The electrophysiological reality of parafoveal processing:
On the validity of language-related ERPs in natural reading

Inaugural-Dissertation

zur

Erlangung des akademischen Grades

eines Doktors der Philosophie (Dr. phil.)

dem

Fachbereich Germanistik und Kunstwissenschaften

der Philipps-Universität Marburg

vorgelegt

von

Franziska Kretzschmar

geb. in Potsdam

Marburg/Lahn 2010

Vom Fachbereich Germanistik und Kunstwissenschaften der Philipps-Universität Marburg als Dissertation angenommen am 9.6.2010

Tag der Disputation: 28.9.2010

Erstgutachter: Prof. Dr. Matthias Schlesewsky
Zweitgutachter: Prof. Dr. Ina Bornkessel-Schlesewsky

Acknowledgements

This thesis wouldn't have been possible without the help and support of a number of people and its contents would be different without the insightful comments of many colleagues. First and foremost, I would like to express my deepest gratitude to my advisors Matthias Schlesewsky and Ina Bornkessel-Schlesewsky. Their keen interest in language, their open-minded view on data and research methods, and, in particular, their continuous support for my PhD project were a constant motivation. Special thanks go to Matthias for discussing virtually everything, even "dead end" hypotheses, with me, for believing in my data when I couldn't and for being available for advice and discussions almost all the time. Special thanks go to Ina for reminding me how fascinating the field of neurolinguistics is and for pointing out the fascinating aspects of my own project. Last but not least, she made it possible for me to extend the time I needed to write up this thesis.

Adrian Staub and Chuck Clifton willingly shared their knowledge about eye movements in reading and language comprehension with me. I am grateful for the patience with which they – nearly instantaneously – answered all my questions and, in many cases, anticipated what I would have asked next.

I hope that the present thesis reflects at least some of the input I received from these people. Any shortcomings are, of course, due to my inability to understand them completely.

I also benefited a lot from my colleagues Petra Schumacher and Markus Philipp. Petra has broadened my knowledge in neurolinguistics and, most importantly, she taught me that less is more. Indeed, this thesis would contain at least 50 more pages if Petra hadn't shaped my style of writing during our collaborative work. My officemate Markus discussed each and every part of this thesis with me and he proved to be a discussion partner with a lot (!) of critical comments on my data interpretations. This and his willingness to help me with every technical, intellectual, or imaginary problem have contributed to the final version of this thesis.

I also want to thank my former colleagues from the University of Marburg and the current members of the neurolinguistics working group at the University of Mainz who all have made my time as a PhD student very agreeable. For their assistance in various stages of data preparation and collection I thank Alexander Dröge, Petra Seipp, Tong Fei, Yu-Chen Hung, and especially Anika Jödicke and Jane Han who collected the data for Experiments 2 – 5. Special thanks are due to Jane Han for proofreading this thesis and for changing my "Franglish" into proper English. The concurrent registration of ERPs and eye movements would not have been possible without the software implemented by Torsten Schenk and other software-related issues would not have been resolved without R. Muralikrishnan. Finally, I enjoyed many lively conversations with Luming Wang, R. Muralikrishnan and Yu-Chen Hung, whose moral support during the last weeks was highly appreciated.

This thesis was generously supported by the DFG graduate program NeuroAct "Neuronal Representation and Action Control", led by Prof. Frank Bremmer and Prof. Karl R. Gegenfurtner. This graduate program also gave me the opportunity to discuss my ideas and findings with a number of different researchers.

Last but not least, I want to acknowledge the support I received from my family and friends. Since my undergraduate studies the "Golden Girls" Jana Wiggers and Katharina Kley have supported me and cheered me up whenever necessary. I'm glad I could count on them all these years. A big thank you to Cord Wilke, Kathrin Spindler, and Ute Conradi, for they have made my stay in Mainz a fairly happy one. Cord, thank you for the joy you took in "kindly" reminding me to continue writing and for also distracting me from it every now and then. Kathrin's personal catering service and her analytic view on major and minor problems have helped me a lot during the last months. I enjoyed many encouraging conversations, fun parties and Schnitzel with Ute and am looking forward to reviving these times. I thank you all for being patient friends while I temporarily turned into a hermit.

Das größte Dankeschön gilt meiner Familie, insbesondere meinen Eltern Bärbel und Klaus Kretzschmar, für ihren unschätzbaren Rückhalt. Ihr habt mich stets bedingungslos bei der Verwirklichung meiner Ziele und Träume unterstützt – egal was es euch in den letzten Jahren meines Umherziehens „in der Weltgeschichte“ abverlangt hat. Ohne Eure Unterstützung hätte ich es nicht so weit geschafft, würde es diese Doktorarbeit nicht geben. Danke.

Contents

List of Tables

List of Figures

Introduction

This thesis is concerned with the question of how input modalities can have an impact on correlates of language processing as measured via event-related potentials (ERPs) in the human electroencephalogram (EEG). Generally speaking, this research goal arose from the problem that every empirical approach to language comprehension (or production) is faced with: A theoretical framework is only as good as the data it is built on. So, in order to derive hypotheses and predictions about the mechanisms in language processing one has to ascertain that the empirical facts on which they rest are both reliable and valid. In the field of psycholinguistics this has led to a steady interest in the application of new methods to track the processes involved when humans produce or comprehend language. Early research had mainly focused on how participants respond to a predefined linguistic stimulus, i.e. it centered on behavioral output. While these data were a reliable source for predictions about human behavior, they were less suited to clarify which processes in the "black box", i.e. the human brain, led to this behavior. In particular, behavioral methods were not able to shed light on one of the most interesting – and still most controversially debated – questions in psycholinguistic research: What are the temporal dynamics of language comprehension and production in the brain? How and when are the different information types in natural languages (e.g., phonology, syntax, semantics, and pragmatics) used in the processing (and planning) of an utterance? With the advent of more sophisticated and powerful computer technologies, it became possible to monitor the neuronal mechanisms that underlie and precede behavioral output.[1] One of these so-called on-line methods is the ERP technique, whose major strength is a very high temporal resolution. Another on-line method widely used

[1] However, these methods still cannot track the neurobiological processes per se. Rather, what is measured in a scalp EEG or in the increased blood flow to a certain brain area is the result of neurobiological processes (i.e., the activation of particular neuronal populations or networks). These processes themselves still elude the currently used online methods.

is the monitoring of eye movements during reading. The latter method has been shown to reflect moment-to-moment cognitive processes with a relatively high temporal resolution despite its behavioral data format (see Rayner, 1998, for a comprehensive review).

Research using these two methods has advanced our knowledge about what drives eye movements in reading or what evokes a particular brain response in language comprehension at the word level, the sentence level and beyond (see Rayner, 1998, and Bornkessel-Schlesewsky & Schlesewsky, 2009a, for overviews). Yet, data from both fields are largely encapsulated, in the sense that findings from one research field are rarely taken into account for the formulation of general hypotheses on language processing in the other. This is most likely due to the apparent methodological paradox. Considering, for instance, the effect of lexical frequency: it has been shown that an infrequent word increases the amplitude of the N400 component, a negative deflection in the EEG signal with a peak latency of about 400 ms post word-onset (Kutas & Federmeier, 2000). In reading, by contrast, the duration of the very first fixation on an infrequent word is increased to, in most cases, no longer than 300 ms (Rayner, 1998; Sereno & Rayner, 2003). Thus, the eyes index a frequency effect considerably earlier than the brain response – even though the brain is thought to trigger the eye movement. This temporal paradox has led some to focus on the question whether ERP components occurring before the N400 have been neglected, and to what extent they may be relevant to the chronology of brain responses (in the visual domain) and eye movements (Sereno, Rayner, & Posner, 1998; Sereno & Rayner, 2003; see Rayner & Clifton, 2009, and Dien, 2009, for further references). However, the focus on early-latency ERP components is only one way to address the time paradox, and it certainly cannot clarify how the N400 might be linked to eye movements in reading (see Dambacher & Kliegl, 2007, for a first attempt to correlate the N400 with fixation durations). For example, it is conceivable that the N400 correlates with increases in the duration of the first fixation on a word or with the probability of refixations on a word – in the absence of first fixation increases. Thus, if one focuses exclusively on early-latency components, the temporal paradox between long-latency ERP components and eye movements in reading cannot be addressed properly, as, for instance, most of the ERP correlates of sentence processing exhibit long latencies (i.e., peak latencies beyond 300 ms post word onset).

The present thesis set out to investigate the reliability of the late endogenous ERP components in the visual domain such as the N400 and particularly the P300. ERPs and eye movements were collected concurrently while participants read whole sentences. ERPs were thus collected under normal reading conditions, rather than with the rapid serial visual presentation (RSVP) method that is common to ERP experiments in the visual domain. With RSVP, the sentence is presented word by word, with each word occurring in the center of a screen and at a fixed presentation rate. Participants are encouraged to suppress blinks and – more importantly – saccadic eye movements, which are typical of a natural reading situation, to avoid muscle artifacts in the EEG. Although the reported experiments are not the first to concurrently register eye movements and ERPs, this thesis represents the first attempt to collect ERPs during reading in an experimentally controlled design that investigates sentence comprehension (see Baccino & Manunta, 2005, and Simola, Holmqvist, & Lindgren, 2009, for single-word approaches and Dimigen, Sommer, Hohlfeld, Jacobs, Engbert, & Kliegl, 2006, for ERPs collected from corpus reading studies).

Natural reading differs from both the auditory modality and the RSVP technique in offering a preview of upcoming parafoveal words in a sentence (i.e., words that are not yet fixated). The availability of this preview yields different temporal dynamics in word recognition. That is, whereas prelexical and lexical information become available to the language comprehension system almost simultaneously with auditory and RSVP methods, natural reading temporally dissociates these levels of word recognition across two fixation positions. Prelexical information such as the orthographic form of a word becomes available before the word is directly fixated (i.e., during the last fixation preceding that word), and processing this prelexical information may even finish before the onset of the first fixation on the word. The first fixation, in turn, gives the language processing system access to lexical-semantic information. This delineation is possible because more visual information is projected onto the retina of the eye than just the fixated (foveal) word form; readers also acquire information from the parafoveal parts of the text, i.e. words adjacent to the current fixation. This is known as the parafoveal preview in reading. Thus, natural reading, particularly in light of the parafoveal preview, differs from both the RSVP technique and the auditory domain in at least three important aspects. First, more than one word can be processed at a time, although the level of processing differs for each. Second, the allocation of attention across foveal and parafoveal words is assumed to be asymmetric. Finally, stimulus information from the parafovea represents a specific type of stimulus degradation. The perceptual processing load for parafoveal words is higher because visual acuity dramatically drops off from the fovea to the parafovea. Hence, in order to fully understand the relationship between ERP components and eye movements in reading, it seems almost inevitable to collect ERPs under natural reading conditions. Since no other input method offers a parafoveal preview, cross-experimental comparisons from separate ERP and eye movement studies cannot properly address the temporal paradox and thus, they remain speculative at best. The concurrent recording of ERPs and eye movements, however, provides the unique opportunity to investigate not only the relationship between ERPs and eye movements, but also the reliability of long-latency ERP waves such as the P300 and the N400.

Since the N400 is probably the most well known language-related ERP wave, it may be surprising that the main focus of this thesis lies on the P300 instead. There are, however, two important reasons for this. On the one hand, the P300 has probably been examined in more experiments than any other ERP component. Since its initial discovery (Sutton, Braren, Zubin, & John, 1965), almost all aspects of its occurrence have been the subject of intense research (see the reviews by Donchin, 1981; Verleger, 1988; Picton, 1992; Polich, 2007).

Several theories about the functions associated with this ERP component have been proposed over the past decades (cf. Verleger, 1988, 1998). Altogether, it seems that, apart from a few contradictory interpretations, a relatively stable depiction of the P300 has arisen from the research of the past decades. It is thus possible to make clear predictions about the circumstances under which a stimulus elicits a P300. The wealth of these data from non-linguistic investigations is highly suggestive with regard to the present experiments, inasmuch as the P300 in language processing should be especially prone to the availability of the parafoveal preview in reading. Specifically, the P300 amplitude and latency have been shown to vary with the amount of attention given to a stimulus, the degree of perceptual degradation, and the whether more than one stimulus has to be processed. Thus, if the parafoveal preview should alter any late components in the EEG, it will most likely be observable in the P300 component.

On the other hand, the psycholinguistic community using ERPs to study language comprehension has remained astonishingly silent about the P300 and its relationship to language processing. The only notable exception to this was the debate about the correlation of the P300 and the P600/ late positivity[2] in studies on syntactic processing in the 1990s. However, the issue of whether the late positivity is an instance of the P300 – and therefore a domain-general component – or a language-specific component does not seem to have been settled and the debate has so far not been revived. There is also a very striking gap in the literature on word processing with regard to the P300. The nature of the N400 priming effect between words has been investigated in countless studies. Yet in most cases, the co-occurring P300 was neglected or regarded as a confounding factor that made the interpretation of the N400 priming effect more difficult. As a matter of fact, there is virtually no systematic attempt to functionally explain the emergence of the P300 in these kinds of experiments (but see Bentin, Mouchetant-Rostaing, Giard, Echallier, & Pernier, 1999; Holcomb, 1988). As will

[2] In the remainder of this thesis I shall use these terms synonymously.

be argued below in more detail, reexamining the P300 from the perspective of a binary decision-making account can explain the occurrence of the same in semantic priming studies, and it also shows that some late positivities, especially those reflecting the well-formedness evaluation of a sentence, are likely members of the P300 family.

Thus, the main research question underlying the first three experiments is if a P300 occurs in natural reading, as this input modality exhibits properties that should affect the P300. To gain maximum comparability with previous findings, the P300 was examined in stimuli that have been consistently shown to elicit it, namely highly constraining antonymy contexts. High-constraint contexts such as *black is the opposite of* ____ lead the language comprehension system to predict one specific word form and associated with it one specific meaning (*white*). The proposed functional interpretation of the P300 in these contexts assumes that the component reflects the successful identification of the word as either the expected antonym or not. The latency of the P300 (i.e., whether it peaks concurrently with the N400 or after it) is assumed to reflect the amount of linguistic processing necessary to identify the word. It is thus assumed that the P300 is related to a decision-making process that makes use of linguistic information, rather than to a particular linguistic operation or domain itself. So, for example, the P300 to any non-antonym (e.g., *black is the opposite of nice*) completing these contexts follows the N400, as a lexical search is necessary before making a decision about the stimulus. With antonyms, both the orthographic and the semantic information confirm the predicted item, so target identification proceeds unimpeded. In these cases, the P300 emerges in the same time window as the N400 for non-antonyms.

To anticipate the main result with respect to the P300 in decision-making, no P300 was found in response to the predicted antonym when ERPs were collected in a natural reading environment. It will be argued that this decision-making process is susceptible to whether the evidence for target identification that is available from a single fixation can allow for a confident decision. The temporal dissociation of prelexical and lexical reading is held responsible for the absence of the P300, as not enough evidence from both linguistic domains can be gathered during a single fixation. The P300 in response to unpredicted non-antonyms, on the other hand, was not affected by this dissociation. Since these words are not expected, their orthographic form is irrelevant for the identification as non-antonyms; target identification proceeds solely based on meaning. The first fixation on the non-antonym, providing semantic information, thus yielded a pronounced foveal P300.

As for the N400 component, the main goal was to investigate its reliability in terms of latency and eliciting fixation, and the extent to which it is accompanied by changes in the eye movement record. It was hypothesized that the dissociation of prelexical and lexical information may revealing important findings for the ongoing debate about the functional relevance of the N400 priming effect. The pattern observed for the N400 component was completely different from the one observed for the P300. The processing of parafoveal information did not eliminate the N400 (as with the P300), but brought about a parafoveal N400 priming effect, in addition to a foveal N400 priming effect. Both priming effects were attributed to different stages in the comprehension of the target word. The parafoveal N400 priming effect seemed to be based on prelexical information, which was influenced by the automatic spread of activation from antonyms to word forms of semantically related concepts. The foveal N400 priming effect appeared to result from the processing of semantic information, so that it reflected the different cloze probability values of antonyms and non-antonyms. Apart from this, neither N400 priming effect evidenced latency shifts. That is, the availability of a parafoveal preview did not induce shorter onset latencies for the N400 component.

These findings bear important implications for research on the P300 and N400, particularly for the assumptions about the architecture of language processing during reading. In order to examine further if these observations on the influence of the parafoveal preview generalize across different linguistic constructions, and to exclude the possibility that the inhibition of the P300 and the occurrence of the parafoveal N400 are mere byproducts of a new attempt to record ERP in reading that has failed altogether, two additional experiments were carried out with stimuli different from the antonym paradigm. These experiments revealed that parafoveal N400 effects can be reliably induced, provided the triggering fixation is located on a word that does not increase processing effort in itself. In the low-constraint contexts used, the parafoveal N400 may also be elicited by semantic information from the upcoming word, suggesting that the acquisition of lexical information from the parafovea interacts with contextual constraint or the predictions maintained by the parser. Moreover, the specific circumstances under which late positivities were observed in one experiment suggested that the P300 can emerge when the language comprehension system is engaged in predicting a particular word on only one linguistic domain, so that the P300 occurs during the fixation that delivers the information relevant to that prediction. However, this P300 was severely reduced in its amplitude and extent of its topographical distribution, indicating that the peculiarities of reading (i.e., partially parallel processing, allocation of attention, and stimulus degradation) can affect the P300 even if the (linguistic) requirements for its emergence are met. Finally, the findings from all experiments indicate that the correlation between eye movements and ERP waves is a fairly delicate one. Modulations of the total reading time measure, for instance, seem to also correlate with non-linguistic processes, e.g. total reading times increased as a function of the ease in decision-making for non-antonyms. Interestingly, the foveal N400 component can correspond to both the increase of the first fixation duration and the likelihood of refixations on the target word, depending on the type of linguistic manipulation (e.g., manipulations in the lexical-semantic domain or in syntactic word order). The parafoveal N400, by contrast, appears to have no corresponding effect in the eye movement record at all.

The thesis proceeds as follows. Chapter I and chapter II give an overview of basic principles and assumptions of the ERP and eye-tracking methodologies. The overviews mainly focus on those aspects that are relevant to the experiments reported in this thesis. The detailed hypotheses derived from this review are summarized in chapter III. Chapter IV summarizes the experiments employing the antonym paradigm. Chapter V contains a general discussion of the major results. Chapter VI introduces the two additional experiments reported in chapter VII, which examine the parafoveal N400 effect in more detail. These experiments have been designed to elicit an N400, which indexes relational processes at the syntax-semantics interface in sentence processing, i.e. semantic processing of verbal arguments beyond the lexical-semantic domain (e.g., Bornkessel-Schlesewsky & Schlesewsky, 2008; Bornkessel-Schlesewsky & Schlesewsky, 2009a, b). The rationale behind this was to extend the range of N400 effects and to thereby increase the generalization of these effects. It was further hypothesized that these stimuli should reliably elicit late positivities of different kinds, so that the resulting data could also be revealing with respect to the relationship between the P300 and the P600/ late positivity in sentence processing.

I Stimulus identification and electrophysiology

1 A brief overview of the P300

This chapter presents a brief description of the EEG/ERP method, explaining the basic principles and underlying assumptions of psychophysiological research. Chapters 1.2 and 1.3 deal with the P300 in psychophysiological and psycholinguistic research, respectively. Let it be said here that a complete review of the P300 literature would easily encompass a thesis of its own. Although focusing on only a subset of the vast data on P300 always, of course, risks the neglection of important facts, special emphasis will be put on perceptual and attentional mechanisms that may underlie the genesis of the P300 component. These mechanisms appear especially relevant when examining the P300 in the context of natural reading. Reviewing the P300 in psycholinguistic research will likewise be confined to those studies that have used experimental designs similar to those used in the present experiments (i.e., semantic priming and syntactic processing).

1.1 The EEG and ERP methodology

1.1.1 Basic concepts and method

The brain is constantly active in order to process the vast amount of information with which human beings are faced – consciously and unconsciously – in every instance of life and, if necessary, to initiate responses to the sensory or cognitive input. The sum of these processes can be measured in the form of electrical activity on the scalp. This measure is called the *spontaneous electroencephalogram* (*EEG*). Dating back to Hans Berger's discovery that the human EEG can be effectively measured (Berger, 1929), researchers have been concerned with the question whether one can track psychological processes via the recording of the

EEG. This was motivated by the simple fact that several separate neuronal responses to specific events are embedded in the summed EEG.

Following Berger's seminal work, scientists mainly focused on the relationship between EEG frequencies and behavior, until in 1965, Sutton and colleagues (Sutton et al., 1965) provided the first evidence that specific components in the EEG, recorded from the surface of the scalp[3], are in fact linked to certain, more or less narrowly defined cognitive events (see Luck, 2005, and Altenmüller & Gerloff, 1999, for short historical overviews). This turn in psychophysiological research was made possible by the availability of advanced recording and analysis techniques, which enabled scientists to filter out so-called *event-related brain potentials* (ERPs) from the spontaneous EEG. Event-related brain potentials are small changes in voltage that occur in response to a particular stimulus (in general between 2 and 20 µV; cf. Altenmüller & Gerlof, 1999; however, the range is typically between 2 and 8 µV for language; cf. Bornkessel-Schlesewsky & Schlesewsky, 2009a). They can be subdivided into *endogenous* and *exogenous* components. While the latter are related to external factors that are perceived by the subject (i.e. they are sensory responses) as they perform an experimental task, the former are linked to internal processes or mental states and exhibit a longer latency than exogenous components. Thus, endogenous components depend entirely on the subject and his or her task performance (cf. Sutton et al., 1965; Luck, 2005).

