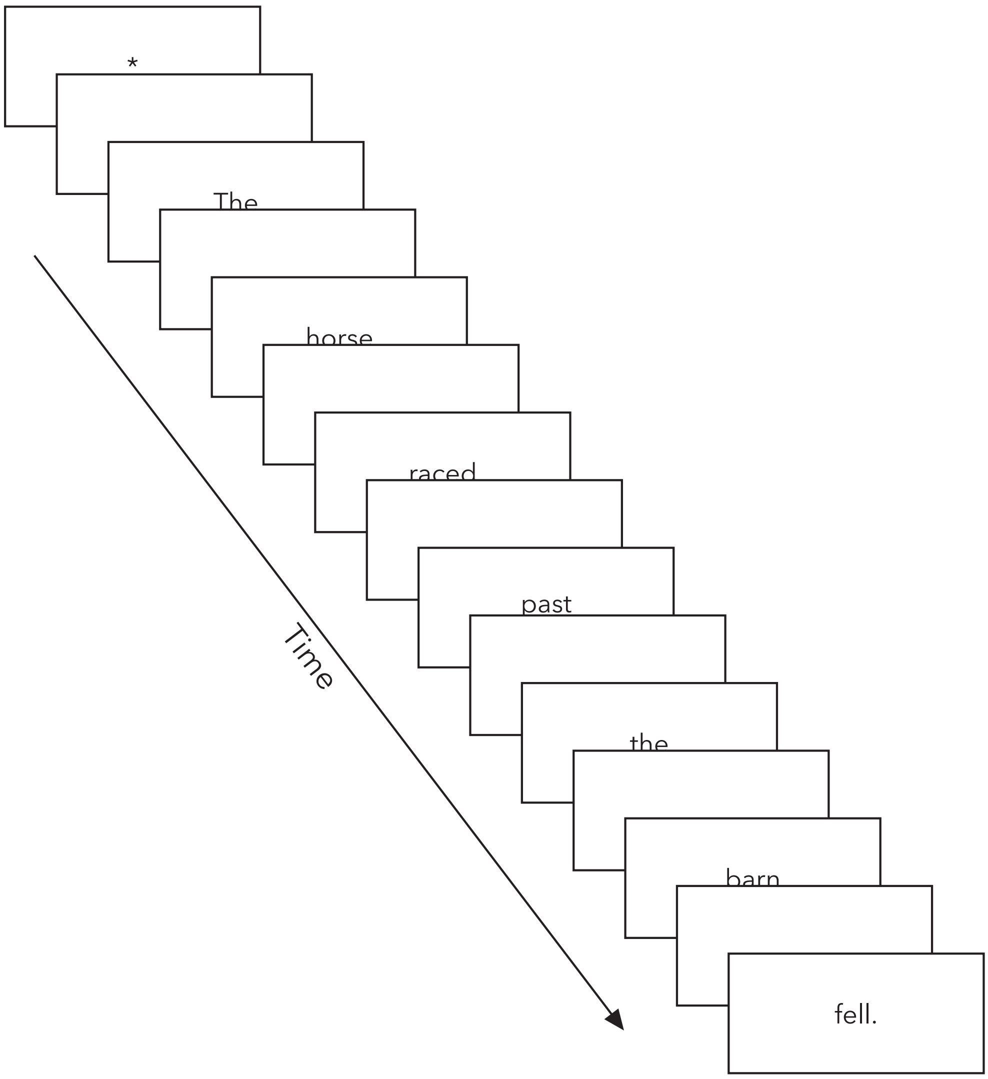

Figure 1.1: Sample trial sequence illustrating stimulus presentation in RSVP, exemplified with a garden-path sentence from Frazier (1978): *The horse raced past the barn fell.*

1.1.1 Event-Related Potentials

In its raw form, the EEG is heavily contaminated by noise, which includes not only the above-mentioned artifacts from eye movements. Besides myogenic artifacts and electromagnetic interference, there is also genuine brain activity that is simply unrelated to the experimental task. It is therefore difficult to gain insights about cognitive processes from unprocessed EEG. *Event-related potentials* (ERPs) are a widely used approach to bringing the EEG into a more manageable format. The logic behind ERPs is that the brain's response to an experimental manipulation should be present whenever participants are exposed to it. In the average of many such expositions, the true signal should therefore emerge, whereas any activity that is not systematically related to the manipulation should cancel out under the assumption that the amplitudes of the background signal not temporally related to the experimenter-controlled event fluctuate randomly around zero. To obtain ERPs, the continuous EEG is segmented relative to certain processing events. In psycholinguistic settings, this is usually the presentation of a word. The result of averaging across these segments is a waveform with a characteristic sequence of deflections. These *evoked potentials* are described and categorized in terms of their polarity (positive- or negative-going), latencies (onset, offset, peak), and scalp topographies.

ERPs have several advantages over other online measures of language processing. First, they do not require participants to engage in an additional task. In contrast, in studies measuring reaction times, participants have to respond to a probe or answer a question (e.g., by pressing a button). Second, ERPs allow the moment-to-moment monitoring of cognitive processes with millisecond precision where behavioral measures like reaction or reading times (e.g., from self-paced reading or eye tracking studies) are assumed to reflect only the time to reach the end state of some mental computation. Third, the interplay of several properties (latencies, polarity, topography) allow the qualitative comparison of different effects, which makes it possible to disentangle effects that are indistinguishable on the basis of reading or reaction times alone (Otten & Rugg, 2005; Rugg & Coles, 1995).

Studies using ERPs have yielded valuable insights into cognitive processes in general and language processing in particular. In a seminal study, Kutas and Hillyard (1980) adopted the oddball paradigm to language processing and had participants read simple seven-word sentences. In the original oddball paradigm, participants were exposed to sequences of soft or loud tones with either one having a higher likelihood of occurrence (Squires, Squires, & Hillyard, 1975). In response to tones with a lower likelihood, Squires et al. observed a pronounced positive peak around 300 ms following its onset: the P300 (or P3a/Pb). By contrast, Kutas and Hillyard observed that unexpected words in final position (e.g., *socks* in *He spread the warm bread with socks* vs. *shoes* in *She put on her high-heeled shoes*) elicited a negative deflection between 200 and 500 ms with a peak around 400 ms. The N400 ("N" for "negative" and "400" for its peak latency) has since been replicated in numerous settings but eludes a clear functional description (for reviews, see Kutas & Federmeier, 2011; Lau, Phillips, & Poeppel, 2008). Regardless of whether the N400 indicates semantic violations or non-compliance with expectations, it is clear that it responds to linguistic manipulations. The N400 has been used as an index of processing difficulties to investigate, inter alia, ambiguity resolution (Hagoort & Brown, 1994), the relationship of semantic and world knowledge (Hagoort, Hald, Bastiaansen, & Petersson, 2004), and the influence of discourse (Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005). Finally, the N400 has also been found to pattern with readers' fixation durations, further adding to its credibility (Dambacher & Kliegl, 2007).

Osterhout and Holcomb (1992) compared the brain responses to the word *to* in *The woman struggled to…* and *The woman persuaded to….* Assuming that, in the spirit of Frazier's (1978) principle *late closure*, readers adopt an active reading of *persuaded*, encountering *to* requires a reorganization of the syntactic tree. In these cases, Osterhout and Holcomb observed the P600, a positive deflection at posterior electrodes with an onset around 600 ms. Hagoort, Brown, and Groothusen (1993) observed a similar positivity in response to subject-verb number agreement violations, the *syntactic positive shift* (SPS). Following these two studies, a volley of consistent findings helped setting up the P600 (or SPS) as the syntactic counterpart of the N400 (e.g. Friederici & Mecklinger, 1996; Osterhout, Bersick, &

McLaughlin, 1997; Osterhout, Holcomb, & Swinney, 1994; Osterhout & Mobley, 1995).

However, several results suggest that the P600 indicates more than reanalysis of syntactic violations. First, the P600 has also been found in well-formed sentences upon completion of long-distance dependencies (Fiebach, Schlesewsky, & Friederici, 2002; Kaan, Harris, Gibson, & Holcomb, 2000; Phillips, Kazanina, & Abada, 2005). It has therefore been proposed that the P600 may also reflect higher structural processing and integration (Gouvea, Phillips, Kazanina, & Poeppel, 2010; Hagoort, Brown, & Osterhout, 1999). Second, the amplitude of the P600 is sensitive to the probability of encountering a violation, which has led to the claim that the P600 may not be a distinct component but rather a member of the P300 family (see, e.g., Coulson, King, & Kutas, 1998; Sassenhagen, Schlesewsky, & Bornkessel-Schlesewsky, 2014). Both are late positivities with a centro-parietal distribution and occur in response to unexpected, intrusive events. Third, the P600 is reduced when syntactic violations are not task-relevant (e.g., Hahne & Friederici, 2002; Osterhout, Allen, McLaughlin, & Inoue, 2002) or when they can be attributed to the speaker's proficiency (Hanulíková, Van Alphen, Van Goch, & Weber, 2012). Finally, Kim and Osterhout (2005) found that sentences like *The hearty meal was devouring the kids* do not elicit an N400 but a P600. This may suggest that readers interpret this sentence as syntactically anomalous, restructuring it to, for example, *The hearty meal was devoured by the kids*. This finding has spawned an ongoing discussion about the "semantic P600" (see, e.g., Bornkessel-Schlesewsky & Schlesewsky, 2008; Brouwer, Fitz, & Hoeks, 2012; Kuperberg, 2007).

Syntactic processing difficulties have also been found to elicit a negativity at frontal portions of the left hemisphere (Neville, Nicol, Barss, Forster, & Garrett, 1991; Osterhout & Holcomb, 1992; Rösler, Pütz, Friederici, & Hahne, 1993). This *left-anterior negativity* (LAN; Coulson et al., 1998) has been found in response to morphosyntactic violations (e.g., verb inflection; Gunter, Stowe, & Mulder, 1997) and in situations in which determiners indicate a syntactically correct but non-canonical (less expected) continuation of a sentence (Rösler, Pechmann, Streb, Röder, & Hennighausen, 1998). Unlike the P600, the LAN is not readily reproducible across languages and experimental tasks. Molinaro, Barber, and Carreiras (2011) argue that, apart from linguistic factors, this is most likely due to the choice of reference electrodes and individual differences. Both claims, however, have recently been refuted (Tanner, 2015).

There is also an active debate about an early variant of the LAN, the *early left-anterior negativity* (ELAN), and whether it is functionally distinct from the LAN (Hahne & Friederici, 2002). It has been argued that the ELAN reflects phrase-structure building and that, assuming a syntax-first model of language processing (Frazier & Fodor, 1978; Friederici, 1995), failure to succeed at this stage blocks later effects like the N400 or LAN. In a recent paper, Steinhauer and Drury (2012) revisit the evidence for Friederici's model and doubt its veracity.

In the context of globally ambiguous linguistic references, several studies have reported sustained frontal negativities starting at around 300 ms (Nieuwland & Van Berkum, 2006, 2008; Van Berkum, Brown, & Hagoort, 1999; Van Berkum, Brown, Hagoort, & Zwitserlood, 2003). These may be related to increased effort in retrieving an antecedent since they have also been observed in gender-mismatching noun-phrase ellipsis (Martin, Nieuwland, & Carreiras, 2012).

To summarize, a number of characteristic brain responses have been identified in relation to language processing. The N400 effect seems to indicate activation of long-term memory and is observed, inter alia, in the context of semantic violations. The P600/SPS is observed when recovery or structural repair are necessary, for example due to grammatical anomalies. Like the P600, the LAN is related to structural processing. It is sensitive to deviations from canonical morphosyntactic marking.

1.1.2 Time-Frequency Representations

The most prominent feature Berger (1929) noted when he discovered that brain activity can be recorded from the unopened skull were rhythmic oscillations of different wavelengths. In his pioneering study, he observed waves of shorter (40–

50 ms, around 10.5 Hz) and longer durations (90–100 ms, 22.25 Hz). Those oscillations would today be categorized as alpha and beta waves. Neural oscillations have been found to be correlated with different states of alertness and cognitive processes, a fact which has been exploited in experimental psychology. The following paragraphs give a brief overview of the types of oscillations that are relevant for language processing.

Alpha waves are neural oscillations between 8 and 13 Hz. What distinguishes them from other oscillations is that their amplitude is reduced in the context of higher processing demands, which is why they are thought to reflect the inhibition of cortical areas that are not relevant for the current task (Klimesch, Sauseng, & Hanslmayr, 2007). Alpha oscillations have been found to be sensitive to memory demands (Başar, Başar-Eroglu, Karakaş, & Schürmann, 2001; Klimesch, 1999, 2012), making them highly relevant for language processing (see below). Another frequency range that is related to memory and language is the theta range between 4 and 7 Hz (Gevins, Smith, McEvoy, & Yu, 1997). Unlike alpha oscillations, theta oscillations increase in magnitude in response to elevated processing demands. Finally, rapid gamma oscillations above 30 Hz have also been proposed to play a role in higher cognition (Herrmann, Munk, & Engel, 2004; Müller, Gruber, & Keil, 2000; Tallon-Baudry & Bertrand, 1999).

There are several approaches for determining the periodic content of the convoluted wave that is measured in the EEG. The *Fast Fourier Transform* (FFT; Cooley & Tukey, 1965) is an established procedure for spectral decomposition. To yield a time-frequency representation (TFR, i.e., to derive the change in spectral composition over time), a *short-term Fourier Transform* (STFT) can be used. Alternatively, the spectral content and its change over time can be estimated using wavelet analysis (Schiff, Aldroubi, Unser, & Sato, 1994) or multitapers (Mitra & Pesaran, 1999). Regardless of how the power spectrum is derived, it must be compared to a baseline to draw reliable conclusions about stimulus-induced changes.

While ERPs are the dominant method in language-related EEG research, a number of valuable insights have been obtained with TFRs (for a more comprehensive review, see Bastiaansen & Hagoort, 2006). For example, Bastiaansen, Van

Berkum, and Hagoort (2002a) investigated changes in the alpha and theta band in online sentence processing. They observed a power decrease in the alpha band and a power increase in the theta band at each word and a slow power increase in the theta band over the course of a sentence. Further investigations found that the theta band is involved in lexical access (Bastiaansen, Oostenveld, Jensen, & Hagoort, 2008, 2005; Maguire, Brier, & Ferree, 2010) and sensitive to syntactic and semantic manipulations (Bastiaansen, Van Berkum, & Hagoort, 2002b; Hald, Bastiaansen, & Hagoort, 2006). The gamma band may also be involved in language processing (Wang, Zhu, & Bastiaansen, 2012). However, perhaps due to its wide frequency range from 30 to 100 Hz, the results are inconclusive or contradictory to results involving other frequency bands (Penolazzi, Angrilli, & Job, 2009).

In sum, the relationship between psycholinguistic manipulations and effects is less clear in TFRs than in ERPs. However, similar frequency ranges respond to language processing and memory-demanding tasks, including the theta and alpha range. The role of the gamma band for cognition in general and in language processing in particular is still subject to debate, but effects in this frequency range are also reported in studies on language processing.

1.2 Eye Movements

As mentioned above, the location of a reader's gaze is used as a dependent variable in psychological experiments. There are several technologies for determining the gaze location, and video-based systems are the most common type in psycholinguistic applications (Rayner, 1998). In these systems, either one (monocular) or both eyes (binocular) are illuminated by an infrared light source to record an image of the pupil and corneal light reflexes. Their relation to gaze locations can be derived after participants have completed a calibration procedure that requires them to fixate points in a grid of known locations. Once this procedure is completed, gaze locations between those points can be interpolated. Although modern video-based systems allow a temporal resolution of up to 2000 Hz, this is rarely required in psycholinguistic studies. The studies reported in the present thesis used a sampling rate of 500 Hz.

The rationale underlying reading studies using eye tracking is that there is a more or less tight coupling of eye movements and psycholinguistic manipulations (Just & Carpenter, 1980; Rayner, 1998; Staub & Rayner, 2007). This assumption is motivated by the way the eyes move during reading. Counter to the subjective reading experience, the eyes do not move smoothly across the text. Rather, they perform short and ballistic movements called *saccades*. During saccades, which typically last between 30 and 50 ms, information uptake is prevented or drastically reduced. Between saccades, the eyes rest for on average 250 ms (Rayner, 1998). During these *fixations*, readers process the visual input. One of the reasons why readers have to move their eyes in reading is that sharp vision can only be attained from a small part of the retina, the *fovea*. It is a matter of debate how much information can be extracted from the *parafovea* that surrounds the fovea and delivers less acute images (for a review, see Schotter, Angele, & Rayner, 2012). Readers also do not necessarily always move their eyes forward in natural reading. On average, 10 to 15% of the saccades are directed backwards. Most of these *regressive saccades* are only a few letters long and probably intended to correct a forward saccade that was too long.

Crucially for psycholinguistic applications, it has been demonstrated that eye movements are affected by low- and high-level factors. For example, word frequency and predictability influence fixation durations (Rayner, 1998) and it has been found that readers make more regressive saccades in response to unexpected or challenging input (Frazier & Rayner, 1982).

Instead of investigating raw gaze locations, researchers have developed a number of dependent variables. The probably most widely-used variable in reading experiments is the duration of the first fixation on a word (*first fixation duration*). For this measure, typically only trials are taken into account where the target word has not been skipped. A variant of first fixation duration is *single fixation duration*, comprising only trials where the target word was not *refixated* (i.e., the first fixation was also the only fixation before leaving the word). Adding up the duration of the first fixation and all refixations during the first pass yields the *gaze duration*. The sum of all fixation durations from sentence onset until a reader leaves the target word to the right are called *regression path duration*. The percentage of trials in which readers make a regression towards earlier material after entering a word is the *regression probability*. Sometimes, readers revisit a word after they already read and left it. The sum of fixation durations during such revisitations is called *rereading time* and the probability with which readers revisit a word is the *rereading probability*.

Eye movements can be analyzed in other meaningful ways, but those are far less common in psycholinguistics. For example, it is fairly common to analyze the landing site of the eyes within a word in oculomotor research. Such analyses are almost entirely absent in the psycholinguistic literature (but see Yan et al., 2014). Also, aggregating reading times at single words is a stark simplification of readers' eye movements, capturing only a fragment of the variance. A method that incorporates both the sequence and duration of fixations is scanpath analysis (von der Malsburg & Vasishth, 2011).

1.2.1 Eye Movements during Sentence Comprehension

Psycholinguistic research using eye tracking far exceeds the scope of this thesis (for reviews, see Clifton, Staub, & Rayner, 2007; Vasishth, von der Malsburg, & Engelmann, 2012). By means of an exemplary psycholinguistic issue, the following section should demonstrate how the study of eye movements contributes to our understanding of online sentence processing and also how their peculiarities must be taken into account.

Some models of sentence comprehension assume that the human language processor (henceforth, *parser*) uses only syntactic information in the first processing stages. In one instantiation of such syntax-first models, Frazier and Fodor (1978) proposed that parsing is initially dictated by a principle called *minimal attachment* (see also Frazier, 1978). According to minimal attachment, the parser initially favors the simplest syntactic integration of new material. In (1a), readers have to decide where to attach the prepositional phrase (PP) *with a stick*. According to minimal attachment, they will interpret the PP as a complement to the verb *poke*. The reason is that attaching the PP to the noun phrase (NP) *the boy* would require the construction of a new node for a complex NP that branches into *the boy* and *with a stick* (see Figure 1.2).

(1) a. John poked the boy with a stick.
    b. John poked the boy with a lisp.

Due to semantic properties of the noun *lisp*, the PP *with a lisp* is eventually not attached to the VP in (1b). The question is whether readers use this information immediately or only in later processing stages. In other words, with reference to Figure 1.2: Do they first build (A) and replace it with (B) or do they immediately build (B), using the semantic information conveyed by the word *lisp*?

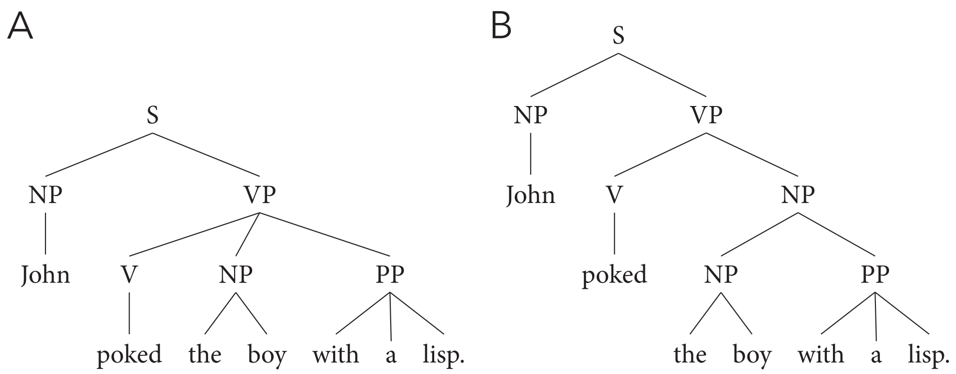

Figure 1.2: Syntax trees representing minimal (A) and non-minimal (B) attachment parses of *John poked the boy with a lisp*.

Rayner, Carlson, and Frazier (1983) used eye tracking to investigate this question. They constructed materials like (2) below where the semantics of the main clause subject either favored a main clause or reduced relative clause interpretation of the following verb.

(2) a. The florist sent the flowers was very pleased.
    b. The performer sent the flowers was very pleased.

If readers use semantic cues for their initial syntactic parse, we should observe stronger signs of processing difficulties at *was* in (2a) than in (2b) because performers are more likely to be sent flowers than florists, and vice versa. Rayner et al. found that semantic information did not affect the initial parse in sentences like (2) but that it was indeed used for the ultimate analysis of sentences like (1). In a follow-up study, Ferreira and Clifton (1986) used verbs like *examine* that place a strict animacy requirement on the subject such that *The evidence examined…* cannot continue with an NP but requires a subject PP like *…by the lawyer*. If readers use such strong semantic constraints in the initial syntactic analysis, they should be garden-pathed at *by* in *The defendant examined by the lawyer…* but not in *The evidence examined by the lawyer…*. Ferreira and Clifton found no such pattern, supporting the claim that semantic constraints do not affect the initial parse.

Trueswell, Tanenhaus, and Garnsey (1994) repeated this study and altered some aspects of the materials. First, they argue that many sentences with inanimate nouns in Ferreira and Clifton's (1986) materials (i.e., *The evidence…*) could actually have been continued as a main clause. Second, Ferreira and Clifton used unreduced relative clauses like *The evidence that was examined…* as a baseline. Drawing upon research on eye movements in reading, Trueswell et al. note that short function words are often skipped. As a consequence, readers may have fixated different parts of the reduced and unreduced relative clauses. Finally, Ferreira and Clifton forced a line break after the main clause verb (*examined*) in both reduced and unreduced sentences, leading to a premature end of line in reduced sentences. Again with reference to eye movement research, Trueswell et al. object that early line breaks can lead to inflated reading times. Consequently, they adjusted these potential flaws and found an immediate influence of animacy on syntactic ambiguity resolution. Participants had less trouble with sentences like *The evidence examined by…* than with sentences like *The defendant examined by…*

The discussion about the time course of syntactic and semantic processing did not end with Trueswell et al. (1994). From the few studies presented in this section, however, it should have become clear that reading is highly sensitive to text properties. To adapt to changing processing demands, readers adjust their reading speed, move their eyes back and read parts of the text again, and skip some words entirely. These facts can be exploited to investigate online sentence processing. However, it is important to note that such studies have to be designed carefully, taking into account how readers move their eyes in reading.

In sum, researchers have garnered important results on language processing using eye tracking. Fixation durations and skipping rates have been shown, inter alia, to vary as a function of word frequency (e.g., Just & Carpenter, 1980; Rayner, 1977; Rayner & Duffy, 1986; Rayner & Raney, 1996) and predictability (Ehrlich & Rayner, 1981). Furthermore, it has been shown repeatedly that fixation durations are sensitive to syntactic manipulations (Pickering & Traxler, 2001; Rayner et al., 1983; Traxler, Pickering, & Clifton, 1998) and that the propensity to make regressive saccades increases in the context of processing difficulties (Ferreira & Clifton, 1986; Frazier & Rayner, 1982; Meseguer, Carreiras, & Clifton, 2002; von der Malsburg & Vasishth, 2013).

1.3 Coregistration

Combining eye movement measurements and ERPs is an obvious way to enjoy the benefits of both methods (for reviews, see Baccino, 2011; Dimigen, 2014; Raney & Rayner, 1993; Sereno & Rayner, 2003). There are two principal ways to go about this (see Kliegl, Dambacher, Dimigen, Jacobs, & Sommer, 2011). One way — arguably the simpler approach — is to record eye movements and EEG in separate sessions. For example, Sereno, Rayner, and Posner (1998) used this approach to determine the time course of lexical access in reading. In a similar study, Dambacher and Kliegl (2007) related fixation durations to the amplitude of the N400, drawing inferences about the lagged word frequency effect. While such an approach makes it possible to relate results from the two methods to each other more directly, it does not solve the problem that readers are prevented to behave normally in RSVP. That is, they can neither spend more time on a word nor skip or regress.

Another way of combining eye tracking and EEG is to record both simultane-

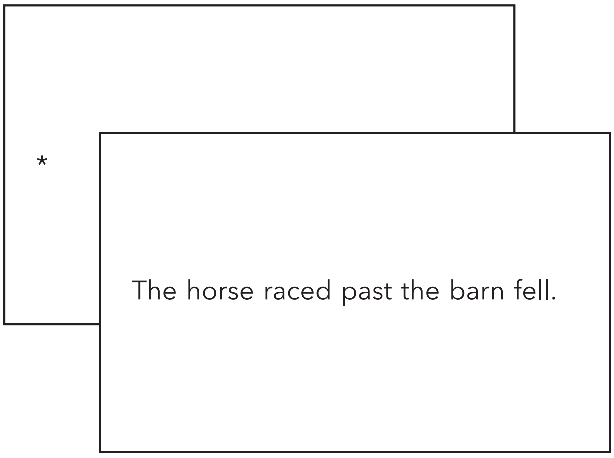

Figure 1.3: Sample trial sequence illustrating stimulus presentation in natural reading, exemplified with a garden-path sentence from Frazier (1978): *The horse raced past the barn fell.*

ously while participants read sentences naturally. Recording the EEG in natural reading entails a number of technical challenges which will be discussed in 1.3.1 below. For sentence processing research, coregistration promises electrophysiological results from much more natural reading situations than RSVP (compare Figures 1.1 and 1.3).

Recent technological advances have made it much easier to conduct such studies, which resulted in a surge of studies using coregistration (e.g., Kamienkowski, Ison, Quiroga, & Sigman, 2012; Kaunitz et al., 2014; Meyberg, Werkle-Bergner, Sommer, & Dimigen, 2015; Simola, Torniainen, Moisala, Kivikangas, & Krause, 2013). However, recording the EEG in natural reading is not an entirely novel method. The earliest investigations used the EOG to determine the onset of saccadic eye movements and were mostly geared towards the neural underpinnings of saccadic eye movements. To date, research on oculomotor processes comprises the bulk of coregistration studies (Bodis-Wollner et al., 2002; Dandekar, Privitera, Carney, & Klein, 2012; Dimigen, Valsecchi, Sommer, & Kliegl, 2009; Graupner, Velichkovsky, Pannasch, & Marx, 2007; Herdman & Ryan, 2007; Kennett, Van Velzen, Eimer, & Driver, 2007; Kurtzberg & Vaughan, 1982; Rämä & Baccino, 2010; Thickbroom, Knezevič, Carroll, & Mastaglia, 1991; Thickbroom & Mastaglia, 1986; Valsecchi, Dimigen, Kliegl, Sommer, & Turatto, 2009; Yuval-Greenberg, Tomer, Keren, Nelken, & Deouell, 2008).

Another prominent topic among studies using coregistration are word recognition and lexical access in reading. Baccino and Manunta (2005) used coregistration to investigate parafoveal-on-foveal effects in reading and found early effects of semantic relatedness between the currently fixated word and a word in the parafovea. Simola et al. (2009) reported similar effects of semantic relatedness that were stronger in the right visual field. Dimigen et al. (2012) used the boundary paradigm in coregistration to investigate this issue. In the boundary paradigm (Rayner, 1975), the word in parafoveal vision is under the experimenter's control. In typical studies using this technique, the word is either identical, somehow related, or unrelated to the eventually displayed word. As soon as the participant's eyes cross an imaginary boundary, the display is switched to its end state. In contrast to Simola et al. (2009), Dimigen et al. (2012) found no influence from semantically related words in the parafovea. In another study using this technique, Hutzler et al. (2013) found that masking the parafoveal word with Xs, a common practice in the field, affects processing of the foveal word. Also related to lexical access, Hutzler et al. (2007) successfully replicated the old/new effect in word recognition with both conventional RSVP and natural reading, providing another proof-of-concept for coregistration.

1.3.1 Technical Challenges

1.3.1.1 Ocular Artifacts

The prime reason why coregistration is not more widely used in experimental psychology is that it is a rather involved procedure. The probably most prominent problem is that the eyeball is an electrical dipole that is positively charged at the cornea (the front) and negatively charged at the retina (the back). As a consequence, rotational movements of the eyeball induce artifacts in the recorded EEG that are typically much larger than the signal originating from brain activity proper. These artifacts are most pronounced at frontal electrodes (Gratton,

1998) but are propagated via volume conduction across the entire scalp (Picton, Van Roon, et al., 2000). There are several approaches for correcting these artifacts mathematically (for reviews, see Brunia, Möcks, Van den Berg-Lenssen, & Coelho, 1989; Croft & Barry, 2000; Delorme, Sejnowski, & Makeig, 2007). In regression-based approaches, propagation coefficients are used to determine the relationship between the EOG and every single EEG channel (Elbert, Lutzenberger, Rockstroh, & Birbaumer, 1985; Gratton, Coles, & Donchin, 1983). The estimated proportion of shared activity is then removed from the EEG channels. The primary concern with these approaches is that they may involuntarily remove cerebral activity. As an alternative, Berg and Scherg (1994) proposed *multiple-source eye correction* (MSEC). In MSEC, a stretch of EEG is recorded during a calibration phase with controlled eye movements and used to estimate the spatial topography of brain activity related to eye movements (see also Ille, Berg, & Scherg, 2002).

A correction procedure that has gained considerable traction in the research community involves blind source separation via *independent component analysis* (ICA; Jung et al., 2000; Makeig, Bell, Jung, & Sejnowski, 1996; Vigario, 1997). In this approach, the EEG is first decomposed into statistically independent components. Artifactual components are then identified, and the EEG is reassembled without them. Whether a component is artifactual can be determined based on its scalp distribution, frequency spectrum, and correlation with the EOG (Iriarte et al., 2003; Li, Ma, Lu, & Li, 2006; Okada, Jung, & Kobayashi, 2007; Rong & Contreras-Vidal, 2006).

These methods yield very convincing results although there is some concern that they may distort certain frequency ranges (Wallstrom, Kass, Miller, Cohn, & Fox, 2004). There are several proposals for automating their application (Ghandeharion & Erfanian, 2010; Joyce, Gorodnitsky, & Kutas, 2004; Li et al., 2006; Mantini et al., 2007; Mognon, Jovicich, Bruzzone, & Buiatti, 2011; Schlögl et al., 2007), and recently, it has been demonstrated that such removal procedures can benefit greatly from the eye movement signal in coregistration studies (Henderson, Luke, Schmidt, & Richards, 2013; Plöchl, Ossandón, & König, 2012). In any case, the ICA must be trained on data, and in most cases, more data leads to better results at the expense of longer run time.

A somewhat problematic property of coregistration studies for ICA is that the artifacts have almost the same time-lock as the signal that is to be isolated. Thus, in the worst case, the artifact removal procedure may also remove all other effects that are time-locked to eye movements. To avoid this, we chose to use the EEG from unrelated sentences to train the ICA. We thereby trained the algorithm on authentic sentence reading EEG without the experimental manipulation. An alternative approach is to use the procedure proposed by Plöchl et al. (2012). Their algorithm removes components that explain a lot of variance during saccades (i.e., when the eyes are moving) but little variance during fixations (i.e., when the eyes are resting). A prerequisite for the application of Plöchl et al.'s approach is that horizontal and vertical gaze location are available as additional channels.

Although ocular artifacts are a severe problem, coregistration has been used without any correction procedures whatsoever. For example, Kretzschmar, Bornkessel-Schlesewsky, and Schlesewsky (2009) and Kretzschmar et al. (2013) obtained satisfying results by restricting their analyses to posterior parts of the scalp which are not as badly contaminated by eye movement artifacts.

1.3.1.2 Overlap

Another problem with ERPs in general is the large degree of overlap between the potentials from two successive fixations: The potential at any given stimulus will be influenced by the potentials from both the preceding and the following stimuli. For three reasons, this is more problematic in natural reading than in RSVP. First, at around 250 ms on average, the interval between two fixations is shorter than the typical interval between two stimuli in typical RSVP (400 to 600 ms). Second, the interval between two fixations is not constant because reading speed is under the reader's control. This is not necessarily a bad thing because some jitter helps in filtering out the overlap of oscillations in higher frequency ranges. However, it also means that deconvolution approaches are even harder to apply (see below).

Finally, the degree of overlap is not independent of the material being read. In the worst case, the overlap of two potentials may therefore vary systematically between two conditions of an experiment.

A number of mathematical solutions have been proposed to deconvolute two overlapping signals (Bardy, Van Dun, Dillon, & McMahon, 2014; Delgado & Özdamar, 2004; Hansen, 1983; Jewett et al., 2004; Wang, Özdamar, Bohórquez, Shen, & Cheour, 2006; Woldorff, 1993). As Dimigen et al. (2011, p. 17) point out, however, a number of prerequisites have to be met for these approaches to work properly. They rely on a high signal-to-noise ratio to minimize the amount of additional noise that accrues through deconvolution and assume either that a waveform without overlap exists to form a deconvolution template or that each stimulus elicits the same brain response. Neither is the case in coregistration. In order to avoid overlap from preceding or following potentials, it is therefore paramount to control the regions around the target word when designing the stimulus materials. If this is done carefully, overlap may not be eliminated but can be assumed to be constant across conditions.

1.3.1.3 Setup

In addition to these more high-level problems, conducting coregistration studies entails a number of practical issues. The first decisions regard the experimental setup, which is typically constrained by what is available in the laboratory. We recorded eye movements and EEG in an electromagnetically shielded and sound-insulated booth, so we had to decide which appliances should be inside and which should be outside of the booth. Despite the potential line noise artifacts, we chose to put the display, eye tracker, and EEG amplifier inside and everything else outside of the booth (see Figure 1.4). To us, the gain in manageability outweighed the slight degradation in signal quality because line noise can be filtered out easily with notch filters and our chief interest was in slow waves that are barely affected. If a study required the analysis of waves around 50 Hz (or, in the U.S., around 60 Hz), avoiding line noise inside the booth would have highest priority, necessitating a reorganization of the experimental setup.