In order to extract ERPs from the ongoing spontaneous EEG, one has to eliminate the activity that is not directly induced by the stimulus. This is usually achieved by averaging the EEG signal of the participants over the number of trials run in the experiment. The critical point here is that averaging is thought to eliminate the noise in the signal (i.e., brain potentials caused by environmental factors or characteristics of subjects, but not by the particular

[3] Another possibility to measure event-related brain activity is single-cell recording. Due to its invasiveness, however, this approach is not very applicable to a wider range of research questions. Research in this domain has usually been based on primate studies or on lesion studies, in which patients already prepared for surgery participated.

stimulus) and to thereby ameliorate the *signal-to-noise ratio*. As the voltage of the spontaneous EEG (between 10 and 100 μV) by far exceeds that of the ERPs, a relatively large number of trials is needed to increase the quality of the signal. As a rule of thumb one can maintain that the smaller the size of the signal, the more trials that will be needed. This is reflected in the component size to some extent, i.e. relatively large components such as the P300 or the N400 require fewer trials (about 30 to 60 trials per condition) than smaller components such as the P100 or the N200, because the former are based on a stronger signal (Luck, 2004). Averaging can in principle be done with respect to the stimulus itself or the behavioral response to it, so that ERPs are classified as stimulus-locked or response-locked potentials.

The resulting components can be classified according to their polarity, latency, and topography. For instance, "P" and "N" indicate a positive or negative deflection of the ERP wave, respectively. Latency is related to the peak amplitude of a component and can be given in milliseconds (e.g. P300) or as an ordinal position (e.g., P3). The latter is advantageous for components that vary considerably in peak latency, as, for instance, the P300. The topography or scalp distribution of a component is useful to distinguish between ERP waves that would be identical in amplitude and latency. For example, the novelty P3 (so-called P3a) and the target-related P300 (so-called P3b or P300) exhibit similar latency and amplitude, while they differ remarkably from each other in their topographies (see chapter 1.2 for a short depiction of the members of the P300 family). Although topographical differences can be used to differentiate distinct ERP components, the topography of ERP waves itself cannot be used to unequivocally determine the underlying neuronal source that generates a certain ERP component (the so-called inverse problem; cf. Bornkessel-Schlesewsky & Schlesewsky, 2009a). Finally, the amplitude (measured in μV) is used to infer quantitative changes in effect size. This latter characteristic is, however, no defining property of ERP components, as opposed to the other three parameters (cf. Luck, 2005, and Bornkessel-Schlesewsky & Schlesewsky, 2009a).

A number of ERP components elicited in the course of language processing have been described and several functional interpretations have been proposed for them (see Bornkessel-Schlesewsky & Schlesewsky, 2009a, Table 16.1, for an overview). While it appeared first that particular ERP waves could be unique markers of certain linguistic domains (e.g., the N400 for the lexical-semantic domain, and the P600 for the syntactic domain), the status quo in electrophysiological research on language processing suggests that this assumption does not hold. It seems that additional analysis methods are required (e.g., fMRI or EEG frequency band analyses) to unravel whether identical ERP waves serve the same or differing functions (see Bornkessel-Schlesewsky & Schlesewsky, 2009a, chapter 16, for more discussion). A more detailed functional classification of the P300 and N400 components that are expected to occur in the present experiments will be given below (see chapters 1.3.1 and 1.3.2).

1.1.2 On the usefulness of ERPs in language comprehension

In the past years, psycholinguistic research has come to a consensus that language comprehension is a highly incremental operation (Bader & Lasser, 1994; Crocker, 1994; Kamide & Mitchell, 1999; Stabler, 1994). New input is rapidly interpreted based on all available information and integrated as the sentence unfolds. Given that this is the case in every-day communication, one needs a method that is capable of tracking these processes in a millisecond-by-millisecond fashion. One might also need a method that can capture qualitative differences if it turns out that linguistic information types are weighted differently during comprehension.

The major advantage of the ERP method in the study of language comprehension is its high temporal resolution and the possibility that qualitatively different processes may be reflected by different components, with the underlying assumption that different neuronal populations may subserve distinct processes. For instance, Kutas & Hillyard (1980) found that sentence-final words that are unexpected in a given sentence context elicit an N400, whereas words whose perceptual features are unexpected engender a positivity peaking at around 560 ms post stimulus onset. The authors took this as evidence that the semantic incongruity of a word is detected by mechanisms different from those that detect physical incongruities (see Allefeld, Frisch, & Schlesewsky, 2005, for further investigations). Similarly, differences in topographic distribution amongst the members of a component family are also indicative with respect to the underlying neuronal generators. It is assumed, for instance, that different topographies of a given component are linked to different generators, whereas identical topographies may be the result of the same, but also of different, generators (e.g., Johnson, 1993; Coulson, King, & Kutas, 1998a). Finally, latency variation can also dissociate different processes in language comprehension. For instance, a positive going ERP wave with a parietal maximum has been reported for syntactic reanalysis (e.g. Osterhout & Holcomb, 1992) and for thematic reanalysis (e.g. Bornkessel, Schlesewsky, & Friederici, 2003). These components are only differentiated by their latency. While the positivity reflecting syntactic reanalysis has an onset latency of about 600 ms post word-onset, the positivity reflecting thematic reanalysis peaks considerably earlier at about 300 ms post word-onset. Thus, the combination of the high temporal resolution and the multidimensionality of the measured ERP component make ERP measures highly appealing when one aims to discover the dynamics of language comprehension.

Behavioral tasks, in which participants are asked, for example, to press a button after they have read or heard a sentence, offer a fairly poor temporal resolution because they can only indicate whether or not a comprehension problem occurred. When such a problem occurred cannot be inferred from the behavioral output, as the response is measured only after the comprehension process is completed. Behavioral measures are also only quantitative in nature as they are only rarely able to reveal (mostly subtle) qualitative differences between two experimental conditions, and they are mostly one-dimensional, as only latency or accuracy information of the response execution can be gathered (one exception to this is the speed-accuracy trade-off (SAT) procedure; cf. McElree, 2006). Similarly, it remains unclear with behavioral responses whether an experimental manipulation alone caused an effect, or whether it interacted with other (more or less controlled) factors in the course of processing.

The major disadvantage of the ERP method is its poor spatial resolution with respect to the brain regions involved in language processing (see above). The steady improvement in technology that combines different methods may soon deliver a solution to this problem (e.g., by the combination of ERPs with fMRI). This is, however, not the only issue; findings from ERP experiments have been challenged on other grounds as well. To avoid artifacts caused by muscle activity and potential changes of the EOG (electrooculogram), participants are usually instructed to avoid eye movements, especially when ERPs are collected in the visual domain. The single words or phrases are then foveally presented on a video monitor. It is crucial that enough time between the individual words is provided to exclude the possibility of overlapping ERP components (Inter-Stimulus Interval; ISI or Stimulus-Onset Asynchrony; SOA). Moreover, it appears that comprehension difficulty generally increases as the ISI decreases. Therefore, visual sentence stimuli usually do not exceed a certain length, because the language comprehension system would then be at risk of collapsing due to memory overload (cf. Mitchell, 2004). This has led some to argue that the RSVP method significantly distorts normal reading processes because sentences are not presented in their entirety and words are presented for a longer time than they would actually be fixated under natural reading conditions (Sereno & Rayner, 2003; Rayner & Clifton, 2009). Based on these aspects, one may speculate that they have led participants to adopt special strategies in processing the stimuli and that language-related ERP components may be somewhat artificial, at least in terms of their (peak) latency.

The present experiments, in which ERPs were collected during natural reading, may help to address these issues. However, it is also evident that one is faced with some methodological problems by combining eye movement measures with ERP recordings. First, typical eye movement experiments present no more than 10 – 15 trials per condition to avoid strategy effects. Furthermore, one cannot increase the temporal intervals between successive words. The average fixation duration on a word is about 250 ms for the first fixation and interword saccades last about 30 – 50 ms (Rayner, 1998). This is an extremely short interval and may well produce overlapping ERP components in investigations into N400 or P300 effects (for the approach taken to solve this problem, see chapter 4.2.2).

1.2 The P300: Determinants of its amplitude and latency

The P300 is an endogenous component that was first discovered in the 1960s (Sutton et al., 1965). Ever since, researchers have investigated the factors that induce P300s in very detail. Most of these investigations have been based on variants of the oddball paradigm, which, in it simplest form, can be defined as follows (Donchin & Coles, 1988). First, participants are presented with a series of discrete stimulus events, which can be discriminated from each other based on their physical properties. Second, the stimuli events are presented in a way that enables participants to assign them to one of two categories: target stimulus vs. non-target stimulus (sometimes a third distractor category is also used; in these cases the paradigm is called the “three-stimulus” oddball task). Third, the categorical distinction is embedded in the experimental task given to the participants, i.e. they are instructed to react to the target stimuli. Fourth, all stimuli are presented in a randomized order (the so-called Bernoulli sequence, i.e. the binominal distribution of the likelihood that one event will occur). Finally, targets and non-targets differ in their probability of occurrence within the experiment.

There has been much debate concerning whether the P300 is a monolithic component or whether it comprises a component family. To date, converging evidence has shown that at least two different components can be distinguished from one another: the P3a and the P3b. This distinction goes back to the seminal work by Squires, Squires, & Hillyard (1975b) who set out to clarify divergent findings in early P300 research. Early investigations showed clear P300s, not only elicited by (a group of) task-relevant stimuli, but also by task-irrelevant stimuli (see Squires, Squires, & Hillyard, 1975b, for references). Squires and colleagues used an oddball design in which the rare auditory stimulus occurred with either a different intensity or a different frequency than the standard stimuli. In both cases, the rare stimulus evoked a positive deflection in the EEG signal, which had a peak latency between 220 and 280 ms and a fronto-central distribution. This component was labeled “P3a” and it emerged regardless of whether the participants attended to the auditory stimuli or ignored them. Contrary to this, a second positive peak slightly thereafter (i.e., with a peak latency after 300 ms) with a parietal maximum was evoked only by the attended rare stimuli; this component was labeled “P3b”.

Both the P3a and P3b are then similarly defined in respective latency and polarity, while they can be clearly dissociated based on their topography and, to some extent, in their functional significance (e.g., Comerchero & Polich, 1999; Sergent, Baillet, & Dehaene, 2005; see Kok, 2001, and Polich, 2007, for reviews)[4]. Some researchers have claimed further that there are other positive ERP waves that appear to be related to the P3a. For instance, Courchesne and colleagues (Courchesne, Hillyard, & Galambos, 1975) have shown that perceptually novel stimuli (i.e., stimuli that are different from both the standard and the target stimuli) elicit what they called a “novelty P3”. Although this positivity also exhibits a fronto-

[4] Throughout this thesis, I shall use the terms P3b and P300 synonymously. Note that some investigators have emphasized that the P3b and the subsequent slow wave may not be easily disentangled (Squires et al., 1975) and in addition, the latency range reported in the P300 literature may result from confounding several P300 waves (e.g., Verleger, 1997; Dien, Spencer, & Donchin, 2004). For the present purposes, I will follow the main approach in assuming that there are basically two clearly distinguishable subcomponents, P3a and P3b.

central distribution like the P3a, it differs from the latter in its longer latency and, more importantly, in its task dependency. In contrast to the P3a, the novelty P3 is only engendered by attended stimuli, and susceptible to habituation effects, i.e., its amplitude is reduced with repeated stimulus representation (see Kok, 2001, and Polich, 2007, for further description and references). In a Go/No-Go task, infrequent non-target (i.e., no-go) stimuli give rise to a "no-go" P300 with a more centro-parietal distribution (e.g., Pfefferbaum, Ford, Weller, & Kopell, 1985; Polich, 2007). Although it is unclear whether these three fronto-central ERP waves are instances of the same component or are entirely distinct (cf. Verleger, Jaskowski, & Wauschkuh, 1994; Polich, 2007; Kok, 2001), they are clearly distinguishable from the target-related P3b.

The following review of the P300 literature will focus on the factors that may influence the P3b in its amplitude and latency and also bear some importance for the interpretation of positive deflections found in ERP research on language comprehension (cf. chapters 1.3.1 and 1.3.2). With that said, the following review is selective and far from a comprehensive look into P300 research. For more general overviews, the reader is referred to, for example, (Donchin, 1981; Donchin & Coles, 1988; Verleger, 1988; Picton, 1992; Kok, 2001; Polich, 2007).

1.2.1 Stimulus probability: Frequency of occurrence

Sutton and colleagues (Sutton et al., 1965) were among the first to report a positive deflection with a peak amplitude at about 300 ms in response to visual or auditory stimuli with a low probability of occurrence. In three experiments, Sutton et al. presented participants with clicks or brief flashes as target stimuli that followed a cueing stimulus indicating the sensory modality of the target. Both stimuli were presented as pairs. In half of these pairs, the cuing stimulus unambiguously indicated the sensory modality of the target, whereas in the other half the cuing stimulus was ambiguous with respect to target modality. In addition, the occurrence of the test stimuli was varied such that one sensory category had a low probability of occurrence (P = .33), a high probability (.66), or both categories were equal in probability (.50). During the 3-5 ms interval between cue and target, participants had to state their guess about the upcoming target. Results showed that across both modalities, the test stimulus with the lower probability engendered a pronounced P300 with a posterior distribution. Interestingly, when ERPs were time-locked to false vs. correct guesses of the stimulus modality in the equal-probability condition, the amplitude of the P300 was significantly more enhanced for false guesses as compared to correct guesses. The authors concluded that the positivity had been evoked by subjects' uncertainty about the test stimuli and they stressed that this pattern was highly influenced by the probability of the test stimulus. The very same pattern has subsequently been replicated in countless studies with various experimental designs (e.g., Comerchero & Polich, 1999; Friedman, Hakerem, Sutton, & Feiss, 1973; McCarthy & Donchin, 1976; Newman, Connolly, Service, & McIvor, 2003; see the reviews in Donchin & Coles, 1988; Polich, 2007; Johnson, 1986, 1993; Verleger, 1988) and is labeled *target probability*.

Target probability is one of the major determinants of the P300, although it does not appear to be a necessary condition (e.g., Squires, Wickens, Squires, & Donchin, 1976; see also Donchin, 1981, and Donchin & Coles, 1988, and references cited therein). It bears an inverse relationship to the amplitude of the P300 in that lower probability scores lead to larger amplitudes. As the definition of the oddball paradigm implies (see chapter 1.2 above), target probability relies on the perceptual distinction between a task-relevant target stimulus and any non-target (standard) stimulus (e.g., the target "x" embedded in a series of "o"). However, it has been shown that target probability is mentally represented as a *category probability*, which may explain why the P300 to targets is elicited even when target and non-target stimuli are equi-probable, or when the probability of some non-target stimulus is actually lower than for the target stimulus. For example, if the relevant task is to discriminate the letter “x” from the other letters in the Roman alphabet, there is no difference in whether it is contrasted with a repeatedly presented non-target stimulus like “o”, or whether it is contrasted with all remaining letters in the alphabet (which are not repeated). In the latter case, the target letter “x” should evoke no P300, if probability alone is critical for the elicitation of the P300, because it occurs more frequently than the remaining (and unrepeated) letters of the alphabet. Yet, this is not the case, the individual probability of occurrence for a non-target stimulus has no impact on the P300. This results from the fact that category probability is dichotomous – one category is task-relevant, the other (i.e., the standard stimuli) is not. In sum, the P300 amplitude appears sensitive to category probability of target stimuli that is weighted against the probability of the category of non-targets. The probabilities of these categories are then relevant with respect to the magnitude of the P300 effect (cf. Courchesne, Hillyard, & Courchesne, 1977; Mecklinger & Ullsperger, 1993).

Stimulus sequence effects are related to probability manipulations, as they reflect a more subjective probability estimate (cf. Donchin & Coles, 1988; Johnson, 1993). Squires and colleagues (Squires, Wickens, Squires, & Donchin, 1976), for example, examined whether the P300 amplitude elicited by a target stimulus changes as a function of the number of non-targets that preceded the target. They presented participants with either high or low tones of varying global target probability; the participants were instructed to count the high-pitched tones. Prior to the critical trial N, the order of the preceding trials varied such that they contained a target or non-target stimulus, ranging from first-order sequences up to fifth-order sequences (e.g., with A indicating the target and B indicating the non-target stimulus, a first-order sequence is the presentation of A, a second-order sequence may be composed of AA or BA, a third-order sequence of AAA, BAA, ABA, or BBA and so forth). The result was that the more non-targets preceding the target tone, the more enhanced was the P300 in amplitude. This also extended to trials in which the probability of the stimuli was equal. Evidently, there is a trial-to-trial variability in amplitude size that is determined by whether the current trial is different from one or more preceding events (see also Luck & Hillyard, 1990).

To sum up, these findings clearly demonstrate that both the globally and the locally determined probability of a stimulus have a significant impact on the amplitude of P300. The underlying psychophysiological states that may or may not enhance P300 amplitude have been a matter of heated debate (cf. Donchin, 1981; Donchin & Coles, 1988, 1998; Verleger, 1988, 1998; Sommer, Leuthold, & Matt, 1998). On the one hand, it has been claimed that rare stimuli are unexpected events that require the revision of the mental situation model (Donchin, 1981; Donchin & Coles, 1988; Coulson, King, & Kutas, 1998). The contrasting view holds that it is not the unexpectedness or the surprise that matters, but rather the awaitedness of a stimulus (Verleger, 1988, 1998; Verleger, Jaskowski, & Wauschkuhn, 1994). That is, subjects become aware that there will occasionally be a target stimulus and eagerly wait for it, because the rare event is very often associated with the experimental task (e.g., to count a tone or to press a button). It is, without a doubt, important to develop models from which specific predictions and hypotheses concerning the cognitive processes underlying the P300 can be derived. In spite of the general need for such systematic models, neither of the two accounts described above can satisfactorily explain the P300 obtained in psycholinguistic studies (cf. chapter 1.3.1 and 1.3.2). For example, in the antonym paradigm, using sentences such as *black is the opposite of white*, it has been found that the best sentence completion in sentences such as *black is the opposite of* __ (i.e., the antonym *white*) elicits a P300 peaking at about 300 ms post word onset. This result strongly supports Verleger's account. However, non-antonyms (e.g., *black is the opposite of nice*) also elicit a P300 with increased latency. This contradicts Verleger's account, while confirming Donchin & Coles's position on the surprise value of stimuli. Therefore, the data reported in the present thesis challenge both models and will not be interpreted on the basis of either model. As will be argued in the following chapters, the emergence of the P300 and its latency shifts in language processing studies can be explained by taking into consideration general decision-making strategies and how they interact with the temporal availability of linguistic information that are necessary to make a decision and, to some extent, with the perceptual quality of the linguistic stimuli.

1.2.2 Stimulus relevance: Allocation of attention and task instruction

All accounts concerning the psychological origin of the P300 highlight that the aforementioned probability effects are not independent variables. They are actually heavily dependent on whether the stimuli presented to the participant are relevant to carry out the experimental task (e.g., Hillyard, Squires, Bauer, & Lindsay, 1971; Courchesne et al., 1975; Squires et al., 1975b; Johnson, 1986). This variable, *task relevance*, has been shown to directly influence the P300 amplitude and the existing data justify the claim that task relevance is in fact the major determinant of the P300 (cf. Johnson, 1986). Specifically, if participants are instructed to attend to only one subset of targets, while ignoring the other stimulus type(s) (both in single-task and dual-task experiments), no P300 is evoked to the ignored stimuli (e.g., Duncan-Johnson & Donchin, 1977; Hillyard, Hink, Schwent, & Picton, 1973; Koivisto & Revonsuo, 2007; Mangun & Hillyard, 1990; Picton & Hillyard, 1974; Proverbio, Esposito, & Zani, 2002; Sheatz & Chapman, 1969; Squires et al., 1975b). Squires and colleagues (1975b), for example, compared the ERP waves for attended auditory target stimuli with ERP waves for unattended stimuli (the task was to ignore them). The authors found that the P300/P3b is only elicited upon the detection of the attended target stimulus, whereas the P3a appears to be independent of attention, as it was engendered by all infrequent stimuli in the experiment. So, it seems that in order to engender a P300, the target stimulus has to be *actively attended* to. Making the stimulus most relevant for the experimental task commonly ensures this. Amplitude modulations also correlate with the amount of attention that is allocated to a target stimulus. Instructing the participants to ignore one type of stimulus while fully focusing on another is just one of the possible ways to distribute attention. A second possibility is to divide attention across two or more target stimuli such that, for example, 80 % of attention is allocated to one stimulus and the remaining 20 % to the other. Mangun & Hillyard (1990), for instance, instructed their subjects to allocate only a certain amount of attention to a visually presented target in either the right or the left visual field. Their data showed that the amplitude of the P300 varied as a function of attention so that increases in attention led to enhanced amplitude.