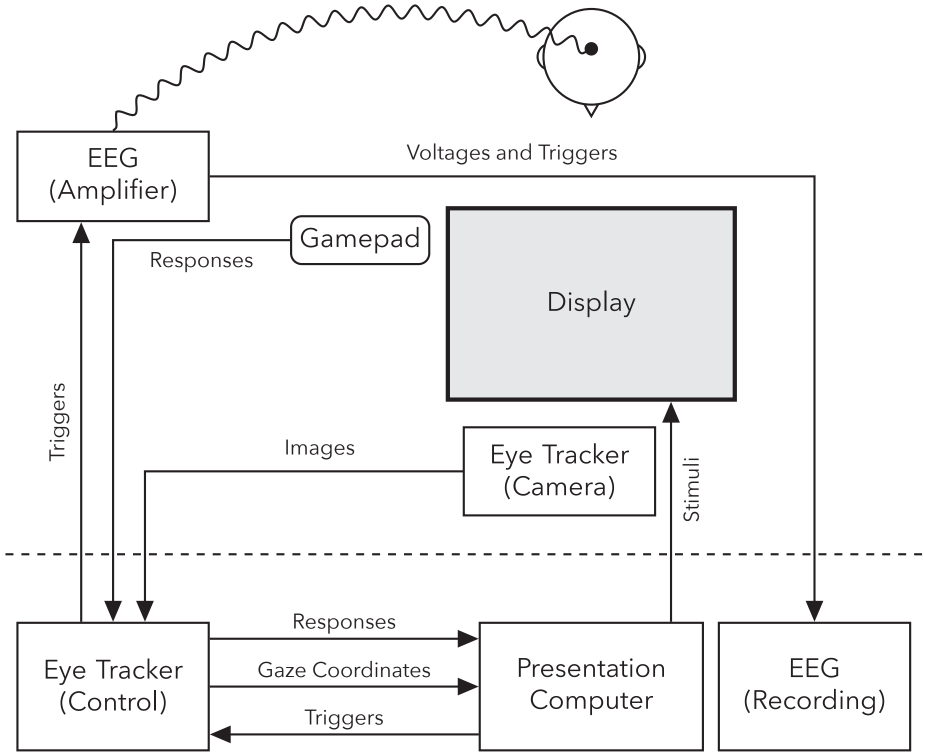

Figure 1.4: Schematic diagram of the experimental setup. Items above the dashed line as well as participants were situated within a electromagnetically shielded and sound-insulated booth.

In coregistration studies, the eye tracker needs to be calibrated and the electrode cap needs to be prepared prior to the experiment. To minimize the time needed before the experiment, participants read written instructions already when the electrode cap was placed on their head and primed with conductive gel. As soon as the electrode cap was ready, participants had typically also finished reading the instructions and could ask clarification questions. We also chose an electrode cap model with pre-installed electrodes (ANT Neuro, Enschede, The Netherlands), eliminating the need to manually connect single electrodes to the amplifier.

1.3.1.4 Synchronizing Timelines

In coregistration experiments, eye movements and EEG are recorded simultaneously but on different computers. One problem that arises from such a configuration is that the clocks in the two data streams will be different, if ever so slightly. Most likely, recording on the two devices will not be started exactly simultaneously, either. An important task is therefore to synchronize the two data streams such that fixations can be used to segment the EEG. We solved this by periodically piping triggers through the eye tracking control computer to the EEG amplifier. These triggers were thus present in both data streams and could be used to determine the difference in clock offsets and slight differences in clock speed. We achieved this with a linear model, estimating the time stamps in the EEG signal from the corresponding time stamps in the eye tracking signal. The intercept and slope of this model were then used to convert a fixation's time stamp to the EEG system's clock.

1.3.1.5 Analysis

The wealth and variety of data is one of the strengths of coregistration. However, this also complicates the analysis considerably. To begin with, it is not always clear where the trigger for the ERP should be when readers browse text freely. For example, if readers skip the target word, there is no fixation relative to which we could segment the EEG. The word must have been processed in some way, however, which raises the question of the appropriate trigger position. Assuming that readers processed the skipped word parafoveally, we should segment relative to the first fixation at the preceding word. For the same reason, we could segment to the last first-pass fixation at the preceding word. Other candidates are, inter alia, the first fixation at the word following the target word and the first fixation at the target word after it was skipped. The correct answer to this question will vary but the problem has to be addressed. Ideally, the materials are designed so that readers are inclined to fixate the target word in the first pass.

Another problem that arises from such highly multidimensional data is to choose the appropriate tools for statistical analysis. In eye movement studies, (generalized) linear mixed-effects regressions are gradually superseding ANOVA as the most popular statistical tool because they can deal with both continuous and categorical data, are capable of modeling crossed random effects of participants and items, and handle missing data well. Analyzing eye movements does not differ between pure eye movement studies and coregistration studies so there is no reason to adopt a different approach.

Analyzing the ERP (or TFR) in coregistration studies is more challenging because we have only vague assumptions about the topographical distribution and the timing of an effect. Even in the case of the well-established N400, it has been demonstrated that the onset can be earlier than what is typically observed in studies using RSVP (Dimigen et al., 2011; Kretzschmar et al., 2009). A common approach to this problem in conventional ERP studies is to slide a moving window across the ERP, perform statistical analyses for each window, and only consider effects that are significant in a number of successive windows. However, there are at least two problems with this approach. First, this requires some form of correction for multiple comparisons. Depending on the correction procedure, small effects may not survive such a procedure. Second, moving a window with a fixed width across the ERP may raise the risk of missing effects that lie on the boundary between two windows. This risk could be eliminated by allowing overlapping windows but this would also aggravate the first problem by increasing the number of comparisons. In essence, all these complications arise from the attempt to convert a multivariate problem into a univariate problem (Maris, 2012) and it has been advised to use alternative methods (Picton, Bentin, et al., 2000).

Drawing on the work of Blair and Karniski (1993) and Karniski, Blair, and Snider (1994), Maris and Oostenveld (2007) developed an elegant approach for the multiple-comparisons problem in exploratory ERP analysis (for a review and simulations, see Groppe, Urbach, & Kutas, 2011a, 2011b). The procedure is as follows. In a first step, statistical tests are performed at every electrode and every time point. Adjacent samples with significant effects are aggregated into clusters and their test statistics are summed up to yield a test statistic for the cluster. This procedure is then repeated a large number of times, each time swapping condition labels within randomly selected participants. The actual significance testing is done at the cluster level by comparing the cluster statistics from the first pass (i.e., the original data) to a distribution that consists of the largest absolute cluster statistics from each iteration. At an α level of .05, absolute cluster statistics in the upper 5% of the distribution can be considered significant.

Choosing the statistical test is within the researcher's degrees of freedom, and it involves a trade-off between statistical power and computational convenience. Fitting linear mixed-effects regressions at every time point and electrode has the benefit of allowing fully-crossed random effects but comes at the expense of running time, because the computational overhead of these models is considerable. In psycholinguistic experiments, the variance between items is typically small such that pair-wise t tests on per-subject averages work reasonably well.

1.4 Open Questions

In the present thesis, three open question about coregistration and its utility for sentence processing research will be addressed. In the following section, we will briefly outline each question's importance, explain how it was addressed, and summarize the main results. In the subsequent three chapters, the details of these studies are presented and discussed.

1.4.1 Sensitivity

So far, coregistration studies have mostly relied on well established effects, evoked by outright violations (Kretzschmar et al., 2009), and by employing large sample sizes in order to deal with a poor signal-to-noise ratio (e.g. Dimigen et al., 2011). Those options are idealized and rarely available in studies which address subtle psycholinguistic questions as, for example, how syntactic dependencies are parsed in regular sentences (i.e., in sentences without obvious violations). First, linguistic manipulations in regular, grammatical sentences typically produce rather small effects in the EEG. Second, it is often challenging to generate a sufficient number of appropriate stimuli for psycholinguistic studies, resulting in a smaller than desirable number of items.

A psycholinguistic topic that continues to receive a lot of attention is the resolution of non-local dependencies. Although it is ubiquitous in sentence processing, many aspects of its workings are still a matter of debate. To name only two open topics, it is not clear how exactly competition between multiple antecedents is decided (Dillon, Mishler, Sloggett, & Phillips, 2013; Sturt, 2003; Vasishth, Brüssow, Lewis, & Drenhaus, 2008; Xiang, Dillon, & Phillips, 2009) and how increased distance affects dependency resolution (Gibson, 2000; Grodner & Gibson, 2005; Hale, 2001, 2003, 2006; Levy, 2008; McElree, Foraker, & Dyer, 2003). The effects in these studies are subtle, and preparing the materials is very effortful. Theoretically, coregistration is particularly well suited to answer these kinds of questions because it allows us to delineate effects that are behaviorally similar by means of an electrophysiological signal. However, given the small effect sizes in previous studies, it is not clear if a poor signal-to-noise ratio or other peculiarities diminish coregistration's value as a tool for sentence processing research.

We conducted three experiments on dependency resolution to clarify if and how different types of linguistic dependencies respond differently to increased distance between their elements. We designed materials that allowed us to manipulate the distance within object-verb dependencies and antecedent-pronoun dependencies without changing the position of their heads. In the first experiment, we found clear locality effects (i.e., higher processing costs in long dependencies) for both dependency types at the word preceding the head in both eye movements and the ERP. Because we also found a confound in our materials, we repeated the experiment with revisited materials. We failed to reproduce the effects from the first experiment and attributed this to the confound of distance and phrase complexity at the pre-target word in the first experiment. To determine whether we would be able to find distance-related effects with a more robust method, we collected data with the same materials in a self-paced reading study. Here, we found locality effects for both dependency types at the word following the target word.

These results were more similar to those in the first experiment. All three experiments and their implications are discussed in detail in Chapter 2.

1.4.2 Neural Oscillations

Previous studies have demonstrated that ERPs recorded in coregistration can be strikingly similar to ERPs from experiments using RSVP (Dimigen et al., 2011; Kretzschmar et al., 2009). It is not clear if this similarity extends to the time-frequency domain. In fact, there is evidence that a similarity of two effects in the time domain does not necessarily entail an analogous similarity in the time-frequency domain (Hagoort et al., 2004; Roehm, Schlesewsky, Bornkessel, Frisch, & Haider, 2004). We replicated Hagoort et al. (2004) using both coregistration and RSVP to find out if oscillatory brain responses in natural and word-by-word reading behave similarly.

In the first experiment using RSVP, we were able to reproduce both the N400 effect in the time domain and the theta power increase in the time-frequency domain that Hagoort et al. (2004) reported. In the second experiment using coregistration, we reproduced the N400 effect but found different effects in the time-frequency domain. In sentences that were incongruent with common world knowledge, the target word elicited a power increase in the delta band and a power decrease in the upper alpha band. These results imply that readers pursue different strategies for building and retrieving memory traces of the parsed materials depending on the presentation mode. For an in-depth discussion, see Chapter 3.

1.4.3 Regressive Saccades

Outside of oculomotor research, coregistration has mainly been used to increase a study's ecological validity. In such studies, the first fixation on a word has replaced the stimulus onset as a time-lock for the ERP, but other than that, the procedure did not differ much from RSVP. Crucially, many facets of eye movements in reading have so far been neglected in coregistration studies on sentence processing. For instance, as pointed out in 1.2, reading is not strictly progressive. Rather, readers frequently make regressive saccades. This is sensitive to text characteristics: Readers are more likely to make a regressive saccade when processing demands are high. It has been observed, however, that readers are also capable of dealing with the same processing difficulties without making a regressive saccade. This has led to the proposal that there is also *covert reanalysis*, that is, readers may reconsider already parsed materials without an overt response (Lewis, 1998). Previous findings from coregistration studies suggest that regressions are indeed associated with the P600, an index of integration difficulties (Dimigen, Sommer, & Kliegl, 2007). We followed up on this question and explored the added benefit of coregistration by analyzing brain responses contingent on whether readers made a regressive saccade (see Chapter 4).

We replicated Hagoort (2003), an RSVP study on syntactic and semantic violations in Dutch. Using German materials, we found very similar results with RSVP: Syntactic violations elicited a P600 in sentence-medial position and a biphasic N400-P600 response in sentence-final position. Semantic violations elicited such a biphasic N400-P600 response in sentence-medial position and an N400 in sentence-final position. In coregistration, the ERPs were qualitatively similar when we analyzed the entire data set. When we constrained our analysis to trials with a violation that triggered a regression from the target word, we found a P600 in sentence-medial position and a biphasic N400-P600 in sentence-final position for both types of violation. By contrast, we found no P600s when the target word did not trigger a regression. Instead, we found sustained centro-parietal negativities for both semantic and syntactic violations in sentence-final position.

These results have a number of implications. They demonstrate that ERP findings from RSVP studies can be reproduced in coregistration and show that regressions co-occur with the P600 effect. This lends support to the notion that both responses indicate recovery processes in reading. The absence of the P600 effect in trials without regression suggests that the parser can pursue a processing strategy where recovery is not attempted.

2

Investigating Expectation and Locality with Concurrent Eye Movement and EEG Recordings

To comprehend an utterance that consists of more than one word, we have to transform the linear sequence of words into a hierarchical structure of conceptual nodes and lexical assignments. An essential and ubiquitous part of this process is to build dependencies between different words and clauses. For instance, readers need to build several such dependencies to understand a simple sentence like (3) below.

(3) The reporter who attacked the senator hoped for a story.

For a complete structural representation of that sentence, readers have to link *the reporter* and *hoped* (main clause subject and verb), *who* and *the reporter* (relative pronoun and antecedent), *who* and *attacked* (relative clause subject and verb), and *attacked* and *the senator* (relative clause verb and object). The dependency between subject and verb in the main clause (*the reporter… hoped*) is non-local: Its elements are not adjacent to each other. In (4), the dependency between relative pronoun and verb in the relative clause (*who… attacked*) is also non-local, which makes the sentence more complex than (3).

(4) The reporter who the senator attacked hoped for a story.

The crucial difference between (3) and (4) lies in the grammatical role of the relative pronoun within the relative clause. In (3), *who* is the subject of the subject-extracted relative clause. By contrast, in (4), *who* is the object of an object-extracted relative clause. It is largely undisputed that subject-extracted relative clauses are easier to process than object-extracted relative clauses (Gibson, Desmet, Grodner, Watson, and Ko, 2005; King and Just, 1991; King and Kutas, 1995; Traxler, Morris, and Seely, 2002; but see the discussion about Chinese in Vasishth, Chen, Li, and Guo, 2014). This advantage could simply be due to subject-extracted relative clauses being more frequent and therefore easier to parse than object-extracted relative clauses (Levy, 2008; Mitchell, Cuetos, Corley, & Brysbaert, 1995). Alternatively, the way in which working memory is taxed in subject- and object-extracted relative clauses may give rise to differences in processing effort. According to such memory-based accounts, it is either the number of syntactic projections (Gibson, 2000) or the distance between head and dependent that determine processing costs (Lewis, Vasishth, & Van Dyke, 2006). While these accounts predict similar patterns for the specific case of subject- and object-extracted relative clauses, their predictions differ for other types of non-local dependencies. Specifically, it is a matter of debate how increased distance affects dependency resolution.

Memory-Based Accounts

Numerous studies have shown that increased distance can have a detrimental effect on dependency resolution (e.g., Bartek, Lewis, Vasishth, & Smith, 2011; Demberg & Keller, 2008; Gibson, 2000; Grodner & Gibson, 2005; Kaan, 2002; McElree et al., 2003; Phillips et al., 2005). Consequently, many models of human language processing incorporate some form of distance metric to accommodate these locality effects (i.e., increased processing costs in long-distance dependencies). For instance, in *dependency locality theory* (DLT; Gibson, 1998, 2000; Gibson et al., 2005), this is implemented in the form of a storage component and an integration component. The storage costs at each point during a parse are determined by the number of syntactic heads that would be necessary to complete the sentence grammatically (see Gibson, 2000, p. 114). Integration costs are given by the number of discourse referents between the two elements of a dependency. The linear distance between two elements of a dependency, however, is not a determinant of integration costs in DLT (Gibson, 2000). Rather, the decisive factor is how many discourse referents are introduced between the two elements of a dependency. Therefore, intervening adverbs and adjectives should not affect dependency resolution whereas an intervening noun phrase should. For example, according to DLT, processing costs at *senator* should not differ between (5a) and (5b) but between (5a) and (5c).

(5) a. The reporter who attacked the senator admitted the error.
    b. The reporter who furiously, viciously, and unexpectedly attacked the senator admitted the error.
    c. The reporter who insulted the congressman and attacked the senator admitted the error.

The activation-based model by Lewis and Vasishth (2005) posits a different relation of retrieval costs and memory. Here, dependencies are resolved on the basis of the head's retrieval cues and corresponding features of the dependent. It is argued that the activation of an item in memory is subject to both decay and interference such that its strength is negatively affected by increased distance between dependent and head. Thus, completing a dependency becomes more difficult in situations where the linear distance between two elements is large and where more items in memory share the same or similar features (Lewis et al., 2006). Both of these *memory-based* models (DLT and activation-based model) predict that increased distance should give rise to increased processing costs.

Expectation-Based Accounts

Konieczny (2000) observed that, in certain configurations, increased distance can lead to anti-locality effects (i.e., decreased processing costs in longer dependencies). In German verb-final clauses like in (6), participants read the verb faster when it was more distant from a preceding object (see also Konieczny & Döring, 2003). Konieczny (2000) argues that readers may anticipate grammatical heads from their arguments: From reading the auxiliary *hat* in (6), they can anticipate a verb.

(6) a. Er hat die Rose hingelegt, und…
He has the rose laid_down, and…
"He laid down the rose, and…"

b. Er hat die Rose auf den Tisch gelegt, und…
He has the rose on the table laid, and…
"He laid the rose on the table, and…"

c. Er hat die Rose auf den kleinen runden Tisch gelegt, und…
He has the rose on the small round table laid, and…
"He laid the rose on the small round table, and…"

In conjunction with a surge of studies supporting the general idea of prediction in language comprehension (for a review, see Kutas, DeLong, & Smith, 2011), this has led to the proposal that the parser can build structure in a top-down fashion, predicting incoming words. In models incorporating this assumption, it is the degree of surprisal (Hale, 2001, 2003, 2006; Levy, 2008) that determines retrieval costs in non-local dependencies. In Levy's (2008) model, surprisal is defined as the log-inverse of a word's conditional probability. For example, in (6), the conditional probability of encountering the verb increases the further we move away from the auxiliary, which in turn decreases surprisal. Following *expectation-based* models, this leads to lower processing costs.

For object-extracted relative clauses like (4), memory- and expectation-based models make similar predictions. According to the DLT, storage costs are higher from keeping the extracted element in memory for a longer time and integration costs are increased by the intervening discourse referent *the senator*. Similarly, the activation-based model ascribes processing difficulties to increased distance and competition from transposing the object. Expectation-based models also predict increased difficulty in object-extracted relative clauses, based on the probabilistic knowledge that most relative clauses in English are subject-extracted (Levy, 2008, p. 1140). Because memory-based models posit increased difficulty due to storage and integration costs or decay and interference, this should become apparent primarily at the verb *attacked*. According to expectation-based models, however, it is the mere presence of an object extraction that causes processing difficulties. This is already clear at *the senator*, and expectation-based models predict that this is where processing difficulties should be observable.

Since storage and integration costs as well as decay and interference are positively correlated with the distance between object and verb, memory-based accounts predict locality effects in object-verb dependencies. However, the expectation of encountering the verb increases monotonically as more constituents are introduced after the object. This is why expectation-based models predict lower processing costs in long object-verb dependencies (i.e., anti-locality effects).

Relating to this issue, Santi and Grodzinsky (2007) investigated *wh* movement and reflexive binding with fMRI. The purpose of their study was to decide between two views on the role of Broca's area in language processing. According to the *specifist* view, it is dealing specifically with displacement in grammatical structures. Alternatively, following the *generalist* view, Broca's area may be involved in any memory-demanding task in sentence processing, including but not limited to syntactic movement. To distinguish these two accounts, Santi and Grodzinsky used materials as in Table 2.1 below. If the role of Broca's area in language processing were specific to syntactic operations, it should be sensitive to increased distance in movement but not binding because only the former is a syntactic operation. If, however, its role were to support any process that taxes memory in language processing, there should be no difference between the two operations. For the materials in Table 2.1, memory-based accounts would predict locality effects for both binding and movement. Expectation-based accounts, however, predict only anti-locality effects for movement because only in movement, the expectations for the second element of the dependency are sharpened as more constituents are introduced.

Santi and Grodzinsky (2007) found that activity in a subregion of Broca's area increased for movement and decreased for binding when the length of a dependency was increased. Based on this finding, they argue that Broca's area is indeed

Table 2.1: Sample sentences from Santi and Grodzinsky (2007).

Binding	
Short	The sister of Kim assumes that Anne loves the **mailman** who burnt **himself**.
Medium	The sister of Kim assumes that the **mailman** who loves Anne burnt **himself**.
Long	Anne assumes that the **mailman** who loves the sister of Kim burnt **himself**.
Movement	
Short	The mailman and the mother of Jim love **the woman** who Kate burnt**_**.
Medium	The mother of Jim loves **the woman** who the mailman and Kate burnt**_**.
Long	Kate loves **the woman** who the mailman and the mother of Jim burnt**_**.

Note. Ungrammatical conditions omitted. The underscore at the end of the filler-gap sentences indicates the gap position.

specific to memory-demanding syntactic processes and that binding and movement must therefore be subject to qualitatively different processes: In *wh* movement, readers start an active search for a gap, which is why movement operations tax working memory already before the dependency is completed (i.e., increasing storage costs in Gibson, 2000). By contrast, there is no active search in binding such that more demands are placed on retrieval (e.g., Gibson, 2000). The pattern reported by Santi and Grodzinsky (2007) is partly in line with memory-based accounts as it shows increasing activation in long-distance dependencies. For the same reason, the results do not support expectation-based accounts that would have predicted decreasing activation in long-distance *wh* movement. Further support for this view comes from Makuuchi, Bahlmann, Anwander, and Friederici (2009), who report that larger distance and larger structural complexity lead to increased activation in Broca's area, albeit in distinct subregions. There is an anti-locality effect in binding which is not explained by either account and which is only briefly discussed by Santi and Grodzinsky.

Although these results are informative about the effect of increased distance in non-local dependencies, Santi and Grodzinsky's study suffers from at least two methodological weaknesses. First, as Matchin, Sprouse, and Hickok (2014) point out, comparing antecedent-reflexive dependencies with *wh* movement is a little like comparing apples and oranges: The filler in *wh* movements initiates an active search for a gap whereas the antecedent in antecedent-reflexive dependencies does not (Fodor, 1978; Frazier, 1987; Frazier & d'Arcais, 1989; Stowe, 1986). Indeed, using materials as in Table 2.2 below, Matchin et al. demonstrate that backwards anaphora elicit the same locality effect as *wh* movement. Since backwards anaphora are not subject to syntactic movement, this refutes Santi and Grodzinsky's claims and rather supports the generalist view.

Table 2.2: Abridged sample sentences from Matchin, Sprouse, and Hickok (2014).

Binding	
Short	Because **she** decorated the cake, **the baker** wowed the customer that made the long order.
Long	Because **she** decorated the cake that was six layers tall, **the baker** wowed the customer.
Movement	
Short	**Which song** did the band play_ poorly at the concert that ended early?
Long	**Which song** did the band that won the contest play_ poorly at the concert?

There are two additional caveats. First, both studies were designed and conducted to clarify the role of a specific brain area in dependency resolution. For this purpose, fMRI is an excellent choice. Because it has poor temporal resolution, however, fMRI does not yield useful insights about the time course of dependency resolution. Second, also in both studies, dependency length was confounded with the position of the target word within the sentence. In (2.2) above, *the baker* and the gap after *play* occur systematically later in the sentence in long dependencies. This is undesirable because any observed locality (or anti-locality) effects could just as well be due to different word positions.

We conducted a series of three experiments on object-verb and antecedent-pronoun dependencies to remedy these shortcomings. To remove the confound of dependency length and word position, we used a distance manipulation in which the position of the dependency head was constant while the position of its dependent was varied. To monitor online dependency resolution in natural reading with high temporal precision, we coregistered eye movements and EEG in two of the experiments. The third experiment used the self-paced reading technique which has been shown to be a reasonable proxy for natural reading (Bartek et al., 2011) with a higher signal-to-noise ratio than what is typically observed in eye movement studies.

The above-mentioned accounts of sentence processing predict distinct effects of distance for these two dependency types. Expectation-based models predict anti-locality effects in object-verb dependencies but no effects in antecedent-pronoun dependencies. Memory-based accounts predict locality effects in both dependency types. Based on Streb, Hennighausen, and Rösler (2004), we further expect a distance-related N400 effect at the target word in both dependency types, which is most readily explained in a memory-based framework. Finally, if the results from Santi and Grodzinsky (2007) generalize to our materials, we would expect a locality effect for object-verb dependencies and an anti-locality effect or no effect of distance for antecedent-pronoun dependencies.

2.1 Experiment 1

2.1.1 Methods

2.1.1.1 Participants

We invited 52 undergraduate students (37 women, 15 men) at the University of Potsdam, using the recruiting software *ORSEE* (Greiner, 2004). They were native speakers of German, right-handed by self-report, and had normal or corrected-to-normal vision. Their age ranged from 19 to 34 years ($M = 25$), and they received course credit or payment for compensation. Due to recording errors, we had to discard the eye tracking data from one participant and the EEG data from three participants. This left eye movement data from 51 participants and EEG data from 49 participants. We restricted our analyses to the data from 49 participants from which we had a complete set of observations (eye movements and EEG).

2.1.1.2 Materials

The experimental materials (Tables 2.3 and 2.4) comprised 120 items with four conditions in a 2×2 design (dependency type $\times$ distance). The sentences consisted of a main clause and two subordinate clauses (see Tables 2.4 and 2.3). They started with a main clause with two coordinated noun phrases in subject position(*Maria and the director*), the main clause verb (*recognize*) and another noun phrase in object position (*the actor*). In object-verb conditions, the main clause was followed by a relative clause with a relative pronoun (*who*), another two coordinated noun phrases as relative clause subject (*the editor and the photographer*), and the relative clause verb (*interview*). In the antecedent-pronoun conditions, a connector (*when*) initiated an ensuing subordinate clause, followed again by the two coordinated noun phrases in subject position, a pronoun (*him*), an adverb to capture spillover (*exclusively*), and the relative clause verb. The target words were the relative clause verb in object-verb dependencies and the pronoun in antecedent-pronoun dependencies. In all conditions, the second clause was followed by another subordinate clause to capture spillover effects and to avoid confounds with sentence wrap-up (*to learn the truth about his marriage.*).

Our goal was to manipulate the distance within a dependency without changing the position of the dependency head. To this end, we increased the complexity of a noun phrase either in the main or subordinate clause. In short conditions, the complex noun phrase occurred before the main clause object, such that it did not intervene between the two parts of the dependency. In long conditions, the second noun phrase within the subordinate clause was complex, leading to greater distance between object and verb or antecedent and pronoun. We used two types of complexity manipulation to introduce more variation. We inserted either an adverb and adjective (*the incredibly talented photographer*) between determiner and noun or nested the noun phrase within another noun phrase (*the daughter of the photographer*). For the analyses, we collapsed these manipulation types.

The experimental sentences were between 18 and 28 words long ($M = 22$). In object-verb dependencies, the target word was on average 9 characters long ($SE = 0.2$). The resulting 480 sentences were distributed across four lists in a latin-square design. Sentences from an unrelated experiment were used as distractor sentences. Those were between 3 and 17 words long ($M = 7$).

Table 2.3: Sample sentences from Experiment 1 with object-verb dependency.

Adverb + Adjective	
Short	Maria und die *unglaublich talentierte* Regisseurin erkennen den Schauspieler, **den** der Redakteur und der Fotograf **interviewen**, um die Wahrheit über seine Ehe herauszufinden.
	(Maria and the *incredibly talented* director recognize the actor, **who** the editor and the photographer **interview** to learn the truth about his marriage.)
Long	Maria und die Regisseurin erkennen den Schauspieler, **den** der Redakteur und der *unglaublich talentierte* Fotograf **interviewen**, um die Wahrheit über seine Ehe herauszufinden.
	(Maria and the director recognize the actor, **who** the editor and the *incredibly talented* photographer **interview** to learn the truth about his marriage.)
Nested Noun Phrase	
Short	Maria und die *Tochter der* Regisseurin erkennen den Schauspieler, **den** der Redakteur und der Fotograf **interviewen**, um die Wahrheit über seine Ehe herauszufinden.
	(Maria and the *daughter of the* director recognize the actor, **who** the editor and the photographer **interview** to learn the truth about his marriage.)
Long	Maria und die Regisseurin erkennen den Schauspieler, **den** der Redakteur und die *Tochter des* Fotografen **interviewen**, um die Wahrheit über seine Ehe herauszufinden.
	(Maria and the director recognize the actor, **who** the editor and the *daughter of the* photographer **interview** to learn the truth about his marriage.)

Object and verb are set in bold face, the intervening material is italicized. Type of complexity (adverb and adjective or nested noun phrase) was a between-item manipulation but is presented within one item here for the purpose of presentation.

2.1.1.3 Procedure

After participants had read the briefing text and signed a consent form, they were seated in a shielded booth approximately 60 cm from a stimulus display with a diagonal length 0f 22 in and a resolution of 1680×1050 pixels. Following electrode cap preparation and eye tracker calibration, the experiment started with five practice trials. Sentences were presented left-justified in the vertical center of the screen in 26 point Arial, stretched across at most two lines. The target word always occurred before the last word on the first line. To proceed to the comprehension question after they had read a sentence, participants had to fixate the bottom-right corner of the display. We chose this procedure to prevent anticipatory eye move-

Table 2.4: Sample sentences from Experiment 1 with antecedent-pronoun dependency.

Adverb + Adjective	
Short	Maria und die *unglaublich talentierte* Regisseurin erkennen **den Schauspieler**, als der Redakteur und der Fotograf **ihn** exklusiv interviewen, um die Wahrheit über seine Ehe herauszufinden.
	(Maria and the *incredibly talented* director recognize **the actor**, when the editor and the photographer interview **him** exclusively to learn the truth about his marriage.)
Long	Maria und die Regisseurin erkennen **den Schauspieler**, als der Redakteur und der *unglaublich talentierte* Fotograf **ihn** exklusiv interviewen, um die Wahrheit über seine Ehe herauszufinden.
	(Maria and the director recognize **the actor**, when the editor and the *incredibly talented* photographer interview **him** exclusively to learn the truth about his marriage.)
Nested Noun Phrase	
Short	Maria und die *Tochter der* Regisseurin erkennen **den Schauspieler**, als der Redakteur und der Fotograf **ihn** exklusiv interviewen, um die Wahrheit über seine Ehe herauszufinden.
	(Maria and the *daughter of the* director recognize **the actor**, when the editor and the photographer interview **him** exclusively to learn the truth about his marriage.)
Long	Maria und die Regisseurin erkennen **den Schauspieler**, als der Redakteur und die *Tochter des* Fotografen **ihn** exklusiv interviewen, um die Wahrheit über seine Ehe herauszufinden.
	(Maria and the director recognize **the actor**, when the editor and the *daughter of the* photographer interview **him** exclusively to learn the truth about his marriage.)

Antecedent and pronoun are set in bold face, the intervening material is italicized. Type of complexity (adverb and adjective or nested noun phrase) was a between-item manipulation but is presented within one item here for the purpose of presentation.

ments to the beginning of the sentence. Participants received regular feedback about their performance on the comprehension questions of the preceding ten trials. After 60, 120, and 180 sentences, they took a short break and relaxed their eyes. An average session lasted between 2.5 and 3 hours.

2.1.1.4 Recording

Eye Movements Gaze position was recorded from the right eye with a desktop-mounted EyeLink 1000 (SR Research, Mississauga, Ontario, Canada). The eye tracker was used in remote mode to allow participants to sit comfortably without a chin rest. The sampling rate was set to 500 Hz and the eye tracker had a reported spatial resolution of 0.01°. It reached an average accuracy of 0.54° in the vertical center of the screen.

We excluded fixations below 20 ms and above 1000 ms from analysis. The remaining fixations were aggregated into the standard fixation measures first fixation duration, gaze duration, and regression probability. First fixation duration is the duration of the first fixation on a word. Gaze duration is the cumulative duration of all fixations on a word during the first pass, that is, from first fixating it until leaving it again. Regression probability is the proportion of trials where readers made a regressive saccade from a word after fixating it but before leaving it to the right. For all these measures, we excluded trials where regions to the right of the word had previously been fixated.

EEG We recorded the EEG from 32 Ag/AgCl electrodes that were mounted in an electrode cap (Advanced Neuro Technology, Enschede, Netherlands) and arranged according to the 10-20 system (Jasper, 1958). Eye movements and blinks were monitored with bipolar electrodes next to the left and right eye as well as below and above the right eye. EEG and EOG were recorded with a sampling rate of 512 Hz and a low-pass filter of 138 Hz. Recordings were referenced to a common average reference and impedances were kept between 5 and 10 kΩ.

The EEG data was preprocessed in BrainVision Analyzer 2 (Brain Products, Munich, Germany) where the signal was first resampled to 500 Hz and filtered (0.3–100 Hz band-pass, 50 Hz notch). We used an ICA (Jung et al., 2000) with a biased variant of the Infomax algorithm to identify and remove eye movement artifacts from the EEG. Following preparatory sphering with a classic PCA, the ICA was trained on EEG from the distractor sentences to obtain reliable estimates from sentence reading without accidentally removing effects that are similarly time-locked to eye movements. We subsequently removed components with a clear frontal or bipolar frontal distribution and segmented the corrected signal from −1000 ms to 2000 ms relative to the first fixation at the pre-target, target, and post-target words. The markers for this segmentation were generated from the eye movement record. In a semi-automatic procedure, segments with muscle artifacts or slow drifts were removed, which led to the loss (across all participants and trials) of 277 trials for the pre-target word (5.4%), 194 trials for the target word (5%), and 213 trials for the post-target word (5.6%) for the post-target word.[1] We performed all further preprocessing and analyses in *R* (R Core Team, 2013).

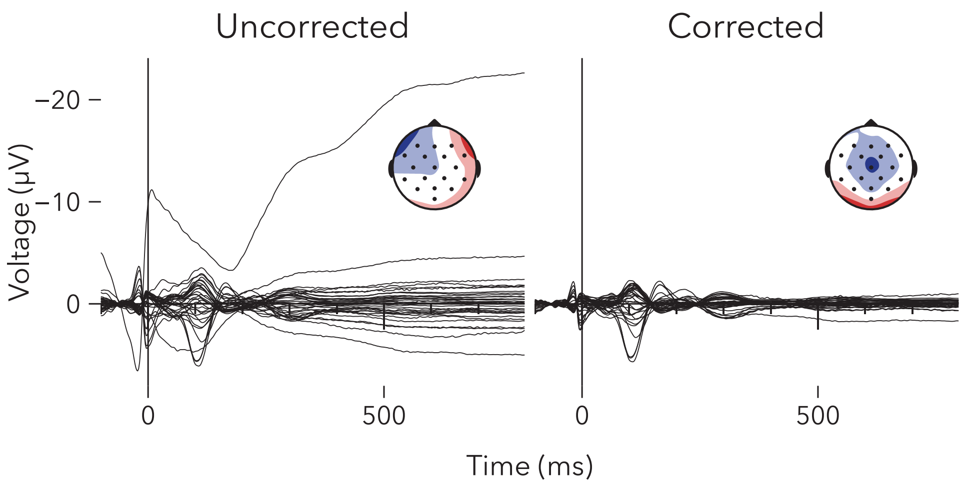

Figure 2.1: Grand-averaged ERP for randomly selected fixations from Experiment 1 before and after artifact correction. Topographic maps show mean amplitude in the first 500 ms following the fixation.

2.1.1.5 Analysis

Behavioral Data Fixation measures and response accuracy were analyzed with linear mixed-effects regressions using the package *lme4* (Bates, Maechler, Bolker, & Walker, 2014) in *R* (R Core Team, 2013). Before analysis, we determined whether a variable had to be transformed to yield normally distributed residuals and which transformation would stabilize variance (Box & Cox, 1964). Following this procedure, we log-transformed duration variables before analysis. For binary variables (first-pass regression yes/no, response accuracy), we used a generalized lin-

[1]Percentages are relative to the number of eligible trials, i.e., trials in which participants did not skip the target word.

ear mixed-effects regression using a binomial distribution with the logit link function.

EEG Prior to aggregation and analysis, each segment was baseline-corrected by subtracting the mean voltage of a 100 ms segment preceding the fixation. Since we had vague expectations but no strong a priori assumptions about the polarity, timing, or distribution of distance-related ERP effects, we decided not to restrict our analyses to a subset of electrodes or time windows. However, a completely exploratory analysis of the EEG would have drastically inflated the Type-I error risk. Cluster-based randomization tests (Maris & Oostenveld, 2007) offer an elegant solution for handling such multiple-comparisons problems. In our implementation of the approach, we performed a pair-wise t test (short vs. long conditions) at each electrode and every time point, separately for object-verb and antecedent-pronoun dependencies. Next, we formed spatio-temporal clusters of significant sample statistics (i.e., $p < .05$), using connected-components clustering (Samet & Tamminen, 1988). The sum of t values within a cluster yielded its test statistic. We then generated a distribution of test statistics by repeatedly running this procedure with randomly swapped conditions, selecting the largest cluster statistic from each of 1000 iterations. This distribution represented the null hypothesis. Cluster statistics that fell in the lower or upper 2.5th percentile of this distribution were considered statistically significant.

2.1.2 Results

2.1.2.1 Comprehension

Response accuracy in experimental trials was 83.8% across conditions. Participants fared best in short antecedent-pronoun dependencies (84.6%), followed by short object-verb (84.1%), long antecedent-pronoun (83.8%), and long object-verb dependencies (82.6%). There were no effects of dependency type or distance.

2.1.2.2 Eye Movements

We included the words preceding (pre-target word) and following the target word (post-target word) in our analyses because the pronoun was skipped in 66.3% of the trials with antecedent-pronoun dependencies (66.6% for short and 66.1% for long dependencies). By contrast, in object-verb dependencies, the target word was skipped in only 6.6% of the trials (6.0% for short and 7.1% for long dependencies). This difference between dependency types was highly significant ($b = 3.71$, 95% CI $[3.53, 3.89]$, $z = 39.75$), but there were no significant effects of distance on skipping rates.

At the post-target word, we observed the opposite pattern: Skipping occurred more frequently in object-verb dependencies (60.7%) than in antecedent-pronoun dependencies (15.7%). The reason might be that readers skipped the target word in so many trials in antecedent-pronoun dependencies, making a fixation on the following word more likely. The difference in skipping rates between dependency types was statistically significant ($b = -2.45$, 95% CI $[-2.59, -2.31]$, $z = -34.34$).

First fixation durations, gaze durations, and regression probability at the pre-target word were increased in long dependencies for both dependency types (see Table 2.5 for the parameter estimates). Gaze durations at the pre-target word were also longer in antecedent-pronoun dependencies.

There were no locality effects at the target word (Table 2.6). In contrast to the pre-target word, first fixation durations and gaze durations at the target word were shorter and regression probability was lower in antecedent-pronoun dependencies.

Like the pre-target word, the post-target word elicited longer gaze durations and higher regression probabilities in antecedent-pronoun dependencies (Table 2.7). There were no main effects of distance but there was a marginal interaction of dependency type and distance in regression probability, suggesting a

Table 2.5: Summary statistics for the mixed-effects regression of first fixation duration as a function of dependency type and distance in Experiment 1. Estimates (b) and 95% confidence intervals (CI) are on the log scale, statistical significance is indicated by bold face.

	b	95% CI	t
Pre-Target Word			
Type	0.01	[−0.01, 0.03]	0.9
Distance	**0.06**	**[0.04, 0.08]**	**6.9**
Type × Distance	0.03	[−0.01, 0.06]	1.5
Target Word			
Type	**−0.20**	**[−0.23, −0.17]**	**−14.9**
Distance	0.02	[−0.01, 0.05]	1.1
Type × Distance	−0.01	[−0.06, 0.04]	−0.3
Post-Target Word			
Type	0.00	[−0.02, 0.02]	0.1
Distance	0.00	[−0.02, 0.02]	0.0
Type × Distance	0.00	[−0.04, 0.04]	0.0

Table 2.6: Summary statistics for the mixed-effects regression of gaze duration as a function of dependency type and distance in Experiment 1. Estimates (b) and 95% confidence intervals (CI) are on the log scale, statistical significance is indicated by bold face.

	b	95% CI	t
Pre-Target Word			
Type	**0.03**	**[0.01, 0.05]**	**2.9**
Distance	**0.04**	**[0.02, 0.06]**	**3.8**
Type × Distance	0.01	[−0.03, 0.05]	0.5
Target Word			
Type	**−0.44**	**[−0.47, −0.41]**	**−31.6**
Distance	0.01	[−0.02, 0.04]	0.8
Type × Distance	−0.01	[−0.06, 0.04]	−0.5
Post-Target Word			
Type	**0.20**	**[0.17, 0.23]**	**13.5**
Distance	0.01	[−0.02, 0.04]	0.5
Type × Distance	0.01	[−0.04, 0.06]	0.5

larger influence of distance in antecedent-pronoun dependencies. This interpretation is supported by the pattern in Figure 2.2.

Table 2.7: Summary statistics for the mixed-effects regression of regression probability as a function of dependency type and distance in Experiment 1. Estimates (b) and 95% confidence intervals (CI) are on the logit scale, statistical significance is indicated by bold face.

	b	95% CI	z
Pre-Target Word			
Type	0.07	$[-0.10, 0.24]$	0.8
Distance	**0.33**	$\mathbf{[0.16, 0.50]}$	**3.8**
Type × Distance	0.13	$[-0.21, 0.47]$	0.8
Target Word			
Type	$\mathbf{-0.45}$	$\mathbf{[-0.68, -0.22]}$	$\mathbf{-3.8}$
Distance	0.07	$[-0.16, 0.30]$	0.6
Type × Distance	-0.05	$[-0.50, 0.40]$	-0.2
Post-Target Word			
Type	**0.66**	$\mathbf{[0.43, 0.89]}$	**5.6**
Distance	0.12	$[-0.11, 0.35]$	1.0
Type × Distance	0.40	$[-0.05, 0.86]$	1.7

2.1.2.3 ERP

Relative to short conditions, long object-verb dependencies engendered two sustained frontal negativities from 330 ms to 794 ms and from 844 ms to 1000 ms (peaks at 530 ms and 984 ms) at the pre-target word (Figure 2.3). Two subsequent posterior positivities from 294 ms to 676 ms and from 680 ms to 924 ms (peaks at 528 ms and 696 ms) accompanied these negativities. At the target word, long object-verb dependencies produced a similar pattern with an early negativity at fronto-polar electrodes (34 to 260 ms, peak at 128 ms) and a positivity at posterior electrodes (86 to 292 ms, peak at 216 ms). There was no such biphasic response to long object-verb dependencies at the post-target word but only a positivity at frontal electrodes (568 ms to 770 ms, peak at 686 ms).

Increased distance in antecedent-pronoun dependencies resulted in a positivity from 448 ms to 1000 ms at parietal electrodes (peak at 718 ms) and a later positivity from 700 ms to 802 ms (peak at 790 ms) at centro-parietal electrodes at

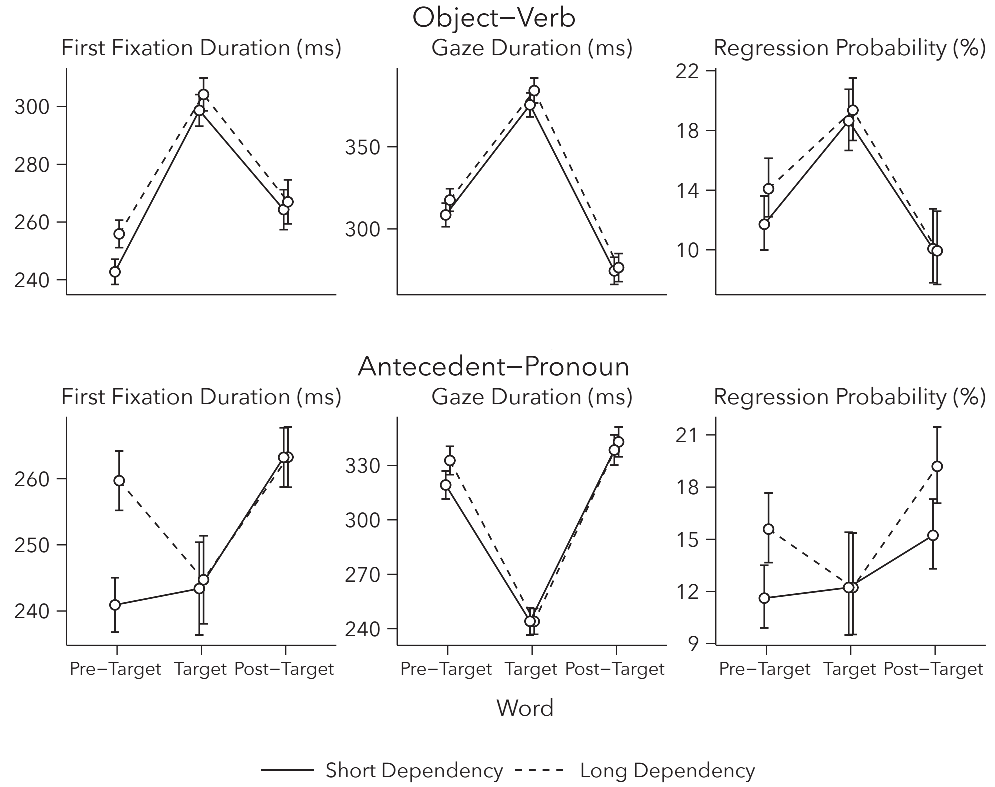

Figure 2.2: First fixation duration, gaze duration, and regression probability in Experiment 1 as a function of dependency type and distance on the pre-target, target, and post-target word (solid lines for short and dashed lines for long dependencies). Error bars are 95% confidence intervals that were computed after removing variance due to different participants and items.

the pre-target word (Figure 2.4). There were no effects of distance in antecedent-pronoun dependencies at the target word but a centro-parietal positivity between 700 and 802 s (peak at 790 ms) at the post-target word.

Regardless of dependency length, the post-target word elicited qualitatively different responses in the two dependency types: A frontal negativity from around 300 ms in object-verb dependencies and a frontal positivity from around 500 ms in antecedent-pronoun dependencies.

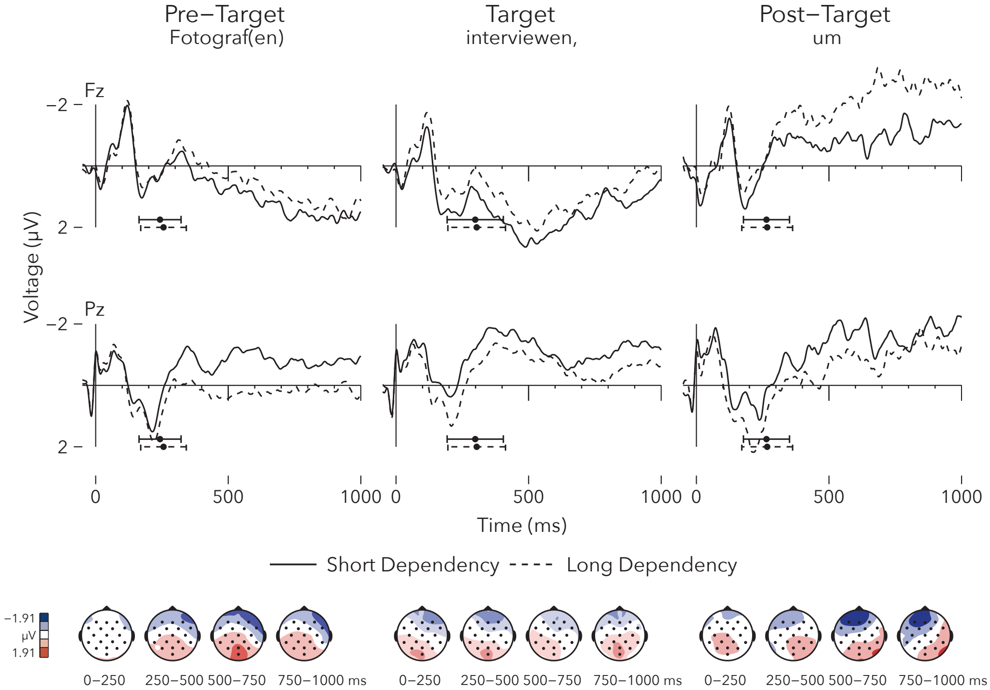

Figure 2.3: Grand-averaged ERP at the pre-target, target, and post-target word in object-verb dependencies for electrodes Fz and Pz in Experiment 1, low-pass-filtered at 30 Hz. Solid lines show amplitudes for short and dashed lines for long dependencies. Horizontal error bars show mean first fixation duration per condition at each word, plus and minus one standard deviation. Topographic maps show amplitude differences (long minus short) in successive time windows of 250 ms length. Averages are based on 2390 observations for the pre-target word (1193 short, 1197 long), 2702 observations for the target word (1356 short, 1346 long), and 1140 observations for the post-target word (566 short, 574 long).

2.1.3 Discussion

A variety of distance-related effects emerged at the pre-target, target, and post-target word. In the eye movement record, greater distance led to elevated first fixation durations, gaze durations, and regression probabilities in both dependency types. In the ERP, object-verb dependencies elicited sustained biphasic effects at frontal and parietal electrodes at the pre-target and target word. Additionally, at the post-target word, long object-verb dependencies elicited a sustained frontal negativity. Long antecedent-pronoun dependencies led to parietal positivities at

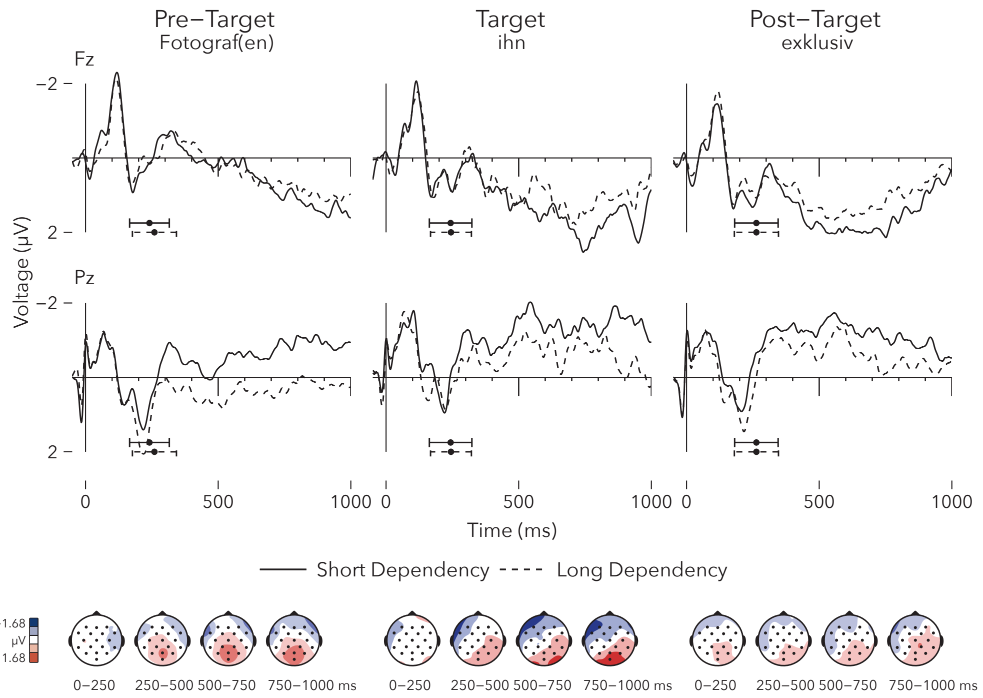

Figure 2.4: Grand-averaged ERP at the pre-target, target, and post-target word in antecedent-pronoun dependencies for electrodes Fz and Pz in Experiment 1, low-pass-filtered at 30 Hz. Solid lines show amplitudes for short and dashed lines for long dependencies. Horizontal error bars show mean first fixation duration per condition at each word, plus and minus one standard deviation. Topographic maps show amplitude differences (long minus short) in successive time windows of 250 ms length. Averages are based on 2446 observations for the pre-target word (1207 short, 1239 long), 981 observations for the target word (486 short, 495 long), and 2439 observations for the post-target word (1224 short, 1215 long).

the pre- and post-target word but no effects at the pronoun itself, which may have been a consequence of the high skipping frequency.

At first sight, these patterns in eye movements and ERP suggest that longer dependencies of both types lead to greater processing effort. This would be in line with memory-based accounts of sentence processing. However, our materials had a problematic confound of distance and phrase complexity at the pre-target word. When the distance within a dependency was long, the noun immediately preceding the second element of the dependency was the head of a complex noun phrase (*the incredibly talented photographer/the daughter of the photogra-*

pher). Conversely, the same noun was the head of a simple noun phrase when the dependency was short (*the photographer*). It has been demonstrated that complex noun phrases are harder to process and that this effect spills over to following words (Hofmeister, 2011). We therefore decided to repeat the experiment with altered materials before interpreting the possibly contaminated effects from Experiment 1.

2.2 Experiment 2

2.2.1 Methods

2.2.1.1 Participants

52 participants (40 female, 12 male) were recruited from the same population as the participants in Experiment 1. Again, all participants were right-handed by self-report, native speakers of German, and had normal or corrected-to-normal vision. Their age ranged from 19 to 43 ($M = 25$), and they were not aware of the purpose of the study. Participants received either course credit or money for compensation and gave their informed consent before the experiment.

2.2.1.2 Materials

The experimental materials from Experiment 1 were altered in one respect for Experiment 2 (Tables 2.8 and 2.9): To deconfound the pre-target word, we moved the intervening material to the first of the two noun phrases in the subordinate clause, leaving a constant region before the target word. We also changed the spillover region in some items such that the sentence never occupied more than one line on the screen. The resulting 480 sentences were again distributed across four lists in a latin-square design and every participant saw each item in only one condition. The 120 experimental sentences in each list were complemented with 120 distractors of similar length and complexity to obscure the goal of the study.

The distractor sentences consisted of a main clause and at least one subordinate clause and were between 10 and 21 words long ($M = 15$).

Table 2.8: Sample sentences from Experiment 2 with object-verb dependency.

Adverb + Adjective	
Short	Maria und die *unglaublich talentierte* Regisseurin erkennen den Schauspieler, **den** der Redakteur und der Fotograf **interviewen**, um die Wahrheit über seine Ehe herauszufinden.
	(Maria and the *incredibly talented* director recognize the actor, **whom** the editor and the photographer **interview** to learn the truth about his marriage.)
Long	Maria und die Regisseurin erkennen den Schauspieler, **den** der *unglaublich talentierte* Redakteur und der Fotograf **interviewen**, um die Wahrheit über seine Ehe herauszufinden.
	(Maria and the director recognize the actor, **whom** the editor and the *incredibly talented* photographer **interview** to learn the truth about his marriage.)
Nested Noun Phrase	
Short	Maria und die *Tochter der* Regisseurin erkennen den Schauspieler, **den** der Redakteur und der Fotograf **interviewen**, um die Wahrheit über seine Ehe herauszufinden.
	(Maria and the *daughter of the* director recognize the actor, **whom** the editor and the photographer **interview** to learn the truth about his marriage.)
Long	Maria und die Regisseurin erkennen den Schauspieler, **den** die *Tochter des* Redakteurs und der Fotograf **interviewen**, um die Wahrheit über seine Ehe herauszufinden.
	(Maria and the director recognize the actor, **whom** the *daughter of the* editor and the photographer **interview** to learn the truth about his marriage.)

Object and verb are set in bold face, the intervening material is italicized. Type of complexity (adverb and adjective or nested noun phrase) was a between-item manipulation but is presented within one item here for the purpose of presentation.

2.2.1.3 Procedure

The presentation procedure was the same as in Experiment 1 but sentences were presented in Arial 16 point so they would fit on one line.

Table 2.9: Sample sentences from Experiment 2 with antecedent-pronoun dependency.

Adverb + Adjective	
Short	Maria und die *unglaublich talentierte* Regisseurin erkennen **den Schauspieler**, als der Redakteur und der Fotograf **ihn** exklusiv interviewen, um die Wahrheit über seine Ehe herauszufinden. (Maria and the *incredibly talented* director recognize **the actor**, when the editor and the photographer interview **him** exclusively to learn the truth about his marriage.)
Long	Maria und die Regisseurin erkennen **den Schauspieler**, als der *unglaublich talentierte* Redakteur und der Fotograf **ihn** exklusiv interviewen, um die Wahrheit über seine Ehe herauszufinden. (Maria and the director recognize **the actor**, when the *incredibly talented* editor and the photographer interview **him** exclusively to learn the truth about his marriage.)
Nested Noun Phrase	
Short	Maria und die *Tochter der* Regisseurin erkennen **den Schauspieler**, als der Redakteur und der Fotograf **ihn** exklusiv interviewen, um die Wahrheit über seine Ehe herauszufinden. (Maria and the *daughter of the* director recognize **the actor**, when the editor and the photographer interview **him** exclusively to learn the truth about his marriage.)
Long	Maria und die Regisseurin erkennen **den Schauspieler**, als der die *Tochter des* Redakteurs und der Fotograf **ihn** exklusiv interviewen, um die Wahrheit über seine Ehe herauszufinden. (Maria and the director recognize **the actor**, when the *daughter of the* editor and the photographer interview **him** exclusively to learn the truth about his marriage.)

Antecedent and pronoun are set in bold face, the intervening material is italicized. Type of complexity (adverb and adjective or nested noun phrase) was a between-item manipulation but is presented within one item here for the purpose of presentation.

2.2.1.4 Recording and Analysis

All eye movement recordings and analyses were performed as in Experiment 1 with the exception that participants were required to rest their chin on a head rest to improve the spatial accuracy of the eye tracker. As a consequence, average accuracy was better than in Experiment 1 with 0.63° overall and 0.37° in the region where the sentences were presented.

Recording, preprocessing, and analysis of the EEG was mostly performed like in Experiment 1, but the EEG was referenced to the left mastoid online and reref-

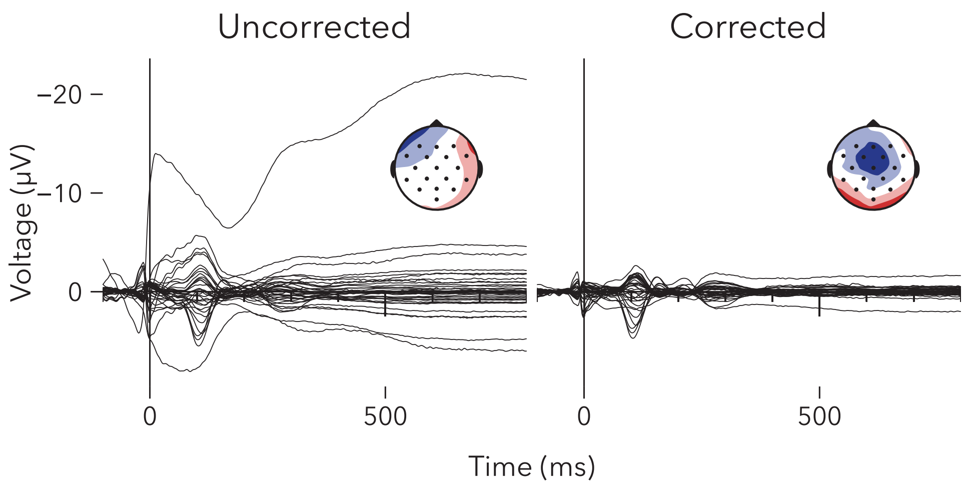

Figure 2.5: Grand-averaged ERP for randomly selected fixations from Experiment 2 before and after artifact correction. Topographic maps show mean amplitude in the first 500 ms following the fixation.

erenced to a common average reference offline. For the pre-target word, 650 trials (11.1%) were discarded because it was skipped and another 62 (1.2)% due to artifacts in the EEG. 1691 observations (28.8%) for the target word were removed due to skipping and 51 (1.2)% due to contaminated EEG. At the post-target word, 1704 trials (29%) were lost due to skipping and 42 (1%) due to contaminated EEG.

2.2.2 Results

2.2.2.1 Comprehension

As in Experiment 1, there were no effects on response accuracy in the comprehension questions. Accuracy was 84.0% across dependency types and lengths. In Experiment 2, participants fared best in long antecedent-pronoun dependencies (85.1%), followed by short object-verb (84.2%), short antecedent-pronoun (83.9%), and long object-verb dependencies (82.9%). As in Experiment 1, neither dependency type nor distance had a significant effect on response accuracy.

2.2.2.2 Eye Movements

As in Experiment 1, there was a substantial difference in skipping rates between the two dependency types (2.9% object-verb vs. 62.8% antecedent-pronoun). As a consequence, we again took the pre- and post-target words into account for analyses.

The eye movement record showed an effect of dependency type in regression probability at the pre-target word ($b = 0.19$, 95% CI $[0.03,\ 0.35]$, $z = 2.36$), the target word ($b = -0.50$, 95% CI $[-0.72,\ -0.29]$, $z = -4.62$), and the post-target word ($b = 0.97$, 95% CI $[0.75,\ 1.19]$, $z = 8.61$): Regression probability was higher in antecedent-pronoun dependencies at the pre- and post-target word but lower at the target word. Gaze durations were shorter at the target word ($b = -0.36$, 95% CI $[-0.39,\ -0.34]$, $t = -26.85$) and longer at the post-target word in antecedent-pronoun dependencies. First fixation durations were also shorter in antecedent-pronoun than in object-verb dependencies at the target word ($b = -0.11$, 95% CI $[-0.14,\ -0.09]$, $t = -8.61$). There were no effects of distance in the eye movement record of Experiment 2.

Table 2.10: Summary statistics for the mixed-effects regression of first fixation duration as a function of dependency type and distance in Experiment 2. Estimates (b) and 95% confidence intervals (CI) are on the log scale, statistical significance is indicated by bold face.

	b	95% CI	t
Pre-Target Word			
Type	−0.01	[−0.03, 0.01]	−1.1
Distance	0.00	[−0.02, 0.02]	0.0
Type × Distance	0.01	[−0.03, 0.05]	0.7
Target Word			
Type	**−0.11**	**[−0.14, −0.08]**	**−8.6**
Distance	−0.01	[−0.04, 0.02]	−0.9
Type × Distance	−0.03	[−0.08, 0.02]	−1.1
Post-Target Word			
Type	0.00	[−0.02, 0.02]	0.3
Distance	0.00	[−0.02, 0.02]	−0.1
Type × Distance	0.00	[−0.04, 0.04]	0.2

Table 2.11: Summary statistics for the mixed-effects regression of gaze duration as a function of dependency type and distance in Experiment 2. Estimates (b) and 95% confidence intervals (CI) are on the log scale, statistical significance is indicated by bold face.

	b	95% CI	t
Pre-Target Word			
Type	−0.02	[−0.04, 0.00]	−1.5
Distance	−0.02	[−0.04, 0.00]	−1.3
Type × Distance	0.01	[−0.04, 0.06]	0.3
Target Word			
Type	**−0.36**	**[−0.39, −0.33]**	**−26.9**
Distance	0.00	[−0.03, 0.03]	−0.2
Type × Distance	−0.02	[−0.07, 0.03]	−0.8
Post-Target Word			
Type	**0.16**	**[0.13, 0.19]**	**11.2**
Distance	0.00	[−0.03, 0.03]	0.2
Type × Distance	−0.01	[−0.06, 0.04]	−0.2

Table 2.12: Summary statistics for the mixed-effects regression of regression probability as a function of dependency type and distance in Experiment 2. Estimates (b) and 95% confidence intervals (CI) are on the logit scale, statistical significance is indicated by bold face.

	b	95% CI	z
Pre-Target Word			
Type	**0.19**	**[0.03, 0.35]**	**2.4**
Distance	−0.05	[−0.21, 0.11]	−0.6
Type × Distance	0.13	[−0.19, 0.45]	0.8
Target Word			
Type	**−0.50**	**[−0.71, −0.29]**	**−4.6**
Distance	0.07	[−0.14, 0.28]	0.6
Type × Distance	0.15	[−0.27, 0.57]	0.7
Post-Target Word			
Type	**0.97**	**[0.75, 1.19]**	**8.6**
Distance	−0.14	[−0.36, 0.08]	−1.2
Type × Distance	0.26	[−0.18, 0.70]	1.2

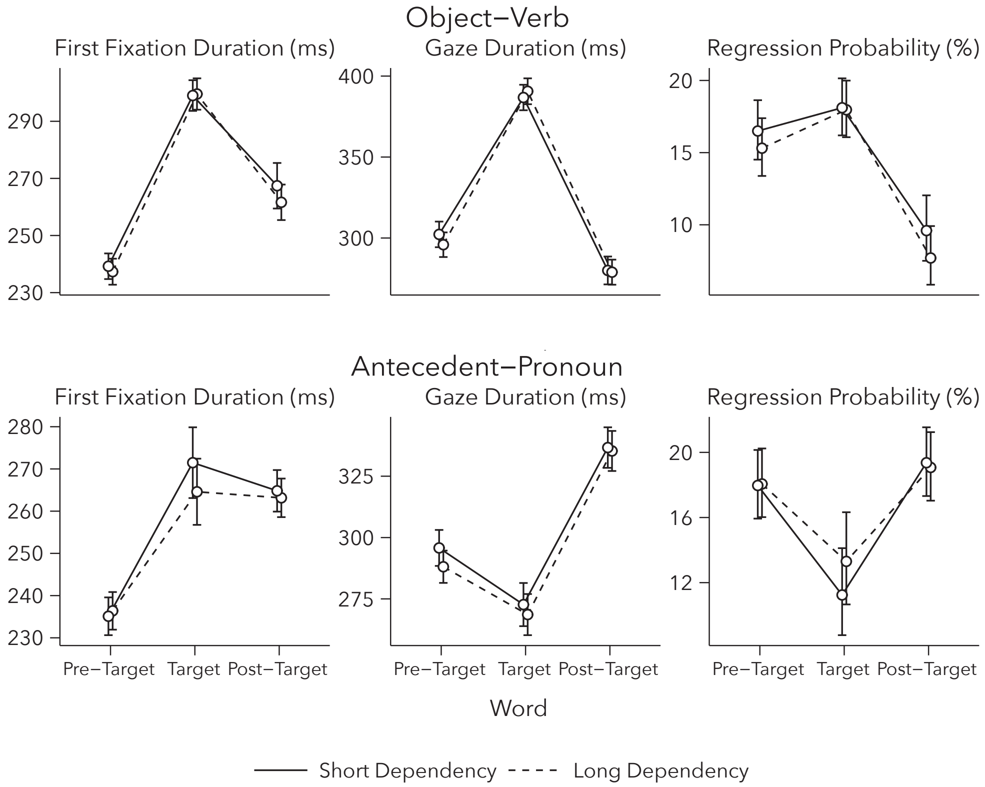

Figure 2.6: First fixation duration, gaze duration, and regression probability in Experiment 2 as a function of dependency type and distance on the pre-target, target, and post-target word (solid lines for short and dashed lines for long dependencies). Error bars are 95% confidence intervals that were computed after removing variance due to different participants and items.

2.2.2.3 ERP

There were no effects of distance in the ERP at the pre-target, target, or post-target word for either dependency type.

2.2.3 Discussion

In Experiment 2, the distance manipulation was no longer confounded with the complexity of the clause immediately preceding the target word. Without this confound, the locality effects from Experiment 1 disappeared entirely. This sug-

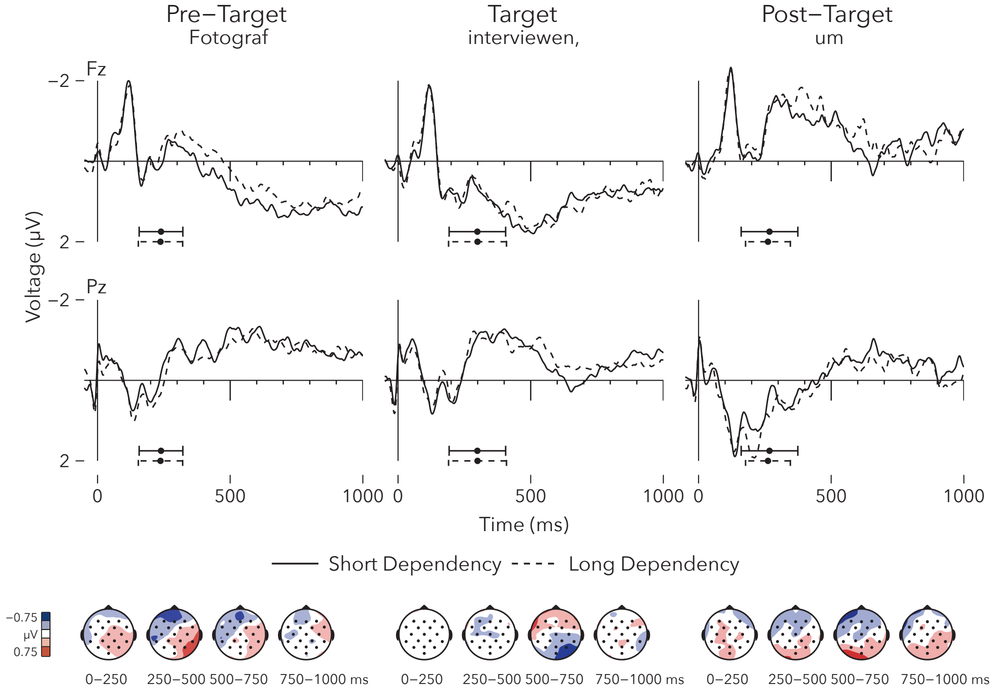

Figure 2.7: Grand-averaged ERP at the pre-target, target, and post-target word in object-verb dependencies for electrodes Fz and Pz in Experiment 2, low-pass-filtered at 30 Hz. Solid lines show amplitudes for short and dashed lines for long dependencies. Horizontal error bars show mean first fixation duration per condition at each word, plus and minus one standard deviation. Topographic maps show amplitude differences (long minus short) in successive time windows of 250 ms length. Averages are based on 2557 observations for the pre-target word (1280 short, 1277 long), 2990 observations for the target word (1489 short, 1501 long), and 1385 observations for the post-target word (683 short, 702 long).

gests that these effects were not driven by dependency length but by the complexity of the noun phrase preceding its head. It is relatively surprising that we did not observe any distance-related effects, considering that locality effects have been reported before (Gibson, 1998; Levy & Keller, 2012; Matchin et al., 2014; Santi & Grodzinsky, 2007). A possible explanation is that, in the context of the rather complex sentences of the present study, the subtle distance manipulation may not have been strong enough to elicit a robust effect. Furthermore, the relatively weak effects in the ERP may have been covered by the large amount of noise in coregistered eye movements and EEG. To rule out the second possibility, we conducted a third experiment using the self-paced reading paradigm.