The positive correlation between P300 amplitude and attention can also be seen in studies investigating the attentional blink phenomenon. In these experiments, two target stimuli (called T1 and T2) are presented sequentially within a rapid stream of stimuli, and both of which the subject is required to detect. When the temporal interval between T1 and T2 is short, the subjects are frequently unable to report the nature of the second target, or even miss it completely (cf. Koivisto & Revonsuo, 2007, 2008a,b; Sergent et al., 2005; Sessa, Luria, Verleger, & Dell'Acqua, 2007; Vogel, Luck, & Shapiro, 1998; Vogel & Luck, 2002). When the second stimulus bears characteristics that elicit a P300 (e.g., low probability of occurrence), it has been found that the positivity is significantly reduced in amplitude (Vogel et al., 1998; Sessa et al., 2007; Koivisto & Revonsuo, 2007) or sometimes suppressed entirely (Koivisto & Revonsuo, 2008a,b; Sergent et al., 2005; Vogel & Luck, 2002). This pattern has been interpreted as reflecting the accumulation of attentional resources for the encoding of the first target. The processing of T2 and its consolidation in visual working memory then suffer from the deprivation of attentional resources due to the temporally overlapping processing of either T1 and T2, or of T2 and the following stimulus (see Sergent et al., 2005, and Vogel & Luck, 2002, for further discussion). When a temporal overlap in processing is avoided, e.g., when T2 is the final stimulus in the stream or when participants are instructed to ignore T1, a P300 in response to T2 can be observed (cf. Vogel & Luck, 2002; Sessa et al., 2007).

Another experimental paradigm in which the correlation between the P300 and attention allocation is clearly visible is the so-called *dual-task paradigm*. Here, participants concurrently perform two more or less similar tasks, and it has been widely demonstrated that the increase in task load results in *interference effects* between the primary task and the secondary task (cf. Luck, 1998). Specifically, it has been shown that the introduction of another task that is ascribed primary importance in the experiment leads to a reduction in the amplitude of the P300 associated with the secondary task, reflecting the need for a reallocation of attention resources from the secondary to the primary task (cf. Hoffman, Houck, MacMillan, Simons, & Oatman, 1985; Isreal, Chesney, Wickens, & Donchin, 1980; Isreal, Wickens, Chesney, & Donchin, 1980; Kramer, Wickens, & Donchin, 1985; Kramer, Sirevaag, & Hughes, 1988; Singhal & Fowler, 2004; Wickens, Kramer, Vanasse, & Donchin, 1983). Isreal and colleagues (Isreal, Chesney et al., 1980), for instance, examined the extent to which attention shared between two tasks may influence the P300. They had participants perform a primary task requiring them to track a moving object with a cursor concurrently with a secondary task in which they had to count high-pitched tones. These high-pitched tones were presented as frequently as low-pitched tones (i.e., they employed the oddball task with a probability of .50 for target and non-target stimuli). The authors found remarkably decreased P300 amplitudes in response to the counted tones in the dual-task condition as opposed to the single-task condition, in which the participants only attended to the count task. However, increasing the difficulty of the primary task had no further influence on P300 amplitude (nor on its latency), even though it clearly impacted reaction times for this task.

The result that the enhanced *task difficulty* for the primary task in the Isreal, Chesney et al. study had no or little influence in P300 amplitude was somewhat unexpected, given that a more difficult task requires more attentional resources. Isreal, Wickens et al. (1980) proposed that the domain of cognitive burden is the crucial case in point here. In a follow-up experiment to their initial study, Isreal and colleagues employed two concurrent tasks which both operated in the perceptual domain. Here, they found a monotonic decrease in P300 amplitude with the increase of difficulty associated with the primary task. They concluded therefore that changes in task difficulty, for tasks that require the execution of motor responses, do not further modulate the P300 amplitude, which would explain why Isreal, Chesney et al. (1980) failed to find clear effects of task difficulty. When task demands are enhanced in the perceptual domain, the P300 amplitude further decreases, because the P300 depends on perceptual encoding processes, but not on motor preparation (see also Wickens et al., 1983; Kramer et al., 1985; Kramer et al., 1988). This difference between perceptual and motor tasks has been replicated in other single-task experiments showing that P300 amplitude is not affected by increasing the difficulty of response-related processes, but is diminished (and P300 latency increased) when the target is perceptually less distinct (cf. McCarthy & Donchin, 1981; Magliero, Bashore, Coles, & Donchin, 1984; but see Christensen, Ford, & Pfefferbaum, 1996, for different findings; see also chapter 1.2.3).

The preceding paragraphs have dealt with the factors that primarily initiate a series of encoding steps, the end of which is marked by the elicitation of a P300. That is, the amount of attention allocated to a stimulus is the central factor, in that attention is a necessary precondition to initiate the processes for stimulus detection. P300 amplitude is considered an index of the amount of allocated attention – under circumstances of both focused and divided attention. P300 latency is interpreted as an index of stimulus evaluation time, in that it varies with the ease of stimulus identification (e.g., McCarthy & Donchin, 1981; Ford, Pfefferbaum, Tinklenberg, & Kopell, 1982; Donchin & Coles, 1988).

With regard to the stimulus evaluation time, there is yet another modulating factor, *task instruction*, which indirectly impacts processing by decoupling stimulus evaluation time from P300 peak latency. The interest researchers have in the way subjects perform a task has arisen from the interest in the timing of mental processes, as indexed by endogenous EEG components, reaction time (RT) and their relation to each other. Since the P300 has been associated with perceptional stimulus evaluation time, one would expect it to peak before the behavioral response, which is the joint outcome of many more processes than those reflected by P300. So, the somewhat paradoxical finding that P300 peak latency occasionally falls behind the point of behavioral response execution needed further clarification (Kutas, McCarthy, & Donchin, 1977). Kutas and colleagues (1977) examined the effects of accuracy vs. speed instructions on the correlation between P300 latency and reaction times. When the subjects were instructed to respond as accurately as possible, P300 latency increased and it was tightly coupled with RT, which followed P300 latency. When instructed to respond as quickly as possible, latencies generally showed earlier onsets and the correlation between RT and P300 was less tight. In addition, there were many instances of fast guesses for which RTs preceded P300 peak latency. Kutas and colleagues argued that this effect may be due to participants trying to accomplish the task by relying on only partial information about the stimulus. The later occurring P300 peak was then taken as evidence that stimulus evaluation processes continued even after a response had been given (see Pfefferbaum, Ford, Johnson, Wenegrat, & Kopell, 1983, for similar results).

In summary, it seems to be a necessary condition that a stimulus is attended to in order to elicit a P300. If a stimulus is ignored or if it is not allocated enough attention, there will be no P300. There is furthermore a positive correlation between attention and P300 amplitude, in that amplitude size is enhanced when more attention is allocated to a target. However, the size of this effect may be reduced by difficulties in stimulus evaluation, as with degraded or difficult-to-discriminate targets. These perceptual factors are discussed in the following chapter. Finally, if task instructions encouraged participants to react on the target on the grounds of incomplete information, P300 latency exceeds reaction time, whereas the reversed relationship between P300 latency and reaction time holds under accuracy instructions.

1.2.3 Stimulus encoding: The role of target discriminability

The role of *target discriminability* or *perceptual load* in target evaluation processes has been investigated from two slightly different starting points. On the one hand, researchers have investigated with auditory signal detection tasks whether the confidence with which participants detect a target stimulus influences the P300 (Hillyard et al., 1971; Hillyard et al., 1973; Kerkhof, 1978; Parasuraman & Beatty, 1980; Paul & Sutton, 1972; Ritter, Simson, & Vaughan, 1972; Squires, Hillyard, & Lindsay, 1973a,b; Squires, Squires, & Hillyard, 1975a). In this line of research, subjects are trained to detect auditory stimuli. Stimuli deviating to some extent from the learned threshold of detection are presented and ERPs are collected for hits, misses, or false alarms. Other investigations, on the other hand, have focused on the perceptual load of the target without referring to confidence in decision-making (e.g., McCarthy & Donchin, 1981; Magliero et al., 1984). In these studies, the perceptual distinctiveness of the visual or auditory target stimulus was varied, e.g. by embedding it in noisy background or by presenting it along with highly similar non-target stimuli. In contrast to the studies above, subjects were not trained prior to the experiment (thus no individual detection threshold was determined a priori) and confidence ratings were not collected together with ERP waves. In doing so, these experiments did not intend to measure the impact of confidence, but rather aimed to study different presentation conditions for untrained subjects. Both of these accounts are discussed separately in the following paragraphs.

Hillyard et al. (1971) used an auditory signal detection task to investigate the correlation between P300 amplitude and perceptual processing. They presented their participants with a continuous noise signal, to which a short tone burst was added in one half of all trials. The task was to decide on a trial-by-trial basis whether a tone had been present or not. The authors then calculated the correlation between P300 amplitude and behavioral accuracy as measured by d' (called perceptual sensitivity in their study) and found that the P300 component was only present for correctly detected targets (i.e., hits), but not for false alarms, misses, or correct rejections. P300 amplitude also increased monotonically with performance accuracy.[5] That is, the higher the accuracy in target detection, the larger the P300 amplitude. This suggests that the P300 amplitude was related to the participants' confidence in their decision about the occurrence of the target tone (see also Parasuraman & Beatty, 1980; Paul & Sutton, 1972). However, this conclusion is undermined when correct rejections and false alarms are considered, as participants are presumably also highly confident about their decisions in these cases (see Squires et al., 1975a, for similar findings regarding correct rejections). To shed light on this issue, Squires et al. (1973b) used a more fine-grained rating scale than Hillyard et al. (1971) and correlated it to P300 amplitude. Their findings suggest that targets detected with high confidence (hits) elicit the most pronounced P300 amplitude, which was marginally different from the P300 to false alarms. The overall trend for hit trials was a decrease in P300 amplitude and an increase in P300 latency, as confidence decreased, while false alarms, misses, and correct rejections did not differ significantly from one another. Apparently, amplitude measures alone cannot clearly capture the processing difference between hits and the other response categories. Kerkhof (1978) suspected that for correctly rejected trials, this might be the consequence of latency variability. As correct rejections are not coupled to the presentation of a stimulus, onset and offset latency for any P300 activity related to the rejection process could be instable across trials, which would then have a converse effect on amplitude measures (because latency jitter reduces amplitude size). In light of this, Kerkhof (1978) adjusted ERP waves according to latency, resulting in P300 to correct rejection trials, which was, however, smaller in amplitude than the P300 to hits. He was also able to show that

[5] There is, however, a decline in amplitude for the highest accuracy values and Hillyard and colleagues interpret this as a learning effect. Since participants received extensive training (more than usual in this type of experiments), their experience with the task may have caused strong anticipation for the target such that P300 amplitude decreased (cf. Hilyard et al., 1971; footnote 24). Consequently, the a priori uncertainty about which event may occur next (i.e., its unexpectedness) is significantly reduced. Alternatively, the authors thought it conceivable that with increasing proficiency, the task became too easy for their participants, so that shifts of attention to irrelevant information minimized amplitudes (see also Picton, 1992).

the negative correlation between P300 latency and confidence rating also held for rejection trials, i.e. a more confident response resulted in shorter P300 latency. In sum, these data revealed that the P300 is elicited once there is a sufficient amount of evidence for confident target identification. If confidence is low and target detection uncertain, P300 amplitude decreases and its latency increases.

Perceptual load as induced by a low degree of perceptual discriminability has been shown to mainly affect the latency of various members of the P300 family such as the P3b and the NoGo-P3 (e.g. Christensen et al., 1996; Combs & Polich, 2006; Ford et al., 1982; Hagen, Gatherwright, Lopez, & Polich, 2006; Katayama & Polich, 1996;[6] McCarthy & Donchin, 1981; Magliero et al., 1984; Luck, 1998; Pfefferbaum et al., 1983; Pfefferbaum & Ford, 1988; Smulders, Kok, Kenemans, & Bashore, 1995), but modulations of P300 amplitude have been reported as well (e.g., Combs & Polich, 2006; Comerchero & Polich, 1999; Kiefer, Marzinzik, Weisbrod, Scherg, & Spitzer, 1998; Kok, 1986; Kok & Looren de Jong, 1980; Kok, van de Vijver, & Rooijakkers, 1985; Kramer et al., 1988; Pfefferbaum & Ford, 1988; Scheffers, Johnson, & Ruchkin, 1991). Only a few studies manipulated perceptual load via the degree of similarity between targets and standard stimuli (Combs & Polich, 2006; Comerchero & Polich, 1999; Kiefer, Marzinzik, et al., 1998; Pfefferbaum et al., 1983), while most lowered the perceptual quality of the stimulus by embedding it into visual or auditory noise (e.g., Ford et al., 1982; Magliero et al., 1984; Pfefferbaum & Ford, 1988; Smulders et al., 1995; Christensen et al., 1996), or by changing its visual shape or intensity (e.g., Kok & Looren de Jong, 1980; Kok et al., 1985; Kramer et al., 1988). The major result is

[6] The aim of this study was to compare 1-, 2-, and 3-tone oddball designs. The 1-tone oddball design showed the shortest P300 latency, suggesting that the recognition of the target is easiest when there is no distractor.

similar to the signal detection experiments reported above, in that latency increases and amplitude decreases monotonically as discriminability decreases.[7]

For instance, McCarthy & Donchin (1981) and the follow-up study by Magliero and colleagues (1984) aimed to disentangle the contributions of perception and response-related processes on P300 and RT. The authors presented the words *left* and *right* on a screen and either embedded them in a noise matrix (a 4- x 6-letter matrix composed of random letter sequences) or in a no-noise matrix (filled with the "#" symbol). Magliero et al. (1984) manipulated the frequency of occurrence of the target word (e.g. *right*) so that it had a probability of (.20) across noise and no-noise trials and the non-target word exhibited the complementary probability (.80). In this typical oddball task, the participants were required to count only the target words. In accordance with the results from McCarthy & Donchin (1981), latency was significantly increased in all noise trials (regardless of target or non-target stimuli) because the perceptual identification and recognition of both the perceived stimulus was disrupted by the existence of distractor letters. P300 latency was only marginally affected by the compatibility between the stimulus and the response requirements (e.g., when a target word such as *left* required a response made with the right hand). In their second experiment, Magliero and colleagues employed a more fine-grained noise manipulation by using different matrix set sizes. For example, the noise size was 1 when there was only one letter surrounding the target word (e.g., A). This set size was then increased to yield three further different degrees of noise: a matrix containing the letters A to D, a second one containing the letters A to G, and a third one with the letters A to Z. Clearly, a small-sized noise matrix will disrupt the identification of the target words *left* and *right* to a much lesser extent than a noise set that contains all the letters of the Latin alphabet (and particularly when it includes letters that are

[7] This assertion about perceptual load has to be modified with respect to the P3a. Task difficulty and noise insertion appear to have a completely reversed effect on P3a amplitude and latency. That is, P3a amplitude is larger and latency is shorter in noise trials (Hagen et al., 2006; Combs & Polich, 2006).

part of the target and/or non-target stimuli themselves). Again, P3 latency was increased for noise trials. This effect was enhanced proportionally with the noise set, i.e. the larger the noise set, the longer the P3 latency.[8]

Some have proposed that the noise matrices in these studies can be so successful in degrading the target stimulus that a visual search is evoked (cf. Verleger, 1997). Other memory/visual search tasks in which participants had to compare a memorized pattern of elements, e.g., an array of digits or symbols, to the stimulus matrix presented (the so-called Sternberg task) confirmed the negative effect of visual search on P300 latency. The general finding obtained in these experiments was that with a larger display set (i.e., with more visual distractors in the stimulus matrix), stimulus evaluation time and consequently P300 latency were prolonged, as more comparisons between the members of the memory set and the display set had to be accomplished (e.g., Brookhuis, Mulder, Mulder, Gloerich, van Dellen, & van der Meere, 1981; Hoffman, Simons, & Houck, 1983; Okita, Wijers, Mulder, & Mulder, 1985; Luck & Hillyard, 1990; Smid, Lamain, Hogeboom, Mulder, & Mulder, 1991; Wijers, Okita, Mulder, Mulder, Lorist, Poisz, & Scheffers, 1987; cf. Verleger, 1997, for more references and discussion).[9]

Stimulus quality can also be changed in the visual domain by presenting stimuli in foveal or extra-foveal vision instead of adding noise. A study by Kok et al. (1985) examined

[8] P3 amplitude modulations were not reported for the second experiment. Visual inspection of the grand averages at Pz, however, suggest that P3 amplitude was reduced for noise trials as opposed to no-noise trials.

[9] The results regarding the impact of display size on P300 amplitude in memory search tasks are less homogeneous. Whereas Brookhuis et al. and Okita et al. (Okita et al., 1985; Exp. 1) reported that P300 amplitude decreases as a function of the number of comparisons that have to be made between memory set and display set, others found amplitude enhancement. One possible explanation for this may lie in the size of the memory set. Brookhuis and colleagues (as well as Okita et al., 1985) used memory sets consisting of up to 4 items, whereas Wijers et al. used one-item memory sets throughout their experiments. Brookhuis et al. argued that a larger memory set, which has to be retrieved from memory and compared to the display set, may lead to a comparatively greater degree of uncertainty in retrieving the memory set. As described above, uncertainty in the identification of the target leads to P300 amplitude reductions.

stimulus degradation in combination with the effects of visual field presentation (see also Kok & Rooyakkers, 1986; Kramer et al., 1988). The critical letter stimuli were presented in upper case or lower case, and they always appeared as pairs on the screen. Subjects were instructed to decide whether the letter pairs consisted of two identical or two distinct letters. The distinctiveness of the stimuli was reduced by altering their shape (i.e., the stimuli were composed of asterisks and two of these were changed in terms of their position). In addition, the letter pairs were either presented in the central visual field (i.e., in foveal vision within 1° of visual angle) or in the left or right visual field (i.e., in parafoveal vision within 2.85° of visual angle). While P3 latency was not reported in this study, Kok and colleagues reported clear decreases in P3 amplitude given a visually degraded stimulus. P3 amplitude was also smaller for all targets in parafoveal vision compared to foveally presented targets. The interaction between stimulus degradation and visual field presentation was such that the amplitude difference between intact and degraded stimuli was larger with foveal presentation, suggesting that the higher visual acuity in the fovea led to better discrimination of stimuli.

In sum, these experiments revealed that both the latency and the amplitude of the P300 are affected by the distinctiveness of the target stimulus. With perceptual degradation, P300 amplitude decreases, most probably because participants are less confident in their decision about whether they have detected the stimulus. P300 latency increases proportionally with decreasing discriminability, because the latency represents the time course of processes that must have been accomplished before the P300 is elicited.

1.2.4 Summary

This brief overview of the P300 has revealed that the target-related P300 is sensitive to a vast amount of different manipulations and therefore may be elicited by numerous different neural sources (Johnson, 1993). Task relevance is the main determinant of P300, as all other manipulations operate on the grounds of a task-relevant target (Johnson, 1986). Specifically, task relevance serves to assure that the target stimulus receives a sufficient amount of attention, which appears to be a necessary precondition for the P300. The probability effect including all its variants is one of the most well studied manipulations. It has been found to reliably influence the size of the P300 amplitude and can thereby reveal how subjects assign experimental stimuli to target and non-target classes of stimuli. With regard to the present experiments, the notions of confidence and perceptual distinctiveness are of major interest. When targets are classified as belonging to either the target or non-target category, a certain level of confidence is involved. If participants are highly confident when making a decision, P300 amplitude is larger and its latency shorter, as compared to a low confidence level. Confidence then appears to reflect the extent to which the collected information provides qualitative and quantitative evidence for categorizing the stimulus. While P300 amplitude is mainly an index of how confident a participant is in making their decision, P300 latency also reflects stimulus evaluation time, i.e. the time it takes to come to that decision (Donchin & Coles, 1988; Pritchard, 1981; but see Verleger, 1997, for a slightly different view). This shows that while the P300 depends on the perceptual properties of the stimulus, it is not elicited by them. Instead, the higher cognitive processes that evaluate these perceptual properties elicit the P300. Given the degradation of perceptual properties, evaluation time is extended, as either more time is required to collect a sufficient amount of evidence, or the search for this evidence is terminated in spite of incomplete information, and confidence in the categorization outcome eventually lowered.

These considerations must be kept in mind with the following review on the P300 in psycholinguistic studies. Results on P300 latency that increased proportionally with the time and effort it takes to evaluate the stimulus suggest that the P300 in linguistic experiments may very likely demonstrate latency shifts, depending on what and how much information is necessary to evaluate the perceived word. That is, the more complex the experimental stimulus, the longer the P300 latency will be in general. As words are more complex than tone pips or arrays of meaningless symbols, there could be longer P300 latencies overall. One could also assume that the evaluation processes for a target word differ with respect to which linguistic domain (e.g., orthography, semantics, or syntax) is sufficient to provide enough evidence for target evaluation. Moreover, the detrimental effects of parafoveal vision on P300 amplitude are especially important with respect the differences between foveal and parafoveal vision in natural reading. There is no doubt that parafoveal information in natural reading is less accurate than foveal information (Rayner, 1998; see also chapter 2.1). Thus, if a significant portion of evidence critical to target evaluation is already provided in parafoveal vision (i.e., before the target word is fixated for the first time), then the P300 in natural reading may be subject to severe visual-perceptual limitations.