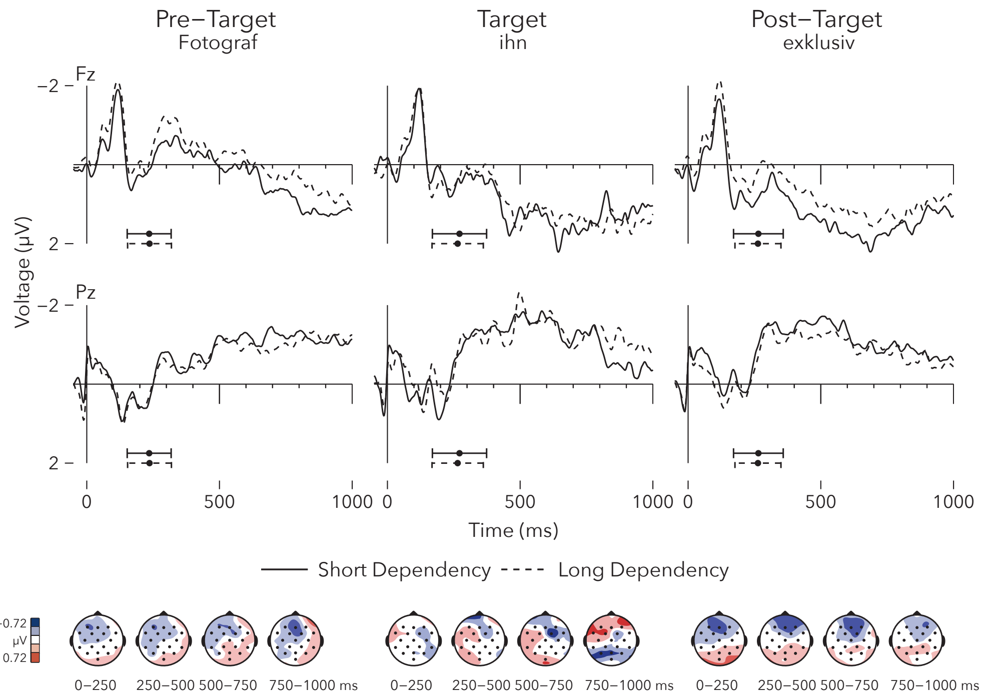

Figure 2.8: Grand-averaged ERP at the pre-target, target, and post-target word in antecedent-pronoun dependencies for electrodes Fz and Pz in Experiment 2, low-pass-filtered at 30 Hz. Solid lines show amplitudes for short and dashed lines for long dependencies. Horizontal error bars show mean first fixation duration per condition at each word, plus and minus one standard deviation. Topographic maps show amplitude differences (long minus short) in successive time windows of 250 ms length. Averages are based on 2611 observations for the pre-target word (1299 short, 1312 long), 1148 observations for the target word (565 short, 583 long), and 2749 observations for the post-target word (1373 short, 1376 long).

2.3 Experiment 3

In Experiment 2, we replicated Experiment 1 with altered materials and no longer observed locality effects in either dependency type. To determine whether this lack of an effect was due to a low signal-to-noise ratio, we repeated the experiment using non-cumulative self-paced reading. Here, readers are forced to read every word, eliminating a large amount of variance from different reading trajectories as well as the large data loss from skipping the pronoun in antecedent-pronoun dependencies.

2.3.1 Methods

2.3.1.1 Participants

63 undergraduates from the University of Potsdam student population participated in Experiment 3 for course credit or payment. They were between 18 and 45 years old ($M = 25$) and had normal or corrected-to-normal vision. None of them had participated in Experiment 1 or Experiment 2.

2.3.1.2 Materials

We used the materials from Experiment 2 but presented some words together on the screen (see next section for the presentation procedure). For example, determiners or possessive pronouns and the respective noun *the policeman* or *his colleague* were presented in one chunk.

2.3.1.3 Procedure

In a non-cumulative, centered version of the self-paced reading paradigm, participants read the sentences at their own pace word by word in the center of a screen; the software *Linger* (Rohde, 2003) was used to present the stimuli and record reading times. Each trial began with a crosshair in the center of the screen. When participants pressed a button, the crosshair was replaced with the first word of the sentence, and upon each successive button press, the next word of the sentence appeared until the end of the sentence was reached. Following every sentence, participants answered a yes/no comprehension question about its content with a response button. The experiment lasted around one hour.

2.3.1.4 Analysis

We removed reading times above 2000 ms because those likely reflect attentional drifts or technical errors. This led to the loss of 0.7% of the data. To further reduce non-normality of the residuals, we log-transformed the remaining reading times for analysis. Reading times at the pre-target, target, and post-target word were analyzed with linear mixed-effects regressions using the package *lme4* (Bates et al., 2014) in *R* (R Core Team, 2013). The initial configuration of each model had a fixed factor for type and distance as well as random factors for participants and items with random slopes for type and distance within each random factor. We reduced each model until it converged.

2.3.2 Results

Reading times at the pre-target word did not differ between dependency types, and there were no effects of distance. We observed a main effect of dependency type at the target word with shorter reading times at pronouns compared to verbs ($b = -0.17$, 95% CI $[-0.18, -0.15]$, $t = -21.86$); this effect was also present at the post-target word ($b = -0.09$, 95% CI $[-0.11, -0.08]$, $t = -13.77$). Additionally, there was a main effect of distance at the post-target word (see Figure 2.9): Long dependencies elicited longer reading times ($b = 0.01$, 95% CI $[0.00, 0.03]$, $t = 2.06$).

2.3.3 Discussion

Experiment 3 showed locality effects for both object-verb and antecedent-pronoun dependencies, which is in support of memory-based accounts of sentence processing. The observed effects were small in magnitude, which may have obscured them in the preceding coregistration studies.

2.4 General Discussion

We conducted three experiments to study the effect of increased distance on dependency resolution. Current models of sentence processing make qualitatively different predictions for the investigated dependencies between antecedent-pro-

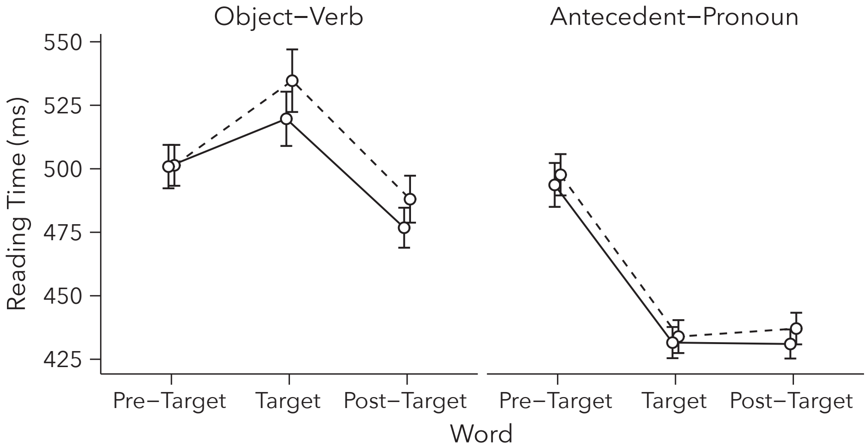

Figure 2.9: Reading times in Experiment 3 as a function of dependency type and distance at the pre-target, target, and post-target word: solid lines for short and dashed lines for long dependencies. Error bars are 95% confidence intervals. The plotted data are partial effects from linear mixed-effects regressions (after removing variance due to different participants and items).

noun and object-verb: Increasing the distance between their elements may lead to slow-downs, speed-ups, or no effect whatsoever. In contrast to previous studies, we manipulated the distance within these dependencies without altering the position of the dependency head within the sentence. We used integrated eye movement and EEG recordings to obtain electrophysiological and behavioral data from natural reading.

The first experiment yielded locality effects in eye movements and ERPs for both object-verb and antecedent-pronoun dependencies. However, these results were tainted by a confound. The complexity of the phrase preceding the dependency head varied systematically between short and long dependencies. In Experiment 2, we used unconfounded materials but failed to find locality effects. This runs counter to a growing body of research reporting locality effects in these or similar configurations (e.g., Gibson, 1998; Levy & Keller, 2012; Levy, Fedorenko, & Gibson, 2013). In Experiment 3, we showed that our materials can also elicit locality effects but that these are small in magnitude. In Experiment 2, they might have been lost in the low signal-to-noise ratio of EEG and eye tracking signals.

These results raise a number of questions. Does the confound at the pre-target region affect antecedent-pronoun and object-verb dependencies differently? Why did we fail to detect locality effects in Experiment 2 when the same materials elicited clear effects in Experiment 3. Can coregistration be used to investigate subtle effects in sentence processing at all? The following discussion offers tentative answers to these questions.

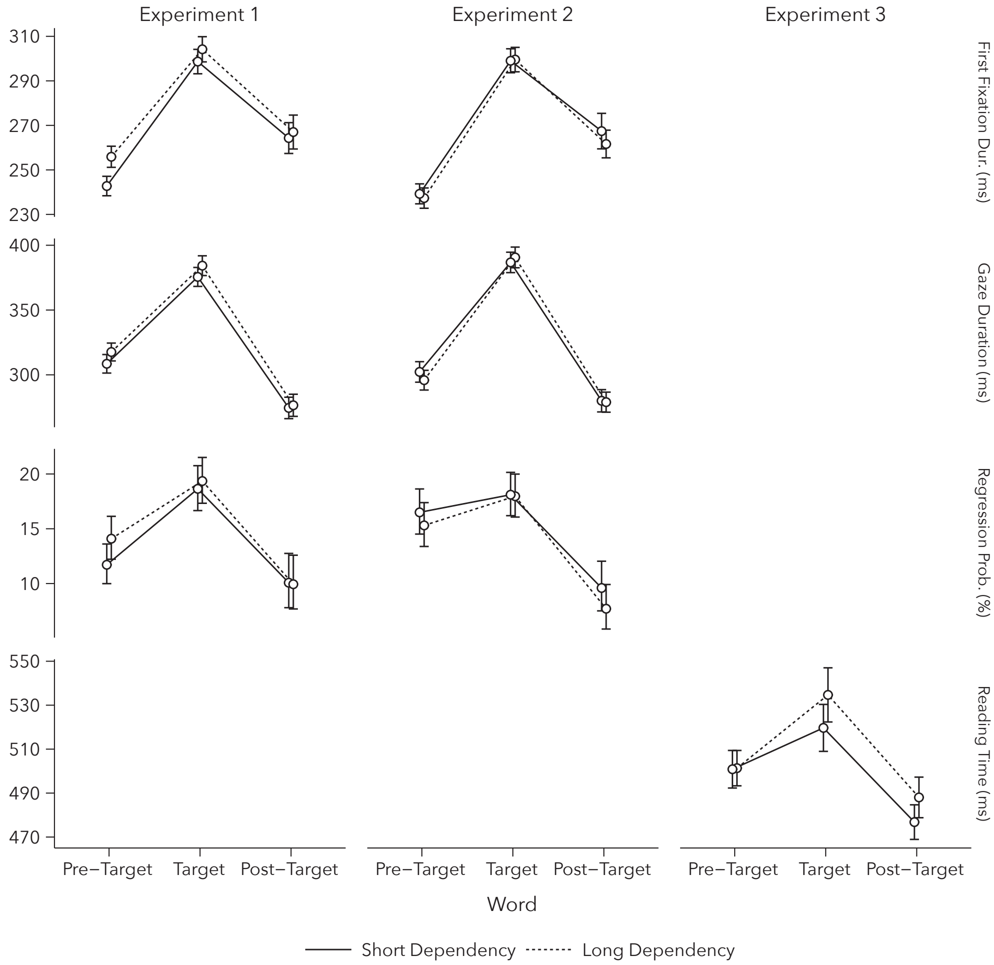

Figure 2.10: Eye movement measures and reading times for object-verb dependencies in Experiment 1, 2, and 3 as a function of dependency type and distance on the pre-target, target, and post-target word: solid lines for short and dashed lines for long dependencies. Error bars are 95% confidence intervals, the plotted data are partial effects from linear mixed-effects regressions (after removing variance due to different participants and items).

The locality effects from Experiment 1 disappeared in Experiment 2 after we had removed a confound at the pre-target word. However, visual inspection of the data suggests that deconfounding the pre-target word had different effects on object-verb and antecedent-pronoun dependencies. Figure 2.10 shows that the general pattern in object-verb dependencies remained the same across the three experiments. In Experiment 1, there was a small locality effect at the pre-target word that spilled over to the target and post-target word. In Experiment 2, this effect went away, such that the fixation measures for long object-verb dependencies dropped slightly below those of short conditions. The overall pattern, however, is highly similar in both experiments and also similar to reading times in Experiment 3.

The picture looks rather different when it comes to antecedent-pronoun dependencies (Figure 2.11). In Experiment 1, we see clear locality effects at the pre-target word that are completely absent in Experiment 2. Unlike in object-verb dependencies, the direction of the changes differs between the dependent variables. In some measures, we observe what may be expected when the pre-target word is less complex: First fixation durations and gaze durations decrease for long conditions at the pre-target word. The locality effect in first fixation duration appears to shift to the target word in Experiment 2, but this visual impression did not translate into statistical significance. The reason may have been that readers spent less time at the now less complex pre-target word and were therefore more likely to show locality effects at the pronoun itself. This did not happen in gaze durations where the pattern at the target and post-target words are virtually identical in Experiments 1 and 2.

In regression probability, results for the short conditions differ between the first two experiments. In Experiment 1, we see locality effects at the pre- and post-target word but no effects at the target word. In Experiment 2, these effects go away, albeit not due to lower regression probability in long conditions. Instead, regression probabilities for short conditions increase to the level of long conditions compared to Experiment 2. This is surprising because we made only minor changes to short conditions. Similar to first fixation durations, the effect seems to have shifted from the pre-target word to the target word, but again, the impression is not supported by statistics. Finally, Experiment 3 shows an entirely different pattern than Experiments 1 and 2, which is presumably due to the fact that readers could not skip the pronoun in self-paced reading.

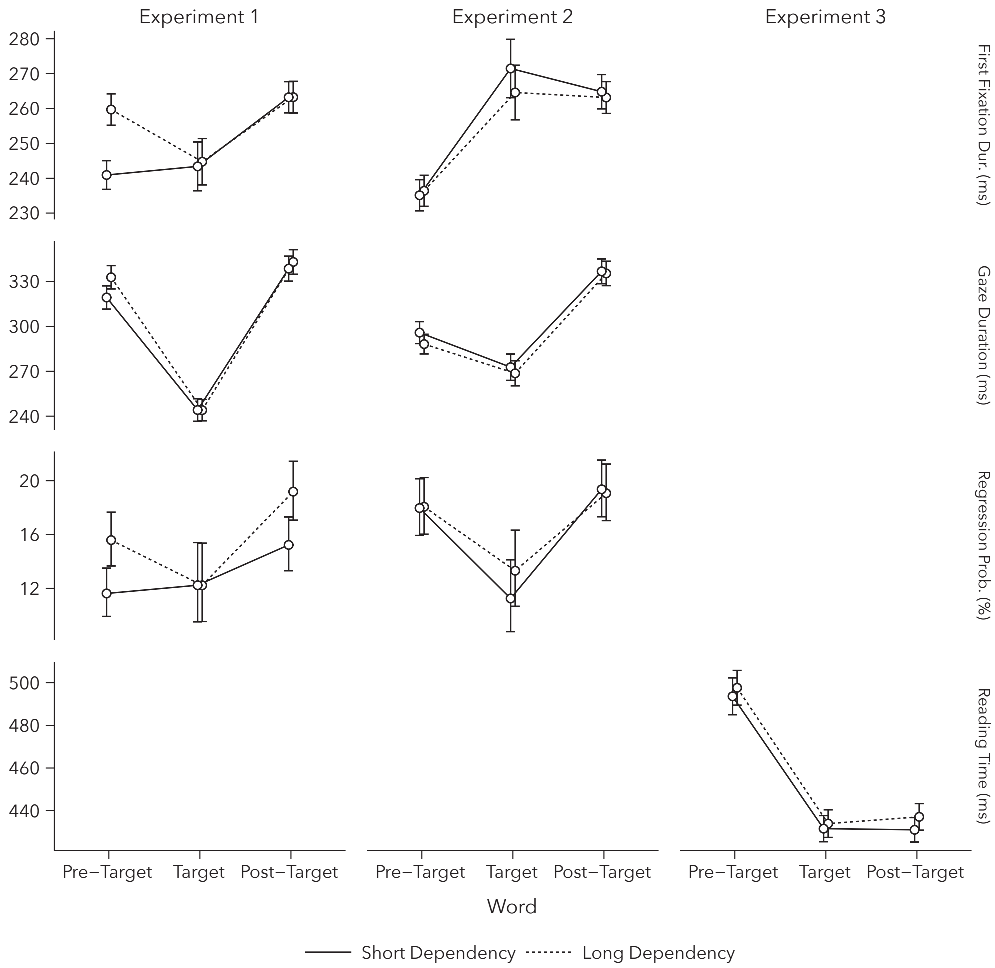

Figure 2.11: Eye movement measures and reading times for antecedent-pronoun dependencies in Experiment 1, 2, and 3 as a function of dependency type and distance on the pre-target, target, and post-target word: solid lines for short and dashed lines for long dependencies. Error bars are 95% confidence intervals, the plotted data are partial effects from linear mixed-effects regressions (after removing variance due to different participants and items).

The most obvious difference between object-verb and antecedent-pronoun dependencies is that readers were much more likely to skip the target word in antecedent-pronoun dependencies. Presumably, they processed the pronoun parafoveally to subsequently skip it. Effects from resolving the dependency would therefore show already at the pre-target word or, as spill-over, at the post-target word. Thus, changing the pre-target word has a larger effect in antecedent-pronoun dependencies than in object-verb dependencies, where most of the processing happens at the target word.

There is another difference between Experiments 1 and 2 that may explain the different results. In Experiment 1, complexity was manipulated symmetrically in the main and subordinate clause: In both main and subordinate clause, the second NP was rendered more complex. In Experiment 2, we varied complexity in an asymmetric fashion. Now, we altered the complexity of the second NP in the main clause and of the first NP in the subordinate clause. There is some evidence that readers rely on superficial sentence properties like order of mention (Gernsbacher & Hargreaves, 1988) and structural parallelism (Grober, Beardsley, & Caramazza, 1978). It is therefore conceivable that the salience of neighboring phrases somehow affects antecedent retrieval. Although this may explain the different results in long conditions, it falls short with regards to the increased regression probabilities in short antecedent-pronoun dependencies in Experiment 2.

Experiment 3 produced locality effects that were in line with a memory-based account of sentence processing. We failed to find these effects in Experiment 2, where we used the same items in coregistration. Several factors may have contributed to this difference. First, as mentioned above, the signal-to-noise ratio is poorer in the signals monitored in coregistration studies than that of self-paced reading. Therefore, assuming that the distance manipulations elicit effects of identical magnitude in both settings, they might come out significant in self-paced reading but not in coregistration.

Second, loosely related to the signal-to-noise ratio, we invited fewer participants for Experiment 2 than for Experiment 3. Considering the lower technical effort as well as shorter session length, it was feasible to collect more data for Experiment 3. However, this complicates the interpretation of any differences between Experiment 2 and Experiment 3. With an estimated effect size of 0.01 and a standard error of 0.01, we obtained a power of 0.54 in Experiment 3. Assuming the same effect size and variance, we would obtain a power of 0.30 for the sample size in Experiment 2. Thus, Experiment 2 may simply have been underpowered.

The remaining question is also the guiding question of the present study: Is coregistration a useful tool for sentence processing research? The answer is not straightforward and depends heavily on the question at hand. It has been shown that coregistration is invaluable for delineating basic memory and recovery mechanisms in sentence processing. However, the experiments in those studies used outright ungrammatical or nonsensical sentences to elicit reliable and large effects that are robust against even large amounts of noise. The effects that are typically observed in studies on, for example, dependency resolution in well-formed sentences, are much smaller and more volatile. As we saw in the present study, those effects do suffer from the low signal-to-noise ratio in coregistration studies. Thus, despite its benefits, coregistration may not be sensitive enough to investigate distance effects in dependency resolution.

Experiments 1 and 2 did not reveal reliable effects in fixation-triggered ERPs related to dependency resolution. The data nevertheless suggest that the method may have sufficient power to disclose hidden processes during sentence comprehension. The ERPs in Figures 2.3 and 2.4 reveal a striking difference between the two dependency types. At the post-target word, the ERPs in antecedent-pronoun and object-verb dependencies developed a qualitatively different time course. In object-verb dependencies, we see a pronounced frontal negativity while in antecedent-pronoun dependencies, we see a frontal positivity. This was neither predicted nor expected, and it might suggest potential future routes of research.

Frontal negativities as the one seen here have been found functionally related to expectation processes as typically realized in the *contingent negative variation* paradigm (Brunia, 2004; Rohrbaugh & Gaillard, 1983; Walter, Cooper, Aldridge, McCallum, & Winter, 1964). In object-verb dependencies, the word that is fixated when the negativity develops is the preposition *um* which introduces a consecutive clause. Thus, the word most likely triggers an expectation for the upcoming clause, which may become manifest in the frontal negativity.

In the antecedent-pronoun condition, the fixated word that triggers the positivity is an adverb. Pronounced positivities have been linked with inhibition processes (as the SPS when the syntax structure has to be repaired or revised, the P300 when an encountered event does not meet the context dependent expectations). A language-related phenomenon is the post-N400 positivity, which appears when words do not fit with a context dependent semantic expectation. For example, DeLong, Quante, and Kutas (2014) report an anterior post-N400 positivity to plausible, low cloze probability sentence-medial words. They argue that this might indicate executive control mechanisms (in pre-frontal cortex) which are needed to override prepotent responses when a memory must be selectively retrieved in the face of other competing memories. The post-target words in the sentence constructions used here in antecedent-pronoun conditions may have some of the features described by DeLong et al. as prerequisites of such a frontal positivity. This observation, although incidental and not tied to immediately comparable conditions, could possibly spark future research to investigate such processing differences which become manifest in distinct ERP and/or distinct fixation patterns (see Chapter 3 of this thesis).

3

Brain Responses to World-Knowledge Violations: A Comparison of Stimulus- and Fixation-Triggered Event-Related Potentials and Neural Oscillations

The contents of this chapter are published as:

Metzner, P., von der Malsburg, T., Vasishth, S., & Rösler, F. (2015). Brain responses to world-knowledge violations: A comparison of stimulus- and fixation-triggered event-related potentials and neural oscillations. *Journal of Cognitive Neuroscience*, *27*(5). doi:10.1162/jocn_a_00731.©

The assessment of eye movements and EEG has helped to advance psycholinguistic research substantially, but both methods have limitations. In eye tracking experiments, participants can read freely and adapt to the characteristics of the text. EEG measures have an excellent temporal resolution and provide information about processes in behaviorally mute epochs. However, for technical reasons, most EEG studies have participants read sentences in a rather unnatural, word-by-word fashion (*rapid serial visual presentation*: RSVP). This allows researchers to avoid a number of problems related to natural reading in EEG, the most prominent of which are artifacts induced by eye movements. Despite this obstacle, combining both methods appears to be an obvious way to overcome their respective weaknesses. Indeed, recent work using ERPs in natural reading has shown that the technical problems can be handled and that effects observed in

RSVP are replicated in natural reading studies (Dimigen et al., 2011; Hutzler et al., 2007; Kretzschmar et al., 2009). The current data show that, despite the general feasibility of the approach, RSVP and natural reading elicit qualitatively different results when it comes to oscillatory brain dynamics.

Most language-related EEG research is based on ERPs. Because single-trial EEG is dominated by noise, participants are exposed to many instances of the same kind of stimulus. Subsequently, the data is averaged across trials for each condition. The underlying assumption is that the brain response to the stimulus is present in each trial while unrelated signals will be canceled out in the averaging process. The resulting effects can be categorized by polarity (negative-going or positive-going), onset, offset, peak latency, and distribution on the scalp. The most widely reported ERP signature in psycholinguistic research is the N400 effect (Kutas & Hillyard, 1980), a negative deflection at centro-parietal electrodes, ranging approximately from 300 to 500 ms. It is sensitive to the effort required to process a word and is typically seen in response to violations involving word meaning. For instance, *socks* in *He spread the warm bread with socks* elicits a larger negativity than *butter* in the same position. The N400 effect is not restricted to semantic violations but can also be observed following statements that do not match common world knowledge (Hagoort et al., 2004; Hald, Steenbeek-Planting, & Hagoort, 2007). The amplitude of the effect also varies as a function of word frequency (Dambacher, Kliegl, Hofmann, & Jacobs, 2006; Van Petten & Kutas, 1990) and predictability (Dambacher et al., 2006; Dimigen et al., 2011).

The N400 effect has been replicated successfully in natural reading situations. Kretzschmar et al. (2009) used graded antonyms like *The opposite of black is white/yellow/nice* and found an N400 effect with a peak around 300 ms for both unpredicted completions (*yellow* and *nice*). Dimigen et al. (2011) found an N400 effect with a centro-parietal distribution and a peak at 384 ms (at electrode Pz) for low-predictability words. Dimigen et al. also reported earlier effects with a similar topography but those were not statistically significant.

Time-frequency representations (TFR) provide an alternative framework for EEG analysis. Like any signal that varies over time, the EEG can be decomposed into a number of oscillations of different wavelengths and phase shifts. This can be done with a fast Fourier transform (FFT; Cooley & Tukey, 1965), wavelet analysis (Schiff et al., 1994), or multitapering (Mitra & Pesaran, 1999). The power of particular frequency ranges in the resulting spectrum is informative about the underlying cognitive processes. For instance, increased theta waves in the range from 4 to 7 Hz and alpha waves in the range from 8 to 13 Hz are often observed in the context of memory processes (Klimesch, 1999).

A number of studies have investigated the spectral signature of the conditions that evoke an N400 effect. Frisch and Schlesewsky (2001) investigated the interaction of grammaticality and animacy using sentences like *Paul fragt sich, welchen Angler der Jäger gelobt hat* (*Paul asks himself which angler [ACC] the hunter [NOM] praised has*) and *Paul fragt sich, welcher Angler der Jäger gelobt hat* (*Paul asks himself which angler [NOM] the hunter [NOM] praised has*), where two nouns are case-marked as subject, rendering the sentence ungrammatical; *Zweig* (*twig*) replaced *Jäger* for the inanimate conditions. They observed an N400 following grammaticality violations only if both noun phrases were animate. Roehm et al. (2004) reanalyzed the data ($N = 16$) from Frisch and Schlesewsky (2001) and reported an N400 effect that was not described in the original paper. It appeared when an inanimate noun phrase with subject case followed an animate noun phrase with object case. Roehm et al. investigated oscillations in the delta and theta band (1–7.5 Hz) for the two N400 effects that looked similar in the time-domain analysis and reported increased power in the upper theta band (6–7.5 Hz) for inanimate vs. animate conditions and increased power in the lower theta band (3.5–5 Hz) for ungrammatical vs. grammatical conditions. Both ungrammatical conditions showed increased power in the delta range (1–3.5 Hz) in comparison to the animate grammatical condition, but the inanimate grammatical condition did not. Roehm et al.'s results thus show that a manipulation that elicits similar effects in the time domain can have entirely different responses in the frequency domain.

Hagoort et al. (2004) used ERP and TFR analyses to investigate the access to semantic knowledge and world knowledge in language processing. They compared the response to an adjective in semantically correct sentences (*The Dutch* *trains are yellow and very crowded*), in semantically ill-formed sentences (*The Dutch trains are sour and very crowded*), and in sentences that were semantically valid but incongruent with world knowledge (*The Dutch trains are white and very crowded*). The ERPs for the different violation types were virtually identical but the traces in the oscillatory response were distinct. Both violations elicited an increase in theta band power, although this effect was stronger in semantic violations, whereas only world knowledge violations led to a marked increase in gamma power. Thus, the superficial similarity of the ERPs across conditions does not necessarily entail identical responses in the TFR.

Hald et al. (2006) conducted another experiment with semantic violations as in Hagoort et al. (2004). They found a stronger power increase at bilateral temporal electrodes in the theta range (3–7 Hz) and an increase in gamma power (around 40 Hz) at right frontal electrodes in semantically ill-formed sentences. In a previous study by Bastiaansen et al. (2005), theta activity at temporal sites has been linked to lexico-semantic access, which is consistent with Hald et al.'s (2006) results.

A number of other studies have investigated the use of world knowledge in online language processing. They cannot be reviewed here due to space restrictions but attest to how reliably these kinds of violations elicit behavioral and electrophysiological effects (see, e.g., Chwilla & Kolk, 2005; Hald et al., 2007; Menenti, Petersson, Scheeringa, & Hagoort, 2009; Rayner, Warren, Juhasz, & Liversedge, 2004; Warren & McConnell, 2007).

Using RSVP, Hagoort et al. (2004) and Roehm et al. (2004) reported varying spectral results in the context of superficially similar ERP effects. This raises the question whether the close correspondence of ERP effects in RSVP and natural reading translates into an analogous relation of oscillatory dynamics in the different presentation modes. We conducted two experiments where participants read world knowledge violations as in Hagoort et al. (2004). In Experiment 1, sentences were presented in a word-by-word fashion. In Experiment 2, participants read the same sentences in a natural reading setting while their eye movements and their EEG were recorded. If the spectral compositions of the EEG in serial presentation and natural reading are comparable, effects in the theta and gamma range as in Hagoort et al. (2004) should emerge in both experiments.

3.1 Experiment 1

3.1.1 Methods

3.1.1.1 Participants

We collected data from 32 self-reportedly right-handed members of the University of Potsdam student population (24 women, 8 men). They had normal eyesight or wore corrective lenses and were between 19 and 49 years old ($M = 26$). They were not told what the study was about. Written consent was collected from all participants and they were compensated with course credit or money.

3.1.1.2 Materials

The experimental material comprised 120 minimal pairs of German sentences with a control condition and a version that was incongruent with common world knowledge. For instance, *Paris is the capital of France* would be a control sentence that is consistent with common world knowledge whereas *Rome is the capital of France* is not. The target word *France* was held constant across conditions and an earlier part of the sentence was manipulated to render the sentence incorrect (*Rome* instead of *Paris*). The sentences had an average length of 7.6 words ($SE = 0.2$) and the target word had an average length of 8.1 characters ($SE = 0.2$). The items were split into two lists in a latin-square design and pseudorandomized. Each participant thus saw only one version of each item. Another 180 items from an unrelated experiment on sentence processing were interleaved with the material for the current study. They were on average 18.1 words ($SE = 0.04$) long.

3.1.1.3 Procedure

Participants signed a consent form at the beginning of the session and were seated in a shielded booth approximately 60 cm from the stimulus display. After the electrode cap was prepared, participants read five practice sentences to familiarize themselves with the procedure. When they had finished the practice trials, they proceeded with the experiment. For Experiment 1, we adopted the presentation procedure described in Hagoort (2003). Sentences were presented word-by-word in the center of a screen with a resolution of 1680×1050 pixels. Each word was presented for 300 ms in 28 point Arial, followed by a blank display that lasted 300 ms. Next, a blank display with a variable duration between one and two seconds preceded a comprehension question that was answered with the press of a button. Another 1150 ms intervened between the response and the onset of the next trial. Every 20 sentences, participants received feedback about their performance in the comprehension questions of the last block. After 90, 180, and 270 sentences, they took a short break. An experimental session lasted for approximately two hours, including preparation and debriefing.

3.1.2 Recording and Analysis

3.1.2.1 Recording

The EEG was recorded from 32 Ag/AgCl electrodes mounted in a 10-20 design (Jasper, 1958) in a shielded electrode cap (Advanced Neuro Technology, Enschede, Netherlands). Eye movements and blinks were monitored with additional bipolar electrodes on the left and right outer canthus and the infraorbital ridges of the right eye. Both EEG and EOG were recorded with a low-pass filter with a cutoff at 138.24 Hz and digitized at a sampling rate of 512 Hz. Recordings were initially referenced to the left mastoid and later converted a common average reference. Impedances were kept between 5 and 10 kΩ at all times.

3.1.2.2 Preprocessing

The EEG data was preprocessed in BrainVision Analyzer (Brain Products, Munich, Germany) where the signal was first resampled to 500 Hz and filtered (0.3–100 Hz band-pass, 50 Hz notch). Eye movements were corrected with an Independent Components Analysis (ICA; Jung et al., 2000), the specifics of which are described in the methods section of Experiment 2. The corrected signal was segmented from −1000 to 2000 ms relative to stimulus onset. Segments with muscle artifacts or slow drifts were discarded, which lead to the loss of 70 trials (1.9%). All further processing steps and analyses were performed in *R* (R Core Team, 2013), where the data were baseline-corrected relative to a 100 ms interval preceding the stimulus for the ERP analyses.

For the spectral analysis, Fourier transformations were performed for overlapping windows of 1000 ms length in 10-ms steps from stimulus onset until 2000 ms following the stimulus. This resulted in 100 frequency spectra for each trial and electrode that were related to a pre-stimulus baseline window of 1000 ms length by calculating the change in power for each frequency in decibels. To avoid spectral leakage at the window edges, each window was filtered by means of a Tukey window (Tukey, 1967) with symmetric Hann functions in the raising and falling part, comprising 10% of the window length. The FFT does not produce power estimates for single time points but for a time span so latencies and durations for spectral effects are reported relative to these windows. The onset of the first time window in which an effect appears is treated as its onset. The duration of an effect is the time from its onset until the onset of the last time window wherein the effect is present. Thus, an effect from 200 ms to 400 ms is an effect that begins in the time window from 200 to 1200 ms and ends in the time window from 400 to 1400 ms.

We analyzed the EEG with a cluster-based random permutation procedure (Maris & Oostenveld, 2007). Although we anticipated an ERP effect with a centro-parietal distribution and a peak around 400 ms, prior research has shown that the timing and topography of an effect can differ slightly between natural reading and serial presentation (Dimigen et al., 2011; Kretzschmar et al., 2009). The cluster-permutation test offers an elegant approach to finding effects without distributional assumptions while controlling the multiple-comparisons problem. The test was implemented as follows. First, for each electrode and time point, a pairwise *t* test of the two conditions (congruent vs. incongruent) was performed with one data point per participant and condition. Next, spatiotemporal clusters of sample statistics above a certain threshold were formed with connected-components labeling (Samet & Tamminen, 1988).[2] The *t*-values of all samples in a cluster were then summed up to yield a cluster statistic. To assess a cluster's significance, conditions were repeatedly and randomly swapped within participants and the cluster-permutation procedure was then performed on the resulting data set. From each of 5000 iterations, the maximal cluster statistic was used to create a distribution representing the null hypothesis. We considered a cluster significant if its cluster statistic fell in the lower 2.5th or upper 97.5th percentile of this distribution.

For power spectra, separate cluster-permutation tests for discrete frequency ranges were performed that were defined as follows: delta (1–3 Hz), theta (4–7 Hz), lower alpha (8–10 Hz), upper alpha (11–13 Hz), beta (14–30 Hz), and gamma (31–70 Hz). Within each frequency range, the average power change for each participant, electrode, time window, and condition entered into the analysis as described above for the ERP. Although it is common practice to relate frequency ranges to each participant's individual alpha frequency, this was not done in the current study. Since Hagoort et al. (2004) used fixed frequency bands such a step would have introduced another difference between the studies, further complicating any comparison.

[2]The choice of this threshold does not directly affect the significance testing because it is identical for the original clusters and the bootstrapping procedure. We used a standard alpha level of .05.

3.1.3 Results

3.1.3.1 Behavioral Data

Participants fared well on the comprehension questions, answering correctly in 92.9% of congruent and 94.7% of incongruent trials. The difference between conditions was not significant in a paired t test on subject averages ($t(31) = -1.83$, $p > .05$). All trials were used for the analysis, regardless of whether or not the response was correct.

3.1.3.2 ERP

The cluster-permutation test showed an increased negativity with a centro-parietal distribution from 238–810 ms (peak at 420 ms) in incongruent trials. Both timing and topography of this cluster are indicative of the anticipated N400 effect. Another cluster in the same time window indicated a positivity at frontal electrodes (322–532 ms, peak at 386 ms) and appeared to be the result of the common average reference. Following this cluster, there were four successive positivities at centro-parietal to parietal electrodes (576–742 ms, peak at 702 ms; 752–942 ms, peak at 906 ms; 1142–1204 ms, peak at 1182 ms; 1322–1380 ms, peak at 1350 ms). Lastly, there were two short-lived late negativities at fronto-polar and frontal electrodes (1150–1224 ms and 1316–1398 ms, peaks at 1190 and 1362 ms; Figure 3.1).

3.1.3.3 TFR

There was one marginally significant cluster in the TFR ($p < .06$). It indicated stronger synchronization in incongruent trials than in congruent trials in the theta range from 0 to 460 ms (peak at 260 ms) at frontal to fronto-central electrodes and was right lateralized. There were no other significant or marginally significant clusters in the TFR (Figure 3.2).

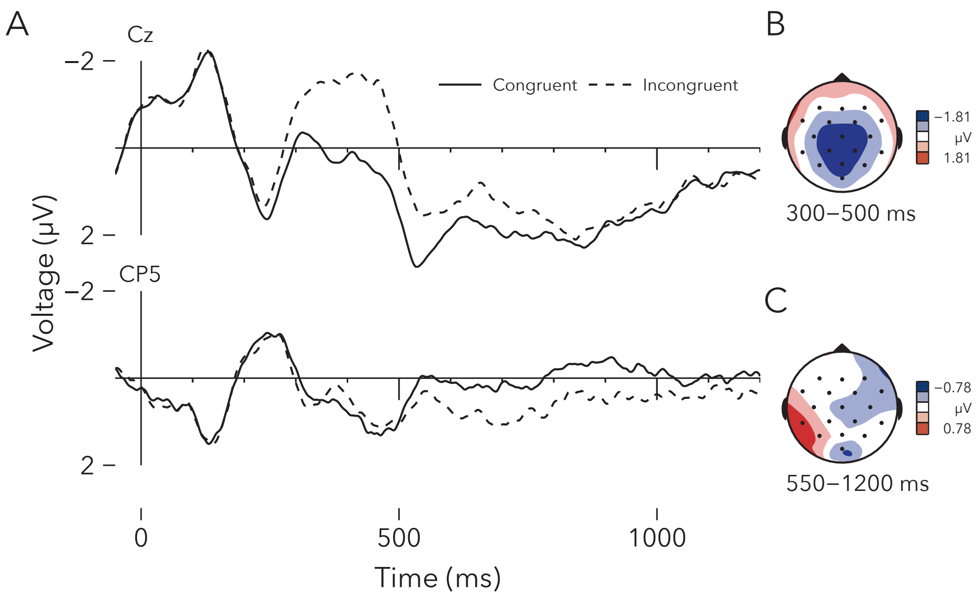

Figure 3.1: (A) Grand-averaged ERP for electrodes Cz (N400) and CP5 (late positivity) in Experiment 1, low-pass-filtered at 30 Hz. Solid lines show congruent amplitudes; dashed lines show incongruent amplitudes. (B) Topographic map of amplitude differences (incongruent minus congruent) in the N400 time window in Experiment 1 (300–500 ms). (C) Topographic map of amplitude differences (incongruent minus congruent) in the time window of the late positivities in Experiment 1 (550–1200 ms).

3.1.4 Discussion

In the ERP, Experiment 1 successfully replicated the N400 effect from Hagoort et al. (2004): There was a relatively larger negativity following world knowledge violations at centro-parietal electrodes with a peak around 420 ms. In the TFR, a power increase in the theta range at frontal electrodes was larger in incongruent than in congruent trials. While this effect was statistically only marginally significant, both the frequency range and direction of the effect are consistent with Hagoort et al. (2004). These results provide a validation of our materials and experimental setup and set the stage for Experiment 2.

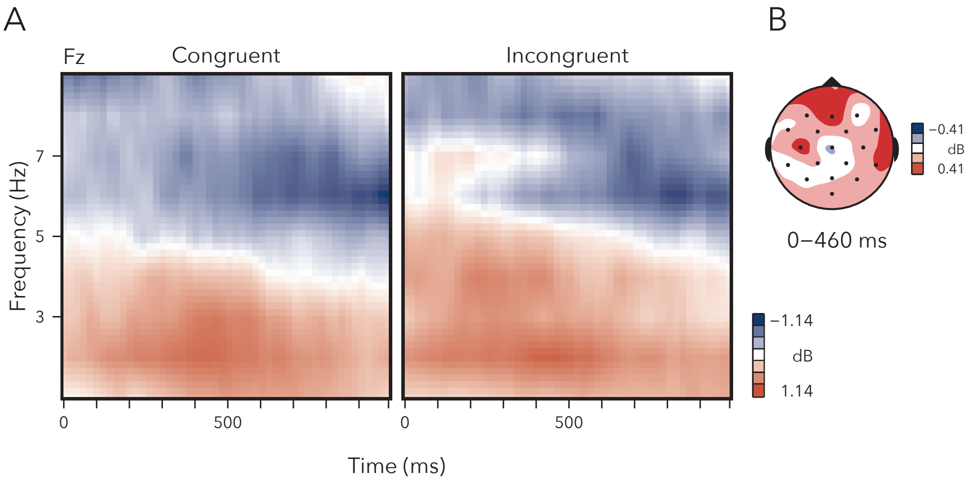

Figure 3.2: (A) Time-frequency plot of power changes in congruent and incongruent trials at electrode Fz in Experiment 1 (frequencies up to 10 Hz, smoothed with bilinear interpolation). (B) Topographic map of differences in power change in the theta range (4 to 7 Hz) in the time window of the theta effect in Experiment 1 (0 to 460 ms).

3.2 Experiment 2

Experiment 1 provided a validation of the German materials and the general setup by partially replicating the N400 effect and the theta power increase in Hagoort et al. (2004). In Experiment 2, we tested the same sentences in a natural reading setting with concurrent eye movement and EEG recordings to answer the question outlined in the introduction.

3.2.1 Methods

3.2.1.1 Participants

Fifty-two participants (37 women, 15 men) were recruited from the University of Potsdam student population (19–34 years, $M = 25$). All participants were right-handed by self-report, native speakers of German, and had normal or corrected-to-normal vision. They had not participated in Experiment 1 and were naive with regard to the aims of the study. They gave written informed consent to the procedure and received either course credit or money for compensation. Because of recording errors, we discarded all data from one participant and the EEG data from another two participants. This left eye tracking data from 51 and EEG data from 49 participants. To obtain a consistent data set for comparing eye movements and EEG, we used only the data from the 49 participants with a complete set of observations (i.e., eye movements and EEG).

3.2.1.2 Materials

The materials were the same as in Experiment 1. The items were presented together with an equal number of items for an unrelated experiment on sentence processing. Due to the nature of the other experiment, those sentences were more complex and longer than the sentences for the present experiment. Every sentence consisted of a main clause and a subordinate clause and had an average length of 22.4 words ($SE = 0.2$).

3.2.1.3 Procedure

Aside from a few details, the procedure was identical to the procedure in Experiment 1. After the electrode cap was prepared, the eye tracker was calibrated. Sentences were not presented word by word in the center of the display but left-justified, vertically centered on a single line in 26 point Arial; the display had a resolution of 1680×1050 pixels. To finish a sentence, participants fixated the bottom right corner of the display. We chose this method over a button press to prevent anticipatory eye movements to the beginning of the sentence. That way, regressions from the end of the sentence could be safely linked to the processing of that region. The sessions in Experiment 2 were slightly longer than in Experiment 1 at around 2.5 hours.

3.2.2 Recording and Analysis

3.2.2.1 Eye movements

Fixational eye movements were recorded with a desktop-mounted EyeLink 1000 (SR Research, Mississauga, Ontario, Canada) in remote mode. This allowed participants to sit comfortably without a chin rest, which reduced myogenic artifacts. Gaze position was sampled at 500 Hz from the right eye with a spatial resolution of 0.01° and an average accuracy of 0.54° in the vertical center of the screen.

We excluded fixations shorter than 20 ms and longer than 1200 ms from analysis, which led to the loss of 2.03% of all fixations. The remaining fixations were aggregated into the standard fixation measures first fixation duration, gaze duration, and regression probability. First fixation duration is the duration of the first progressive fixation on a word (i.e., coming from the left). Gaze duration is the duration of all fixations on a word from the first progressive saccade until the eye leaves the word again. Regression probability denotes the probability to make a regressive saccade from a word immediately after entering it with a progressive saccade (i.e., before leaving it to the right).

Fixation measures were analyzed with linear mixed-effects models using the package *lme4* (Bates, Maechler, Bolker, & Dai, 2013) in *R*. The target word (*France*) and the word preceding it (*of*) were analyzed if and only if they received a progressive fixation. Trials without a progressive fixation on any of these words were discarded. Before analysis, it was determined whether a variable had to be transformed in order to afford normality of the residuals and which transformation would stabilize variance (Box & Cox, 1964). Following this procedure, all eye movement analyses were performed on log-transformed duration variables. Binary responses (regression: yes or no) were analyzed using the logit link in a generalized linear mixed-effects model. All models were fit with varying intercepts and slopes for each fixed factor, including a correlation term, unless there was a failure to converge or the correlation estimated implied a degenerate variance covariance matrix; in such cases, the model was simplified until the model converged and had no degeneracy.

3.2.2.2 EEG

The EEG was recorded with the same setup as in Experiment 1 but measured to the common average reference already online. Like in Experiment 1, an ICA was used to identify brain activity related to horizontal and vertical eye movements. The ICA used a biased variant of the Infomax algorithm and was trained on the filler sentences. By means of this, it was ensured that eye movements from sentence reading featured in the training data but effects that were time-locked to fixational eye movements in the target sentences were not systematically removed. A classic PCA was used for preparatory sphering. All channels were included except for the mastoid electrodes. Components with a fronto-polar or bipolar frontal distribution were identified and removed from the signal because those components were assumed to represent vertical and horizontal eye movements, respectively. The components were related to the HEOG and VEOG by ascertaining that activity in those channels was minimized through the procedure. As another post-hoc validation, we computed the ERPs and TFRs of the signal before correction and submitted them to the same analyses as the corrected signal. We found neither N400 effects nor theta or gamma increases in this data. The success of the correction procedure is illustrated in Figure 3.3. After correction, trials with severe artifacts were removed from the data, which lead to the loss of 154 trials (3.1%). In an offline procedure, the first progressive fixation on a word was identified from the eye tracking data, and its time stamp was aligned with the EEG data with the help of synchronization markers at the beginning and end of each trial. In 18.5% of the experimental trials, the target word did not receive a progressive fixation. All following segmentation and preprocessing was the same as in Experiment 1.

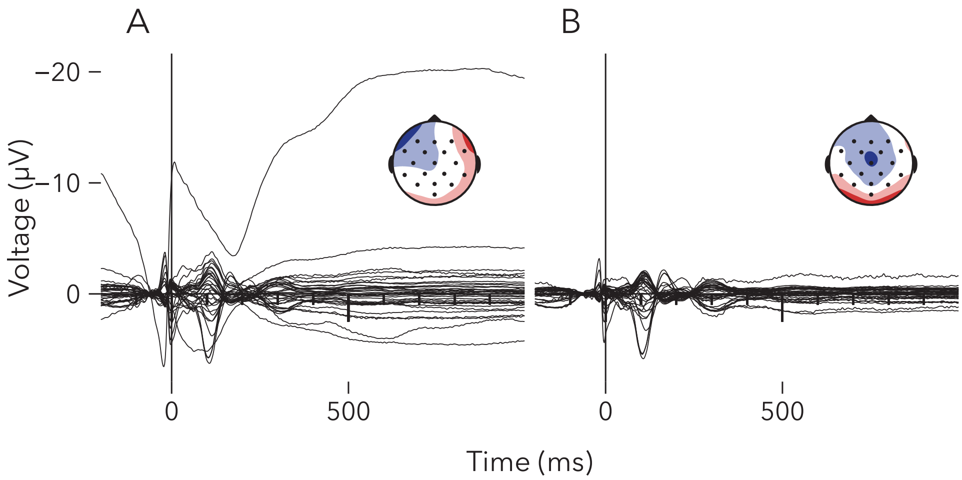

Figure 3.3: Grand-averaged ERP for randomly selected fixations from Experiment 2 before (A) and after (B) artifact correction. Topographic maps show mean amplitude in the first 500 ms following the fixation.

3.2.3 Results

3.2.3.1 Behavioral Data

Participants scored high on the comprehension questions with 92.7% accuracy in congruent and 95.0% in incongruent trials. The difference in accuracy between conditions was significant in a paired t test ($p < .05$), with lower accuracy in congruent trials.

3.2.3.2 Eye Movements

The eye movement record for the target word exhibited clear effects of world knowledge incongruence (see Figure 3.4). Participants had longer first fixation durations (265 vs. 279 ms, $b = 0.03$, $SE = 0.01$, $t = 4.75$), gaze durations (351 vs. 379 ms, $b = 0.05$, $SE = 0.01$, $t = 5.02$), and regression probabilities (35.8 vs. 39.0%, $b = 0.10$, $SE = 0.04$, $z = 2.58$) in incongruent trials.

Kretzschmar et al. (2009) reported no effect in the eye movement record for

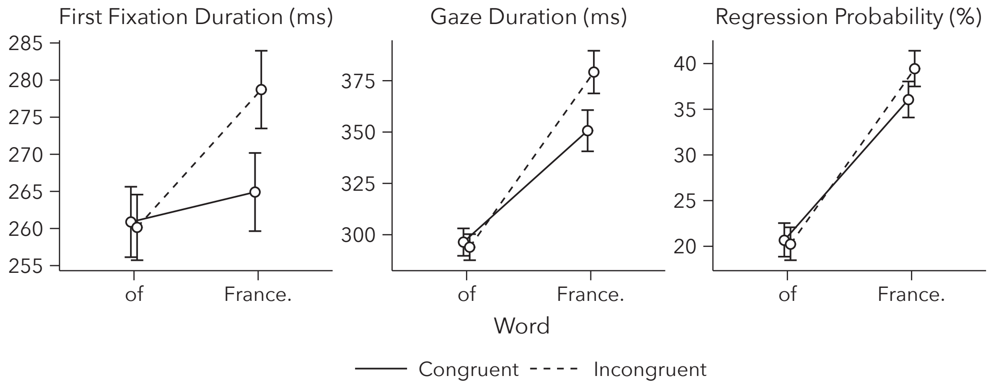

Figure 3.4: First fixation duration, gaze duration, and regression probability on the pre-target (*of*) and target word (*France*) in Experiment 2 as a function of world knowledge congruence: solid lines for congruent and dashed lines for incongruent words. Error bars are 95% confidence intervals.

the pre-target word. To check whether this was also the case in the present study, eye movements on the pre-target word were also analyzed. As Figure 3.4 shows, there were no effects on the pre-target word in the fixation record, and the fixation measures diverged only after the target word.

3.2.3.3 ERP

The ERP time-locked to the first fixation on the target word contained a significant cluster from 222 to 514 ms, peaking at 378 ms (Figure 3.5). It indicated a relative negativity in incongruent trials with a right lateralized, occipito-parietal distribution. As in Experiment 1, a positivity at frontal to fronto-central electrodes accompanied the N400 effect. It ranged from 318 to 626 ms (peak at 476 ms). A second positivity started at 692 ms and lasted until the end of the analysis window at 1400 ms (peak at 1382 ms, centered around CP1).

As in Kretzschmar et al. (2009), there were effects in the ERP on the pre-target word. Similar to the ERP on the target word, there was a centro-parietal negativity from 334 to 826 ms and a simultaneous positivity at frontal electrodes. Both effects occurred slightly later than on the target word with peak latencies of 608 ms and 658 ms, respectively. However, parafoveal preview was not controlled (e.g., via

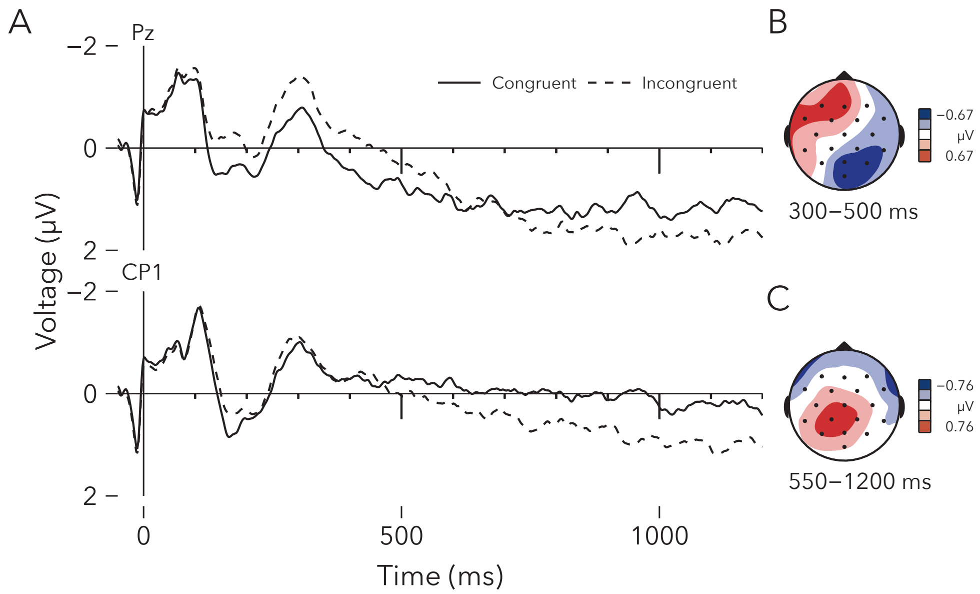

Figure 3.5: Panel A: Grand-averaged ERP for electrodes Pz and CP1 in Experiment 2, low-pass-filtered at 30 Hz. Solid lines show congruent, dashed lines show incongruent amplitudes. Panel B: Topographic map of amplitude differences (incongruent minus congruent) in the time window of the occipito-parietal negativity in Exp. 2 (200 to 500 ms). Panel C: Topographic map of amplitude differences (incongruent minus congruent) in the time window of the late positivity in Exp. 2 (700 to 1200 ms). The electrode from the respective waveform plot is underlined in the topographic map.

a boundary paradigm) and the EEG from pre-target and target word may have overlapped due to the short time interval between the fixation on the pre-target and the target word (295 ms on average). In other words, the effects observed on the pre-target word may actually have been in response to the target word, not the pre-target word. Thus, any interpretation of the effects on the pre-target word would be purely speculative.

3.2.3.4 TFR

There were significant TFR effects in the delta range and in the upper alpha range. No other frequency range showed significant differences.

3.2.3.5 Delta Range (1–3 Hz)

Power in the delta range increased at central sites in time windows from 0 to 990 ms (peak at 500 ms) following the first fixation on the target word (Figure 3.6). This synchronization was significantly larger in incongruent trials than in congruent trials. The center of this spatiotemporal cluster was at electrode Cz.

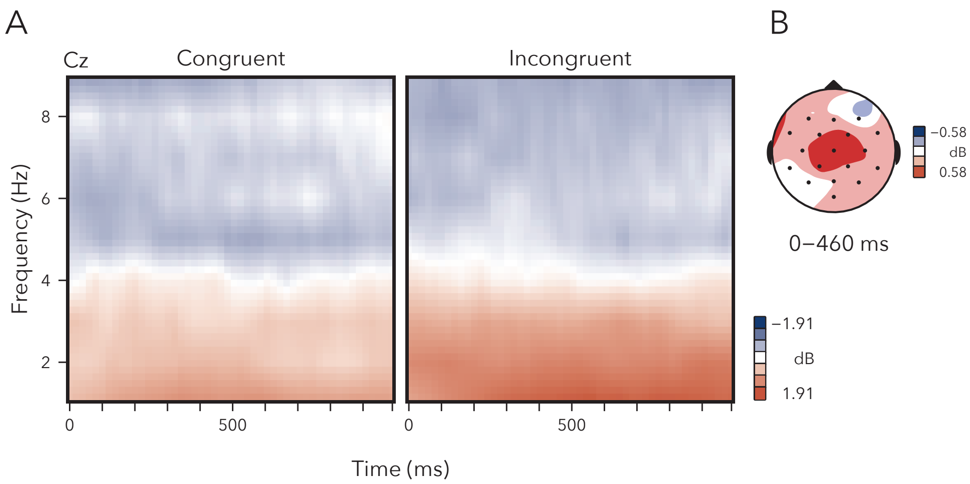

Figure 3.6: (A) Time-frequency plot of power changes in congruent and incongruent trials at electrode Cz in Experiment 2 (frequencies up to 8 Hz, smoothed with bilinear interpolation). (B) Topographic map of differences in power change in the delta range (1–3 Hz) in the time window of the delta effect in Experiment 2 (0 to 990 ms).

3.2.3.6 Upper Alpha Range (11–13 Hz)

A second cluster indicated desynchronization in the upper alpha range at occipito-parietal electrodes around POz (Figure 3.7). It started immediately after the fixation on the target word, lasted until 760 ms post fixation (peak at 520 ms), and was larger for incongruent than for congruent trials.

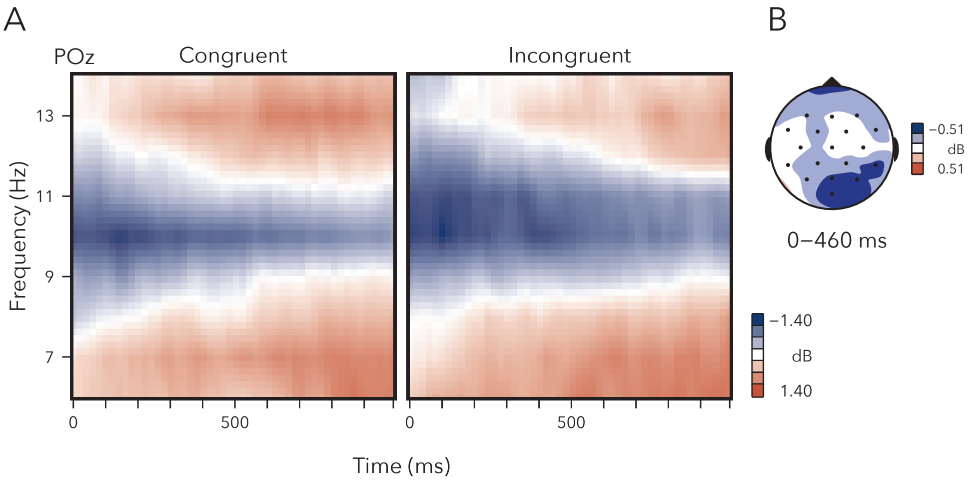

Figure 3.7: (A) Time-frequency plot of power changes in congruent and incongruent trials at electrode POz in Experiment 2 (frequencies between 6 and 14 Hz, smoothed with bilinear interpolation). (B) Topographic map of differences in power change in the alpha range (11–13 Hz) in the time window of the alpha effect in Experiment 2 (0–760 ms).

3.2.3.7 Single-Frequency Analysis

The specific choice of 1–3 Hz for the delta and 3–7 Hz for the theta band may have been responsible for the presence of a delta effect and the lack of a theta effect. To rule out this possibility, cluster-permutation tests were performed for single frequencies in the range from 1 to 7 Hz. This yielded clusters at 1, 2, and 5 Hz. The effects in the delta range had a central to fronto-central distribution, the effect at 5 Hz ranged from frontal to centro-parietal electrodes. It is due to a small power decrease in congruent trials and an equivalent power increase in incongruent trials. There were no other significant single-frequency clusters.

3.2.4 Discussion

Experiment 2 replicated the N400 effect in the time domain from Exp. 1 and Hagoort et al. (2004). In addition, we found a late positivity at centro-parietal electrodes that resembled the late positive shift reported by Hald (2003).[3] In the TFR, we found effects in the delta and upper alpha range but no effects in the theta or gamma range, which is at odds with prior results. We discuss potential explanations for the different results in Experiment 1 and 2 below.

3.3 General Discussion

We conducted two experiments to compare event-related potentials and the oscillatory dynamics of the EEG in word-by-word presentation and natural reading. In both experiments, participants read single sentences that were either congruent (e.g., *Paris is the capital of France*) or incongruent with common world knowledge (*Rome is the capital of France*). Such a manipulation elicited an N400 effect and increased activity in the theta and gamma range in prior studies using RSVP (Hagoort et al., 2004; Hald et al., 2007).

Experiment 1 used a word-by-word presentation paradigm and successfully replicated the N400 effect in the ERP found by Hagoort et al. (2004). In Experiment 2, the analyses of eye movements and ERP also revealed the anticipated patterns. The target word in incongruent trials led to increased first fixation durations, gaze durations, and regression rates. This is consistent with the results from Dambacher and Kliegl (2007), who observed the same pattern in fixation durations in an experiment on word frequency and predictability with eye movements and EEG obtained from different participants. Additionally, in line with Hagoort et al. (2004) and Hald et al. (2007), there was an increased negativity at parietal electrodes with a peak around 400 ms following incongruent words in both Experiment 1 and Experiment 2. This lends further support to the assumption that event-related potentials from RSVP and natural reading yield similar results.

Experiment 1 confirmed the power increase in the theta range from Hagoort et al. (2004). In Experiment 2, the TFR analysis revealed two different effects.

[3]Hagoort et al. (2004) analyzed the EEG data from Hald (2003), who analyzed a longer time window and reported a biphasic response comprising an N400 and a late positivity. Hagoort et al. focused their analysis on a shorter time window and therefore report only a monophasic N400.

Compared to a pre-fixation baseline, power in the delta range increased and power in the upper alpha range decreased following the first fixation on the target word. Both effects were significantly larger in incongruent trials. This is at odds with prior results and requires an explanation.

The effect in the delta range is difficult to interpret because, with few exceptions, synchronization in the delta range has not yet been reported in language-related studies. As noted above, Roehm et al. (2004) described increases in evoked and whole delta power at electrode Pz in grammaticality violations. Roehm, Bornkessel-Schlesewsky, and Schlesewsky (2007) found increased delta power in three experiments. Using graded antonyms like Kretzschmar et al. (2009), they reported an N400 and a P600 in the ERP. In the TFR, they found increased delta and theta power in the N400 time window and increased delta power in the P600 time window. In the second experiment, word pairs were presented without a sentence context. The second word (with the first word being *black*) could be a valid antonym (*white*), a related word (*yellow*), an unrelated word (*nice*), or a pseudoword. In comparison to valid antonyms, all word pairs elicited an N400 but no P600. In the TFR, unrelated words elicited more power in the lower theta range than valid antonyms. Pseudowords induced larger delta power in comparison with both valid antonyms and unrelated words. In the third experiment, Roehm et al. found increased delta power in object-initial versus subject-initial clauses and the opposite pattern in the theta range. While the manipulations in these four experiments are different from the world knowledge violations in the present study, the results show that delta power is sensitive to higher cognitive processes such as language processing.

In cognitive domains outside of language processing, the delta range has received somewhat more attention (for a review, see Harmony, 2013). Harmony, Alba, Marroquín, and González-Frankenberger (2009) reported increased delta power at frontal electrodes in the no-go condition of a go/no-go task and linked increased delta activity with the inhibition of movement. Moreover, Knyazev (2007) concluded that power changes in the delta and alpha range are inversely related but ascribed inhibitory processes to the alpha range. A simultaneous increase in delta power and decrease in alpha power is the very pattern found in Experiment 2 where readers may have inhibited a planned eye movement and engaged in memory retrieval upon encountering an unexpected noun. Thus, there is evidence for an involvement of the delta range in sensorimotor control as well as higher cognitive processes such as language processing.

In contrast to synchronization in the delta range, desynchronization in the alpha range and synchronization in the theta range are well-described oscillatory brain responses (Klimesch, 2012). Both have been observed in the context of memory-demanding tasks (Klimesch, 1999; Klimesch, Schimke, & Schwaiger, 1994) and appear to play a role in language processing (e.g., Roehm, Klimesch, Haider, & Doppelmayr, 2001). Roehm et al. (2001) investigated the involvement of the two frequency ranges in language processing while participants read sentences in four chunks. The crucial manipulation was whether or not they had to name the superordinate concept for a probe word in the penultimate chunk (e.g., "bird" for "sparrow"). In trials with that additional task, Roehm et al. observed a stronger power decrease in the upper alpha band at occipital and frontal electrodes; the theta band did not show qualitative differences between the two conditions. Roehm et al. concluded that theta oscillations reflect domain-general working memory processes whereas the upper alpha band is sensitive to linguistic processes. This may also be the reason why activity in the theta band was more prominent in Exp. 1 where sentences were presented word by word. This presentation mode may have imposed a higher load on the working memory system than natural reading.

Additional evidence for an involvement of theta and alpha oscillations in language processing comes from a study on open- and closed-class words (Bastiaansen et al., 2005). In prior studies using ERPs, open-class words elicited a larger N400 than closed-class words (King & Kutas, 1995; Van Petten & Kutas, 1991) and a frontal negative shift (Brown, Hagoort, & ter Keurs, 1999). Bastiaansen et al. (2005) reported a stimulus-evoked power decrease in the alpha and beta band as well as a power increase in the theta band that were larger following open-class words. The theta increase was strongest at left temporal and occipital electrodes and the alpha desynchronization at right occipito-temporal electrodes. The topography of the alpha effect is particularly interesting because it matches with the distribution of the upper alpha effect in the present study. Because the alpha decrease in Bastiaansen et al. (2005) did not vary as a function of word class, they ascribed it to general sensory processing of incoming information. However, their figures show that the alpha decrease was more widespread and slightly larger at occipito-temporal sites following open-class words. The reason for this quantitative difference may be that the search for a specific entry is actually more effortful in the considerably larger set of open-class words in the mental lexicon. It also matches the current results and those of Roehm et al. (2001) where more effortful processing led to decreased alpha activity at occipital electrodes.

Hagoort et al. (2004) reported synchronization in the theta range around 5 Hz and in the gamma range between 35 and 40 Hz but no effects in the alpha band. Whereas their theta effect was markedly stronger in semantic violations but also present in world-knowledge violations, the gamma effect was restricted to the latter. From this divergence, Hagoort et al. concluded that the two violation types are treated differently on a neuronal level. The theta effect is likely to reflect relatively more effortful memory access in semantically incongruent trials. The nature of the gamma effect, however, is less clear. Hagoort et al. note that gamma oscillations have been ascribed to integrative processes in both local and distributed neural networks but do not explain why this would be more important in world knowledge violations than in semantic violations.

The experimental designs of Hagoort et al. (2004) and the present study are almost identical, so it is not clear why we did not find an increase in the gamma range. Note that the current materials and Hagoort et al.'s (2004) differ in at least three aspects. First, the current studies were run in German. Second, the target word was the same in congruent and incongruent sentences of the current study, whereas an earlier word in the sentence was varied. In contrast, Hagoort et al. had varied the target word itself. Finally, the target word was sentence-final in 90 of 120 items in the current study whereas Hagoort et al. (2004) made sure that this was never the case. Wrap-up effects at the end of the sentence may have introduced a higher noise level, which may have attenuated effects in the gamma range. However, because Experiment 1 replicated the N400 effect and the power increase in the theta range despite all these differences, it is not clear why they would lead to a selective attenuation or deletion of the gamma effect.

Another potential source of the differences lies in the way to compute power changes in the EEG. In the current study, we obtained power spectra with windowed FFTs. The choice of an FFT over wavelet analysis is not likely to have a substantial influence on the results because both approaches are formally equivalent (Bruns, 2004). Due to the requirements of the FFT, however, we used the same baseline window from 1000 ms preceding fixation onset to fixation onset for all frequencies. Hagoort et al. (2004) used a wavelet transform and analyzed power changes relative to a baseline window from 150 ms preceding stimulus onset to stimulus onset. For wavelets with slower core frequencies that did not fit into that window, it was extended to the left and right.[4] To check whether the different baselines may be responsible for the diverging results, we repeated all analyses with a baseline window from −575 to 425 ms. This resulted in slight changes in the timing of the TFR effects but had no qualitative impact on the results. Thus, the difference in baselines is apparently not responsible for the lack of a theta or gamma increase in Experiment 2.

The specific choice of 1–3 Hz for the delta range and 4–7 Hz for the theta band may also have obscured effects around the boundary between the two ranges. To exclude this possibility, we performed a post-hoc analysis of Experiment 2 where we submitted power changes for single frequencies to the clustering algorithm instead of averages for predefined frequency bands. We found single-frequency effects at 1, 2, and 5 Hz. Crucially, there were no effects around the boundary between delta and theta range (i.e., at 3 and 4 Hz). The effect at 5 Hz indicated desynchronization in congruent trials and synchronization in incongruent trials. This is not consistent with Hagoort et al. (2004), who observed theta power increases with varying magnitude in all conditions. Thus, the choice of frequency

[4] According to correspondence with one of the authors.

bands was apparently not the reason for the lack of an effect in the theta range, either.

Given that Experiment 1 replicated the N400 effect and the theta increase in Hagoort et al. (2004), it appears more plausible that the different results in Experiment 2 are due to the distinct reading situations. In auto-paced word-by-word presentation, participants cannot skip words or make regressive saccades. By contrast, in natural reading situations, short and highly predictable words are skipped and regressive saccades occur frequently when readers experience processing difficulties (Rayner, 1998). A possible consequence of the different demands in natural reading and RSVP is that the sentence comprehension system adapts resource allocation as well as encoding and retrieval strategies to the reading situation. A speculative account of the different oscillatory responses in RSVP and natural reading goes as follows. In experimental settings with word-by-word presentation, the sentence parser most likely builds rich representations of the currently processed material, which leads to relatively easy retrieval from working memory in the event of unanticipated or mismatching input. This memory access is possibly reflected by a power increase in the theta range. For a natural reading situation it can be assumed that the encoding is less elaborate because earlier parts of the sentence can be reread if processing difficulties arise. The increased delta and decreased alpha activity might reflect the inhibition of progressive eye movements and increased attention (see Harmony, 2013).

The increased theta power in RSVP may also be a task effect rather than a feature of language processing. An implication is that listening studies should show similar results as they share at least two features with word-by-word presentation: Earlier material cannot be reheard, and the pace is not under the listener's control. In contrast to RSVP, listeners have to segment a continuous auditory stream. In an MEG study, Wang et al. (2012) reported an N400m as well as power decreases in the alpha and beta range in response to semantic violations. These results do not support the view that the theta increases in RSVP are solely because of readers' inability to control the presentation speed and revisit earlier material. In Kretzschmar et al. (2013), participants read longer stretches of text on different media naturally while their eye movements and brain potentials were recorded. Their mean cumulative fixation duration on a page and the absolute power in the theta range were positively correlated. Thus, theta power increases are not limited to word-by-word presentation but also occur in natural reading, which further challenges the view of theta power increases as task effects.