1.3 The P300 in language processing

The following chapter aims to sketch out the circumstances under which the P300 may be observed in psycholinguistic experiments. Typically, this component has been investigated in relation to the two most prominent language-related ERPs, the N400 and late positivity or P600. Since N400 and P600 are brain responses to linguistic stimuli that can be grouped along the distinction between semantics and syntax in a number of cases, research to date has centered around two questions. Dating back to the early endeavors by Kutas and Hillyard (Kutas & Hillyard 1980a, b; 1984a,b), researchers have examined whether semantically and/or contextually unexpected words can be categorized as infrequent events which elicit the P300. As will be discussed in chapter 1.3.1, which deals with semantic priming studies, this is the case only in very specific contexts because experimentally defined probability does not mirror probability in linguistic terms. Rather, semantic or contextual incongruency leads to an increased N400 amplitude. On the other hand, the enormous variability in the latency of the P3b component has led others to contend that the late positivity/P600 that is found in response to syntactic irregularities may be a member of the P300 family, and thus a domain-general rather than a language-specific component (cf. chapter 1.3.2).

1.3.1 P300 and word identification

As the P300 component has traditionally been regarded as being associated with the unexpectedness or improbability of an event, the question soon arose whether unexpected words would also give rise to this component, as did simple tones or visual objects in a number of oddball experiments before. Many researchers used a priming procedure in which either a single word or an entire sentence made the occurrence of one word more probable than any other. The first endeavor in this regard was a series of experiments published by Kutas and Hillyard in 1980 (Kutas & Hillyard, 1980; see also Kutas & Hillyard, 1984a). This study set out to investigate ERPs in response to violations of readers' expectations. Assuming that an improbable event violating participants' expectancies consistently elicits a P300 component (e.g., Donchin, 1981, see chapter 1.2.1), Kutas and Hillyard investigated whether the same holds for complex stimuli that require more controlled cognitive processes than do the stimuli in oddball tasks. They presented simple sentences such as those given in (1) – (3) (underlines added) and instructed their participants to read the sentences for comprehension in order to answer comprehension questions after the experimental session.

(1) It was his first day at <u>work</u>.

(2) He spread the warm bread with <u>socks</u>.

(3) She put on her high heeled <u>shoes</u>.

The first example sentence served as the control condition in three experiments, in two of which the semantic congruency of the sentence-final word was altered in 25 % of all sentences (ranging from a moderate to a strong violation of contextual appropriateness). In the third experiment, only the letter size of the otherwise semantically congruous final word was changed on 25 % of all trials. In all cases, the semantically incongruous or physically deviant words were low-probability stimuli that were clearly distinct from the other terminal words. Consequently, subjects should have established two categories of stimuli in these experiments.

The results were fairly clear-cut. Semantically incongruous words did not elicit a P300-like component, eliciting an N400 instead. Physically deviant word forms, by contrast, engendered a late positive complex with three principal peaks and the latest of these had a centro-parietal maximum (termed P560). The P300 was apparently sensitive to the probability of the physical properties of a terminal word, while in terms of the semantic congruency, the probability effect did not initiate the processes that correlate with the P300. This dichotomy was fairly robust in subsequent research, such that the linguistic manipulation elicited either an N400 or a P600 to semantic congruency or syntactic reanalysis, respectively, whereas the physically deviant event showed an enhanced amplitude for the P300 (cf. Allefeld, Frisch, & Schlesewsky, 2005; Kutas & Hillyard, 1984a; Osterhout, McKinnon, Bersick, & Corey, 1996). Studies investigating semantic congruity without any physical tasks replicated the N400 for semantically incongruent words, in the absence of any concomitant P300 activity (e.g., Camblin, Gordon, & Swaab, 2007; Connolly, Phillips, Stewart, & Brake, 1992; Ditman, Holcomb, & Kuperberg, 2007; Neville, Kutas, Chesney, & Schmidt, 1986; Nobre & McCarthy, 1994).

Strikingly, there was no P300 associated with semantically incongruent words, even though they had (i) a lower experimentally defined probability in the Kutas & Hillyard study and (ii) are in themselves less expected and should thus have a lower probability of occurrence within the respective linguistic context (independent of experimental probability). Based on this finding, one can agree with Kutas & Hillyard’s (1980a) conclusion that the cognitive operations indexed by the N400 must be different from those reflected in the P300. Kutas and Hillyard are also correct in stating that language comprehension is a continual context updating process that predicts upcoming words and revises its hypotheses in the case of conflicting evidence. This fits well with Donchin' s (1981) proposal of context updating as the psychological correlate of P300 activity and one is tempted to willingly accept the original hypothesis tested by Kutas & Hillyard, which states that P300 should necessarily be elicited by contextually incongruous words when they bear a low experimentally defined probability. However, this implicit bias in their argumentation is erroneous for two reasons. Firstly, it equates external probability as defined a priori by the experimenter with linguistic probability. Manipulating the frequency of occurrence for any given target in an oddball experiment is fundamentally different from external probability manipulations of linguistic stimuli. Words are by far more complex than typical stimuli in oddball experiments because they are always used and perceived in context (outside the experimental session), so it must be questioned how (and why) the a priori defined experimental probability translates into subjects' mental discourse models.[10] There is no compelling reason to assume that an external probability applied to words embedded in context will have similar or identical outcomes as the probability manipulation in oddball experiments, in which stimuli have no meaning outside the experimental setting. Secondly, the hypothesis tested by Kutas & Hillyard overlooks that a binary decision between target stimulus and non-target stimulus has to be made, so the P300 is first and foremost an index of successfully completed identification of the target stimulus (for review, see Johnson, 1986; Pritchard, 1981).

Specifically, semantic incongruity as introduced by Kutas and Hillyard should actually be a striking deviation of the oddball design in which a target category with a set size of one is alternated with a limited number of non-target categories (e.g., in 2-stimuli or 3-stimuli

[10] Consider, for example, the following two mini-discourses. Whereas example (1) is implausible, (2) is semantically correct because the mental discourse model set up by the name *Pippi Langstrumpf* licenses that one particular girl is physically strong enough to carry a horse. The linguistic probability of occurrence is thus unaffected by experimental probability.
(1) Once upon a time there lived a girl in a small village. One day, the girl carried a horse.
(2) Once upon a time there lived a girl called Pippi Langstrumpf in a small village. One day, the girl carried a horse.

oddball tasks). On closer inspection, neither the contextually appropriate word nor the inappropriate words are *linguistically* defined as belonging to a set of limited size in that study. That is to say, in a context sentence such as (2), there are various words that would provide a meaningful continuation, and virtually an infinite number of inappropriate words. To be clear, a sentence fragment such as *he spread the warm bread with…* can be completed with any member of an admittedly small set of congruent words such as *Marmite*, *peanut butter*, *jam*, or any other food that has the property of being spread on bread. Yet, under these circumstances, it is almost impossible for the language comprehension system to make a binary decision, for each member of the congruent set may equally well serve as the terminal word. There is thus no single, precisely defined target stimulus, as is necessary for a binary decision. Put another way, if several words could be the relevant target stimulus completing the sentence fragment, it is not clear which equi-probable target will be taken into consideration for the binary decision. One might assume that the language comprehension system decides between all target candidates in parallel, but this is very unlikely as target set size increases. From the perspective of binary decision-making, one would thus not expect a P300 to incongruous words in the kind of sentences exemplified in (2) above because the basic criteria for its elicitation are not met.[11] What this account also entails is that there might be P300 activity in highly constraining contexts in which the set of appropriate continuations is reduced to only one possible target word. A binary decision would become feasible in these cases. In fact, the contexts in which linguistic stimuli elicit P300s clearly allow for a binary

[11] As for physically deviant words, the binary decision is based on the discrimination of simple visual information such as letter shape or size. Because it is not necessary to extract the meaning of the words, target probability can fully exert its influence for physically deviant word forms and thus engender a P300. Although the physical deviation was not relevant for the comprehension task in that study, all manipulations were implemented at the sentence-final word. So, participants may have interpreted the physical deviation as a task-relevant cue (see chapter 1.3.2 for similar arguments concerning the Osterhout et al. 1996 study). Further support for this claim stems from the Allefeld et al. (2005) experiment. They used a similar experimental design in terms of semantic incongruity and physical deviance, but exchanged the comprehension task with a probe task, which inherently required a physical matching process. The pattern reported by Kutas & Hillyard was replicated in Allefeld et al.'s study.

decision. One of these contexts is referred to as a high-constraint context where the cloze probability of a particular word is so high that a binary decision is surely based on it, the other one, lexical decision, is in fact an extreme example for binary decision-making for when participants judge the lexicality of a letter string, either outcome of the binary decision (word versus nonword) engenders a P300.

The difference between high constraining and low constraining contexts is measured by cloze probability, which is defined via the proportion of trials with which a word is used to complete a sentence fragment. A word with a high cloze probability is considered highly predictable from the preceding context, whereas a word with a low cloze probability is less predictable and may also at times represent an implausible completion of the sentence. With an experimental design similar to the one reported above, Kutas & Hillyard (1984b) collected ERPs from terminal words in simple sentences. They manipulated the cloze probability of the terminal words and embedded them in either high constraining or low constraining sentences (i.e., the mean values of cloze probability were in general higher for the former than for the latter). The results showed a P300 for the final word only when a high-constraint sentence was completed with a highly probable word (e.g., *he mailed the letter without a* <u>*stamp*</u>), whereas words with low cloze probability showed graded N400 amplitude effects across low-constraint and high-constraint sentences, with an inverse correlation between cloze probability and N400 amplitude. This seems to support the claim made above that semantic incongruity as implemented in their earlier study (cf. examples 1 – 3) does not elicit a P300 even though contextually anomalous words are unexpected, because the sentences used were not constraining enough to substantially reduce the set of appropriate terminal words.

The resulting question as to whether the contextual constraint or the cloze probability of the terminal word is the predominant factor driving the effect described above has not been addressed directly in electrophysiological research on language. The data from Kutas & Hillyard (1984b) together with the findings from Federmeier & Kutas (1999) remain inconclusive as to whether cloze probability, the presence of lexical associates, or an interplay of both factors is more influential in narrowing down the set of possible terminal words that fit the sentence. Federmeier & Kutas (1999) investigated to what extent high-constraint and low-constraint contexts may facilitate the processing of terminal words that were (i) the best completion to a sentence, (ii) unexpected words semantically related to the best completion or (iii) unexpected words not related to the expected word. They found increased N400 amplitudes for the two unexpected words, whereas the best completion (i.e., the expected word) showed no N400 activity. Instead, the expected word elicited what they termed a late positivity between 300 and 500 post-onset of the word. This positivity is difficult to interpret as P300 based on several aspects of the data. First, there are results showing that each word in a sentence gives rise to an N400 component, the amplitude of which, however, also decreases as the sentences unfolds and more contextual information aids word processing (Kutas, Van Petten, & Besson, 1988; Van Petten & Kutas, 1990). Although the amplitudes for all conditions in the Federmeier & Kutas study were generally more positive in high-constraint contexts, best completions showed no different pattern in low versus high constraining contexts, suggesting that in either case they may have benefited more from enriched contextual information than the two unexpected completions. If there had been a reduced N400 to best completions, measuring at the sentence-final position might have eliminated this effect. Second, the positive wave did not exhibit the well-defined peak typical of the P300, even though it had a maximum at central-parietal sites (Federmeier & Kutas, 1999).

It is interesting to note that in the Federmeier & Kutas study, the average cloze probability for best completions was 0.89 in the high-constraint context and 0.58 in the low-constraint context. The probability values ranged from 0.17 to 1.0 across both context types. Thus, one could conjecture that even a modest decline in average cloze probability significantly hinders binary decision, which is reflected in a reduction of the P300 amplitude. This would suggest that cloze probability is more important than contextual constraint.

However, this conclusion is confounded by the divergence in stimulus materials across both studies. Federmeier & Kutas excluded lexical associates of the critical word in their target sentences, whereas the example stimuli by Kutas & Hillyard (1984b) involved lexical associates. It appears possible that context yielded stronger facilitation in the study of Kutas & Hillyard because of the lexical association between words. Support for this claim could stem from their medium-constraint context/high cloze probability condition (e.g., *she locked the valuables in the* <u>*safe*</u>) that showed a reduced P300 amplitude relative to the high-constraint context/high cloze probability condition. This means that a definite binary decision during language comprehension is only possible if cloze probability and lexical association converge in strongly hinting at one particular target. Nevertheless, because the correlation between cloze probability, lexical associates, and the P300 has not been studied in detail, there remains some speculation about the exact relation between them.

It is important that cloze probability should not be taken prematurely as an equivalent to target probability in the oddball task. If this were the case, one might see an inverse probability effect in language comprehension with the stimulus exhibiting the highest value of cloze probability eliciting a P300 component (cf. Verleger, 1988, and his notion of awaitedness). The alternative hypothesis proposed here is that, given the importance of the binary decision process, cloze probability actually serves to single out the task-relevant target stimulus. As any linguistic context offers far more possible target and non-target candidates than the oddball task, the cloze probability mechanism seems to be a fairly effective way to define the target a priori.

The importance of an a priori defined target becomes even more evident in experiments using a lexical decision task. Volunteers in this kind of experiment read or hear (pairs of) letter strings and are required to judge whether they are real words or not. The lexical decision can be made for each word in the stream or, especially when strings are presented pair-wise, for the target word only (which appears second in the word pair). Typically, the aim of these studies is to find electrophysiological indices of semantic priming, that is, the facilitation in word processing when a target word follows a semantically related prime word. The major finding across all studies is that semantic priming results in the reduction of the N400 amplitude for the target word.

Most of the investigations focused on the strength of semantic relatedness required to induce semantic priming (e.g., Bentin, 1987; Chwilla, Kolk, & Mulder, 2000; Hagoort, Brown, & Swaab, 1996; Hill, Ott, & Weisbrod, 2005; Kiefer, Weisbrod, Kern, Maier, & Spitzer, 1998; Kutas & Iragui, 1998; Weisbrod, Kiefer, Winkler, Maier, Hill, & Roesch-Ely, & Spitzer, 1999). Some investigated differences across modalities (e.g., Anderson & Holcomb, 1995; Holcomb & Anderson, 1993; Holcomb & Neville, 1990) and hemispheric differences for language-related ERPs including the N400 (Kiefer et al., 1998; Nobre & McCarthy, 1994), or aimed at clarifying the influence of general processing constraints on the N400 such as the allocation of attention, task instruction, or stimulus degradation (e.g., Bentin, Kutas, & Hillyard, 1993, 1995; Chwilla, Brown, & Hagoort, 1995; Holcomb, 1993; Kounios & Holcomb, 1992; McCarthy & Nobre, 1993; Roehm, Bornkessel-Schlesewsky, Rösler, & Schlesewsky, 2007). Consequently, much is known about the semantic priming effect as reflected in the N400. For instance, although the N400 component was also elicited for unattended words, N400 priming effects only emerged when prime and target words were actively attended to (Bentin et al., 1993, 1995; McCarthy & Nobre, 1993; but see Rolke, Heil, Streb, & Hennighausen, 2001 for results showing that missed prime words within the attentional blink can elicit small priming effects) or were substantially enhanced when the prime word had to be actively attended to (Holcomb, 1988). Moreover, degrading verbal stimuli by altering intensity or by overlaying a matrix of visual noise on the respective target word generally increased N400 latency, but did not prevent or interact with the semantic priming effect (Holcomb, 1993).

In contrast to the mass of data and theories on the processing nature underlying the N400, comparatively few efforts have been made to systematically investigate and explain the concomitant positivity that is engendered by words and non-words alike (but see Bentin, Mouchetant-Rostaing, Giard, Echallier, & Pernier, 1999; Bentin, McCarthy, & Wood, 1985; Holcomb, 1988). This explanatory gap is astonishing because the occurrence of the P300 in psycholinguistic studies varies in accordance with most of the determinants described in chapters 1.2.1 through 1.2.3. First, the P300 in lexical decision tasks is prone to the deprivation of attentional resources as in any other task. For example, Holcomb (1988) investigated whether the allocation of attention to the prime in a prime-target pair had a reliable influence on the recognition of either word. His findings showed that prime words, where no lexical decision was required, only elicited a small but reliable positivity when their meaning had to be processed attentively. The P300 to target words was larger when the participants had to attend to the meaning of both target and prime (i.e. the condition in which more attention was allocated to the semantic relation between prime and target). Furthermore, McCarthy & Nobre (1993) examined the interaction between semantic priming and spatial selective attention. In this experiment, participants were to fixate on the center on the screen and focus their attention on words either presented in the right visual field or in the left visual field. The task was to detect exemplars of an a priori defined semantic category. Words in the other visual field could be ignored. In accordance with prior findings from psychophysiological studies on the allocation of attention, they found a P300 for target words only if they were attended to. Semantically appropriate exemplars in the unattended field yielded no P300 activity, and no N400 semantic priming or repetition priming effects.

Second, stimulus degradation during language comprehension also seems to affect the P300. Bentin and colleagues (1995) aimed to investigate the semantic priming effect under dichotic listening conditions. The P300 was completely absent for all targets with dichotic listening (as was the N400 priming effect to unattended words). Provided that words in the attended and unattended channel were presented simultaneously, this seems to suggest that dichotic listening introduces a strong noise factor that does interfere with semantic processing, and appeared to dramatically impede the processes relevant for the elicitation of the P300. A reason may be that words are not perceived as discrete linguistic units under dichotic listening conditions. A similar point can be made for the results obtained by Anderson & Holcomb (1995) who compared how different time intervals between the onset of the prime and the onset of the target (stimulus onset asynchrony; SOA) across the visual and auditory modalities influence semantic priming. When the words were presented auditorily, there was no ERP wave that could be reliably linked to P300 activity in the two short SOA conditions (with a duration of either zero or 200 ms), whereas the long SOA (800 ms in duration) evidenced a P300 sensitive to word identification. The N400 semantic priming effect also became larger as the SOA increased. This finding replicates Bentin et al.'s (1995) results in that short SOAs apparently lead to full or partial dichotic listening of prime and target. This shows that the short SOA exceeded the threshold at which discrete linguistic units were still recognizable. The resulting indeterminacy of the linguistic units was detrimental to the P300. Contrary to this, when prime and target were visually displayed to the left and right of a fixation mark and, there were clearly visible positivities for related and unrelated words in the two short SOA conditions (i.e., a P300 in response to the lexical identification of the stimulus). Semantic priming was also evident as an increased N400 for unrelated words, peaking slightly earlier than the P300 component. Apparently, simultaneous presentation in the visual domain has a less detrimental effect on P300 amplitude than the concurrent presentation of two spoken words. During simultaneous presentation, visual-spatial position, as opposed to audio-spatial cues, may have been a more effective cue aiding to selectively attend to the target word since they unambiguously indicated the word boundaries of prime and target, yielding discrete visual objects.

Finally and most importantly, the P300 is elicited by words and non-words because both categories are task-relevant target stimuli in a lexical decision task, and P300 amplitude and latency also interact with the ease of the lexical decision.[12] For instance, Bentin and colleagues (1985) were among the first to show that all items in a lexical decision task exhibit a P300. Their data also indicated that both P300 latency and amplitude differed between words and non-words. In particular, latency was earlier and amplitude more positive for target words that followed a semantically related prime, compared to all other items.[13] The sensitivity of P300 latency and amplitude to semantic relatedness has been subsequently replicated in other studies. Holcomb (1988) reports the P300 to semantically related target words peaks earlier than to words that were not preceded by a prime, or to words that were semantically unrelated to the prime. Chwilla and colleagues (1995) compared semantic priming effects in a lexical decision task, and in what they called a physical task in which participants had to respond to upper- or lower-case words. Both the N400 and the P300 emerged in the brain waves associated with the lexical decision task, but due to a high degree of component overlap, Chwilla et al. gave no functional interpretation of the P300. As for the physical task, there was again a pronounced P300, which, in addition, varied in its amplitude as a function of stimulus probability. Critically, there was no significant N400 priming effect associated with this task, instead the probability effect for the P300 amplitude completely depended on the categorization of target words as being either related or unrelated to the

[12] The target stimulus in a typical oddball experiment is usually the one that is relevant for the present task. Accordingly, each letter string for which a lexical decision is to be made becomes a target. This emphasizes the importance of task relevance in the elicitation of P300 (cf. Johnson, 1986). As will be further described below, the different stimulus types in lexical decision experiments can be distinguished according to P300 latency, and to some extent, amplitude variability.

[13] The authors reported no P300 latency effects (but reliable amplitude effects) after the ERPs to words and non-words were reaveraged to only include trials with normalized reaction times (RT). Although P300 latency has been shown to covary with RT, this correlation was reduced when verbal stimuli were to be identified under speed instruction (cf. Kutas et al., 1977). Since the participants in the Bentin et al. study were given explicit speed instructions, the latency adjustment adopted by Bentin and colleagues may be called into question based on this weaker correlation between speeded responses and P300 latency.

prime word.[14] As Chwilla and colleagues noted, this may suggest that the semantic relatedness feature may have been used to categorize stimuli. Finally, Holcomb & Neville (1990) found that pseudowords elicited both an increased N400 (compared to words) and a P300 following the N400 component. This positivity was less pronounced than the monophasic P300 to illegal non-words. Although the amplitude difference between pseudowords and non-words could be explained with the lower probability of non-words as opposed to the class of word-like items (including pseudowords), as suggested by Holcomb and Neville, it cannot account for the lack of N400 activity for non-words. The authors argued instead that non-words are rejected as words based on prelexical properties, whereas pseudowords possess prelexical properties that can initiate lexical search.