In summary, the current results confirm the general feasibility of recording EEG in natural reading by replicating Hagoort et al.'s (2004) N400 effect. The time-frequency analysis shows, however, that oscillatory brain dynamics are qualitatively different in natural reading and serial presentation. This may reflect differences in how representations are constructed and retrieved from memory in the two presentation modes. Further experiments are necessary to delineate the differences and similarities of stimulus- and fixation-triggered brain responses.

4

The Importance of Reading Naturally: Evidence from Combined Recordings of Eye Movements and Electric Brain Potentials

When we listen to speech, we process words in the order in which they are uttered, and we have little control over their rate. Reading is different. In reading, we can look at every word for as long as we wish, and we are not forced to read the words sequentially. Eye tracking research has demonstrated that we make ample use of this freedom: Words that are difficult to integrate with the evolving interpretation of a sentence are typically fixated longer, and more frequently trigger leftward eye movements (regressions). If a word is easy to process, or when it can be guessed from the context, we may not look at it at all (Rayner, 1998).

Researchers in psychology and psycholinguistics often use a presentation format where this freedom to navigate the sentence is taken away from the reader. One example is auto-paced word-by-word presentation (otherwise known as *rapid serial visual presentation*, RSVP). In this form of reading, one word is shown at a time and each word is presented for a fixed duration. For example, word-by-word presentation has been used in research investigating event-related brain potentials (ERPs) during reading (Hagoort et al., 2004; Kutas & Hillyard, 1980; Osterhout & Holcomb, 1992). The motivation for using this presentation mode is that electric

---

This chapter has been submitted for publication in a peer-reviewed journal and was still under revision when the thesis was published. Please cite the final version.

potentials generated by eye movements in natural reading would contaminate the recordings of the electroencephalogram (EEG).

One assumption made by experiments using word-by-word presentation is that, despite the highly constrained form of reading, comprehension accuracy is largely unaffected. This is a reasonable assumption because the rate of speech and the order or words during listening comprehension is also beyond the comprehender's control.

However, word-by-word reading takes away the ability to make regressive eye movements, which are crucial markers of processing difficulty in sentence comprehension (Clifton et al., 2007). Indeed, a recent study has shown that interfering with the reader's ability to pick up visual information during regressions causes comprehension accuracy to fall (Schotter, Tran, & Rayner, 2014). Schotter and colleagues masked words immediately after they were read for the first time. This masking led to lower comprehension accuracy compared to a natural reading condition.

So what is the difference between word-by-word versus natural reading with respect to comprehension accuracy and the ERP response? In natural reading, is there a difference in the ERP response when a regression occurs versus when it does not? We address these questions by directly comparing sentence comprehension difficulty during word-by-word presentation and in natural reading while recording EEG signals. We show that comprehension improves in natural reading compared to word-by-word presentation and that regressive eye movements reveal the strategic choices made by the comprehension system. Thus, natural reading furnishes important information about sentence comprehension processes that cannot be uncovered with word-by-word presentation.

In order to systematically modulate comprehension difficulty, we adapted a design by Hagoort (2003) and had participants read German sentences containing words that violated either grammar (syntax) or common world knowledge (semantics). We included both syntactic and semantic violations because both have been studied extensively in psycholinguistic research and their ERP correlates are relatively well understood.

An example of a syntactic violation in German is as follows. At the start of a sentence and in the absence of any other information, the feminine-marked determiner *Die* ("the") raises an expectation for a feminine-marked noun. If a masculine noun is encountered instead, this should be a surprise to the reader. At this stage, the human sentence comprehension system can react in one of several ways. One option is to initiate a recovery attempt by registering the gender mismatch and rejecting the resulting structure as ungrammatical; such a syntactic recovery attempt generally elicits a P600 effect (e.g., Osterhout & Holcomb, 1992). Alternatively, no recovery may be initiated, either because the violation is not detected (due to lack of attention, etc.), or because the comprehension system builds a partially well-formed representation, treating the sentence with a violation as "good enough" (Ferreira & Patson, 2007).

A semantic violation can be triggered by using an adjective like *neugierig*, ("inquisitive"). This word raises an expectation for a noun representing an animate referent. Readers should be surprised if the next word is an inanimate-referring noun like *Bauernhof* ('farm'). As in the syntactic violation, this type of violation should either result in the recognition of an anomaly, or in mistakenly treating the adjective-noun combination as acceptable. Such semantic anomalies are known to trigger an N400 effect (e.g., Kutas & Hillyard, 1980).

Thus, both syntactic and semantic violations generally trigger recovery processes. Such processes are part of a broader class of recovery process in sentence comprehension, a prominent example of which is triggered by "garden-path" sentences. Given a sentence fragment such as *The lawyer examined…*, native speakers of English expect a sentence in which the lawyer is doing the examining and most readers will have a strong expectation that the next constituent will be the object of *examined*, as in *The lawyer examined the evidence*. However, the sentence could also continue with *…by the nurse was ill*, that is, with a reduced relative clause. This would be ungrammatical under the favored interpretation of *examined* as the main verb of the sentence. The dashed expectation that results from a reduced relative continuation leads to a search for alternative syntactic structures. This search is associated with longer fixation durations and higher rates of regressions in eye tracking studies (Braze, Shankweiler, Ni, & Palumbo, 2002; Clifton et al., 2007; Frazier & Rayner, 1982) and with a centro-parietal negativity (Hopf, Bader, Meng, & Bayer, 2003) and/or the P600 effect (Gouvea et al., 2010; Osterhout & Holcomb, 1992) in ERP studies.

A close correspondence has been observed between the recovery process in garden-path sentences and outright ungrammatical structures: At the earliest moments of processing, the same recovery processes are believed to be initiated in both types of sentences. For example, Hopf et al. (2003) compared ungrammatical and garden-path sentences in German using ERPs and reported a negativity in the 300–500 ms range, with a similar onset latency, amplitude, and centro-parietal scalp distribution in both types. Similarly, Gouvea et al. (2010) demonstrated that in English, both garden-path sentences and syntactic violations trigger a P600. Both papers conclude that similar recovery processes are triggered when the anomaly is detected. Of course, in garden paths, the end result should generally be a grammatical structure, whereas in a syntactic violation or a semantic anomaly, the end result should be a recognition of ungrammaticality/anomaly. Nevertheless, as Gouvea et al. (2010) and Hopf et al. (2003), have demonstrated, in the first moments of detecting the violation/anomaly, the recovery process has an ERP response similar to that of the garden path.

In sum, syntactic violations and semantic anomalies are a good choice for investigating recovery processes in reading. They can easily be manipulated in an experimental configuration, are well-studied in the ERP and eye tracking literature, and trigger a class of recovery process that is of great importance in sentence processing research. Accordingly, we had participants read sentences with words that were inconsistent with their currently maintained expectation. Violating this expectation was supposed to either trigger a recovery process that would lead to a recognition of ungrammaticality (syntactic violation) or anomaly (semantic violation), or to lead to a misjudgment of the sentence as well-formed.

Like the study that inspired our design (Hagoort, 2003), we were also interested in the effect of the position of the anomaly within a sentence. When we start to hear or read a sentence, we may be less certain about the identity of the upcoming word or part of speech than when we reach the end of the sentence. For example, when we hear *The deteriorating* ..., we may not be as sure about the identity of an upcoming noun compared to when we hear *The experienced actor played a difficult* .... One possible outcome of such a change in certainty as a function of word position is that a violation of expectation later on in a sentence could result in a greater surprise and easier detection of the violation compared to an earlier position in the sentence. This is what Hagoort found for the semantic violation in his study. He found a larger N400 effect in sentence-final position, compared to effects seen in the sentence-medial condition. Hagoort suggested that this larger amplitude may have to do with "the strength of the semantic constraints increas[ing] towards the end of the sentence" (p. 894). In other words, with increasing information about the sentence, the prediction regarding the upcoming word becomes sharper. Such a sharpened expectation account would predict higher grammaticality judgment accuracy in sentence-final position. Although Hagoort did not report accuracy as a function of position in his study, it is likely that no such effect was found since accuracy across conditions was close to 100%. This high accuracy was probably because the experimental task was relatively easy: A single sentence, followed by the judgment task.

An alternative possible effect of word position on linguistic violations is that, due to higher certainty towards the end of a sentence, the reader pays *less* attention to a violation. The second possibility can be seen as the modulation of the comprehender's strength of belief about the message: in early parts of the sentence the reader may have weaker prior beliefs about the content and may be therefore more willing to attend to the sentence; but in later parts they may have formed a much stronger belief, so strong that a violation does not have enough weight to sway that belief. A clear prediction of such an attenuation in attention is reduced accuracy in detecting the violation/anomaly. Since the acceptability judgment accuracies in Hagoort's study were close to 100% for all conditions, there was not much evidence in favor of this position.

The effect of sentence position is therefore a potentially important factor for investigating whether and how the comprehension system detects an anomaly. We therefore included syntactic or semantic violations in a sentence-medial versus sentence-final position. In order to have a baseline, the sentence-medial and sentence-final conditions each had a control sentence that had no violation. In other words, we had a 2×3 factorial design: position (sentence-medial vs. sentence-final) and violation type (control, syntactic, semantic). Hagoort's study had an additional condition with a combined syntactic and semantic violation. We did not have such a condition in order to avoid complexity in the experimental design and because, unlike Hagoort, we were not primarily interested in the joint effect of syntactic and semantic violations but rather in the differences between recovery processes in word-by-word presentation and natural reading.

We conducted two experiments with the same items. One was a typical reading study using word-by-word presentation. In the second experiment, we allowed participants to read the sentences on a computer screen at their own pace, while concurrently recording their electroencephalogram and their eye movements. This approach has been made possible by recent advances in signal processing and computation that enable the removal of eye movement artifacts from EEG recordings (Dimigen et al., 2011; Makeig et al., 1996). The analysis focused on known markers of processing difficulty in the EEG (N400 and P600 effects) and on readers' performance in the task probing comprehension of the text. Having an electrophysiological *and* behavioral signal allowed us to analyze brain potentials contingent on the reading strategies observed using eye tracking.

As discussed above, our main goal was to investigate whether experiment mode matters in reading studies. Either we would see identical effects in word-by-word presentation and natural reading, or there would be theoretically important differences. Given that our study modified a design by Hagoort (2003), it is possible to make relatively specific predictions for the effect of position on syntactic and semantic violations. In sentence-medial position, Hagoort found that semantic violations elicited an N400 effect and that syntactic violations elicited a P600. The amplitude of the P600 in the syntactic violation was larger than in the semantic violation. In sentence-final position, an N400 was seen in both syntactic and semantic violations. In semantic violations, the amplitude of the N400 was larger in sentence-final position than in the sentence-medial position.

Based on Hagoort's results, we expected (relative to the control condition) a P600 for syntactic violations and an N400 for semantic violations in both presentation modes. P600 effects in response to semantic violations have been reported by some authors (e.g., Kim & Osterhout, 2005), but they are believed to occur only under specific linguistic constraints involving semantic role reversal that are not given in our design (for a discussion, see Bornkessel-Schlesewsky & Schlesewsky, 2008). In the natural reading mode, we expected higher regression probability in violations compared to the control in both sentence-medial and sentence-final position. Higher regression probability is a classic marker of recovery processes in eye tracking research (Braze et al., 2002; Frazier & Rayner, 1982). We had no clear predictions about the association between ERP components and regressive eye movements.

We also expected that sentence-medial conditions would show a different pattern of effects than sentence-final conditions, due to the greater predictability of the final word in the latter. Considering Hagoort's results, we expected an N400 sentence-finally for both violation types. In addition, in the natural reading mode, we expected an overall higher regression probability sentence-finally than sentence-medially. This is because a higher regression probability is a well-known consequence of sentence-final wrap-up processes (Rayner, Kambe, & Duffy, 2000).

4.1 Materials and Methods

4.1.1 Participants

We tested 72 students at the University of Potsdam, Germany (18 male, 54 female, mean age 25 years). The software *ORSEE* was used for participant recruitment (Greiner, 2004). The number of participants was determined before the start of the study. Twenty-four participants were randomly selected to read word-by-word (the same number of participants was tested in Hagoort, 2003). The remaining 48 participants read the text naturally. We tested twice as many participants in the natural reading condition because we expected a lower signal-to-noise ratio in that condition based on previous studies that were conducted in our lab (see Chapter 3 or Metzner et al., 2015).

4.1.2 Design

The experiment was conducted in German, and there were six conditions. The sentences contained a noun that introduced either a syntactic, a semantic, or no violation; violations occurred in the middle or at the end of the sentence (see Table 4.1 for example sentences). Sentences with syntactic violations had a noun whose grammatical gender was different from that of its determiner (*the*$_{FEM}$ *deteriorating farm*$_{MASC}$). Sentences with semantic violations had a noun that was incongruent with the preceding adjective given commonsense knowledge (*the*$_{MASC}$ *inquisitive farm*$_{MASC}$).

To control adjectives for frequency, we created clusters using *k*-means clustering on logarithmic frequency (extracted from the lexical database dlexDB, Heister et al., 2011) and paired only adjectives from the same cluster. Sentence-medial adjectives had an average log-frequency of 1.10 (baseline sentences and syntactic violations: $M = 1.14$, $SE = 0.02$; semantic violations: $M = 1.02$, $SE = 0.02$). In sentence-final violations, the adjective had an average log-frequency of 0.83 (baseline sentences and syntactic violations: $M = 0.86$, $SE = 0.02$; semantic violations: $M = 0.75$, $SE = 0.02$). Adjectives within an item differed by no more than two characters in length (sentence-medial: $M = 9.02$, $SE = 0.16$ in baseline and syntactic violations, $M = 9.04$, $SE = 0.16$ in semantic violations; sentence-final: $M = 9.52$, $SE = 0.15$ in baseline and syntactic violations, $M = 9.52$, $SE = 0.15$ in semantic violations). To avoid an influence of grammatical gender, we balanced the number of male, female, and neutral gender within each position and condition.

Table 4.1: Sample set of sentences with English translation. Manipulated word is italicized and the target noun is in boldface. The determiner's gender is indicated by a subscript (MASC = masculine, FEM = feminine, NEUT = neuter). 'Syn' and 'Sem' in parentheses denote whether the violation was syntactic or semantic.

Sentence-Medial			
Der$_{\text{MASC}}$	verfallene	**Bauernhof**$_{\text{MASC}}$ braucht eine Renovierung. Er wird…	
Die$_{FEM}$	verfallene	**Bauernhof**$_{\text{MASC}}$ braucht eine Renovierung. Er wird…	(Syn)
Der$_{\text{MASC}}$	*neugierige*	**Bauernhof**$_{\text{MASC}}$ braucht eine Renovierung. Er wird…	(Sem)
The$_{\text{MASC}}$/*The*$_{FEM}$ deteriorating/*inquisitive* **farm**$_{\text{MASC}}$ needs a renovation. It is being…			
Sentence-Final			
Der erfahrene Star spielt die$_{\text{FEM}}$	schwierige	**Rolle**$_{\text{FEM}}$. Er überzeugt…	
Der erfahrene Star spielt *das*$_{NEUT}$	schwierige	**Rolle**$_{\text{FEM}}$. Er überzeugt…	(Syn)
Der erfahrene Star spielt die$_{\text{FEM}}$	*elektrische*	**Rolle**$_{\text{FEM}}$. Er überzeugt…	(Sem)
The experienced star plays the$_{\text{FEM}}$/*the*$_{NEUT}$ difficult/*electric* **role**$_{\text{FEM}}$. He convinces…			

4.1.3 Apparatus

The EEG was recorded using a shielded electrode cap with 32 Ag/AgCl electrodes (Advanced Neuro Technology, Enschede, Netherlands) mounted following a variant of the 10-20 layout. Bipolar electrodes were placed on the left and right outer canthus and the infraorbital ridges of the right eye to record the electrooculogram. Recording was at a sampling rate of 512 Hz and with an anti-aliasing low-pass filter at 138 Hz. Impedances were kept below 5 kΩ and recordings were referenced against the left mastoid. After the experiment, the recordings were rereferenced to linked mastoids. Eye movements were recorded with an EyeLink 1000 (SR Research, Mississauga, Ontario, Canada) with a sampling rate of 1000 Hz, a spatial resolution of 0.01°, and an average accuracy of 0.32° in the area where the sentences were presented (0.53° overall).

4.1.4 Procedure

Participants sat in a dimly lit, electromagnetically shielded, and sound-insulated booth. The eyes were approximately 60 cm from the presentation screen, which had a diagonal length of 22 in. An experimental session began with ten practice trials to familiarize participants with the procedure. In sessions with word-by-word presentation, each trial began with a fixation dot in the center of the screen. After 1000 ms, the dot was removed and the sentence was displayed word by word in the same position. Every test sentence was presented together with a follow-up sentence to avoid end-of-trial effects in the final region of interest. Each participant read 360 test/follow-up sentence pairs randomly interspersed with 180 similar pairs of distractor sentences. Each word was presented for 300 ms, followed by a 300 ms inter-stimulus interval. The final words of the two sentences were presented together with a period. After the last word of the second sentence was presented, based on the procedure described by Hagoort (2003), a blank display with a pseudo-randomly varying interval between one and two seconds was shown. This was followed by a judgment task. Also following the design by Hagoort (2003), participants were prompted by a row of asterisks to decide whether or not the sentences they had just read were well-formed and to respond accordingly with a button press. Following the response, a blank display of 1150 ms preceded the onset of the next trial.

In sessions testing natural reading, trials began with a fixation dot in the vertical center at the left edge of the display. As soon as the participant had stably fixated the dot, it disappeared and the entire sentence appeared on the screen, offset by 80 px to the right in order to induce a saccade to the first word of the sentence. To end the sentence presentation and to proceed to the judgment task, participants had to fixate the lower right corner of the screen.

In both presentation modes, participants were encouraged to take a short break every twenty trials. During these breaks, they received feedback about their performance on the judgment task. After 240, 360, and 480 trials, participants had to take longer breaks to relax their neck muscles and eyes.

4.1.5 Data Preprocessing

A velocity-based saccade detection algorithm was used to detect saccades and fixations in the raw eye tracking data (Engbert & Kliegl, 2003). Fixations shorter than 20 ms and longer than 1000 ms were removed. This led to the loss of 0.1% of all fixations. The following eye tracking measures were calculated for the target word (the noun): first fixation duration, gaze duration, and regression probability. First fixation duration is the duration of a first fixation on a word when it is entered from the left (i.e., during first pass). Gaze duration is the cumulative duration of all fixations during first pass on a word from the first incoming saccade until the first outgoing saccade. Regression probability is estimated by dividing the number of trials with a leftward saccade (after first entering a word from the left and before reading a word to its right) by the total number of trials; only regressions during first pass are considered as they are widely assumed to index early processing events (Vasishth et al., 2012).

The EEG data were preprocessed using BrainVision Analyzer 2 (Brain Products, Munich, Germany). We first resampled the signal to 500 Hz and filtered it with a bandpass filter of 0.3 through 70 Hz (both at 48 dB/oct) and a notch filter at 50 Hz. We then identified artifacts generated by the eye movements during reading using independent components analysis. The Infomax algorithm was used for training on all distractor sentences. Components with a frontal or bipolar frontal distribution were removed from the signal such that variance in the eye electrodes was minimized (see Figure 4.1). From the data corrected in this manner, we removed epochs with other artifacts, as head movements or slow drifts in a semi-automatic procedure. This resulted in the loss of 79 sentence-medial (1.8%) and 52 sentence-final trials (1.2%) in the word-by-word data, and 229 sentence-medial (2.7%) and 150 sentence-final trials (1.7%) in the natural reading data.

In the word-by-word data, the time-lock for the ERP analysis was the onset of the target word. In the natural reading data, the time-lock was the time when a reader's gaze first landed on the target word. Trials in which the target noun was skipped in the first pass were not considered for the ERP analysis, which led

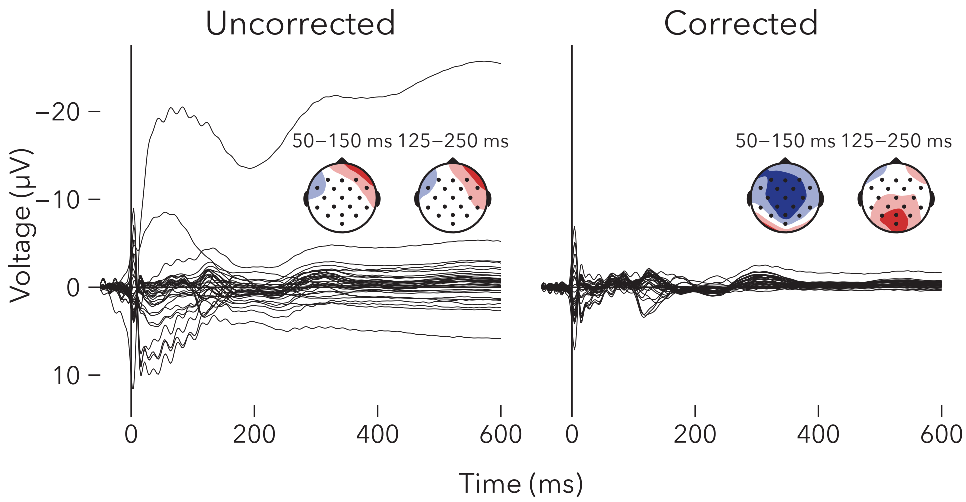

Figure 4.1: Grand-averaged ERP for randomly selected fixations before and after artifact correction. Topographic maps show mean amplitude in the N1 and P2 time window.

to the loss of 444 sentence-medial (5.1%) and 562 sentence-final trials (6.5%) in natural reading sessions. Epochs starting 200 ms preceding the onset of the target word and ending 1000 ms later were exported to *R* (R Core Team, 2013) for further processing. Prior to the statistical analysis, the data were baseline-corrected by subtracting the average amplitude in a 100 ms interval preceding the time-lock.

4.1.6 Analysis

Response accuracy in the judgment task was analyzed with two logistic mixed-effects regressions as implemented in the *R* package *lme4* (version 1.1.7, using the BOBYQA algorithm for optimization; Bates et al., 2014). Accuracy was treated as a categorical variable (correct: 1, incorrect: 0). The first model investigated the effect of presentation mode on judgment accuracy. This model included violation type, violation position in the sentence, presentation mode (word-by-word or natural reading) as fixed factors, as well as all two- and three-way interactions of these factors. Violation type was coded using a treatment contrast with the control condition as the baseline. Violation position and modality (the latter a between-participant factor) were coded using sum contrasts. The second model investigated the effect of regressive eye movements on judgment accuracy and used only data from natural reading sessions. The predictors were the same except that modality was replaced by a predictor indicating whether or not a regression had occurred in the respective trial (sum contrast). Again, all possible interactions were included.

The models had a full variance-covariance specification for the random effects (intercepts and slopes for the within-participant fixed effects, with correlation estimated). In cases where a model did not converge, we dropped the random effect with the least variance and refit the model. This procedure was repeated until the model converged. An effect was considered statistically significant at $\alpha = 0.05$ if the corresponding absolute t or z statistic was greater than 1.96.

We analyzed eye movement measures at the noun with linear and logistic mixed-effects models. The continuous variables first fixation duration and gaze duration were log-transformed to obtain approximately normally distributed residuals. The occurrence of regressions was treated as a categorical variable (regression: 1, no regression: 0) and modeled using the logit link function. Fixed effects were violation type, violation position, and their interactions. The contrast coding and procedure for determining the random effects structure were the same as described above.

ERPs were analyzed with nonparametric cluster-based randomization tests (Maris & Oostenveld, 2007), which offer an elegant solution for the multiple-testing problem that often arises in the analysis of ERP data. We implemented the procedure as follows. First, paired t-tests were performed for the mean amplitude at each time point and electrode with one value per participant and condition. Next, samples with a test statistic that was significant at an α of .05 were clustered using connected-component labeling (Rosenfeld & Pfaltz, 1966). All test statistics within a cluster were then summed up to yield a cluster statistic. To assess each cluster's significance, we generated a distribution representing the null hypothesis by means of a randomization procedure: In each of 1000 iterations, condition labels were first randomly swapped within participants. With data randomized in this manner, clusters were formed as described above and the largest cluster statistic was entered into the distribution. Clusters found in the original data were considered significant if their test statistic fell in the lower 2.5th or upper 97.5th percentile of this distribution. Note that, depending on the threshold for the individual t-tests, these clusters can capture long-lasting effects whose distribution on the scalp changes over time. For example, a positivity that peaks at posterior electrodes around 700 ms after stimulus onset may be connected with an earlier or later positivity at frontal electrodes. For further details of this cluster-based randomization approach, see Maris and Oostenveld (2007).

To investigate the relation between regressive eye movements and ERP effects, we split the data for the natural reading sessions in two subsets: one with trials in which a first-pass regression occurred on the target word and one in which no such regressions occurred. These subsets were then analyzed individually using the procedure described above. Since the rate of regressions was too low in the baseline condition to conduct statistical tests, we compared ERP data from violation trials with regressions to all baseline trials taken together (irrespective of whether a violation had occurred).

4.2 Results

4.2.1 Judgment Accuracy

Figure 4.2 shows participants' performance in the judgment task. We fit two linear mixed models: One model combined the data from word-by-word presentation and natural reading to investigate the effect of mode (a between-participants factor), the other model investigated the effect of regressions on accuracy. See Table 4.2 and Table 4.3 for the parameter estimates from the two models.

Mean accuracy for target items was 87%, showing that participants were attending to the task. However, both linear mixed models showed that violations had highly significant effects on accuracy: Accuracy was lower when the sentence had a syntactic or a semantic violation (compared to the control condition). Both models also showed that accuracy was on average lower when the violation occurred in sentence-final vs. -medial position (marginally significant in the model testing presentation mode).

The model testing presentation mode (Table 4.2) showed that accuracy was significantly improved when the participants read sentences naturally instead of word-by-word. For syntactic violations, a three-way interaction showed a lower accuracy on average when the violation occurred in final position, and when sentences were presented word-by-word. The model set up to test the effect of regressions showed that accuracy was higher in sentences with violations when a regression occurred.

Table 4.2: Summary statistics for the mixed-effects regression of response accuracy as a function of violation type, position, and presentation mode. Estimates (b) and 95% confidence intervals (CI) are on the logit scale, statistical significance is indicated by test statistics in bold face (z).

	b	95% CI	z
Syntax	**−0.54**	**[−0.94, −0.14]**	**−2.6**
Semantics	**−1.72**	**[−2.12, −1.32]**	**−8.4**
Position	−0.31	[−0.67, 0.05]	−1.7
Mode	**0.51**	**[0.14, 0.88]**	**2.7**
Syntax × Position	−0.28	[−0.71, 0.15]	−1.3
Semantics × Position	**−0.86**	**[−1.31, −0.41]**	**−3.7**
Syntax × Mod.	−0.07	[−0.75, 0.61]	−0.2
Sem. × Mod.	−0.01	[−0.68, 0.66]	0.0
Pos. × Mod.	−0.29	[−0.74, 0.16]	−1.3
Syn. × Pos. × Mod.	**1.87**	**[1.33, 2.41]**	**6.8**
Sem. × Pos. × Mod.	0.05	[−0.43, 0.53]	0.2

4.2.2 Eye Tracking Data

Figure 4.3 shows mean fixation durations and 95% CIs for the eye tracking measures (first fixation duration, gaze duration, regression probability). The parameter estimates from the linear mixed model analysis are in Table 4.4.

Consistent with earlier research (e.g., Braze et al., 2002), the eye movement data showed that, compared to the control condition, readers slowed down at the target word when a syntactic or semantic violation occurred. This effect was found

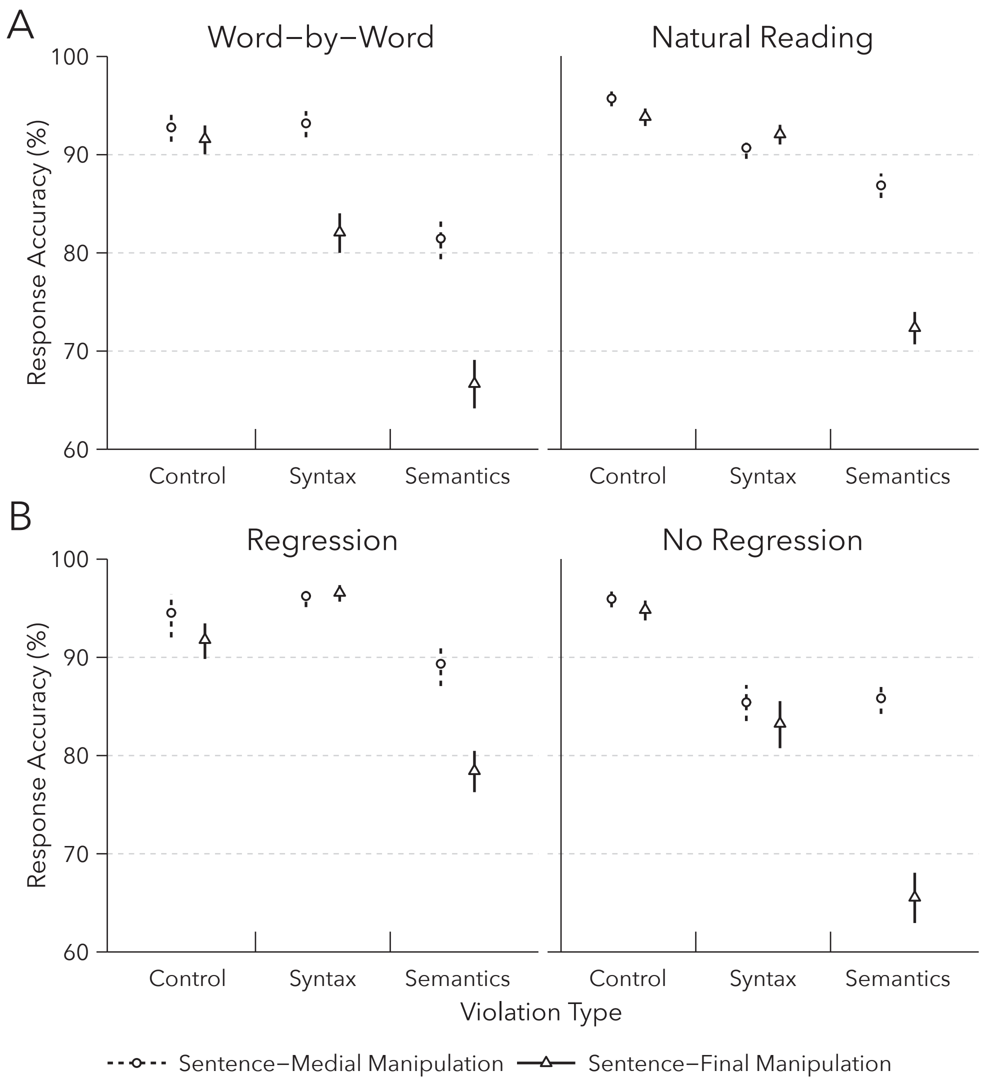

Figure 4.2: Accuracy in the judgment task in word-by-word presentation and natural reading (A) and within natural reading sessions, in trials with regression and without regression (B). Dashed lines show sentence-medial manipulations and solid lines sentence-final manipulations. The bars denote 95% confidence intervals. Accuracy was better when the text was read naturally than when it was read word-by-word. The benefit of natural reading was mainly driven by trials in which the eyes regressed. When they did not regress,the accuracy was similar to that observed during word-by-word reading.

Table 4.3: Summary statistics for the mixed-effects regression of response accuracy in natural reading sessions as a function of violation type, position, and the occurrence of regressions. Estimates (b) and 95% confidence intervals (CI) are on the logit scale, statistical significance is indicated by test statistics in bold face (z).

	b	95% CI	z
Syntax	**−0.70**	**[−1.27, −0.13]**	**−2.4**
Semantics	**−1.86**	**[−2.42, −1.30]**	**−6.5**
Position	**−0.57**	**[−1.09, −0.05]**	**−2.1**
Regression	−0.29	[−0.70, 0.12]	−1.4
Syn. × Pos.	0.51	[−0.10, 1.12]	1.6
Sem. × Pos.	**−0.93**	**[−1.54, −0.32]**	**−3.0**
Syn. × Reg.	**2.00**	**[1.51, 2.49]**	**7.9**
Sem. × Reg.	**0.93**	**[0.48, 1.38]**	**4.1**
Pos. × Reg.	−0.27	[−0.92, 0.38]	−0.8
Syn. × Pos. × Reg.	0.59	[−0.22, 1.40]	1.4
Sem. × Pos. × Reg.	0.28	[−0.47, 1.03]	0.7

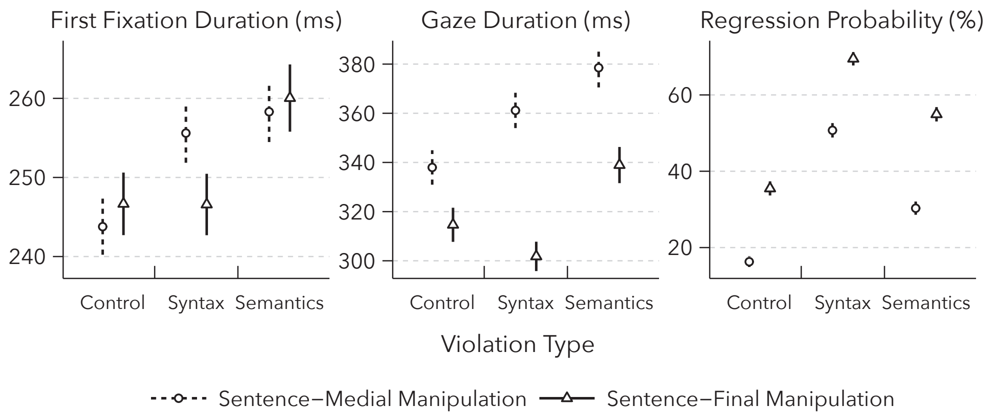

Figure 4.3: First fixation duration, gaze duration, and regression probability at the noun in sentence-medial and sentence-final items. The bars denote 95% confidence intervals. Dashed lines show sentence-medial manipulations and solid lines sentence-final manipulations.

Table 4.4: Summary statistics for the mixed-effects regression of first fixation duration, gaze duration, and regression probability in natural reading sessions as a function of violation type, position, and regression. Estimates (*b*) and 95% confidence intervals (CI) are on the log scale for first fixation and gaze duration and on the logit scale for regression probability, statistical significance is indicated through bold face.

	b	95% CI	*t*/*z*
	First Fixation Duration		
Syntax	**0.02**	**[0.01, 0.03]**	**2.39**
Semantics	**0.06**	**[0.05, 0.07]**	**7.16**
Position	0.00	[−0.02, 0.02]	−0.14
Syntax × Position	**−0.04**	**[−0.06, −0.02]**	**−2.60**
Semantics × Position	−0.01	[−0.02, 0.00]	−0.45
	Gaze Duration		
Syntax	**0.03**	**[0.02, 0.04]**	**2.43**
Semantics	**0.10**	**[0.09, 0.11]**	**10.08**
Position	**−0.08**	**[−0.11, −0.05]**	**−2.64**
Syntax × Position	**−0.10**	**[−0.12, −0.08]**	**−4.52**
Semantics × Position	−0.03	[−0.05, −0.01]	−1.77
	Regression Probability		
Syntax	**1.75**	**[1.66, 1.84]**	**18.68**
Semantics	**0.90**	**[0.85, 0.95]**	**17.24**
Position	**1.13**	**[1.01, 1.25]**	**9.50**
Syntax × Position	**−0.27**	**[−0.40, −0.14]**	**−2.02**
Semantics × Position	0.07	[−0.03, 0.17]	0.73

in first fixation duration and gaze duration. The gaze duration data also showed that the target word was read faster when it was in sentence-final position. An interaction between syntactic violation and position was seen in all measures. In duration data, a syntactic violation led to shorter fixations times sentence-finally compared to sentence-medially. The first-pass regression data showed the opposite pattern to that seen in the fixation durations: A syntactic violation led to higher regression probability sentence-finally compared to sentence-medially. The shorter fixation durations sentence-finally were due to regressive eye movements cutting the first pass short. This association in the sentence-final conditions was established by fitting a linear mixed model that examined the effect of regres-

Table 4.5: Summary of ERP results in the word-by-word reading and natural reading experiments.

	Word-by-Word Presentation		Natural Reading	
Position	Syntax	Semantics	Syntax	Semantics
Medial	P600	N400/P600	P600	N400/P600
Final	N400/P600	N400	N400/P600	N400/P600

Regression-Contingent Analysis			
Regression	Position	Syntax	Semantics
Yes	Medial	P600	P600
	Final	N400/P600	N400/P600
No	Medial	–	–
	Final	Sustained Negativity	Sustained Negativity

sion (as a binary predictor) on fixation durations. In both first fixations and gaze duration, a regressive eye movement led to significantly shorter durations.

4.2.3 Event-Related Potentials

All ERP results for the word-by-word study, the natural reading study, and the regression-contingent analyses in the natural reading study are summarized in Table 4.5. We discuss these below. For simplicity, instead of saying "relative negativities and positivities compared to control sentences", we use the abbreviations negativities and positivities.

4.2.3.1 Word-by-Word Presentation

In the word-by-word presentation experiment, we found results similar to those of Hagoort (2003). Syntactic violations in sentence-medial position led to a P600-like centro-parietal positivity from 544 to 1000 ms (peak at 812 ms, $p < .001$). In sentence-final position, an N400-like effect from 312 to 526 ms (peak at 428 ms, $p < .05$) was followed by a late centro-parietal positivity from 570 to 1000 ms (peak at 688 ms, $p < .001$).

Semantic violations in sentence-medial position elicited a short-lived centro-parietal negativity from 476 to 520 ms (peak at 504 ms, $p < .01$) and a centro-parietal positivity from 684 to 1000 ms (peak at 836 ms, $p < .001$). In sentence-final position, a sequence of four centro-parietal negativities from 128 to 656 ms together constituted an N400 effect (peaks at 170, 230, 358, and 536 ms; all $p < .001$).

4.2.3.2 Natural Reading

In the ERP recorded during natural reading (Figure 4.4), syntactic violations in sentence-medial position elicited a P600-like centro-parietal positivity from 478 to 1000 ms (peak at 828 ms, $p < .001$); in sentence-final position, a similar centro-parietal positivity from 590 to 1000 ms (peak at 776 ms, $p < .001$) was preceded by an N400-like occipito-parietal negativity from 130 to 434 ms (peak at 216 ms, $p < .01$).

Semantic violations led to an N400/P600 response in both sentence-medial and sentence-final position. Sentence-medially, it comprised an occipito-parietal negativity from 230 to 372 ms (peak at 268 ms, $p < .05$) and a centro-parietal positivity from 760 to 1000 ms (peak at 980 ms, $p < .01$). Sentence-finally, both effects occurred slightly earlier from 104 to 522 ms (peak at 360 ms, $p < .001$) and from 696 to 988 ms (peak at 842 ms, $p < .01$).

4.2.4 Regression-Contingent Analysis

The eyes regressed in 60% of the trials with sentences including syntactic violations and in about 43% of the trials with sentences including semantic violations. As shown in Figure 4.3, regression probabilities were higher in violations than control sentences, and higher in final position compared to medial position. To examine the ERP in trials with and without regressions, we split the data into two subsets: one in which regressions occurred during first pass, and one in which no first-pass regressions occurred from the noun. We will refer to these as the regression and no-regression trials, respectively. On these two subsets, we applied similar statistical analyses as on the full data-set (see Figure 4.4).

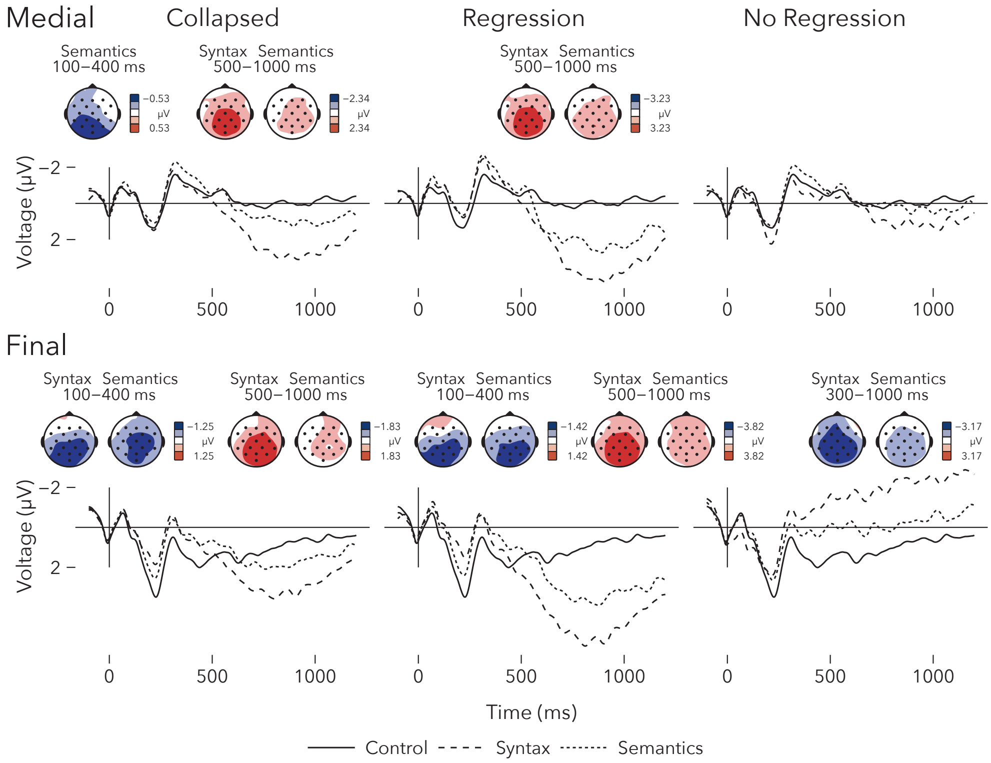

Figure 4.4: Grand-averaged ERPs from natural reading sessions at electrode Pz for control sentences (solid line), syntactic violations (dashed line), and semantic violations (dotted line) at the noun in sentence-medial and sentence-final position. Isovoltage maps show the topographic distributions of amplitude differences (violation minus control) from 100 to 400 ms (N400), from 500 to 1000 ms (P600), and from 300 to 1000 ms (sustained negativities). The collapsed data (comprising trials with and without regressions) largely replicate earlier results by Hagoort (2003). However, the separate analyses for trials with and without regressions show that P600 effects are strongly associated with regressions. In trials without regressions, we instead found sustained negativities that were canceled out by the stronger P600 effects in the analysis of the full data set. Separating the trials using the behavioral signal thus revealed two qualitatively different brain responses to syntactic and semantic violations.

4.2.4.1 Regression Trials

Considering trials in which a first-pass regression occurred at the target word, both violation types elicited a P600 effect in sentence-medial conditions. In syntactic violations, it ranged from 290 to 1000 ms (peak at 828 ms, $p < .001$) and in semantic violations from 540 to 1000 ms (peak at 848 ms, $p < .001$). In sentence-final position, both violations elicited an N400/P600 response. In syntactic violations, a centro-parietal negativity from 24 to 378 ms (peak at 164 ms, $p < .05$) was followed by a centro-parietal positivity from 244 to 1000 ms (peak at 868 ms, $p < .001$). In semantic violations, an occipito-parietal negativity from 98 to 392 ms (peak at 360 ms, $p < .05$) was followed by a centro-parietal positivity from 412 to 1000 ms (peak at 842 ms, $p < .001$).

4.2.4.2 No-Regression Trials

Considering trials in which no first-pass regression occurred at the target word, neither violation type showed any ERP effects in sentence-medial conditions. In sentence-final conditions, both violation types elicited a sustained, centro-parietal negativity. In syntactic violations, a single effect from 310 to 1000 ms (peak at 586 ms, $p < .001$) represented the sustained negativity. In semantic violations, it comprised two disjoint effects from 336 to 646 ms (peak at 592 ms, $p < .01$) and from 652 to 774 ms (peak at 692 ms, $p < .05$).

4.3 Discussion

Our main goal was to establish whether any differences exist between reading studies using the word-by-word presentation method and natural reading. To this end, we compared judgment accuracy and ERP responses in the two presentation modes. In the natural reading study, we also performed regression-contingent analyses of judgment accuracy and ERP responses. We discuss the implications of each of these.

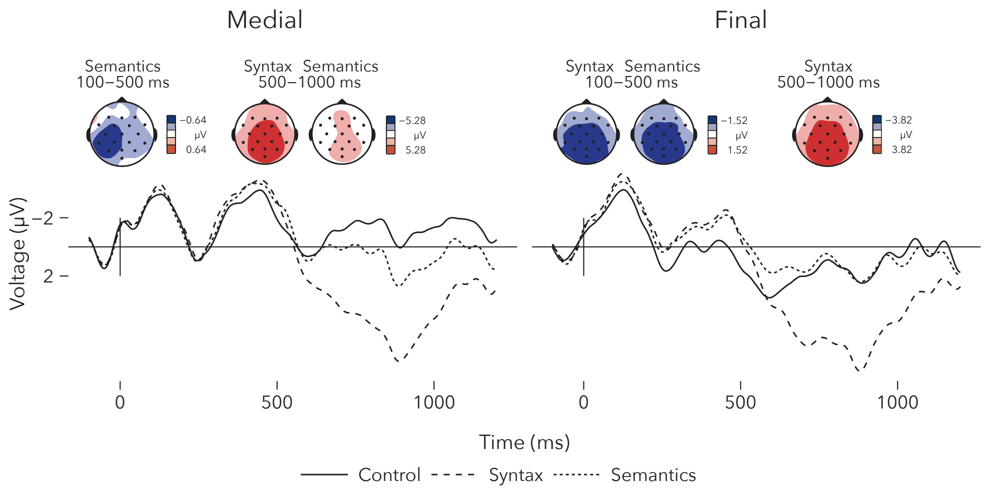

Figure 4.5: Grand-averaged ERPs from word-by-word sessions at electrode Pz (low-pass-filtered at 10 Hz, negativity plotted upwards) for control sentences (solid line), syntactic violations (dashed line), and semantic violations (dotted line) at the target word in sentence-medial and sentence-final position. Bicubic spline-interpolated isovoltage maps above the waveforms show the topographic distributions of mean amplitude differences (violation minus control) from 100 to 500 ms (N400) and from 500 to 1000 ms (P600).

The judgment accuracy data provide strong evidence that comprehension is, on average, better when sentences are read naturally than when they are presented word-by-word. The ability to revisit earlier material, only available in natural reading, seems to be the key to explaining this difference. If the eyes make a regressive saccade, comprehension improves substantially. If the eyes do not make a regressive saccade, accuracy in the violation conditions is as low or lower than that observed during word-by-word presentation. The latter result suggests that the freedom to control viewing times alone does not improve comprehension in these sentences. In the current setting, the added benefit of natural reading really is the ability to revisit earlier material.

Table 4.5 shows that the ERP responses in word-by-word presentation and natural reading were remarkably similar, but also that there were two differences. First, in the semantic violation which occurred sentence-finally, word-by-word presentation showed an N400 effect, whereas the reading study showed both an N400 and a P600. Second, in the natural reading study, a sustained negativity was seen in both syntactic and semantic violations in no-regression trials of the sentence-final conditions. The P600 in the natural reading study is not too surprising given that P600 effects have been observed in other settings testing semantic violations (Kim & Osterhout, 2005). However, this effect shows that semantic P600 effects are not limited to semantic violations involving thematic role reversal.

At first sight, the sustained negativity might look like a reflex of wrap-up effects that occur at the end of a sentence. But if that were the case, we would see a similar sustained negativity in the sentence-final region in the sentence-*medial* conditions. Instead, in sentence-medial conditions, we see no effects in the sentence-final region when we compare regression and no-regression trials. Thus, it is unlikely that the sustained negativity reflects only wrap-up effects. This sustained negativity, found only in the sentence-final conditions when no regression occurred, suggests that the sentence-final wrap-up process interacts with the recovery process. One possibility is that the sustained negativity reflects the "good-enough" toleration strategy discussed earlier. The weakness of this explanation is that we should have seen a sustained negativity in the no-regression trials for the sentence-medial conditions, as well. Instead, we see no effect. It is worth investigating in future work, perhaps with higher statistical power, whether a sustained negativity is also generated in medial positions.

What new insights about the comprehension process could be uncovered using natural reading? The regression-contingent analyses suggest an answer. Syntactic violations did not just elicit stronger P600 effects, they also prompted regressive eye movements more often than semantic violations. This suggests that regressions may be linked to the P600 effect and thus to recovery processes. The idea that regressions are associated with P600 effects also receives support from a corpus study showing that backward-directed eye movements in reading are accompanied by a P600-like effect that is larger if the regression is longer (Dimigen et al., 2007). This study used simple sentences suggesting that even when the material is easy to process, misinterpretations may occur and may demand recovery procedures similar to those triggered by our material. In our experiment, P600 effects occurred when the reader carried out a regressive saccade from the mismatching word to earlier words, and no P600 effects were found in no-regression trials. Along with the fact that judgment accuracy was lower in no-regression trials, the absence of a P600 suggests that recovery was initiated less often or not at all in these trials. The regression-linked P600 suggests that regressive eye movements support the recovery process in sentence comprehension. When faced with processing difficulty, the sentence comprehension system may carry out a recovery process or alternatively resort to tolerating the violation/anomaly. When a recovery process is started, regressive eye movements improve its probability of success, leading to a correct detection of ungrammaticality. When the comprehension system resorts to a toleration strategy, readers tend to not regress. Similar to this finding, earlier research on the processing of garden-path sentences also showed that readers engage in strategic choices when faced with words that challenge the currently maintained interpretation. Von der Malsburg and Vasishth (2011, 2013) used a novel analysis method to investigate regression scanpaths in response to garden-pathing and found that readers differ strongly in how they apply different reading strategies and that this variability is partly explained by individual differences in working memory resources.

The word-by-word presentation format did not reveal this variability in processing strategies because this format does not provide us with a behavioral signal that would allow us to analyze subsets of the ERP data separately. Thus, the ERP signal was potentially aggregated over two sets of processing strategy: recovery and the toleration strategy. In word-by-word presentation, we also found P600 effects when a violation occurred. That is, in a situation where regressions were impossible, readers may have deployed a *covert reanalysis* strategy (Lewis, 1998). Covert reanalysis refers to a recovery process that is initiated without moving the eyes leftwards. However, this covert recovery was not deployed often in word-by-word reading or it was less efficient, as can be seen from the lower accuracy in the judgment task during word-by-word presentation compared to the natural reading study. Clearly, the toleration strategy was used more often.

Thus, although word-by-word reading and natural reading can yield comparable ERP results (Dimigen et al., 2011; Kretzschmar et al., 2009; but see Metzner et al., 2015), coregistration using natural reading reveals two qualitatively different reading strategies (attempted recovery vs. toleration of the inconsistency) that were not detected in data obtained using word-by-word presentation alone. Coregistering the EEG signal in addition to fixation data also reveals that one function of regressive eye movement in reading is to support the recovery process when comprehension difficulties arise.

5

Summary and Conclusions

The goal of the present thesis was to assess the advantages and limitations of integrated eye movement and EEG recordings for the study of sentence comprehension. In a series of three studies totalling to six experiments, we investigated three areas systematically to reach this goal. In the following sections, each study and its contribution will be summarized.

5.1 Subtle Effects in Coregistration

Studies investigating ERPs in sentence processing research have traditionally used word-by-word presentation with a high signal-to-noise ratio or relied on large effects from ungrammatical or nonsensical sentences. Can we use ERPs to study sentence processing when we have neither? When readers read sentences naturally, they will skip the target word at least in some trials and exhibit parafoveal effects. If this further lowers the signal-to-noise ratio, can we still investigate linguistic questions with more subtle effects?

We conducted a series of three experiments on the effect of distance in dependency resolution to find answers for these questions. The experiments were also designed to delineate two classes of models of sentence processing. For the object-verb and antecedent-pronoun dependencies used in these experiments, memory-based models predict a positive correlation of distance and processing costs. Expectation-based models predict no effect of distance on pronoun resolution and a negative correlation between distance and processing costs in object-verb dependencies.

Two experiments used coregistration and produced mixed results. In the first experiment, we found locality effects in both dependency types, but we also identified a confound at the word preceding the target word in that it was the head of a complex noun phrase in long and of a simple noun phrase in short dependencies. In the second experiment, where we used deconfounded materials (i.e., the pre-target word was the head of a simple noun phrase in both long and short dependencies), we found no locality effects. We concluded that the locality effects in the first experiment were at least in part caused by the different complexity of the pre-target noun phrases. Nonetheless, the results of the first experiment support memory-based accounts. The more complex noun phrase presumably imposed a higher load on working memory than a simple noun phrase. In these conditions, we see locality effects. The results of these two experiments suggest that distance and phrase complexity interact in dependency resolution.

Because the lack of locality effects was not in line with prior research, we repeated the second experiment (with always a simple noun phrase preceding the target) with a simpler method that has been shown to reflect locality effects. With self-paced reading, we observed small but significant locality effects for both dependency types, supporting memory-based models of sentence processing.

The contributions of the first study are twofold. First, it adds further support to memory-based accounts of sentence processing, sharpening our picture of how people transform linear sequences of words into hierarchical structures. Second, the study highlights limitations of coregistration that cannot easily be overcome, potentially dampening the field's expectations to a more realistic level.

5.2 Neural Oscillations in Natural Reading

While most ERP studies on language processing are focused on the time domain, an increasing number of studies investigate the time-frequency traces of cognitive processes. A consistent finding in such studies is that the frequency bands that respond to memory-demanding tasks are also sensitive to linguistic operations. Some studies have also shown that two manipulations that elicit similar effects in the time domain do not necessarily lead to analogously similar effects in the time-frequency domain. A good example is the study by Hagoort et al. (2004), who investigated brain responses to semantic violations and violations of common world knowledge. Both engendered an N400 effect in the time domain but the spectral responses were distinct.

Effects from RSVP and coregistration studies have been shown to align nicely in the time domain (Dimigen et al., 2011; Kretzschmar et al., 2009). We conducted a study with two experiments to investigate if that analogy extends to the spectral correlates of these effects. In the first experiment, we replicated Hagoort et al. (2004) with German sentences and reproduced the N400 effect in the time domain and the theta power increase in the time-frequency domain. In a coregistration experiment with the same materials, world knowledge violations no longer elicited a theta power increase but a power increase in the delta band as well as a power decrease in the upper alpha band.

Theta power increases are linked with processes involving working memory, encompassing encoding and retrieval of memory traces. That theta power increases in reading are restricted to word-by-word reading suggests that readers build more elaborate representations of the currently processed materials when they cannot revisit earlier parts of the sentence. Simultaneous delta power increases and alpha power decreases have been observed in the context of go/no-go tasks where participants have to actively suppress a motor response. This is reminiscent of what readers have to do when they encounter unexpected input: suppress an already programmed saccade to either linger on the word or make a regressive saccade.

Thus, the second study of the present thesis contributes important insights about memory encoding and retrieval in different reading modes. The data underlying those insights could only be obtained with coregistration, which highlights its methodological importance. The study also shows that, despite many analogies between word-by-word and natural reading, there are important differences in the neural processes that accompany behavioral responses.

5.3 Regressive Saccades and Recovery

During reading, the eyes do not always move forward. In most cases, readers make regressive saccades to correct an overshoot (i.e., they moved their eyes a bit too far to the right). However, readers also frequently make regressive saccades when they are experiencing high-level processing difficulties. Why exactly readers do this is not entirely clear and still a matter of debate (Frazier & Rayner, 1982; Meseguer et al., 2002; Mitchell, Shen, Green, & Hodgson, 2008; von der Malsburg & Vasishth, 2011, 2013). We investigated the brain responses in the context of regressive saccades to obtain insights on the neural processes involved in backwards-directed eye movements in reading.

To this end, we replicated a study by Hagoort (2003) in German because it showed robust effects and allowed us to investigate structural and semantic anomalies in different parts of a sentence. Using RSVP, we found very similar results as Hagoort (2003): Sentence-medially, syntactic violations elicited a P600 effect and semantic violations a biphasic N400-P600 effect. Sentence-finally, syntactic violations elicited such a biphasic N400-P600 effect and semantic violations an N400 effect.

Coregistration yielded very similar results with an additional biphasic N400-P600 effect in sentence-final semantic violations. When we split the data contingent on whether readers made a regressive saccade in response to a violation, we found a rather surprising pattern. When readers made a regressive saccade, we found P600 effects in sentence-medial position and biphasic N400-P600 effects in sentence-final position for both violation types. Without a regressive saccade, readers showed no ERP effects in response to sentence-medial violations but large, sustained negativities for sentence-final violations.

This study demonstrates again that results from RSVP studies can be reproduced using coregistration. Importantly, it also shows how coregistration can go beyond what could be achieved with RSVP. The regression-contingent analyses suggest that regressions are strongly linked with the P600 effect. This supports the notion that these responses are a sign of recovery or repair processes in reading. The fact that the ERP is qualitatively different when readers do not make a regressive saccade suggests that readers can pursue different processing strategies that do not involve recovery.

The third study takes explicit advantage of the strengths of coregistration. By analyzing fixation-triggered brain responses with respect to how readers moves their eyes following the fixation, it goes beyond what had been done with coregistration before. Perhaps as a counterpoint to the first study, this study shows the immense potential of coregistration for shedding light on how cognitive processes and eye movements in reading are guided by text properties.

5.4 Outlook

We have shown that coregistration has certain limitations, but also that it offers a lot of potential for the study of human language processing. The present thesis should therefore be thought of as a stepping stone for future investigations, pointing out obstacles and possible solutions. To conclude, a few promising studies are proposed that seem worthwhile but are beyond the scope of the present thesis.

Dependency Resolution

The study on dependency resolution in Chapter 2 ended without clear conclusions, largely due to a low signal-to-noise ratio in the coregistration experiments. There are probably ways for improving the study and obtaining more convincing results. The distance manipulation in the first study was rather subtle in relation to the overall sentence length. Compared to (7a), the distance between object and verb is increased by two words (*unglaublich talentierte*) in (7b). This may be sufficient in a different context, but the sentences in (7) are 23 words long.

(7) a. Maria und die unglaublich talentierte Regisseurin erkennen den Schauspieler, den der Redakteur und der Fotograf interviewen, um die Wahrheit über seine Ehe herauszufinden.

b. Maria und die Regisseurin erkennen den Schauspieler, den der unglaublich talentierte Redakteur und der Fotograf interviewen, um die Wahrheit über seine Ehe herauszufinden.

Two things can be done to improve this ratio: make the distance manipulation stronger or make the sentences shorter. In fact, if we do both, the above example could look like in (8).

(8) a. Die unglaublich talentierte Assistentin der Regisseurin erkennt den Schauspieler, den der Redakteur und der Fotograf interviewen, um die Auflage zu steigern.

b. Die Regisseurin erkennt den Schauspieler, den die unglaublich talentierte Assistentin des Redakteurs und der Fotograf interviewen, um die Auflage zu steigern.

Those are still very complex sentences but the ratio between distance manipulation and overall length looks a lot better. The first noun phrase in the main clause was omitted, the distance manipulation was made stronger, and the spillover phrase was shortened. The sentence is still 21 words long but the distance manipulation now consists of four words, including a simple nesting.

To improve the signal-to-noise ratio, such an experiment would need a sufficiently large sample size. Using the effect size from Experiment 3 in Chapter 2 as a gauge, around 200 participants would be necessary to obtain a power of 0.8, which is not feasible in a coregistration study. However, the required sample size decreases considerably when the effect size is increased by 50% ($N \approx 90$) or even 100% ($N \approx 50$). Thus, both stronger manipulations and larger sample sizes are necessary to elicit noticeable effects and measure them with reasonable certainty.

Parafoveal Preview

The quality and amount of information that readers can extract from regions that are not yet in foveal vision are the subject of a lively scientific debate (for a review, see Schotter et al., 2012). In this area, researchers have widely adopted the boundary paradigm (see Chapter 1). This experimental technique makes it possible to control precisely what type of information is available to readers in the parafovea and record the resulting eye movements. Potentially, coregistration enables researchers to also study the brain responses to stimuli that are only available in parafoveal vision. Dimigen et al. (2012) already explored this possibility with word list reading and found distinct preview effects in the ERP. However, it may prove problematic that there is no operational solution for the overlap problem in coregistration studies: Whatever effect is measured at word N with a longer latency (i.e., >250 ms) may just as well be the effect of word N+1, showing up as potential overlap at word N. Nevertheless, using the boundary paradigm in a coregistration study with reading of continuous text is without doubt a promising endeavor.

Visual World

In visual world studies, participants listen to sentences while they look at scenes with several objects. While the sentence unfolds, a participant's gaze typically shifts to the object that is currently salient. For example, while listening to *John poked the boy with a stick*, people may initially look at the picture representing John and then shift their attention to the picture of a boy. These scenes are typically designed in a way that allows researchers to decide between different theories based on participants' gaze preferences. In the John-poke-boy example from Chapter 1, there may be pictures of a boy with a stick, a boy without a stick, and a separate stick. The order in which listeners focus these objects is informative about their parse of the sentence.

Obtaining electrophysiological data in such experiments can be informative about the cognitive processes that trigger attention shifts. For example, a visual world study with strong implications for theories of online sentence processing is Altmann and Kamide (1999). They showed that participants use semantic properties of the verb to focus attention at likely referents. In a scene with, among other things, a boy and a cake, readers reliably direct their gaze at the cake upon hearing *The boy will eat…* Conducting such a study with a coregistration setup could shed light on the exact processes at work. What is the brain response when there are multiple or no possible referents? How are agreement violations registered and how does that interact with gaze preference? These and other questions could and should be answered with coregistration.

Refixations, Rereading, and Scanpaths

In Chapter 4, we investigated brain responses in natural reading contingent on whether readers made a regressive saccade. There are many more eye movement phenomena in reading that are worth investigating. For example, do the brain responses to the first fixation on a word and to subsequent refixations differ? This could inform us about whether the word is processed anew at each new fixation. Likewise, do readers show the same or similar brain responses when they reread a word? This is related to the question of whether readers actually reread a sentence when they move their eyes back to the beginning of the sentence upon encountering processing difficulties. Finally, instead of distinguishing between trials with and without regression, one could analyze trials contingent on the quality of readers' scanpaths. This would potentially enable us to delineate the cognitive processes that precede different eye movement trajectories in reading.

References

Altmann, G. T. M. & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. doi:10.1016/S0010-0277(99)00059-1

Baccino, T. (2011). Eye movements and concurrent event-related potentials: Eye fixation-related potential investigations in reading. In S. P. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 857–870). Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780199539789.013.0047

Baccino, T. & Manunta, Y. (2005). Eye-fixation-related potentials: Insight into parafoveal processing. *Journal of Psychophysiology*, *19*(3), 204–215. doi:10.1027/0269-8803.19.3.204

Bardy, F., Van Dun, B., Dillon, H., & McMahon, C. M. (2014). Deconvolution of overlapping cortical auditory evoked potentials recorded using short stimulus onset-asynchrony ranges. *Clinical Neurophysiology*, *125*(4), 814–826. doi:10.1016/j.clinph.2013.09.031

Bartek, B., Lewis, R. L., Vasishth, S., & Smith, M. R. (2011). In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1178–1198. doi:10.1037/a0024194

Başar, E., Başar-Eroglu, C., Karakaş, S., & Schürmann, M. (2001). Gamma, alpha, delta, and theta oscillations govern cognitive processes. *International Journal of Psychophysiology*, *39*(2–3), 241–248. doi:10.1016/S0167-8760(00)00145-8

Bastiaansen, M. C. M., Oostenveld, R., Jensen, O., & Hagoort, P. (2008). I see what you mean: Theta power increases are involved in the retrieval of lexical semantic information. *Brain and Language*, *106*(1), 15–28. doi:10.1016/j.bandl.2007.10.006

Bastiaansen, M. C. M., Van Berkum, J. A., & Hagoort, P. (2002a). Event-related theta power increases in the human EEG during online sentence processing. *Neuroscience Letters*, *323*(1), 13–16. doi:10.1016/S0304-3940(01)02535-6

Bastiaansen, M. C. M., Van Berkum, J. A., & Hagoort, P. (2002b). Syntactic processing modulates the θ rhythm of the human EEG. *NeuroImage*, *17*(3), 1479–1492. doi:10.1006/nimg.2002.1275

Bastiaansen, M. C. M., Van der Linden, M., ter Keurs, M., Dijkstra, T., & Hagoort, P. (2005). Theta responses are involved in lexical-semantic retrieval during language processing. *Journal of Cognitive Neuroscience*, *17*(3), 530–541. doi:10.1162/0898929053279469

Bastiaansen, M. & Hagoort, P. (2006). Oscillatory neuronal dynamics during language comprehension. *Progress in Brain Research*, *159*, 179–196. doi:10.1016/S0079-6123(06)59012-0

Bates, D., Maechler, M., Bolker, B., & Dai, B. (2013). lme4: Linear mixed-effect models using S4 classes [Software]. R package.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4 (Version 1.1-7) [Software].

Berg, P. & Scherg, M. (1994). A multiple source approach to the correction of eye artifacts. *Electroencephalography and Clinical Neurophysiology*, *90*(3), 229–241. doi:10.1016/0013-4694(94)90094-9

Berger, H. (1929). Über das Elektrenkephalogramm des Menschen. *European Archives of Psychiatry and Clinical Neuroscience*, *87*(1), 527–570.

Blair, R. C. & Karniski, W. (1993). An alternative method for significance testing of waveform difference potentials. *Psychophysiology*, *30*, 518–525. doi:10.1111/j.1469-8986.1993.tb02075.x

Bodis-Wollner, I., Gizycki, H., Avitable, M., Hussain, Z., Javeid, A., Habib, A., … Sabet, M. (2002). Perisaccadic occipital EEG changes quantified with wavelet analysis. *Annals of the New York Academy of Sciences*, *956*(1), 464–467. doi:10.1111/j.1749-6632.2002.tb02856.x

Bornkessel-Schlesewsky, I. & Schlesewsky, M. (2008). An alternative perspective on “semantic P600” effects in language comprehension. *Brain Research Reviews*, *59*(1), 55–73. doi:10.1016/j.brainresrev.2008.05.003

Box, G. E. P. & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B (Methodological)*, *26*, 211–252.

Braze, D., Shankweiler, D., Ni, W., & Palumbo, L. C. (2002). Readers’ eye movements distinguish anomalies of form and content. *Journal of Psycholinguistic Research*, *31*(1), 25–44. doi:10.1023/A:1014324220455

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, *1446*, 127–143. doi:10.1016/j.brainres.2012.01.055

Brown, C. M., Hagoort, P., & ter Keurs, M. (1999). Electrophysiological signatures of visual lexical processing: Open- and closed-class words. *Journal of Cognitive Neuroscience*, *11*(3), 261–281. doi:10.1162/089892999563382

Brunia, C. H. M. (2004). Slow potentials in anticipatory behavior. *Journal of Psychophysiology*, *18*(2), 59–60. doi:10.1027/0269-8803.18.23.59

Brunia, C. H., Möcks, J., Van den Berg-Lenssen, M. M., Coelho, M., et al. (1989). Correcting ocular artifacts in the EEG: A comparison of several models. *Journal of Psychophysiology*, *3*(1), 1–50.

Bruns, A. (2004). Fourier-, Hilbert-and wavelet-based signal analysis: Are they really different approaches? *Journal of Neuroscience Methods*, *137*(2), 321–332. doi:10.1016/j.jneumeth.2004.03.002

Chwilla, D. J. & Kolk, H. H. J. (2005). Accessing world knowledge: Evidence from N400 and reaction time priming. *Cognitive Brain Research*, *25*(3), 589–606. doi:10.1016/j.cogbrainres.2005.08.011

Clifton, C., Jr., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In *Eye movements: A window on mind and brain* (pp. 341–372). Oxford: Elsevier. doi:10.1016/B978-008044980-7/50017-3

Cooley, J. W. & Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, *19*, 297–301. doi:10.1090/S0025-5718-1965-0178586-1

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, *13*(1), 21–58. doi:10.1080/016909698386582

Croft, R. & Barry, R. (2000). Removal of ocular artifact from the EEG: A review. *Neurophysiologie Clinique*, *30*(1), 5–19. doi:10.1016/S0987-7053(00)00055-1

Dambacher, M. & Kliegl, R. (2007). Synchronizing timelines: Relations between fixation durations and N400 amplitudes during sentence reading. *Brain Research*, *1155*, 147–162. doi:10.1016/j.brainres.2007.04.027

Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research*, *1084*, 89–103. doi:10.1016/j.brainres.2006.02.010

Dandekar, S., Privitera, C., Carney, T., & Klein, S. A. (2012). Neural saccadic response estimation during natural viewing. *Journal of Neurophysiology*, *107*(6), 1776–1790. doi:10.1152/jn.00237.2011

Delgado, R. E. & Özdamar, Ö. (2004). Deconvolution of evoked responses obtained at high stimulus rates. *The Journal of the Acoustical Society of America*, *115*(3), 1242–1251. doi:10.1121/1.1639327

DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, *61*, 150–162. doi:10.1016/j.neuropsychologia.2014.06.016

Delorme, A., Sejnowski, T., & Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage*, *34*, 1443–1449. doi:10.1016/j.neuroimage.2006.11.004

Demberg, V. & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *109*(2), 193–210. doi:10.1016/j.cognition.2008.07.008

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, *69*(2), 85–103. doi:10.1016/j.jml.2013.04.003

Dimigen, O. (2014). *Co-registration of eye movements and EEG during active vision* (Doctoral dissertation, Humboldt-Universität zu Berlin, Berlin, Germany).

Dimigen, O., Kliegl, R., & Sommer, W. (2012). Trans-saccadic parafoveal preview benefits in fluent reading: A study with fixation-related brain potentials. *NeuroImage*, *62*(1), 381–393. doi:10.1016/j.neuroimage.2012.04.006

Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., & Kliegl, R. (2011). Coregistration of eye movements and EEG in natural reading: Analyses and review. *Journal of Experimental Psychology: General*, *140*(4), 552–572. doi:10.1037/a0023885

Dimigen, O., Sommer, W., & Kliegl, R. (2007). Long reading regressions are accompanied by a P600-like brain potential. Evidence from simultaneous recordings of eye movements and ERPs. In *Proceedings of the 14th European Conference on Eye Movements*. Potsdam, Germany.