Similar results concerning lexical relatedness were obtained when comparing different degrees of lexical relatedness between prime and target (e.g., Chwilla et al., 2000; Hill et al., 2005; Kiefer et al., 1998; Weisbrod et al., 1999). In such experiments, participants had to discriminate words from pseudowords. Semantic relatedness was varied in a threefold manner: prime and target words were directly related (e.g., *leg – arm*), indirectly related (*lion – stripes*, i.e. mediated priming), or unrelated (*anchor – man*). The results of these studies are particularly intriguing with respect to which kind of linguistic information the P300 is

[14] This effect resulted from the presentation of two lists that differed in target word probability. In one list, related words occurred with a probability of .80, in the other with a probability of .20. The P300 amplitude was larger to low-probability unrelated words in the first list, and correspondingly larger to the low-probability related target words in the second list. Note that the results for the non-words used in that experiment are less clear. In the lexical decision task, non-words led to an increased N400 amplitude. The subsequent time range in which non-words seemed to show a somewhat more positive-going ERP wave than the word stimuli was not examined statistically. In the physical task, the P300 to non-words was less positive compared to words, and especially, to unrelated words exhibiting a lower probability. Note furthermore that Chwilla et al. gave no examples for the non-word stimuli, so that it remains unclear whether the authors used illegal non-words or pseudowords.

reflecting. The largest P300 was registered for directly related targets, while N400 amplitude increased with increasing semantic distance between prime and target.[15]

These latency and amplitude effects are more obvious in cases where semantic relatedness gives rise to full predictability (i.e., cloze probability of one), as with antonyms. Each of the few studies on antonym processing reported a pronounced P300 for antonyms, but no clearly visible N400 (Bentin, 1987; Kutas & Iragui, 1998; Roehm et al., 2007). Roehm and colleagues further aimed to unravel the temporal overlap of N400 effects and P300 by investigating to what extent the P300 in the processing of antonyms depended on task instruction, contextual factors, and stimulus type. They compared antonyms (*black – white*) with both within-category (i.e., semantically related words such as in *black – yellow*) and across-category violations (i.e., semantically unrelated words such as in *black – nice*) in three experiments, comprising either a sentence context with an antonymy judgment (*X is the opposite of Y*), word pairs with a lexical decision task, or word pairs with an antonymy judgment. With the antonymy judgment, the authors found a P300 for antonyms which peaked slightly earlier than the N400 for the two non-antonym conditions, regardless of whether words were embedded in a sentence or presented in pairs. In addition, related words differed significantly from unrelated words in showing a reduced N400 amplitude. This could be taken as support for the claim that semantic preactivation has a facilitative effect on within-category violations of semantic expectancies (Federmeier & Kutas, 1999). All non-antonyms in the study elicited a late P300 following the N400 epoch. The P300 amplitude did not differ between the non-antonym conditions at parietal electrode sites with an antonymy judgment for word pairs; it was, however, more pronounced for unrelated target words in a sentence context.

[15] The Chwilla et al. (2000) study did not directly test the differences in P300 amplitude across experimental conditions. Visual inspection, however, suggests a trend similar to the one reported.

This amplitude difference is surprising because there should be no difference within the set of non-antonyms, because both types of non-antonyms are equally unexpected. A functional explanation along the lines of task or stimulus complexity is ruled out because the positivity for the two non-antonym conditions was undistinguishable at parietal electrodes in Roehm et al.'s second experiment using word pairs and wholly undifferentiated across all electrode sites in their third experiment using a lexical decision task (see below). If the unrelated word per se required more processes, thereby increasing task complexity, one should see the amplitude modulation independent of context manipulation. At present, it appears safe to presume that the unusual pattern for non-antonyms in sentence context is probably connected to component overlap brought about by particular circumstances in the experiment. Alternatively, one could assume that the P300 to related target words is reduced in sentence context because lexical relatedness interferes with the cloze probability effect. These words are unexpected given the preceding sentence context, but they also bear some semantic association with the expected word. This would imply that the decision-making on related target words is less certain, yielding amplitude reduction (cf. chapter 1.2.2).

Different results appeared when participants were instructed to discriminate words from pseudowords. Semantically related words and antonyms showed the well-established N400 semantic priming effect, i.e. the N400 amplitude was reduced relative to unrelated target words. There was, however, no concomitant P300 for the antonym condition. All word conditions, regardless of semantic relation to the prime, showed a small-sized P300 following the N400 instead. Pseudowords not only elicited an N400 with an amplitude similar to the one for unrelated target words, but also a pronounced P300 that was significantly larger in amplitude than the one for word stimuli.[16] The P300 result is in accordance with other studies,

[16] Note that the analyses reported by Roehm et al. (2007) did not include an explicit comparison between words and pseudowords. Based on visual inspection of the ERP waves (cf. Figs. 2 and 3), it appears that the P300 for words peaked at the same time as the P300 for pseudowords, the only difference being that the amplitude of the word-P300 was smaller.

which also found enhanced P300 to pseudowords (cf. Hill et al., 2005; Holcomb & Neville, 1990, based on visual inspection). A likely explanation for this is that pseudowords are more complex stimuli because they provide contradictory information in that they conform to orthographical and phonotactical rules of a language, without having a lexical entry. They thus require increased processing efforts in order to reconcile contradictory information before they can be recognized as an instance of non-words; generally, stimulus complexity has been shown to enhance P300 amplitude (see Johnson, 1986, for review).

There is one final result in the study from Roehm and colleagues that deserves mentioning. In comparing patterns of antonym processing with and without sentence context, Roehm and colleagues noted that the P300 for unrelated non-antonyms presented in isolation was reduced in size, as compared to the P300 for those words in sentence context. The authors then subdivided their participants into two groups, one exhibiting a P300 and another without it. Strikingly, the occurrence of the P300 to unrelated non-antonyms correlated with the elicitation of the P300 to antonyms. For the group that did not elicit a late P300 to unrelated non-antonyms, there was also no P300 antonyms, and vice versa. Crucially, there was no group-specific difference in the reported N400 effects. This finding is particularly interesting in light of the debate as to whether P300 and the late positivity indexing grammatical processes are functionally related or not (see chapter 1.3.2).

Let us now turn to the major implications that these studies bear for the present series of experiments. Whereas the majority of the studies reported here found that the P300 followed the N400 effects (which either reflected semantic priming effects or lexical search for

---

This contradicts other studies, which found a P300 for semantically related target words that peaked slightly earlier than the N400 to unrelated target words (cf. Hill et al., 2005; Holcomb, 1988). It seems this striking difference cannot be explained in terms of task differences. A double lexical decision task was used in both the Roehm et al. and the Hill et al. study. The SOA had a duration of 800 ms in the Roehm et al. experiment, which is only 50 ms shorter than the SOA used by Holcomb (1988). However, the data from Roehm and colleagues only exhibit a P300 amplitude modulation, whereas the other investigations also reported latency (and amplitude) variations. Thus, the reason for the present dissociation remains unclear.

pseudowords) because lexical access was necessary before the perceived stimulus could be categorized as a word or non-word, some studies reported the lack of clear N400 activity for the target words strongly related to the primes. This effect for strongly related words seems to be reducible to the specific task instructions in those studies. For instance, Bentin instructed his participants to "think of the antonym" (Bentin, 1987, p. 311) to a given prime word, while Holcomb (1988) encouraged participants to evaluate the semantic relationship between prime and target as this could help them to carry out the lexical decision task. Evidently, the antonym judgment in the Roehm et al. study also strengthened the semantic relationship between prime and target (see also Chwilla et al., 2000, Exp. 1 versus Exp. 2). Recall that antonyms elicited an N400 and a subsequent P300 with a lexical decision task, whereas the early P300 without a concomitant N400 was only present with an antonymy judgment task. It seems the lack of N400 activity, in the presence of a very early and pronounced P300, correlates with the attentive processing of the prime and the anticipation of a specific upcoming target.

The presence or absence of a P300 in the studies investigating word identification processes can thus be predicted mainly by task relevance (see Hagoort et al., 1996, who found no P300 when subjects were told to simply listen to a series of words without a task) and, to some extent, stimulus degradation. This is very much in line with the findings sketched in chapter 1.2.2 and 1.2.3. When prime and target words are presented in a noise-free environment (i.e., the stimuli are not degraded, so that they can be perceived as discrete linguistic units), the actual task – be it a lexical decision or an antonymy judgment – is intrinsically tied to a binary decision on the perceived input. On the one hand, the P300 follows the N400, which reflects either lexical access or integration, and this seems to suggest that the P300 is not engendered by lexical processing. On the other hand, there are circumstances under which the P300 occurs concurrently with the N400. This latency shift is in fact the most remarkable finding, as one would expect that the P300 *always* follows the

N400, provided it indeed reflects the completed identification of the stimulus as being either a word or not. Only after lexical access, the decision on the lexical status of the perceived stimulus can be made. So, what causes the latency shift of the P300 and which factors may lead to the dramatic decrease in N400 amplitude for highly related targets when N400 and P300 epochs overlap? One may contend that this is an artifact due to component overlap, but this does not explain how lexical relatedness might influence P300 amplitude (Chwilla et al., 1995; Hill et al., 2005), nor the similarity with non-words that entail no N400 activity, since such non-words can be recognized based on their deviant orthographic and phonological form (cf. Bentin et al., 1999; Holcomb & Neville, 1990).

The hypothesis put forward and tested in experiments 1 through 3 is that the P300 as reported above in fact reflects the completion of target identification. The review on P300 literature has revealed that P300 latency correlates with stimulus evaluation time (e.g., McCarthy & Donchin, 1981; Pritchard, 1981). Consequently, target identification may be based on different information types, as shown by P300 latency. According to this hypothesis, the identification of a target does not necessarily entail lexical access. When the N400 precedes the P300, lexical-semantic processing certainly takes place before the target is successfully identified. This seems to be the general pattern for unpredicted targets or words that are unrelated to the prime word. However, when the P300 and N400 overlap temporally, or when there is no detectable N400 activity, the P300 may reflect identification based on other stages of word recognition such as pre-lexical information. Specifically, orthographic or phonological information could be sufficient to recognize predictable targets or targets highly related to a prime.

This hypothesis in partly motivated by the processing pattern of non-words, which are rejected at a prelexical stage. Like highly predictable or semantically related words, they show no N400 activity, but a pronounced P300 in the N400 time window.[17] Although it has only rarely been investigated whether pre-lexical information is activated for highly predictable words, in addition to semantic concepts (but see Dominguez, de Vega, & Barber, 2004), a recent EEG investigation suggests that this is in fact the case (Laszlo & Federmeier, 2009). In this study, the processing of expected words ending a short sentence was compared with that of unexpected items (containing an equal proportion of unrelated words, pseudowords, and non-words), which were either orthographic neighbors of the expected word or not. The authors reported significantly reduced N400 amplitudes for unexpected items that were orthographic neighbors of the expected words, and thus concluded that specific predictions in language comprehension can activate orthographic information and semantic concepts.

So far, it has been argued that the variation in P300 latency is due to task instructions which encourage participants to attend to both the prime and the target word or to contextual constraints that differentiate predictable from unpredictable target words. Under these circumstances, successful word recognition, as reflected in the P300, can be grounded on pre-lexical information. This is a quite general account for the data and says little about the underlying linguistic processes, because the overall pattern gives little indication for how the mechanisms underlying semantic priming motivate the P300 latency shift. It has been a matter of debate whether semantic priming is a consequence of the automatic spread of activation, or the prediction of a specific target to which attention is allocated (e.g., Collins & Loftus, 1975; Neely, 1977, 1991; Posner & Snyder, 1975; for more recent discussions see Federmeier & Kutas, 1999; Laszlo & Federmeier, 2009). The spread of activation from one word to its associates is commonly assumed to differ from specific predictions in that it is an automatic and fast process that does not require attention, and results in facilitative effects (cf. Collins &

[17] This by no means implies that lexical processing is stopped with the emergence of the P300 to highly predictable or semantically related target words. It has been argued that stimulus processing continues even after the P300 has been elicited (e.g., Kutas et al., 1977).

Loftus, 1975; Neely, 1977). Predictions, by contrast, are assumed to be controlled processes that take time to build up, require intentional processing, and may result in both a facilitation for related concepts and an inhibition of unrelated concepts (cf. Neely, 1977, 1991; Posner & Snyder, 1975).

Although it appears attractive to link the P300 to one of these two possible generators of semantic priming, the previously found data prohibit such a relation. Studies of semantic priming effects show that spreading activation is mostly present at short SOAs and rapidly decays with longer SOAs, whereas predictions are more dominant with a long SOA (cf. Neely, 1977). However, the P300 latency shifts described above cannot be grouped along the lines of short vs. long SOAs. Holcomb (1988), for example, found the P300 to related target words to peak within the same time window as the N400, when the interstimulus interval (ISI) was 800 ms (see also Bentin, 1987, for the P300 emerging with long SOA). The concomitant occurrence of P300 and N400 has also been found with an intermediate ISI (cf. Roehm et al., 2007, who employed an ISI of 400 ms) and with a short ISI (cf. Weisbrod et al., 1999; Kiefer et al., 1998). However, a study that directly compared the impact of short versus long SOAs on semantic priming found the early P300 present only for the short SOA, with a significant reduction for the long SOA (cf. Hill et al., 2005). Thus, this P300 latency shift may be induced by both an automatic spread of activation and controlled predictions.[18] This highlights the necessity of a new method of stimulus presentation that can disentangle spreading activation from predictions – without imposing too great demands on participants, as would be the case, for instance, for RSVP with very short SOA or ISI. The dissociation of

[18] Alternatively, one may regard each component as being associated with one process. From this perspective, the P300 correlates with spreading activation, while the N400 correlates with the predictions made by the language processing system. This is sensible when taking into account that spreading activation results in stronger facilitation for highly related concepts. The identification of these targets may thus be accomplished earlier. However, the consistent finding that spreading activation and predictions depend on the time between prime and target speaks strongly against this account such that it will not be pursued here. Note that the findings from Experiment 2 reported below also do not support this account.

spreading activation and controlled predictions will thus be investigated in Experiment 2 in which ERPs were collected during natural reading.

These considerations on what exactly the language processing system predicts and on whether spreading activation and/or prediction affect the P300 give rise to an alternative interpretation of the P300 – particularly for predictable target words. Instead of being elicited by prelexical information alone, the emergence of the P300 could depend on when the decision on the perceived word becomes possible. When both a specific word form and a specific meaning are predicted (as, for example, in the case of antonyms) both linguistic information types are necessary to successfully identify the word as the expected target word. This predicts that the P300 to highly expected words should behave differently from the one to unexpected words when information about word form and meaning become available to the language processing system at different points in processing. As chapter II will show, prelexical and lexical information are delivered at different positions in natural reading, allowing for a temporal dissociation of these information types.

In conclusion, it appears that the P300 is consistently elicited by the stimuli in word recognition studies when the experimental task enforces a binary decision. Contextual congruency, however, offers an inappropriate experimental setting to elicit the P300 with linguistic stimuli, as there is no a priori defined target. Although the set of congruent words may be largely reduced in linguistic terms, there are still too many candidates to allow for a binary decision. On the other hand, there are high-constraint contexts that make only one target available such as lexical decision experiments distinguishing words from non-words and antonymy judgment tasks requiring the differentiation between antonyms and non-antonyms. Such investigations appear to tap into those mechanisms of expectancy and probability that are decisive for the emergence of the P300. The semantic nature of the task thereby determines when the P300 will be elicited. In a lexical decision task, words are discriminated from pseudowords by their meaning, so the P300 peaks just after the N400 for both of these items. Non-words are recognized more quickly based on prelexical information, so that there is a well-shaped P300 with no visible N400 activity. When the semantic relation between prime and target is highlighted either by task instruction or by linguistic context (e.g., when antonym processing is investigated in sentence context), pre-activation of the expected target substantially reduces stimulus recognition time, and the P300 to these expected targets occurs concurrently with the N400 to targets which are not pre-activated or expected.

The studies reported in the present section have dealt with the processing of single words. The next section will focus on studies investigating sentence comprehension, with particular reference to syntactic processing. In these studies, (morpho-)syntactic information is the crucial part of the word that allows for successful target identification. As shown below, such investigations replicate the major conclusion of the present chapter: The P300 in language comprehension experiments is not elicited by manipulating the experimental probability of a linguistic stimulus, but rather when task instructions lead participants to expect an a priori defined target.

1.3.2 The identity hypothesis: Is the P600 a P300 or not?

Some researchers argued that the P600 in psycholinguistic studies investigating syntactic processing could be a member of the P300 family and, consequently, should not be considered a language-specific component (Coulson et al., 1998a; Coulson, King, & Kutas, 1998b; Gunter, Stowe, Mulder, 1997; Gunter & Friederici, 1999). This assumption is based on two observations. First, P300 latency increases proportionally with the time it takes to evaluate a stimulus, so it is probably increased for words, which are more complex than typical stimuli in oddball experiments. Second, the distribution of both the P300 and the P600 has a posterior maximum. While distinct scalp topographies strongly suggest different underlying neuronal generators, similar topography may also be the product of one and the same neuronal source (see Coulson et al., 1998a, for discussion). Given these characteristics, it was further presumed that the P600 should prove sensitive to some of the factors that evoke the P300 (P3b), if it is indeed a P300 family member. Researchers testing these assumptions have mostly employed the probability effect by manipulating the overall probability of occurrence for a particular experimental condition (Coulson et al., 1998a; Gunter et al., 1997; Osterhout, McKinnon, Bersick, & Corey, 1996) or by presenting a proportion of the critical stimuli in a different font (e.g., all letters in upper case; Gunter & Friederici, 1999; Osterhout et al., 1996). For instance, Coulson and colleagues (Coulson et al., 1998a) compared the processing of two types of morphosyntactic violations (pronoun case violations, e.g. *the plane took *we to paradise and back*, and subject-verb agreement violations, e.g., *every Monday he *mow the lawn*) with nonanomalous sentences, also manipulating the probability of occurrence of the ungrammatical sentences. Ungrammatical structures thereby constituted either 20% or 80% of the materials within a presentation block. The data revealed main effects of grammaticality (suggesting a P600 to morphosyntactic violations) and probability (indicating a P300 to improbable stimuli). As would be expected if the P600 and P300 shared identical neuronal sources, the scalp topographies of the positivities elicited by ungrammatical stimuli and by improbable items did not statistically differ from one another. Besides, the authors found several interactions of the factors suggesting that they cannot have purely additive effects (which would provide evidence of the independence of P600 and P300). The probability effect was much more pronounced for trials containing a morphosyntactic violation relative to grammatical sentences. Moreover, both grammaticality and probability effects for the pronoun case violation were stronger than for subject-verb agreement errors, presumably because the former are more salient to native speakers of English. Based on these results, the authors concluded that the P600 is very likely a member of the P300 family, and that it reflects the occurrence of the “relatively rare linguistic event of ungrammaticality” (Coulson et al., 1998a, p. 45; see Gunter et al., 1997, for a similar conclusion).

These data contradicted the findings from Osterhout et al. (1996), who compared the processing of grammatical structures using sentences with subject-verb agreement violations (e.g., *the doctors believes…*) and sentences in which the correctly inflected verb was presented in upper-case letters. Their experiments evidenced that upper-case words elicited a pronounced P300 that peaked earlier, had a different scalp distribution, and exhibited a larger amplitude, as opposed to the P600 elicited by the ungrammatical sentences. The P300 to uppercase words also proved to be more affected by changes in task instructions (acceptability judgment vs. passive comprehension task) and probability of occurrence. Specifically, the P300 to the font manipulation was more pronounced with an acceptability judgment than when the sentences were simply read for comprehension. The late positivity for agreement violations showed only a modest effect in this direction. The probability manipulation had an impact on the upper-case condition only, with larger P300 amplitude for the improbable conditions.[19] Furthermore, when the physical and the agreement violations were combined in one condition, the corresponding P300 was larger in amplitude than either violation type alone, indicating that both violation types had additive, and thus independent, effects. These differences in all relevant characteristics of ERP components (i.e., time course, scalp distribution, and amplitude size) led Osterhout and colleagues to conclude that the P600 engendered by agreement violations is not a P300. Nevertheless, the authors emphasized that the important issue is "the *degree* of neural or cognitive similarity between two effects, rather than whether the syntactic positive shift is or is not a P3b or slow wave" (Osterhout et al., 1996, p. 521; highlighting in original), given that the P300 is sensitive to a multitude of different sources. From this perspective, the authors argued, the P600 and P300 might share a neuronal source that is sensitive to task relevance. However, this does not entail complete

[19] Similarly, Gunter et al. (1997) argued that the overall probability of an upper-case word in the stimuli of Osterhout et al. (1996) is remarkably lower than low-probability ungrammatical sentences, when each individual word within a sentence is regarded as a single event. This may have boosted the probability effect for these items, rendering its comparison to the syntactic anomaly highly asymmetrical.

identity of the two components, and the critical issue is then to disentangle the sources that are shared from those that are not.