Dimigen, O., Valsecchi, M., Sommer, W., & Kliegl, R. (2009). Human microsaccade-related visual brain responses. *The Journal of Neuroscience*, *29*(39), 12321–12331. doi:10.1523/JNEUROSCI.0911-09.2009

Ehrlich, S. F. & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*(6), 641–655. doi:10.1016/S0022-5371(81)90220-6

Elbert, T., Lutzenberger, W., Rockstroh, B., & Birbaumer, N. (1985). Removal of ocular artifacts from the EEG — A biophysical approach to the EOG. *Electroencephalography and Clinical Neurophysiology*, *60*(5), 455–463. doi:10.1016/0013-4694(85)91020-X

Engbert, R. & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research*, *43*, 1035–1045. doi:10.1016/S0042-6989(03)00084-1

Ferreira, F. & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*(3), 348–368. doi:10.1016/0749-596X(86)90006-9

Ferreira, F. & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, *1*(1–2), 71–83. doi:10.1111/j.1749-818X.2007.00007.x

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language*, *47*(2), 250–272. doi:10.1016/S0749-596X(02)00004-9

Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry*, *9*(3), 427–473.

Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies* (Doctoral dissertation, University of Connecticut, Storrs, CT).

Frazier, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language & Linguistic Theory*, *5*(4), 519–559. doi:10.1007/BF00138988

Frazier, L. & d'Arcais, G. B. F. (1989). Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language*, *28*(3), 331–344. doi:10.1016/0749-596X(89)90037-5

Frazier, L. & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291–325. doi:10.1016/0010-0277(78)90002-1

Frazier, L. & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178–210. doi:10.1016/0010-0285(82)90008-1

Friederici, A. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and Language*, *50*(3), 259–281. doi:10.1006/brln.1995.1048

Friederici, A. D. & Mecklinger, A. (1996). Syntactic parsing as revealed by brain responses: First-pass and second-pass parsing processes. *Journal of Psycholinguistic Research*, *25*(1), 157–176. doi:10.1007/BF01708424

Frisch, S. & Schlesewsky, M. (2001). The N400 reflects problems of thematic hierarchizing. *NeuroReport*, *12*(15), 3391–3394. doi:10.1097/00001756-200110290-00048

Gernsbacher, M. A. & Hargreaves, D. J. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language*, *27*(6), 699–717. doi:10.1016/0749-596X(88)90016-2

Gevins, A., Smith, M. E., McEvoy, L., & Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice. *Cerebral Cortex*, *7*(4), 374–385. doi:10.1093/cercor/7.4.374

Ghandeharion, H. & Erfanian, A. (2010). A fully automatic ocular artifact suppression from EEG data using higher order statistics: Improved performance by wavelet analysis. *Medical Engineering & Physics*, *32*(7), 720–729. doi:10.1016/j.medengphy.2010.04.010

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1). doi:10.1016/S0010-0277(98)00034-1

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain. Papers from the first Mind Articulation Project symposium* (pp. 95–126). Cambridge, MA: MIT Press.

Gibson, E., Desmet, T., Grodner, D., Watson, D., & Ko, K. (2005). Reading relative clauses in English. *Cognitive Linguistics*, *16*(2), 313–353. doi:10.1515/cogl.2005.16.2.313

Gouvea, A., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes*, *25*(2), 149–188. doi:10.1080/01690960902965951

Gratton, G. (1998). Dealing with artifacts: The EOG contamination of the event-related brain potential. *Behavior Research Methods, Instruments, & Computers*, *30*(1), 44–53. doi:10.3758/BF03209415

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*(4), 468–484. doi:10.1016/0013-4694(83)90135-9

Graupner, S., Velichkovsky, B. M., Pannasch, S., & Marx, J. (2007). Surprise, surprise: Two distinct components in the visually evoked distractor effect. *Psychophysiology*, *44*(2), 251–261. doi:10.1111/j.1469-8986.2007.00504.x

Greiner, B. (2004). An online recruitment system for economic experiments. In K. Kremer & V. Macho (Eds.), *Forschung und wissenschaftliches Rechnen. Beiträge zum Heinz-Billing-Preis 2003* (Vol. 63, pp. 79–93). GWDG-Berichte. Göttingen, Germany.

Grober, E., Beardsley, W., & Caramazza, A. (1978). Parallel function strategy in pronoun assignment. *Cognition*, *6*(2), 117–133. doi:10.1016/0010-0277(78)90018-5

Grodner, D. & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science*, *29*(2), 261–290. doi:10.1207/s15516709cog0000_7

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011a). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, *48*(12), 1711–1725. doi:10.1111/j.1469-8986.2011.01273.x

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011b). Mass univariate analysis of event-related brain potentials/fields II: Simulation studies. *Psychophysiology*, *49*(12), 1726–1737. doi:10.1111/j.1469-8986.2011.01272.x

Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology*, *34*(6), 660–676. doi:10.1111/j.1469-8986.1997.tb02142.x

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. doi:10.1162/089892903322370807

Hagoort, P. & Brown, C. (1994). Brain responses to lexical ambiguity resolution and parsing. In C. Clifton Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 45–81). Lawrence Erlbaum Associates, Inc.

Hagoort, P., Brown, C. M., & Osterhout, L. (1999). The neurocognition of syntactic processing. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 273–316). Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198507932.003.0009

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, *8*(4), 439–483. doi:10.1080/01690969308407585

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*(5669), 438–441. doi:10.1126/science.1095455

Hahne, A. & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research*, *13*(3), 339–356. doi:10.1016/S0926-6410(01)00127-6

Hald, L. A. (2003). *The integration of semantic versus world knowledge during on-line sentence comprehension* (Doctoral dissertation, The University of Arizona, Dissertations Abstracts International, DAI-B 63/12, 6114.).

Hald, L. A., Bastiaansen, M. C. M., & Hagoort, P. (2006). EEG theta and gamma responses to semantic violations in online sentence processing. *Brain and Language*, *96*(1), 90–105. doi:10.1016/j.bandl.2005.06.007

Hald, L. A., Steenbeek-Planting, E. G., & Hagoort, P. (2007). The interaction of discourse context and world knowledge in online sentence comprehension. Evidence from the N400. *Brain Research*, *1146*, 210–218. doi:10.1016/j.brainres.2007.02.054

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of NAACL 2001* (pp. 1–8). Pittsburgh, Pennsylvania. doi:10.3115/1073336.1073357

Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, *32*(2), 101–123. doi:10.1023/A:1022492123056

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, *30*(4), 643–672. doi:10.1207/s15516709cog0000_64

Hansen, J. C. (1983). Separation of overlapping waveforms having known temporal distributions. *Journal of Neuroscience Methods*, *9*(2), 127–139. doi:10.1016/0165-0270(83)90126-7

Hanulíková, A., Van Alphen, P. M., Van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, *24*(4), 878–887. doi:10.1162/jocn_a_00103

Harmony, T. (2013). The functional significance of delta oscillations in cognitive processing. *Frontiers in Integrative Neuroscience*, *7*(83). doi:10.3389/fnint.2013.00083

Harmony, T., Alba, A., Marroquín, J. L., & González-Frankenberger, B. (2009). Time-frequency-topographic analysis of induced power and synchrony of EEG signals during a go/no-go task. *International Journal of Psychophysiology*, *71*(1), 9–16. doi:10.1016/j.ijpsycho.2008.07.020

Heister, J., Würzner, K., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). dlexDB – Eine lexikalische Datenbank für die psychologische und linguistische Forschung. *Psychologische Rundschau*, *62*(1), 10–20. doi:10.1026/0033-3042/a000029

Henderson, J. M., Luke, S. G., Schmidt, J., & Richards, J. E. (2013). Co-registration of eye movements and event-related potentials in connected-text paragraph reading. *Frontiers in Systems Neuroscience*, *7*(28). doi:10.3389/fnsys.2013.00028

Herdman, A. T. & Ryan, J. D. (2007). Spatio-temporal brain dynamics underlying saccade execution, suppression, and error-related feedback. *Journal of Cognitive Neuroscience*, *19*(3), 420–432. doi:10.1162/jocn.2007.19.3.420

Herrmann, C. S., Munk, M. H. J., & Engel, A. K. (2004). Cognitive functions of gamma-band activity: Memory match and utilization. *Trends in Cognitive Sciences*, *8*(8), 347–355. doi:10.1016/j.tics.2004.06.006

Hofmeister, P. (2011). Representational complexity and memory retrieval in language comprehension. *Language and Cognitive Processes*, *26*(3), 376–405. doi:10.1080/01690965.2010.492642

Hopf, J., Bader, M., Meng, M., & Bayer, J. (2003). Is human sentence parsing serial or parallel? Evidence from event-related brain potentials. *Cognitive Brain Research*, *15*(2), 165–177. doi:10.1016/S0926-6410(02)00149-0

Hutzler, F., Braun, M., Võ, M. L., Engl, V., Hofmann, M., Dambacher, M., … Jacobs, A. M. (2007). Welcome to the real world: Validating fixation-related brain potentials for ecologically valid settings. *Brain Research*, *1172*, 124–129. doi:10.1016/j.brainres.2007.07.025

Hutzler, F., Fuchs, I., Gagl, B., Schuster, S., Richlan, F., Braun, M., & Hawelka, S. (2013). Parafoveal X-masks interfere with foveal word recognition: Evidence from fixation-related brain potentials. *Frontiers in Systems Neuroscience*, *7*(33). doi:10.3389/fnsys.2013.00033

Ille, N., Berg, P., & Scherg, M. (2002). Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *Journal of Clinical Neurophysiology*, *19*(2), 113–124. doi:10.1097/00004691-200203000-00002

Iriarte, J., Urrestarazu, E., Valencia, M., Alegre, M., Malanda, A., Viteri, C., & Artieda, J. (2003). Independent component analysis as a tool to eliminate artifacts in EEG: A quantitative study. *Journal of Clinical Neurophysiology*, *20*(4), 249–257. doi:10.1097/00004691-200307000-00004

Jasper, H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, *10*(2), 371–375.

Jewett, D. L., Caplovitz, G., Baird, B., Trumpis, M., Olson, M. P., & Larson-Prior, L. J. (2004). The use of QSD (q-sequence deconvolution) to recover superposed, transient evoked-responses. *Clinical Neurophysiology*, *115*(12), 2754–2775. doi:10.1016/j.clinph.2004.06.014

Joyce, C. A., Gorodnitsky, I. F., & Kutas, M. (2004). Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. *Psychophysiology*, *41*(2), 313–325. doi:10.1111/j.1469-8986.2003.00141.x

Jung, T., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, *111*(10), 1745–1758. doi:10.1016/S1388-2457(00)00386-2

Just, M. A. & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354. doi:10.1037/0033-295X.87.4.329

Kaan, E. (2002). Investigating the effects of distance and number interference in processing subject-verb dependencies: an ERP study. *Journal of Psycholinguistic Research*, *31*(2), 165–193. doi:10.1023/A:1014978917769

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*(2), 159–201. doi:10.1080/016909600386084

Kamienkowski, J. E., Ison, M. J., Quiroga, R. Q., & Sigman, M. (2012). Fixation-related potentials in visual search: A combined EEG and eye tracking study. *Journal of Vision*, *12*(7), 1–20. doi:10.1167/12.7.4

Karniski, W., Blair, R. C., & Snider, A. D. (1994). An exact statistical method for comparing topographic maps, with any number of subjects and electrodes. *Brain Topography*, *6*(3), 203–210. doi:10.1007/BF01187710

Kaunitz, L. N., Kamienkowski, J. E., Varatharajah, A., Sigman, M., Quiroga, R. Q., & Ison, M. J. (2014). Looking for a face in the crowd: Fixation-related potentials in an eye-movement visual search task. *NeuroImage*, *89*, 297–305. doi:10.1016/j.neuroimage.2013.12.006

Kennett, S., Van Velzen, J., Eimer, M., & Driver, J. (2007). Disentangling gaze shifts from preparatory ERP effects during spatial attention. *Psychophysiology*, *44*(1), 69–78. doi:10.1111/j.1469-8986.2006.00470.x

Kim, A. & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*(2), 205–225. doi:10.1016/j.jml.2004.10.002

King, J. W. & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*, 580–602. doi:10.1016/0749-596X(91)90027-H

King, J. W. & 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. doi:10.1162/jocn.1995.7.3.376

Kliegl, R., Dambacher, M., Dimigen, O., Jacobs, A. M., & Sommer, W. (2011). Eye movements and brain electric potentials during reading. *Psychological Research*, *76*(2), 145–158. doi:10.1007/s00426-011-0376-x

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews*, *29*(2–3), 169–195. doi:10.1016/S0165-0173(98)00056-3

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617. doi:10.1016/j.tics.2012.10.007

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: The inhibition-timing hypothesis. *Brain Research Reviews*, *53*(1), 63–88. doi:10.1016/j.brainresrev.2006.06.003

Klimesch, W., Schimke, H., & Schwaiger, J. (1994). Episodic and semantic memory: An analysis in the EEG theta and alpha band. *Electroencephalography and Clinical Neurophysiology*, *91*(6), 428–441. doi:10.1016/0013-4694(94)90164-3

Knyazev, G. G. (2007). Motivation, emotion, and their inhibitory control mirrored in brain oscillations. *Neuroscience & Biobehavioral Reviews*, *31*(3), 377–395. doi:10.1016/j.neubiorev.2006.10.004

Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, *29*(6), 627–645. doi:10.1023/A:1026528912821

Konieczny, L. & Döring, P. (2003). Anticipation of clause-final heads: Evidence from eye-tracking and SRNs. In *Proceedings of the ICCS/ASCS: Joint International Conference on Cognitive Science*. Sydney, Australia.

Kretzschmar, F., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). Parafoveal versus foveal N400s dissociate spreading activation from contextual fit. *NeuroReport*, *20*(18), 1613–1618. doi:10.1097/WNR.0b013e328332c4f4

Kretzschmar, F., Pleimling, D., Hosemann, J., Füssel, S., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Subjective impressions do not mirror online reading effort: Concurrent EEG-eyetracking evidence from the reading of books and digital media. *PloS ONE*, *8*(2), 1–11. doi:10.1371/journal.pone.0056178

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23–49. doi:10.1016/j.brainres.2006.12.063

Kurtzberg, D. & Vaughan, H. G., Jr. (1982). Topographic analysis of human cortical potentials preceding self-initiated and visually triggered saccades. *Brain Research*, *243*(1), 1–9. doi:10.1016/0006-8993(82)91115-5

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190–207). Oxford, United Kingdom: Oxford University Press. doi:10.1093/acprof:oso/9780195395518.003.0065

Kutas, M. & 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. doi:10.1146/annurev.psych.093008.131123

Kutas, M. & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. doi:10.1126/science.7350657

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920–933. doi:10.1038/nrn2532

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. doi:10.1016/j.cognition.2007.05.006

Levy, R. P. & Keller, F. (2012). Expectation and locality effects in German verb-final structures. *Journal of Memory and Language*, *68*(2), 199–222. doi:10.1016/j.jml.2012.02.005

Levy, R., Fedorenko, E., & Gibson, E. (2013). The syntactic complexity of russian relative clauses. *Journal of Memory and Language*. doi:10.1016/j.jml.2012.10.005

Lewis, R. L. (1998). Reanalysis and limited repair parsing: Leaping off the garden path. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 247–285). Dordrecht: Kluwer Academic Publishers. doi:10.1007/978-94-015-9070-9_8

Lewis, R. L. & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375–419. doi:10.1207/s15516709cog0000_25

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454. doi:10.1016/j.tics.2006.08.007

Li, Y., Ma, Z., Lu, W., & Li, Y. (2006). Automatic removal of the eye blink artifact from EEG using an ICA-based template matching approach. *Physiological Measurement*, *27*(4), 425–436. doi:10.1088/0967-3334/27/4/008

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

Maguire, M. J., Brier, M. R., & Ferree, T. C. (2010). EEG theta and alpha responses reveal qualitative differences in processing taxonomic versus thematic semantic relationships. *Brain and Language*, *114*(1), 16–25. doi:10.1016/j.bandl.2010.03.005

Makeig, S., Bell, A. J., Jung, T., & Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. In D. Touretzky, M. Mozer, & M. Hasselmo (Eds.), *Advances in neural information processing systems 8* (pp. 145–151). Cambridge, MA: MIT Press.

Makuuchi, M., Bahlmann, J., Anwander, A., & Friederici, A. D. (2009). Segregating the core computational faculty of human language from working memory. *Proceedings of the National Academy of Sciences*, *106*(20), 8362–8367. doi:10.1073/pnas.0810928106

Mantini, D., Perrucci, M. G., Cugini, S., Ferretti, A., Romani, G. L., & Del Gratta, C. (2007). Complete artifact removal for EEG recorded during continuous fMRI using independent component analysis. *NeuroImage*, *34*, 598–607. doi:10.1016/j.neuroimage.2006.09.037

Maris, E. (2012). Statistical testing in electrophysiological studies. *Psychophysiology*, *49*(4), 549–565. doi:10.1111/j.1469-8986.2011.01320.x

Maris, E. & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. doi:10.1016/j.jneumeth.2007.03.024

Martin, A. E., Nieuwland, M. S., & Carreiras, M. (2012). Event-related brain potentials index cue-based retrieval interference during sentence comprehension. *NeuroImage*, *59*, 1859–1869. doi:10.1016/j.neuroimage.2011.08.057

Matchin, W., Sprouse, J., & Hickok, G. (2014). A structural distance effect for backward anaphora in Broca's area: An fMRI study. *Brain and Language*, *138*, 1–11. doi:10.1016/j.bandl.2014.09.001

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, *48*(1), 67–91. doi:10.1016/S0749-596X(02)00515-6

Menenti, L., Petersson, K. M., Scheeringa, R., & Hagoort, P. (2009). When elephants fly: Differential sensitivity of right and left inferior frontal gyri to discourse and world knowledge. *Journal of Cognitive Neuroscience*, *21*(12), 2358–2368. doi:10.1162/jocn.2008.21163

Meseguer, E., Carreiras, M., & Clifton, C., Jr. (2002). Overt reanalysis strategies and eye movements during the reading of mild garden path sentences. *Memory & Cognition*, *30*(4), 551–561. doi:10.3758/BF03194956

Metzner, P., von der Malsburg, T., Vasishth, S., & Rösler, F. (2015). Brain responses to world-knowledge violations: A comparison of stimulus- and fixation-triggered event-related potentials and neural oscillations. *Journal of Cognitive Neuroscience*, *27*(5). doi:10.1162/jocn_a_00731

Meyberg, S., Werkle-Bergner, M., Sommer, W., & Dimigen, O. (2015). Microsaccade-related brain potentials signal the focus of visuospatial attention. *NeuroImage*, *104*, 79–88. doi:10.1016/j.neuroimage.2014.09.065

Mitchell, D. C., Cuetos, F., Corley, M. M. B., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research*, *24*(6), 469–488. doi:10.1007/BF02143162

Mitchell, D. C., Shen, X., Green, M. J., & Hodgson, T. L. (2008). Accounting for regressive eye-movements in models of sentence processing: A reappraisal of the selective reanalysis hypothesis. *Journal of Memory and Language*, *59*(3), 266–293. doi:10.1016/j.jml.2008.06.002

Mitra, P. P. & Pesaran, B. (1999). Analysis of dynamic brain imaging data. *Biophysical Journal*, *76*(2), 691–708. doi:10.1016/S0006-3495(99)77236-X

Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, *48*(2), 229–240. doi:10.1111/j.1469-8986.2010.01061.x

Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, *47*(8), 908–930. doi:10.1016/j.cortex.2011.02.019

Müller, M. M., Gruber, T., & Keil, A. (2000). Modulation of induced gamma band activity in the human EEG by attention and visual information processing. *International Journal of Psychophysiology*, *38*(3), 283–299. doi:10.1016/S0167-8760(00)00171-9

Neville, H., Nicol, J., Barss, A., Forster, K., & Garrett, M. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, *3*(2), 151–165. doi:10.1162/jocn.1991.3.2.151

Nieuwland, M. S. & Van Berkum, J. A. (2006). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. *Brain Research*, *1118*, 155–167. doi:10.1016/j.brainres.2006.08.022

Nieuwland, M. S. & Van Berkum, J. A. (2008). The interplay between semantic and referential aspects of anaphoric noun phrase resolution: Evidence from ERPs. *Brain and Language*, *106*(2), 119–131. doi:10.1016/j.bandl.2008.05.001

Okada, Y., Jung, J., & Kobayashi, T. (2007). An automatic identification and removal method for eye-blink artifacts in event-related magnetoencephalographic measurements. *Physiological Measurement*, *28*(12), 1523–1532. doi:10.1088/0967-3334/28/12/006

Osterhout, L., Allen, M., McLaughlin, J., & Inoue, K. (2002). Brain potentials elicited by prose-embedded linguistic anomalies. *Memory & Cognition*, *30*(8), 1304–1312. doi:10.3758/BF03213412

Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition*, *25*(3), 273–285. doi:10.3758/BF03211283

Osterhout, L. & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785–806. doi:10.1016/0749-596X(92)90039-Z

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 786–803. doi:10.1037/0278-7393.20.4.786

Osterhout, L. & Mobley, L. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, *34*(6), 739–773. doi:10.1006/jmla.1995.1033

Otten, L. J. & Rugg, M. D. (2005). Interpreting event-related brain potentials. In T. C. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 3–16). Cambridge, MA: MIT Press Cambridge, MA.

Penolazzi, B., Angrilli, A., & Job, R. (2009). Gamma EEG activity induced by semantic violation during sentence reading. *Neuroscience Letters*, *465*(1), 74–78. doi:10.1016/j.neulet.2009.08.065

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, *22*(3), 407–428. doi:10.1016/j.cogbrainres.2004.09.012

Pickering, M. J. & Traxler, M. J. (2001). Strategies for processing unbounded dependencies: Lexical information and verb–argument assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1401–1410. doi:10.1037/0278-7393.27.6.1401

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., … Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*(2), 127–152. doi:10.1111/1469-8986.3720127

Picton, T. W., Van Roon, P., Armilio, M. L., Berg, P., Ille, N., & Scherg, M. (2000). The correction of ocular artifacts: A topographic perspective. *Clinical Neurophysiology*, *111*(1), 53–65. doi:10.1016/S1388-2457(99)00227-8

Plöchl, M., Ossandón, J. P., & König, P. (2012). Combining EEG and eye tracking: Identification, characterization and correction of eye movement artifacts in electroencephalographic data. *Frontiers in Human Neuroscience*, *6*(278). doi:10.3389/fnhum.2012.00278

R Core Team. (2013). R: A language and environment for statistical computing [Software]. Vienna, Austria: R Foundation for Statistical Computing.

Rämä, P. & Baccino, T. (2010). Eye fixation-related potentials (EFRPs) during object identification. *Visual Neuroscience*, *27*(5–6), 187–192. doi:10.1017/S0952523810000283

Raney, G. E. & Rayner, K. (1993). Event-related brain potentials, eye movements, and reading. *Psychological Science*, *4*(5), 283–286. doi:10.1111/j.1467-9280.1993.tb00565.x

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, *7*(1), 65–81. doi:10.1016/0010-0285(75)90005-5

Rayner, K. (1977). Visual attention in reading: Eye movements reflect cognitive processes. *Memory & Cognition*, *5*(4), 443–448. doi:10.3758/BF03197383

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. doi:10.1037/0033-2909.124.3.372

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 358–374. doi:10.1016/S0022-5371(83)90236-0

Rayner, K. & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), 191–201. doi:10.3758/BF03197692

Rayner, K., Kambe, G., & Duffy, S. A. (2000). The effect of clause wrap-up on eye movements during reading. *The Quarterly Journal of Experimental Psychology: Section A*, *53*(4), 1061–1080. doi:10.1080/713755934

Rayner, K. & Raney, G. E. (1996). Eye movement control in reading and visual search: Effects of word frequency. *Psychonomic Bulletin & Review*, *3*(2), 245–248. doi:10.3758/BF03212426

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1290–1301. doi:10.1037/0278-7393.30.6.1290

Roehm, D., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2007). The internal structure of the N400: Frequency characteristics of a language related ERP component. *Chaos and Complexity Letters*, *2*(2), 365–395.

Roehm, D., Klimesch, W., Haider, H., & Doppelmayr, M. (2001). The role of theta and alpha oscillations for language comprehension in the human electroencephalogram. *Neuroscience Letters*, *310*(2–3), 137–140. doi:10.1016/S0304-3940(01)02106-1

Roehm, D., Schlesewsky, M., Bornkessel, I., Frisch, S., & Haider, H. (2004). Fractionating language comprehension via frequency characteristics of the human EEG. *NeuroReport*, *15*(3), 409–412. doi:10.1097/00001756-200403010-00005

Rohde, D. (2003). Linger: A flexible platform for language processing experiments (Version 2.88) [Software]. Cambridge, MA: Massachusetts Institute of Technology.

Rohrbaugh, J. W. & Gaillard, A. W. K. (1983). Sensory and motor aspects of the contingent negative variation. In A. W. K. Gaillard & W. Ritter (Eds.), *Tutorials in event-related potential research: Endogenous components* (pp. 269–310). North-Holland Publishing Company. doi:10.1016/S0166-4115(08)62044-0

Rong, F. & Contreras-Vidal, J. L. (2006). Magnetoencephalographic artifact identification and automatic removal based on independent component analysis and categorization approaches. *Journal of Neuroscience Methods*, *157*(2), 337–354. doi:10.1016/j.jneumeth.2006.04.024

Rosenfeld, A. & Pfaltz, J. L. (1966). Sequential operations in digital picture processing. *Journal of the ACM*, *13*(4), 471–494. doi:10.1145/321356.321357

Rösler, F., Pechmann, T., Streb, J., Röder, B., & Hennighausen, E. (1998). Parsing of sentences in a language with varying word order: Word-by-word variations of processing demands are revealed by event-related brain potentials. *Journal of Memory and Language*, *38*(2), 150–176. doi:10.1006/jmla.1997.2551

Rösler, F., Pütz, P., Friederici, A. D., & Hahne, A. (1993). Event-related brain potentials while encountering semantic and syntactic constraint violations. *Journal of Cognitive Neuroscience*, *5*(3), 345–362. doi:10.1162/jocn.1993.5.3.345

Rugg, M. D. & Coles, M. G. H. (1995). The ERP and cognitive psychology: Conceptual issues. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind: Event-related brain potentials and cognition* (pp. 27–39). Oxford: Oxford University Press.

Samet, H. & Tamminen, M. (1988). Efficient component labeling of images of arbitrary dimension represented by linear bintrees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *10*(4), 579–586. doi:10.1109/34.3918

Santi, A. & Grodzinsky, Y. (2007). Working memory and syntax interact in Broca's area. *NeuroImage*, *37*(1), 8–17. doi:10.1016/j.neuroimage.2007.04.047

Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, *137*, 29–39. doi:10.1016/j.bandl.2014.07.010

Schiff, S. J., Aldroubi, A., Unser, M., & Sato, S. (1994). Fast wavelet transformation of EEG. *Electroencephalography and Clinical Neurophysiology*, *91*(6), 442–455. doi:10.1016/0013-4694(94)90165-1

Schlögl, A., Keinrath, C., Zimmermann, D., Scherer, R., Leeb, R., & Pfurtscheller, G. (2007). A fully automated correction method of EOG artifacts in EEG recordings. *Clinical Neurophysiology*, *118*(1), 98–104. doi:10.1016/j.clinph.2006.09.003

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, & Psychophysics*, *74*(1), 5–35. doi:10.3758/s13414-011-0219-2

Schotter, E. R., Tran, R., & Rayner, K. (2014). Don't believe what you read (only once): Comprehension is supported by regressions during reading. *Psychological Science*, *25*(6), 1218–1226. doi:10.1177/0956797614531148

Sereno, S. C. & Rayner, K. (2003). Measuring word recognition in reading: Eye movements and event-related potentials. *Trends in Cognitive Sciences*, *7*(11), 489–493. doi:10.1016/j.tics.2003.09.010

Sereno, S. C., Rayner, K., & Posner, M. I. (1998). Establishing a time-line of word recognition: Evidence from eye movements and event-related potentials. *NeuroReport*, *9*(10), 2195–2200. doi:10.1097/00001756-199807130-00009

Simola, J., Holmqvist, K., & Lindgren, M. (2009). Right visual field advantage in parafoveal processing: Evidence from eye-fixation-related potentials. *Brain and Language*, *111*(2), 101–113. doi:10.1016/j.bandl.2009.08.004

Simola, J., Torniainen, J., Moisala, M., Kivikangas, M., & Krause, C. M. (2013). Eye movement related brain responses to emotional scenes during free viewing. *Frontiers in Systems Neuroscience*, *7*(41), 1–16. doi:10.3389/fnsys.2013.00041

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, *38*(4), 387–401. doi:10.1016/0013-4694(75)90263-1

Staub, A. & Rayner, K. (2007). Eye movements and on-line comprehension processes. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 327–342). Oxford: Oxford University Press.

Steinhauer, K. & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language*, *120*(2), 135–162. doi:10.1016/j.bandl.2011.07.001

Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, *1*(3), 227–245. doi:10.1080/01690968608407062

Streb, J., Hennighausen, E., & Rösler, F. (2004). Different anaphoric expressions are investigated by event-related brain potentials. *Journal of Psycholinguistic Research*, *33*(3), 175–201. doi:10.1023/B:JOPR.0000027961.12577.d8

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, *48*(3), 542–562. doi:10.1016/S0749-596X(02)00536-3

Tallon-Baudry, C. & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, *3*(4), 151–162. doi:10.1016/S1364-6613(99)01299-1

Tanner, D. (2015). On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement. *Cortex*, *66*, 149–155. doi:10.1016/j.cortex.2014.04.007

Thickbroom, G. W., Knezevič, W., Carroll, W. M., & Mastaglia, F. L. (1991). Saccade onset and offset lambda waves: Relation to pattern movement visually evoked potentials. *Brain Research*, *551*(1–2), 150–156. doi:10.1016/0006-8993(91)90927-N

Thickbroom, G. W. & Mastaglia, F. L. (1986). Presaccadic spike potential: Relation to eye movement direction. *Electroencephalography and Clinical Neurophysiology*, *64*(3), 211–214. doi:10.1016/0013-4694(86)90167-7

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, *47*. doi:10.1006/jmla.2001.2836

Traxler, M. J., Pickering, M. J., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, *39*(4), 558–592. doi:10.1006/jmla.1998.2600

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*(3), 285–318. doi:10.1006/jmla.1994.1014

Tukey, J. W. (1967). An introduction to the calculations of numerical spectrum analysis. In B. Harris (Ed.), *Spectral analysis of time series* (pp. 25–46). New York: Wiley.

Valsecchi, M., Dimigen, O., Kliegl, R., Sommer, W., & Turatto, M. (2009). Microsaccadic inhibition and P300 enhancement in a visual oddball task. *Psychophysiology*, *46*(3), 635–644. doi:10.1111/j.1469-8986.2009.00791.x

Van Berkum, J. A., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, *41*(2), 147–182. doi:10.1006/jmla.1999.2641

Van Berkum, J. A., Brown, C. M., Hagoort, P., & Zwitserlood, P. (2003). Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology*, *40*(2), 235–248. doi:10.1111/1469-8986.00025

Van Berkum, J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443–467. doi:10.1037/0278-7393.31.3.443

Van Petten, C. & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition*, *18*(4), 380–393. doi:10.3758/BF03197127

Van Petten, C. & Kutas, M. (1991). Influences of semantic and syntactic context on open- and closed-class words. *Memory & Cognition*, *19*(1), 95–112. doi:10.3758/BF03198500

Vasishth, S., Brüssow, S., Lewis, R. L., & Drenhaus, H. (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science*, *32*(4), 685–712. doi:10.1080/03640210802066865

Vasishth, S., Chen, Z., Li, Q., & Guo, G. (2014). Processing chinese relative clauses: Evidence for the subject-relative advantage. *PLoS ONE*, *8*(10), 1–14. doi:10.1371/journal.pone.0077006

Vasishth, S., von der Malsburg, T., & Engelmann, F. (2012). What eye movements can tell us about sentence comprehension. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(2), 125–134. doi:10.1002/wcs.1209

Vigario, R. N. (1997). Extraction of ocular artefacts from EEG using independent component analysis. *Electroencephalography and Clinical Neurophysiology*, *103*(3), 395–404. doi:10.1016/S0013-4694(97)00042-8

von der Malsburg, T. & Vasishth, S. (2011). What is the scanpath signature of syntactic reanalysis? *Journal of Memory and Language*, *65*(2), 109–127. doi:10.1016/j.jml.2011.02.004

von der Malsburg, T. & Vasishth, S. (2013). Scanpaths reveal syntactic underspecification and reanalysis strategies. *Language and Cognitive Processes*, *28*(10), 1545–1578. doi:10.1080/01690965.2012.728232

Wallstrom, G. L., Kass, R. E., Miller, A., Cohn, J. F., & Fox, N. A. (2004). Automatic correction of ocular artifacts in the EEG: A comparison of regression-based and component-based methods. *International Journal of Psychophysiology*, *53*(2), 105–119. doi:10.1016/j.ijpsycho.2004.03.007

Walter, W. G., Cooper, R., Aldridge, V., McCallum, W., & Winter, A. (1964). Contingent negative variation: An electric sign of sensorimotor association and expectancy in the human brain. *Nature*, (203), 380–4.

Wang, L., Jensen, O., Van den Brink, D., Weder, N., Schoffelen, J., Magyari, L., … Bastiaansen, M. (2012). Beta oscillations relate to the N400m during language comprehension. *Human Brain Mapping*, *33*(12), 2898–2912. doi:10.1002/hbm.21410

Wang, L., Zhu, Z., & Bastiaansen, M. (2012). Integration or predictability? A further specification of the functional role of gamma oscillations in language comprehension. *Frontiers in Psychology*, *3*, 1–12. doi:10.3389/fpsyg.2012.00187

Wang, T., Özdamar, Ö., Bohórquez, J., Shen, Q., & Cheour, M. (2006). Wiener filter deconvolution of overlapping evoked potentials. *Journal of*

Neuroscience Methods, *158*(2), 260–270. doi:10.1016/j.jneumeth.2006.05.023

Warren, T. & McConnell, K. (2007). Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic Bulletin & Review*, *14*(4), 770–775. doi:10.3758/BF03196835

Woldorff, M. G. (1993). Distortion of ERP averages due to overlap from temporally adjacent ERPs: Analysis and correction. *Psychophysiology*, *30*, 98–119. doi:10.1111/j.1469-8986.1993.tb03209.x

Xiang, M., Dillon, B., & Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain and Language*, *108*(1), 40–55. doi:10.1016/j.bandl.2008.10.002

Yan, M., Zhou, W., Shu, H., Yusupu, R., Miao, D., Krügel, A., & Kliegl, R. (2014). Eye movements guided by morphological structure: Evidence from the Uighur language. *Cognition*, *132*(2), 181–215. doi:10.1016/j.cognition.2014.03.008

Yuval-Greenberg, S., Tomer, O., Keren, A. S., Nelken, I., & Deouell, L. Y. (2008). Transient induced gamma-band response in EEG as a manifestation of miniature saccades. *Neuron*, *58*(3), 429–441. doi:10.1016/j.neuron.2008.03.027