Arguments in favor of each position have been discussed in detail by Osterhout & Hagoort (1999) and Coulson et al. (1998b) and will therefore not be reiterated here. There are, however, two aspects in this line of research that somewhat weaken the arguments on either side. Consider, for instance, the probability manipulation. Ungrammatical strings are probably less frequent in every day communication than grammatical strings, but it is not clear why experimental probability should exert any notable additional influence on language-related components. The probability manipulation in oddball experiments served indirectly to make the (rare) target, to which participants were supposed to respond, more salient. This is a crucial difference to psycholinguistic experiments in which subjects usually perform a task after every trial. For example, an acceptability judgment is made for every sentence type, regardless of whether its remaining characteristics are assumed to elicit a P600. One may therefore ask what the exact function of the probability manipulation is and what the (in)sensitivity of the P600 to the probability of occurrence actually implies (beside its supposed relation to the P300). The mixed evidence for the effectiveness of the probability manipulation in studies on syntactic processing is not really compelling for either proponents or opponents of the identity hypothesis stating that the P600 is a member of the P300 family; if anything, it demonstrates that the probability of occurrence does not reliably induce a P300 in psycholinguistic experiments – this corroborates the findings from the lexical domain (cf. chapter 1.3.1 and the discussion of the data from Kutas & Hillyard, 1980). In other words, one cannot exclude with utmost certainty that domain-general processes weigh experimental probability differently with linguistic stimuli. This is because the domain-general processes receive concurrent information from language-specific, task-relevant operations.

If the P600 elicited by grammatical violations is indeed a P300, indexing the low frequency of ungrammatical events, it is surprising that lexically infrequent words have been reported to elicit a robust N400 effect at all sentence positions (except at the final position in a sentence where they appear to elicit a late positivity) (Van Petten & Kutas, 1991; see Kutas & Federmeier, 2000, for review on the N400). Thus, frequency of occurrence or experimental probability does not seem to be a major factor in P600 generation. The unanimously accepted statement that the P300 is generated by numerous determinants (Johnson, 1993) needs to be taken more seriously, by investigating, for instance, whether the P600 is related to other (probability-independent) processes known to elicit a P300 (e.g., template matching; cf. Pritchard, 1981, for a collection of functional interpretations regarding the P300). This claim is, as such, a reformulation of Osterhout et al.'s statement cited above; but this reformulation becomes particularly relevant in view of recent attempts to reconcile the emergence of the P600 in fairly divergent linguistic contexts, many of which do not contain an ungrammatical event (Bornkessel-Schlesewsky & Schlesewsky, 2009b). As argued below, a P600 reflecting the well-formedness evaluation of a clause may well be considered identical to the P300.

A related concern is the comparison of syntactic violations with deviations in regular font type (Osterhout et al., 1996). Osterhout & Hagoort (1999), in defense of Osterhout et al.'s design, state correctly that a control condition eliciting an undisputed P300 is necessary. Nevertheless, Osterhout and colleagues were able to find a P300 to the font manipulation in their physical condition in which the verb was written in upper case. This is striking, given that the task was to judge the acceptability of the sentences, and that the physical properties of the verb were therefore entirely task irrelevant, which should have prohibited any significant P300 activity.[20] Moreover, the uppercase verbs were still grammatical and should have yielded high accuracy ratings. Yet, the behavioral data of all their experiments clearly show that the acceptability judgments for the physical condition were not higher than the judgments for the ungrammatical sentences. One explanation for this unexpected result may lie in the

[20] Their subjects were instructed to judge the sentences based on semantic coherence and grammaticality. Both of these dimensions are clearly independent of the font size manipulation.

fact that upper case font is a rather unfamiliar pattern in the English alphabetic script (cf. e.g., Juhasz, Liversedge, White, & Rayner, 2006, for evidence from readers' eye movements).[21] This speculation is supported by Osterhout et al.'s finding that uppercase verbs elicited a P3a, which usually indexes the recognition of infrequent and deviant, but task irrelevant, stimuli (cf. Courchesne et al., 1975; Polich, 2007). Provided that the upper case manipulation enables readers to initially identify these stimuli as unfamiliar, it could cause the P3a. The later occurring P3b to these stimuli may then be explained as reflecting the overall lower experimental probability of case deviations (see footnote 19) and that participants may have mistakenly interpreted this as a task-relevant cue, since only verbs carried information about syntactic or physical anomalies. It is important to note that, with the exception of their second experiment (in which the filler sentences served to induce different experimental probability values for the two violation types), Osterhout and colleagues used entirely nonanomalous fillers. Due to this and the fact that there was no linguistic violation other than agreement errors in Osterhout et al.'s study, the subjects in these experiments could have responded to any violation at the finite verb position by judging the sentence as unacceptable. The P300 to uppercase words then exhibited an earlier peak latency because the physical violation was easier to recognize than morphosyntactic violations, which required full word recognition first (instead of simple pattern matching). One may further contend that the inherent differences in task complexity were responsible for the different degrees of sensitivity the P300 and P600 exhibited to task manipulations such as probability and task instruction.

[21] In most studies, alternating case is used to explore the role of familiarity of orthographic patterns. However, the study from Perea & Rosa (2002) implies that uppercase words are also disadvantaged under certain circumstances. One may also argue that effects of case deviations are highly dependent on the script under investigation. For example, upper case is a productive means in German alphabetic script to distinguish nominals from other word categories, whereas the application of upper case is much more limited in written English. Thus, one could contend that the presentation of verbs in upper case, as in the Osterhout et al. study, is particularly unfamiliar in English.

Overall, it seems that a different experimental design is called for, one in which the P300 and P600 can be compared under normal linguistic processing conditions, i.e. without relying on physical parameters of the stimulus. Such an approach could, for instance, make use of how P300 latency varies with the ease of linguistic processing. The data reported by Haupt and collaborators (Haupt, Schlesewsky, Roehm, Friederici, & Bornkessel-Schlesewsky, 2008) suggest that the comparison of syntactically ambiguous structures, which are reanalyzed towards the dispreferred reading, with their ambiguous controls, which require no reanalysis since the initial reading is confirmed, may prove to be a viable approach. The authors set out to investigate the comprehension of subject-object ambiguities in German (cf. example 4).

(4) Example stimuli from Haupt et al. (2008)[22]

(a) subject-initial structures (preferred reading, no reanalysis)

… dass Bertram Surferinnen geärgert hat.

… that Bertram$_{\text{NOM/ACC/DAT.SG}}$ surfers$_{\text{NOM/ACC/DAT.PL}}$ annoyed$_{\text{ACC}}$ has$_{\text{SG}}$

'… that Bertram annoyed surfers'

(b) object-initial structures (dispreferred reading, reanalysis due)

… dass Bertram Surferinnen geärgert haben.

… that Bertram$_{\text{NOM/ACC/DAT.SG}}$ surfers$_{\text{NOM/ACC/DAT.PL}}$ annoyed$_{\text{ACC}}$ have$_{\text{PL}}$

'… that surfers annoyed Bertram'

The resolution of this syntactic ambiguity elicited a biphasic N400/P600 pattern for object-initial sentences such as (4b), with the N400 assumed to reflect reanalysis and the P600

[22] In German, the subject agrees with the verb in number and person, and it is marked with the nominative case. In examples (4a,b), the two pre-verbal NPs are fully ambiguous with respect to case assignment. When encountering the first case-ambiguous NP in a clause, the language comprehension system adopts a subject-initial reading; so, *Bertram* is initially assigned the subject function in (4a) and (4b). However, since the syntactic subject in German agrees with the verb, this initial parse has to be revised in the case of (4b) because the plural inflection on the verb requires the plural NP *Surferinnen* to be the subject. See Haupt et al. (2008) for further discussion and references. Abbreviations: NOM – nominative case; ACC – accusative case; DAT – dative case; SG – singular; PL – plural.

assumed to reflect a wellformedness evaluation of the structure (see Bornkessel-Schlesewsky & Schlesewsky, 2009b, for further explanation of the well-formedness P600). The data also showed an interesting effect of task instruction. One group of participants was to perform an acceptability judgment in the first of two experimental sessions and a comprehension question in the second, while a second group was presented with these tasks in reverse order. Interestingly, the verbs in the unmarked subject-initial structures gave rise to a P300, peaking between 400 and 500 ms post word onset, when the acceptability task was performed (cf. Haupt et al., 2008, Figure 6). The P300 found in subject-initial sentences extended to both the N400 and the P600 time windows. In the object-initial condition, there was no clearly visible P300 peak, but the ERP waves in the N400 and P600 time windows were significantly more positive going with the acceptability judgment.[23] Thus, one may speculate that the P300 to subject-initial sentences peaked earlier and was more visible in terms of amplitude because no reanalysis occurred (and, accordingly, no component overlap in the N400 time window). In the object-initial structures, the reanalysis process (as indexed by the N400) very likely preceded the evaluation process, so that the positivity engendered by the marked object-initial structures peaked after the N400 component. This suggests that the P600 reflecting wellformedness evaluation and the P300 reflecting general stimulus evaluation are closely connected, at least when task relevance and evaluation time for the stimulus are considered. The wellformedness positivity could then be considered a member of the P300 family.

Evidently, the conjecture about the functional equivalence of the wellformedness P600 and the P300 needs corroboration in future research, for example by using different types of ambiguous clauses (cf. Mecklinger, Schriefers, Steinhauer, & Friederici, 1995, who reported a P345 in response to object-initial relative clauses in German and judged it to be a P300).

[23] The P600 time window additionally exhibited an interaction between task and group which was due to the fact that the positivity was more positive under acceptability instructions (compared to comprehension instructions) only when this session preceded the session with the comprehension task.

Altogether, the heterogeneous findings from sentence-level investigations of the P300 suggest that a systematic reevaluation of word recognition studies and sentence-level experiments is necessary to fully understand whether or not linguistic stimuli engender a target-related P300. To give an idea of some of the experimental findings reported here, the combined results from Experiments 2 and 4 indicate that functionally different P300 types are affected differently by specific characteristics of input modality.

1.3.3 Summary

The data from studies on semantic priming suggest that the P300 is reliably elicited by linguistic stimuli when they allow for a binary decision. The most important precondition to be met is that the target category must not include more than one member, since a binary decision is impossible if more than one possible target has to be taken into consideration. A binary decision can be made, for example, in a lexical decision task, where words have to be discriminated from non-words. The decision on the lexical status of a letter string is dichotomous, as lexical properties of the string can only either be associated with a lexical entry or not. Since the actual meaning is irrelevant to the lexical decision, the precondition of the binary decision is perfectly met. A binary decision between the target category and the non-target category can also be made with certain semantic tasks such as the antonymy judgment. Here, the decision is made with respect to whether or not the target word is the antonym of the preceding prime word. In both the lexical decision task and the antonymy judgment task, the P300 emerges for every stimulus, but its onset latency varies according to the processes that have to be accomplished before the decision can be made. The P300 to antonyms, for example, occurs within the N400 time window because the perceived letter string perfectly matches the language processing system's predictions regarding word form and meaning. However, the P300 follows the N400 component for any non-antonym, since lexical access is necessary to determine that the perceived word is in fact not the expected antonym. In other words, the P300 in these semantic priming studies follows the actual linguistic processing. So, the P300 to antonyms occurs within the N400 time window because the processing of the highly predictable antonym proceeds smoothly. However, it follows the N400 in the case of non-antonyms because they are unpredicted, leading to the well-known cloze probability effect that enhances the N400 amplitude. Only after the meaning of these unexpected words is retrieved can the decision on their relation to the first antonym (the prime) be made. The P300 is thus considered an index of domain-general processes that make use of linguistic processing, but it does not reflect a specific linguistic operation.

While the reliability of experimental probability is severely cast into doubt in psycholinguistics experiments, the reported data provide further converging evidence for task relevance as the critical determinant of the P300. For example, the different onset latencies of the P300 to antonyms and non-antonyms are only visible with an antonymy judgment. When the same stimuli are used in a lexical decision task, there is no such variation in onset latency. The P300 in experiments investigating the processing of syntactic anomalies or syntactic ambiguities is just another example: its amplitude is enhanced with an acceptability judgment as compared to a simple comprehension task.

Overall, the existing data from semantic priming studies appear to reveal a much more stable pattern regarding the P300 and its antecedent conditions, whereas there is no consensus on the status of the P300 in the literature on syntactic processing. The first three experiments reported in this thesis will therefore employ the antonym paradigm to investigate the P300 in more detail. It was argued in the preceding section that the P300 to highly predictable antonyms might depend on when prelexical orthographic and lexical-semantic information become available in the processing of the antonym. It was further suggested that the temporal dynamics in natural reading might differ from both listening and RSVP, in that orthographic and lexical processing is temporally dissociated in the former modality, but not in the latter two. The goal of the next chapter is therefore to give a brief overview of the temporal dynamics in natural reading.

II Reading as an input modality

2 A brief overview of eye movements

2.1 Basic definitions

Visual information is gathered via a complex interplay of eye movements and periods in which the eye seems to be relatively stable. Eye movements are generally divided into one group that compensates for movements of the object and a second group that serves to compensate for body or head movements. Saccades and smooth pursuit movements belong to the first group. *Saccades* are rapid eye movements with a peak velocity of up to 1000°/ s (Joos, Rötting, & Velichkovsky, 2003). Their function is to move the fovea (see below) to the next object of interest. The saccadic system is spatially biased insofar as the distance between target and fovea determines the velocity of saccades (Kandel, Schwartz, & Jessel, 2000). An important characteristic of the saccadic system is that vision is suppressed during, and to some extent immediately before and after, a saccade (cf. Joos et al., 2003; Rayner, 1998). This phenomenon is called saccadic suppression. As saccades are not well suited to keep a moving object in foveal vision, the *smooth pursuit* system computes how fast a target is moving and keeps the fovea focused on the moving object. Consequently, smooth pursuit movements usually have a maximum velocity of 100°/s (Kandel et al., 2000).

Eye movements belonging to the second group compensate body and head movements. They are summed as *vestibular reflexes* that are activated when the eyes rotate in order to maintain the direction of vision and thereby a stable retinal image (Rayner, 1998). Specifically, the vestibular ocular reflex (VOR) keeps the object stable on the retina while we move our head, while the optokinetic reflex (OKR) compensates for slow movements of objects that would lead to a new image on the whole retina. Finally, the *vergence system* is used to adjust the eyes to objects moving in depth (Kandel et al., 2000).

The relatively stable periods between two eye movements that are necessary for visual information uptake, are termed *fixations*. The eyes are not entirely still in these periods, as fixations are usually characterized by so-called fixational eye movements, i.e. microsaccades, drifts, and nystagmus. During a fixation, the nystagmus (or tremor) is a constant miniature movement that is thought to prevent the neural habituation effects that cause image fading (Rayner, 1998). The eyes may also make small, slow drifts, which are sometimes related to noise in the oculomotor system. Microsaccades correct these small drifts, by bringing the eyes back to their original location. However, the issue whether microsaccades serve further functions as well has not been settled (cf. Horowitz, Fine, Fencsik, Yurgenson, & Wolfe, 2007; Laubrock, Engbert, Rolfs, & Kliegl, 2007; Horowitz, Fencsik, Fine, Yurgenson, & Wolfe, 2007).

Although the entire visual field of the eyes covers 100° (Joos et al., 2003), the effective visual field is much smaller. Visual information is first projected onto the retina, which can be subdivided into three adjacent regions that differ remarkably with regard to visual acuity. The highest acuity is obtained within the fovea (*fovea centralis*), i.e. within approximately 1-2° of visual angle. Acuity then drops off in the *parafoveal* region (3-5° of visual angle) and the *peripheral* region. Note that this decline in visual acuity affects visual search and reading tasks alike (Rayner, 1998). In reading, the visual field from which relevant information can be extracted is called the *perceptual span* (McConkie & Rayner, 1975; Henderson, Pollatsek, & Rayner, 1989). The size of the perceptual span is a function of the writing system, specifically, characteristics such as reading direction and orthographic density. The perceptual span is asymmetric in reading direction, i.e. it extends further to the right visual field for left-to-right writing systems, whereas its extension is larger to the left visual field for right-to-left reading systems (Rayner, 1998). For alphabetic writing systems such as English or German, it is often assumed that the perceptual span covers 4 character positions to the left of the fixation point and up to 15 positions to the right of the fixation and that approximately 3 letters may fall into foveal vision (cf. Rayner, 1998; Rayner & Pollatsek, 1989; Staub & Rayner, 2007). In a logographic writing system such as Chinese, the span extends only from 1 character to the left of the fixation point to a maximum of 2-3 characters to the right because linguistic information is spatially condensed (e.g., Chen & Tang, 1998; Inhoff & Liu, 1998; Rayner, 1998). In contrast to these divergences across writing systems, the *parafoveal preview* of upcoming words, which can then be selected as the next saccade goal, is a property shared by all writing systems.

2.2 Eye movements in reading

2.2.1 *When* and *where* decisions in reading

The eye movements relevant in reading are saccades and fixations; saccades move the fovea across the (usually immobile) text and visual information is gathered during a fixation. There is a remarkable variability across individuals as well as within one individual with respect to reading patterns (Rayner, 1998), but some generalizations are possible nonetheless. For example, the average fixation duration on a word is about 200-250 ms and the average saccade length ranges from 7 to 9 letters for alphabetic scripts (Sereno & Rayner, 2003; Staub & Rayner, 2007). The duration of an average saccade in reading is about 30 ms (Rayner, 1998). A short description of some basic eye movement parameters, which are often examined in eye movement research, is given in Table 1. The multitude of different eye movement parameters is one major reason why monitoring readers' eye movements in reading has become one of the dominant methods of data collection for psycholinguists. Eye movements in reading offer the high temporal resolution necessary to investigate which linguistic processes occur and when they occur during visual word recognition. Some experimental manipulations can affect fixation times, whereas others impact on saccadic measures, so eye movements are a multidimensional measure to track cognitive processes during reading. The relationship between cognitive processes during reading and eye movements is captured via the *eye-mind span* assumption, which posits that eye movements, particularly fixation durations, reflect on-going mental processes in a moment-to-moment manner (Rayner & Sereno, 1994). The eye-mind span hypothesis does not postulate that all cognitive processes are accomplished during the current fixation, as parafoveal preview benefits (cf. chapter 2.3.1) and spillover effects (i.e., processing costs spill over from word *n* to the following word *n* + 1) seriously challenge this claim. It does assume a tight coupling of eye movement behavior and cognitive processing, in that processing disruption due to the currently fixated word should directly influence the same fixation on it.

When eye movements are not used as a tool to investigate language comprehension, but are considered the dependent variable, the main research focus is usually on the *when* and *where* decision in reading, i.e. when the eyes move on and where they land (cf. Rayner, 1998; Rayner & Pollatsek, 1989). Both saccade execution and fixation duration have been shown to be sensitive to linguistic factors, with some linguistic factors primarily affecting saccade execution and others mainly affecting reading time.

Eye movement parameter	Definition
	Saccadic eye movements
Progressive saccade	Rapid eye movement following the canonical reading direction in a writing system
Regressive saccade	Rapid eye movement opposite to canonical reading direction
First-pass regressions out	Percentage of trials on which a word or region of text was left with a regressive saccade
Regressions into a region	Percentage of trials on which there was a regressive saccade into a region of text
Word skipping	A word is not fixated during first-pass reading
	Non-cumulative measures of fixation duration
Single fixation	One and only fixation on a word/ region of text
First fixation	The first of multiple fixations on a word (including single fixations)

Cumulative measures of fixation duration	
Gaze duration (first pass time for multi-word regions)	The sum of all first-pass fixations on a word (or region of text), before the eyes leave the word
Go past time	The sum of all first-pass fixations and any time spent before the word (or region of text), before the eyes leave the word (following the canonical reading direction)
Total time	The sum of all fixations on a word, including refixations after the eyes have already left the word
Descriptive terms to differentiate early and late processing time	
First-pass reading	The initial sequence of fixations on a portion of text (following the canonical reading direction)
Second-pass reading	The repeated sequence of fixations on a portion of text (including regressive interword saccades)

Table 1. Basic definitions of eye movement parameters in reading.

For example, orthographic complexity is one (but not the only) critical factor that determines the ease or difficulty of visual word recognition in alphabetic writing systems. When spaces between words in a sentence are removed, readers experience some difficulty in recognizing words because of the lack of clear-cut word boundaries. Accordingly, their progressive saccades tend to be shorter than average saccade length, and land slightly before the optimal viewing position on a word (which lies somewhat to the left of the centre of the word) that guarantees perceptually best information uptake (see e.g. Perea & Acha, 2009; Pollatsek, & Rayner, 1982; Rayner & Pollatsek, 1996; Rayner, Fischer, & Pollatsek, 1998). Word length is another example of how spatial information alters saccadic parameters. Very short words (with less than four letters) are very unlikely to be fixated during first-pass reading because they are already processed on the immediately preceding fixation. That is, very short words are skipped because they have already been recognized in the parafovea (Drieghe, 2008; Drieghe, Rayner, & Pollatsek, 2005; Kliegl, Grabner, Rolfs, & Engbert, 2004; Kliegl, Nuthmann, & Engbert, 2009). Thus, there is much support for the assumption that spatial orthographic information guides saccades and the selection of saccade targets.

This suggests that one main point behind eye movement control in reading is to recognize upcoming words and select one of these as the next saccade target, thereby providing new information for linguistic processing. This input-driven view on the role of eye movement control predicts that saccadic parameters should differ as a function of the writing system, especially its orthographic complexity, under investigation. Accordingly, writing systems with a higher orthographic complexity should generally exhibit a different pattern than less complex ones. This is indeed what one finds with the logographic script of Chinese. Although the average fixation duration is similar to alphabetic scripts, the perceptual span is smaller and readers make smaller saccades (about 2-3 characters in length) because words are encoded in a smaller spatial area as compared to alphabet languages. The spatial density of linguistic information is thus higher than for alphabetic scripts (e.g., Chen, Song, Lau, Wong, & Tang, 2003; Inhoff & Liu, 1998).

Non-orthographic factors such as syntactic complexity may also affect saccadic parameters. In these cases, it is useful to distinguish progressive saccades from the regressive saccades targeting previously read words. Several studies have provided evidence that syntactic reanalysis, i.e. the resolution of syntactic ambiguities towards the unpreferred reading, leads to a higher proportion of regressive saccades (e.g., Frazier & Rayner, 1982; Mitchell, Shen, Green, & Hodgson, 2008), as compared to control sentences with the preferred reading.

More importantly for the present experiments, fixation durations are sensitive to all kinds of linguistic manipulations, i.e. word-level factors (e.g., orthography, phonology, frequency of word meaning, word predictability), sentence-level and discourse-level factors (e.g., syntactic ambiguity, referential linking). Fixation durations generally increase when words are presented in an unusual print format or embedded in a deviating visual environment (e.g., alternating case, words surrounded by random symbols; cf. Perea & Achea, 2009; Pollatsek, Raney, Lagasse, & Rayner, 1993). Fixation durations are also likely to increase when a word contains atypical or unfamiliar letter clusters (Lima & Inhoff, 1985; White, 2008). Furthermore, it has been shown that the fixation duration on a word correlates inversely with the word's frequency of occurrence and predictability from prior context. The less frequent or predictable a word (or a particular word meaning for homonyms), the longer the fixation duration (e.g., Calvo & Meseguer, 2002; Duffy, Morris, & Rayner, 1988; Kliegl et al., 2004; Sereno, Pacht, Rayner, 1992). Prolonged fixation times also occur with syntactic reanalysis and, for example, words that are incongruent with the preceding discourse (e.g., Camblin et al., 2007; Frazier & Rayner, 1982; Staub, Rayner, Pollatsek, Hyönä, & Majewski, 2007).

2.2.2 Word identification and semantic priming

There is robust electrophysiological evidence that word recognition is facilitated when a word is preceded by a congruent context or a lexically associated prime word; a decrease in N400 amplitude indexes this facilitation (see chapter 1.3.1 above). Corresponding results on the semantic priming effect have also been found in eye movement studies, showing that fixation time on the target word following a lexically associated prime is reduced. If the preceding context makes the target word highly predictable, there are both a reduction in fixation time and an increase in the probability of being skipped (e.g. Ashby, Rayner, & Clifton, 2005; Balota, Pollatsek, & Rayner, 1985; Calvo & Meseguer, 2002; Ehrlich & Rayner, 1981; Kliegl et al., 2004; Morris, 1994; Pynte, New, Kennedy, 2008; Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006; Rayner & Well, 1996; Vainio, Hyönä, & Pajunen, 2009).

Morris (1994), investigating lexical relatedness, found that the first fixation and the gaze duration on a target word decreased when it was preceded by lexically related words, provided these formed a context that was congruent with the target word (e.g., *the barber trimmed the mustache this morning* vs. *the person saw the mustache this morning*, where *mustache* is the target word). Based on the detailed data collected in two experiments, Morris was also able to show that lexical relatedness between words alone cannot produce facilitated word recognition in natural reading; lexical relatedness and contextual congruency must converge to produce this effect. This finding has been replicated, showing that lexical relatedness in reading yields stronger effects when the prime words occur in a context that additionally primes the target word (Calvo & Meseguer, 2002). This suggests that intralexical priming may not reliably influence eye movement parameters by itself, but is in need of extra support from contextual predictability (see, however, Vainio et al., 2009, for findings of lexical predictability). The following paragraphs therefore focus on contextual predictability in reading.

Rayner and Well (1996) monitored readers' eye movements when reading sentences with target words of high, medium, and low predictability (e.g., *the woman took the warm cake out of the oven/ stove/ pantry*). The results were fairly clear. Low-constraint targets (*pantry*) yielded longer reading times in the first fixation, gaze duration, and total time measures. There was no difference between high-constraint (*oven*) and medium-constraint targets (*stove*) in these measures. High-constraint targets did differ from medium and low-constraint targets in that they were skipped more frequently than the targets in the other conditions, which had the same skipping probability. The authors concluded that a high degree of contextual predictability is necessary to produce word-skipping effects, whereas a moderately constraining context is sufficient to reduce fixation durations. They presumed further that more parafoveal information is obtained for highly predictable words than for unpredictable words, which is likely to increase their skipping probability.

This latter assumption was first proposed by Balota and colleagues (Balota et al., 1985). They set out to investigate how parafoveal information is processed when the context makes the occurrence of a particular target more probable than a semantically related, but less predictable, word. Balota et al. examined sentences like those given in Table 2. The sentence fragment made the target *cake* far more predictable than *pies*. The parafoveal preview for each of these target words was manipulated by presenting the same word in the parafovea (*cake – cake*), a visually similar non-word (*cahc – cake*), a semantically related word (*pies – cake*), a visually dissimilar non-word (*picz – cake*), or a contextually inappropriate word (*bomb – cake*).

Sentence Frame **Since the wedding was today, the baker rushed the wedding ______ to the reception.**				
identical preview	visually similar	semantically related	visually dissimilar	anomalous
Parafoveal previews for the high-predictable word *cake*				
cake	*cahc*	*pies*	*picz*	*bomb*
Parafoveal previews for the low-predictable word *pies*				
pies	*picz*	*cake*	*cahc*	*bomb*

Table 2. Example stimuli from Balota et al. (1985).

The eye-contingent boundary paradigm was used to replace the initially presented parafoveal preview with the correct target word (see Rayner, 1975, for first application of this eye-contingent display technique). The invisible boundary was between the second-to-last and last letters of the precritical word, e.g. between *n* and *g* in *wedding* above (see Table 2). When the readers made a saccade toward the target word and their eyes crossed this boundary, the parafoveal preview was replaced with the correct target word. Since the display change was accomplished during the saccade, readers were not aware of it (due to saccadic suppression, see chapter 2.1). So, the initial sequence of fixations on the target word gives insights into how much parafoveal preprocessing of the previews took place.

One major finding of this study was that the orthographic characteristics of the word in parafoveal preview influenced the duration of the first fixation. Specifically, fixation time increased when a visually dissimilar item was initially presented in the parafovea (as reflected in the mean fixation time of the semantically related, the visually dissimilar, and the anomalous conditions), as opposed to visually similar items (as reflected in the mean fixation time for the identical and visually similar preview conditions). The predictability of the correct target word did not appear to influence the first fixation. Identical and visually similar non-word previews for highly predictable target words also increased the probability of the target word being skipped. This suggests that basic orthographic information is used not only to locate the next saccade goal but also to initiate early word recognition processes, which yields a processing facilitation once the target is fixated.

The gaze duration measure, on the other hand, did reveal that predictability and orthographic similarity interacted, indicating that probably more than orthographic information from parafoveal previews is processed under specific circumstances. Specifically, the benefit of orthographically similar parafoveal previews (as reflected in the mean fixation time of the identical and visually similar conditions) was greater for high-predictable words than for their low-predictable counterparts. Also, the comparison between identical and visually similar previews for high-predictable target words (*cake* – *cake* vs. *cahc* – *cake*) revealed that the latter led to slightly increased gaze duration. Numerical trends further revealed that the semantically related condition for low-predictable target words (*cake* – *pies*) increased gaze duration relative to the visually dissimilar conditions (*cahc* – *pies*), whereas neither condition differed for high-predictable target words (*pies* – *cake* vs. *cahc* – *cake*). This led the authors to conclude that lexical processing is progressed to a large degree when the highly predictable target word was used as the semantically related preview for a low-predictable word (see also Table 2 above). Since lexical processing was already well on its way when the display replaced the semantically related preview (i.e., the high-predictable word) with the low-predictable target, this processing routine had to be cancelled. This processing disruption was seen as a result of the inconsistency between the information acquired from the parafovea and information from the fovea. The same argument clearly holds for the contrast between identical and visually similar previews for high-predictable words. The facilitation from identical previews resulted from the fact that lexical processing of the identical preview could continue when the target was fixated, since the foveal information did not contradict the previously processed information. The visually similar preview, in contrast, contradicted the orthographic input from the foveal word.

Overall, Balota and colleagues conclude that orthographic information is acquired from the parafoveal region of text and, more importantly, that parafoveal preprocessing is greater when words are highly constrained by the preceding context. The authors remain somewhat indecisive with respect to whether the parafoveal processing of high-predictable previews might include semantic processing. Their arguments seem to suggest that it is very limited, if present at all. However, the length of target words in their study ranged from 4 to 8 letters (with about 5 letters on average), so one cannot definitely exclude the possibility that part of the reported effects are indeed driven by full parafoveal recognition of short targets. As a matter of fact, the skipping data across all items suggest that a significant part of word information must have been processed parafoveally, so that the high-predictable target is skipped over. If indeed more than just orthographic information from the parafovea is acquired, the first fixation on the following target word may not be an adequate index of parafoveal preprocessing. There is an on-going debate as to whether parafoveal processing entails semantic processing, although it is generally agreed upon that the modulation of the fixation times on the pre-critical word provides the best testing ground for possible effects of semantic preprocessing in the parafovea. The next chapter will therefore focus in more detail on parafoveal processing.

2.3 Allocation of attention in reading: serial or parallel?

2.3.1 Parafoveal preview benefit and parafoveal-on-foveal effects

This chapter aims to give a short overview of the current data on parafoveal processing and the issue of whether parafoveal processing is limited to prelexical orthographic processing, as suggested in the Balota et al. (1985) study. The depth of processing in the parafovea has important implications for studying ERPs in natural reading. If the data collected so far unequivocally supported semantic processing in the parafovea, one would be likely to find parafoveally induced ERP correlates of lexical processing. However, parafoveal semantic processing also entails that several words are processed in parallel, so that attention is presumably allocated to more than one word at a time in these cases. This in turn should have converse effects ERP components collected during reading, since attention deprivation negatively affects the N400 priming effect and the P300 (see chapter 1.3.1 above).

There is general agreement that words are best perceived when they are in foveal vision, i.e. directly fixated, because the high visual acuity allows for complete information extraction from the fixated word. Conversely, the drop of visual acuity in the parafovea somewhat limits information uptake. Despite this restriction on information availability, the parafoveal preview of upcoming words appears to be a major determinant of unimpeded reading. For instance, reading when only the currently fixated word is available is much slower than when the next word or the initial sequence of its letters is visible (Rayner, Well, Pollatsek, & Bertera, 1982). One obvious reason is that a saccade to the next word has to be programmed – the oculomotor system is in need of a target location, which is selected from the parafovea. Also, when useful information from a word is obtained parafoveally, the first fixation on that word is about 20-50 ms shorter than when an incorrect preview is presented (see Rayner, 1975, for first demonstration and Rayner, 1998, for further references). This is called the *parafoveal preview benefit* in reading and several studies have indicated that the parafoveal acquisition of prelexical orthographic and phonological codes is most likely responsible for this benefit (see e.g. Ashby, Treiman, Kessler, & Rayner, 2006; Chace, Rayner, & Well, 2005, for phonological preview effects and Rayner et al., 1982; Lima & Inhoff, 1985; Rayner, Balota, & Pollatsek, 1986; White, 2008, for orthographical preview effects; see Rayner, White, Kambe, Miller, & Liversedge, 2003, for further references).

Interestingly, the parafoveal preview benefit does not appear to be a uniform phenomenon. It is not present for every parafoveal word that is perceived, nor insensitive to the wider linguistic context. The first claim is of importance to models of eye movement control in reading that make different assumptions about how attention is allocated across words (see chapter 2.3.2). If attention is distributed across several words, there may be parafoveal preview benefits from more than just the word to the right of the current fixation. Yet, many studies failed to find a parafoveal preview benefit for word n+2, i.e. two words beyond the currently fixated word (Kliegl, Risse, & Laubrock, 2007; Rayner, Juhasz, & Brown, 2007; but see Wang, Inhoff, & Radach, 2009, for different results). It was reasoned that the lack of these preview effects could be due to either a strictly serial allocation of attention, or the fact that the preview benefit can only be obtained from the saccade target (McDonald, 2006). To test these competing explanations, McDonald (2006) carried out an experiment in which the boundary technique was used to initiate a display change that occurred either one or two saccades prior to fixation on word n+1. In the mid-word boundary condition the boundary triggering the display change was set somewhat to the left of the center of word n, whereas in the control condition (post-word boundary) the boundary was set after the final letter of word n. Since word n was always a 9- or 10-letter word, refixations were very likely and thus, the mid-word boundary most often fell between the first fixation on that word and the following refixation. The main result was that there was a robust preview benefit for the post-word boundary, with shorter gaze and first fixation durations for word n+1, whereas the numerically visible preview in the mid-word boundary condition was statistically unreliable. This finding led McDonald to conclude that a preview benefit is obtained from word n+1 only when it is the saccade goal rather than from word n+1 per se (or, more generally, from all words in the perceptual span).

With respect to the latter claim made above, the data from Balota et al. (1985) already suggested that the parafoveal preview depends on contextual constraints, such that more parafoveal information from the upcoming word is acquired, when it is congruent with the preceding context. Henderson and Ferreira (1990) further investigated how the difficulty of foveal processing interacts with the acquisition of parafoveal information. They manipulated both the availability of useful parafoveal information (identical word preview vs. visually similar or dissimilar non-words) and foveal load (implemented as lexical frequency or syntactic complexity in two separate experiments) using a boundary technique. While they did not find any influence of the parafoveal word *n*+1 on the first fixation or gaze duration on word *n*, their data clearly indicated that the properties of word *n* affected the parafoveal preview of word *n*+1. Specifically, when word *n* was a low-frequency word, there was no difference as to whether an identical or visually (dis)similar parafoveal preview was presented. However, when word *n* was a high-frequency word, the identical and visually similar previews led to shorter gaze durations on word *n*+1 than the visually dissimilar preview. The same pattern of results was replicated when word *n* increased foveal processing load by disambiguating a syntactically ambiguous sequence towards the dispreferred reading. The authors concluded that little or no parafoveal information is acquired under high foveal load conditions.

The previous paragraph has dealt with the impact foveal processing has on parafoveal processing. A more controversial matter in current eye movement research is whether parafoveal processing can also influence foveal processing, and whether these effects are prelexical or lexical in nature (see Rayner et al., 2003, for an overview). The distinction between prelexical and lexical parafoveal-on-foveal effects is important because only the latter is relevant to the question whether more than one word can undergo lexical (semantic) processing at a time, or how attention is allocated during reading, as attention serves to bind all word features to a cohesive word representation (see Reichle, Pollatsek, & Rayner, 2006, for discussion). The debate over parafoveal-on-foveal effects is closely tied to computational models of eye movements in reading. The majority of the collected data has been taken as evidence supporting one particular class of computational models, while disconfirming the other. According to serial attention shift (SAS) models, usually exemplified by the E-Z Reader model (see Reichle, Warren, & McConnell, 2009, for a recent version of the model), attention in reading is strictly serial, so that only foveal words are processed lexically as only they receive focused attention. According to guidance by attention gradient (GAG) models, usually exemplified by the SWIFT model (Engbert, Nuthmann, Richter, & Kliegl, 2005; Richter, Engbert, & Kliegl, 2006), attention is distributed gradually within the perceptual span, so that lexical processing is enabled for more than one word at a time (see chapter 2.3.2 for further details and Reichle, Rayner, & Pollatsek, 2003, for a comparison of different models; for discussion of serial versus parallel accounts of attention allocation see Reichle, Liversedge, Pollatsek, & Rayner, 2009).

While there are data indicating that unusual or infrequent orthography can induce prelexical parafoveal-on-foveal effects (Starr & Inhoff, 2004; White, 2008), the data with regard to lexical parafoveal-on-foveal effects are less consistent. Some have reported lexical parafoveal-on-foveal effects, which show that higher-level semantic properties of the to-be-fixated words (word *n*+1 or *n*+2) have a significant influence on fixation times of the currently fixated word (word *n*) (e.g., Kennedy, 2008; Kennedy & Pynte, 2005; Kliegl et al., 2006; Kliegl et al., 2007; Kennedy, 2000; Pynte et al., 2008). Others, however, have reported the complete absence of such effects (e.g., Angele, Slattery, Yang, Kliegl, & Rayner, 2008; Drieghe, Rayner, & Pollatsek, 2008; Rayner et al., 2007).[24] The overall picture becomes

[24] These data are based on the English script, an alphabetic writing script. The reported explanations thus hold for the alphabetic writing system only. Note that different findings have been reported for the logographic writing system in Chinese, which suggests that parafoveal semantic processing may differ as a function of the writing system under investigation (e.g., Yan, Richter, Shu, & Kliegl, 2009; Yang, Wang, Xu, & Rayner, 2009). Although there is a growing interest in the collection of eye movement data from Chinese, there is virtually no systematic comparison across writing systems that uses similar stimuli to ensure maximum comparability. Thus, data from Chinese reading is used to undermine generalizations drawn from data on alphabetic writing systems, without a careful comparison of script-based similarities and differences (e.g., in terms of visual complexity). This approach

increasingly complex when considering that most of the parafoveal-on-foveal effects were found in corpus studies, whereas experimental designs failed to produce any. Moreover, the issue of mislocated fixations has not yet been settled either. According to Drieghe and colleagues (Drieghe et al., 2008), parafoveal-on-foveal effects may result from foveal processing. Specifically, they provided experimental evidence that saccades targeting the critical word fell short of and landed on the pre-critical word, thereby producing such parafoveal-on-foveal effects. Their results were challenged, however, because the effect magnitude in their study was far beyond the effect size that is usually reported for parafoveal-on-foveal effects (about 10 ms) (Kennedy, 2008).

Furthermore, some of the experimental studies on parafoveal-on-foveal effects compared the processing of words with that of illegal non-words (e.g. Kliegl et al., 2007; Angele et al., 2008). In light of the electrophysiological finding that non-words, in contrast to pseudowords, can be recognized by their deviant prelexical information (cf. Holcomb & Neville, 1990), generalizations about lexical parafoveal-on-foveal effects appear somewhat unwarranted when based on data from non-words. If parafoveally perceived non-words increase the current fixation duration, this disadvantage for non-words could occur because the language processing system registers that some of the orthographic input does not conform to that language's grapheme-to-phoneme mapping rules. Semantic processing is no necessary precondition to this effect. There is also no clear pattern as to when pseudowords are used instead of non-words. Starr & Inhoff (2004) examined to what extent orthographic information acquired from the parafovea influences the current fixation. They replaced the final word in a sequence of three associated words (e.g., *steaming bath water*) with an illegal non-word (e.g., *bvkln*), a pseudoword, (e.g., *wafar*), or a contextually incongruent word (e.g., *night*). The first fixation and gaze duration measures on the second word (*bath*) registered

presupposes implicitly that the underlying mechanisms in reading are identical across writing systems, which, I believe, is a claim that still awaits empirical investigation.

increased fixation durations for non-words as opposed to the identical preview of the target word (the control condition). Pseudowords showed reliable increases in fixation durations relative to the controls only for a subset of trials – these trials were chosen to exclude, among other things, the possibility that mislocated fixations on the target word could account for the parafoveal-on-foveal effects. While these findings demonstrate that unexpected or deviant orthographic information has an impact on the current fixation, the influence of the lexical status of parafoveal words is less clear, as only the analysis on a subset of trials revealed reliable effects for pseudowords. However, the general pattern can still be accounted for by considering the strong lexical association between the three critical words. For instance, when the word *bath* in the sequence *steaming bath water* is reached, the language comprehension system has most likely adopted very specific predictions about the next word, including word form (see e.g., Laszlo & Federmeier, 2009, for similar findings from ERPs). A deviant word form in the parafovea is then evaluated as to whether it matches this prediction or not. Both pseudowords and non-words are equally unpredicted and therefore yield processing costs which can arise solely from the orthographic mismatch. The effect for pseudowords may then be smaller (or even absent) because they share more initial letters with the predicted word form (see Balota et al., 1985, for similar explanations).[25]

So, while there is compelling evidence that parafoveal information is actively acquired during reading, and that it interacts with the allocation of attention and the immediate

[25] Note that contextually incongruent words yielded no reliable effects, which this account failed to predict. However, these items produced a remarkably different result than either non-words or pseudowords; for example, when the parafoveal preview was an incongruent word, fixation time on the target was in fact shorter than when the parafoveal preview was identical with the replaced target. Experiment 2 revealed that unpredicted and unrelated target words (e.g., *black is the opposite of nice*) yielded shorter gaze durations on the pre-target region (*the opposite of*). This seems to suggest that contextually incongruent words entail a fundamentally different influence on eye movement than any other deviant form. One cannot decide on the grounds of these two studies whether lexical processing is necessarily involved in the eye movement results. Nonetheless, the electrophysiological data from Experiment 2 demonstrate that the eye-movement results are associated with orthographic processing, without lexical processing.

linguistic context, there is currently no consensus regarding the existence of lexical parafoveal-on-foveal effects in reading. Prelexical information is the linguistic information type most commonly reported in the literature as being extracted from the parafovea. There is, however, evidence that the combined measurement of ERPs and eye movements in reading can reveal clear electrophysiological parafoveal-on-foveal effects that are absent in the eye-movement record (cf. Baccino & Manunta, 2005; Simola et al., 2009). Importantly, these electrophysiological parafoveal-on-foveal effects are lexical in nature, which makes Experiment 2 particularly interesting. Similar to the stimuli used in the Balota et al. (1985) and Starr & Inhoff (2004) studies, the antonym paradigms provides a high-constraint context that should facilitate the acquisition of parafoveal information. However, while one can more or less confidently conclude that electrophysiological correlates of parafoveal processing should be detectable, particularly for the last fixation before the critical item is reached, one cannot predict with the same confidence whether these effects are lexical or prelexical in nature and whether the eye-movement record shows corresponding results.

2.3.2 Allocation of attention in reading

As indicated in the previous chapter, the question of whether information processing is serial or parallel is tightly linked to the issue of attention allocation, which is controversially debated by proponents of serial attention (SAS) models and guidance by attention gradient (GAG) models of eye movements in reading. In the following, the basic assumptions of the E-Z Reader model (an SAS model) and the SWIFT model (a GAG model) concerning the allocation of attention will be described. The diverging core assumptions should lead to different predictions regarding the processing routines in reading. Note that neither model explicitly attempts to account for ERP data, and that the ERP data obtained from the experiments reported in this thesis are not taken to invalidate the core assumptions of these models. Rather, the rationale of the present chapter is to derive predictions regarding the possible allocation of attention in reading and implications for ERP components. The preceding chapters have shown that eye movement data are too heterogeneous and do not provide compelling evidence in favor of either serial or parallel allocation of attention in reading, rendering predictions, based on empirical findings, for the present experiments rather ad hoc. However, predictions about the P300 and N400 in reading can be derived from each model, as they are coherent frameworks that allow the formulation of very specific, yet opposite predictions (cf. chapter III). As the model assumptions on the allocation of attention are most important, the descriptions below are narrowed down to this aspect of each model.

The E-Z Reader model basically assumes that attention in reading is allocated in a strictly serial manner, and that saccade programming and execution are decoupled from shifts of attention. During early visual processing, low-spatial information (used to select the next saccade target) and high-spatial information is gathered. The high-spatial information is passed on to the lexical processor, where it undergoes two distinct processing stages. During the first stage (L_1), orthographic information is processed to accomplish the "familiarity check" (Reichle et al., 2003, p. 452) for a word form. The initial encoding stage is extended when, for example, infrequent or unfamiliar orthographic forms are perceived (e.g., White, 2008), or when the layout of the text makes it rather difficult to detect word units (e.g., when the physical contrast between text and background is reduced; cf. Reingold & Rayner, 2006; Drieghe, 2008). The initiation of the L_1 stage triggers the oculomotor programming of the saccade to the next target word. When L_1 indicates that full identification of the next saccade goal is about to occur (e.g., when a word is highly predictable from the prior context), the saccade programming for this word can be cancelled and redirected to a new saccade target, so that the initial goal is skipped. The second stage in word encoding (L_2) is initiated after the completion of L_1, and involves lexical-semantic processing. This occurs with foveal inspection, as only directly fixated words are within the focus of attention. Attention is also assumed to be the crucial precondition for the establishment of a whole-word representation

(Reichle et al., 2006). The focus of attention remains on the fixated word until lexical access is completed. Attention shifts to the next word (i.e., the next saccade goal) only shortly before saccade execution, which constitutes the parafoveal preview benefit. As a consequence, if the foveal word induces processing costs, attention will shift considerably later to the next word, as the completion of L_2 takes more time, and less information is acquired from parafoveal preview (resulting in spillover effects on word n+1). Lexical processing thus depends essentially on focused attention and attention is strictly serial, inasmuch as only one word (or put more generally, one discrete object) can reasonably be processed under focused attention conditions at a time.

This model accounts for prelexical parafoveal-on-foveal effects by postulating the parallel processing of two or more words within the early L_1 stage of word recognition. This is not regarded as weakening the claim of strictly serial attention because such parafoveal orthographic processing is assumed to be pre-attentive. Specifically, the physical properties of parafoveal words are assumed to "pop out" (if deviant in some way) and lead to unattended processing (e.g., Reichle et al., 2006, p. 6). On this account, reading equals focused attention on fixated words, while parafoveal words receive little or no attention. This implies that components such as the P300 or the N400 semantic priming effect, which depend on attention to the stimulus, are less likely to emerge for parafoveal words.

In contrast to the serial E-Z Reader model, the SWIFT model assumes that more than one word is attended to in reading and that a maximum of four words can be processed in parallel, the majority of which fall within the attentional gradient to the right of the fixation (Engbert et al., 2005). Attention is thus distributed across several words in the perceptual span. The words are only distinguished by their degree of activation, which determines the depth of processing. Specifically, the model presupposes a dynamic activation field, conceived of as a map of visual salience, so that words vary as to the level of activation. Linguistic processing is defined as a lexical processing rate, which is a function of both the distance of the word from the current fixation and its progress in one of the two stages of lexical processing. On the one hand, processing speed is limited by visual acuity restrictions, so that processing is faster with more accurate visual information, i.e. the lexical processing rate for a word is higher the closer it is to the current point of fixation. On the other hand, activation varies according to the processing in each of the two assumed processing stages. That is, activation for a word increases during the first phase of preprocessing and decreases during the second phase of lexical completion. The activation of a word tends to zero before its processing has begun and when lexical processing is completed. Since saccade targets are presumably more salient than currently fixated words, the lexical processing rate (*viz*., activation) for them is higher than for the currently fixated word or, more generally, for words that currently undergo lexical completion. These stages of preprocessing and lexical completion are not identical to parafoveal and foveal information processing. Rather, preprocessing is thought to primarily define possible saccade targets, which may fall into the foveal or parafoveal regions of the perceptual span. Both prelexical and lexical parafoveal-on-foveal effects are conceivable within this model, which follows from the assumption that attention is distributed across several words and that these can be processed – at any processing stage – in parallel.

Furthermore, saccades are decoupled from word identification processes by virtue of their autonomous programming. The only direct instance where word identification may interfere with saccade execution is when increased foveal load inhibits saccade programming – if the lexical completion stage of a word signals difficult processing, the execution of a saccade to the next word can be delayed. Another indirect influence of word properties on saccade programming involves the way a word’s predictability determines its likelihood of being the next saccade goal. The processing rate for highly predictable words declines during the stage of preprocessing, as long as it has not yet been fixated. The level of activation thus rises at a slower rate than with unpredictable words. Importantly, upcoming words compete with one another to be the next saccade goal. This competition is resolved by activation levels, with the more activated words representing the likely saccade goals. Highly predictable words are therefore less likely to be chosen as the next saccade target, and more likely to be skipped. When, by contrast, the highly predictable word is at the stage of lexical completion, its processing rate increases according to its predictability, which yields full recognition faster than with low-predictable words. In sum, these assumptions imply that parafoveal words can evoke ERP components related to linguistic processing, provided that their level of activation is sufficiently high. However, predictions concerning high-predictability words, like antonyms in the antonym paradigm, are less clear-cut. If the parafoveal processing of antonyms overlaps to a large degree with the proposed stage of preprocessing, they will be less activated than unpredicted words completing the sentence fragment *X is the opposite of Y* and accordingly, they should show no particular ERP effect. If, however, the stage of lexical completion has already begun during parafoveal processing, the increased lexical rate in this processing stage should facilitate the emergence of parafoveal ERP effects to antonyms.

In sum, both models differ fundamentally in their core assumptions about (i) how attention is allocated in reading and (ii) when lexical processing takes place. These distinctions are of relevance for the comprehension of the antonym stimuli because they make different predictions regarding the emergence of ERP effects in response to parafoveal words. While the serial E-Z Reader model does not predict any significant electrophysiological correlates to parafoveal processing, the SWIFT model clearly does predict such effects. Under some circumstances, some of these effects may be expected particularly for highly predictable antonyms.

2.4 Summary

To sum up the preceding sections, it is evident that low-level factors (i.e., spatial properties of text) appear to mainly influence the *where* decision in reading, while some higher level linguistic factors (e.g., orthography, semantics, syntax) affect both the *where* and *when* decisions in reading. The varying impacts of linguistic information on the *when* and *where* decisions also suggests that linguistic information is passed on to the oculomotor system in rather different ways, depending on the actual type of information. Studies on eye movements in reading have revealed some very interesting characteristics which separate natural reading from both the RSVP method and the auditory modality. Table 3 shows the important differences between the three presentation methods in order to clarify how reading differs most from the other two methods.

	Allocation of attention	**Information processing**	**Input speed (wpm)[a]**	**Stimulus quality**
Auditory modality	focused	serial	moderate	high
Rapid serial visual presentation (RSVP)	focused	serial	slow	high[b]
Natural reading			fast	
Parafovea	divided	parallel		low
Fovea	focused or divided[c]	serial or parallel[c]		high

Table 3. Characteristic differences between input modalities. [a] words per minute; [b] with foveal presentation; [c] see the text for explanation.

In most cases, the average duration of the first fixation on a word extends to no more than 250 ms, which is remarkably shorter than what is allowed in RSVP experiments. Even if the presentation of a word is adjusted according to the average fixation duration it would receive, the use of SOAs and ISIs in RSVP clearly exceeds the average inspection time during reading. For example, the duration of an ISI of 200 ms is considerably longer than the average duration of a saccade between two adjacent words, which lasts about 30 ms. Word skipping is another case in point here, as all words are presented in an RSVP design, even the ones that might be skipped in reading. This contributes greatly to the faster input speed in natural reading. Average fixation duration is also shorter than spoken input, given that the average length of a syllable is about 200 ms (cf. Poeppel, Idsardi, & van Wassenhove, 2008, and Rayner & Clifton, 2009, for references). Thus, information processing is faced with considerably faster input speed in reading than in any other modality. The parameter of stimulus quality yields the same implications as described above. Stimuli in unimpaired listening and in foveal vision (either in RSVP or natural reading) offer a fairly high degree of perceptual acuity, whereas parafoveal stimuli in reading are a marked exception to this.

The RSVP method also resembles the auditory modality more than natural reading when the remaining two factors are considered. In the auditory modality and with RSVP, attention is focused on the currently perceived stimulus and only one stimulus is perceived at a time (leaving dichotic listening and the insertion of visual or auditory white noise aside). This essentially renders information processing serial. Natural reading, on the other hand, differs from listening and the RSVP method in each of these dimensions, mainly due to the parafoveal preview. This preview of upcoming words, which is necessary for saccade programming, allows the division of attention across foveal and parafoveal words in reading. This causes information processing in reading to be, to some noticeable extent, more parallel than in listening or in RSVP. Such parallel processing follows from the parallel information extraction from both parafoveal and foveal parts of the retina. Importantly, the literature on parafoveal processing in reading suggests that the parallel information acquisition may not extend to all levels of linguistic processing. There is converging evidence that prelexical orthographic processing is partly parallel, as orthographic features of several words can be processed at a time. However, parafoveal semantic processing is a highly debated topic because of the heterogeneous empirical data concerning the serial vs. parallel nature of lexical processing. A related issue concerns the allocation of attention to foveal and parafoveal words. While some contend that lexical processing is serial and that focused attention on the fixated word is necessary for full word identification, others assume that lexical processing also occurs under conditions of divided attention and is therefore distributed across several words in parallel. These considerations are of major importance for hypotheses concerning the elicitation of P300 or N400 components in natural reading.

III Interim summary and hypotheses

Based on the preceding chapters, it is possible to derive predictions for the following experiments (1 through 3). These, however, partially diverge with respect to when ERPs are collected during natural reading (Experiment 2), and when they are collected with auditory or rapid serial visual presentation (RSVP) (Experiments 1 and 3, respectively). The first main result concerning the P300 in language processing was that it (i) was found consistently when the task relevant stimulus was a priori defined, and (ii) peaked after enough information had been gathered to enable stimulus identification in accordance with task instructions. These two findings stem from lexical processing studies, so experiments 1 through 3 employed the antonym paradigm in a sentence context, allowing for comparison with previous findings. This paradigm is also suitable to examine the second major finding concerning the P300 in lexical processing: that its latency varies with the amount of information necessary to identify the stimulus. Two different scenarios were then hypothesized. Either the P300 to expected words peaks in the N400 time window when prelexical information such as orthographic or phonological codes render target identification possible, or because evidence from several linguistic domains (e.g., orthography and semantics) is collected at once, quickly meeting the predictions concerning the expected word form and meaning. Thus, the P300 in language processing is not bound to a particular linguistic level (e.g., lexical-semantic processing), but depends on the time the subject has to evaluate a stimulus. This predicts that the P300 should be affected by modality-inherent difference in the time course of information delivery. That is, when prelexical information is dissociated from semantic information (as is the case in natural reading), both scenarios make different predictions about the occurrence of the P300.

However, all prior studies reporting P300 latency variation do not provide truly compelling evidence in favor of either one of these hypotheses, because the presentation methods used cannot offer a temporal resolution fine enough to show exactly when prelexical and lexical information is accessed. With the RSVP method, for instance, words are displayed on the screen one after the other so that participants can easily encode it at once, and the sentence is encoded in a serial manner. A similar point can be made for auditory presentation. Furthermore, although ERPs offer a temporal resolution high enough to infer pre-lexical processing from early-latency ERPs (see Bentin et al., 1999; see Grainger & Holcomb, 2009, for review), these early components do not reflect the decision-making process that the P300 indicates. Natural reading, in contrast, appears to be a good candidate to test the two central claims of the present thesis; pre-lexical and lexical processing of a word are spatially delineated, with pre-lexical processing occurring mostly before the critical word is fixated. Consequently, the P300/N400 pattern described for highly related and unrelated words (see chapter 1.3.1) could diverge in reading, such that each component is elicited by different fixation positions relative to the target word.

The dissociation of the biphasic pattern using fixation position can elucidate further the nature of the N400 component in lexical processing. Many studies have tested whether the reduced N400 for semantically related words results from the automatic spread of activation, or from controlled and specific predictions of words and thus from contextual integration (e.g., Brown & Hagoort, 1993; Federmeier & Kutas, 1999; Laszlo & Federmeier, 2009). Behavioral findings from reaction time studies have led to the general agreement that spreading activation is a fast acting process that is independent of the allocation of attention. Moreover, the facilitation resulting from this interlexical mechanism is rapidly decaying. Interestingly, not only semantic concepts can be primed, but also orthographic representations (cf. Collins & Loftus, 1975). Specific predictions, in contrast, need time to be built up (cf. Neely, 1977; 1991). As the language processing system is conceived of as generating a set of expected words, specific predctions depend on the allocation of attention and may also entail the inhibition of unexpected concepts. In contrast to priming through spreading activation, expectancy-induced priming is thought of as a postlexical integration effect (e.g., Kutas &

Federmeier, 1999; Weisbrod et al, 1999). In reading, one may observe parafoveal and foveal N400 effects assuming that lexical processing begins with parafoveal preprocessing. Expectancy-induced priming should then be detectable in foveal vision only, whereas priming due to spreading activation may also be induced by orthographic preprocessing in the parafovea, engendering a parafoveal N400 effect.

The hypotheses for each of the first three experiments were derived based on these considerations.

Experiment 1: the P300 in the auditory modality

The first experiment is designed to replicate the visually evoked ERP findings from Roehm et al. (2007) in the auditory domain. Based on previous systematic comparisons between the visual and auditory modalities, the pattern of results should be identical to the corresponding experiment by Roehm and colleagues. That is, a biphasic P300/N400 pattern is expected, with antonyms eliciting a P300 and non-antonyms giving rise to a graded N400 effect depending on semantic relatedness (related < unrelated). In addition, the non-antonym conditions are expected to elicit a P300 that peaks following the N400. Differences in onset and offset latencies could occur with latencies starting earlier and being overall longer in the auditory modality (cf. Holcomb & Neville, 1990). However, such variations would be due to modality-inherent differences, rather than to altered comprehension strategies.

Experiment 2: the P300 in natural reading

It has been shown that pre-lexical and lexical processing in natural reading can be dissociated based on fixation position – in particular, that pre-lexical information is already acquired in parafoveal vision (cf. chapter 2.2.2). Concerning the antonym condition in the present experiments, it was hypothesized that (i) they elicit a P300 in the N400 time window since orthographic information is sufficient to evaluate these words, or (ii) that the predictions of a specific word form and a specific meaning must be fulfilled at the same time. From this, one can infer two contrasting hypotheses about the P300.

The first hypothesis is that the P300 previously found for strongly related prime-target pairs may translate into a parafoveal P300 occurring in response to parafoveal orthographic processing. That is, a P300 is elicited by the last fixation prior to the critical antonym, provided the parafoveally perceived orthographic information is sufficient for a decision. This would yield a situation in which there is no P300 for the first fixation on the antonym because the processes relevant to P300 elicitation would already be finished during parafoveal processing. Antonyms should engender the least amount of foveal N400 activity, as they are the only expected continuation for the preceding sentence fragment. However, one must also take two important characteristics of parafoveal information, which could counteract a parafoveal P300, into account. First, perceptual stimulus quality depends on retinal eccentricity and parafoveal input is always reduced in acuity (see chapter 2.2.3). Thus, the perceptual load induced by parafoveal stimuli should be higher than by foveal stimuli, and this in turn is likely to reduce the amplitude of the parafoveal P300 when compared with the P300 from Experiments 1 and 3 using auditory modality or RSVP. Second, parafoveal information of an upcoming word is processed simultaneously with the currently fixated word and therefore, parafoveal processing receives less attentional resources. This may also reduce P300 amplitude in natural reading.

The second hypothesis posits that there should be no P300 at any fixation position in Experiment 2, since orthographic and semantic information are processed consecutively. Specifically, one can presuppose that orthographic information is acquired from the parafovea and that a large portion of orthographic preprocessing is accomplished before the antonym is first fixated. This follows from the fact that more parafoveal information is acquired for highly predictable words (Balota et al., 1985) and that linguistic processing continues during saccade execution (Irwin, 1998). The language processing system makes two very specific predictions, one about the word form and one about the meaning, and both must be met so that the upcoming word can be identified as the expected antonym. When, however, word form information and semantic information are delivered at different time points that do not allow for processing overlap, the system receives only partial evidence for the decision process. In other words, parafoveal orthographic forms are used to meet the word form prediction, but do not provide any information relevant for the semantic prediction. Information acquired in foveal vision can meet the semantic prediction, but it fails to provide concurrent evidence for word form prediction, i.e. the orthographic evaluation process has already been completed, so word from evaluation does not take place simultaneously with semantic evaluation. Therefore, decision-making at either fixation point does not allow for the unequivocal identification of the upcoming antonym with regard to both linguistic domains. As the P300 is reduced in amplitude when categorization is based on a low-confident decision, it may be completely absent in the proposed scenario. This extreme case is possible because linguistic stimuli are far more complex than the simple tone bursts in auditory discrimination experiments. Due to the inherent complexity of the linguistic stimulus, meeting the predictions in every linguistic domain may be particularly important for the decision process. When the system thus only acquires incomplete information to evaluate the antonym target, the P300 diminishes. Note that, alternatively, the outcome of the parafoveal decision-making process could also be added to the foveal categorization process, which should lead to a visible P300 to antonyms. Thus, the appearance of the P300 to antonyms is a test case for whether these categorization processes are additive or whether they depend on cascaded linguistic processing.

With these considerations in mind, it is conceivable to find N400 semantic priming effects for antonyms, uncontaminated by the co-occurrence of a P300. When the antonym's orthographic information is perceived parafoveally, it could be primed by spreading activation, thus reducing the amplitude of the parafoveal N400 relative to non-antonyms. As

N400 amplitude has been shown to be rather insensitive to stimulus degradation in the visual domain (Holcomb, 1993), these aspects of parafoveal input should not influence a parafoveal N400 effect. As antonyms are the only words that can reasonably complete the sentence, their integration into the preceding sentence context should also be facilitated, as opposed to both non-antonym conditions. The amplitude of the foveal N400 should therefore be reduced in comparison to non-antonyms.

Regarding the two non-antonym conditions, the predictions differ slightly according to whether the unexpected non-antonym is semantically related to the prime word or not. Semantically related non-antonyms may show a reduced parafoveal N400 amplitude, provided that the automatic spread of activation from the prime also leads to their orthographic form being recognized more easily than with unrelated non-antonyms. If these priming effects are much stronger for antonyms than for related non-antonyms, there might also be amplitude differences between these two conditions. If there is no such automatic pre-activation for semantically related targets at all, the parafoveal N400 components for related and unrelated non-antonyms should not differ from each other, but differ significantly from the antonyms. In addition, both non-antonym conditions should elicit a foveal N400 effect, which indicates the enhanced processing effort and integration cost of semantic information from non-antonym targets, since non-antonyms do not fit the preceding sentence context. In other words, the foveal N400 to unexpected non-antonyms should be sensitive to their cloze probability, thereby replicating the inverse correlation between N400 amplitude and cloze probability. The foveal N400 should then be followed by a P300 indicating that the targets have been categorized as non-antonyms. The P300 to non-antonyms is not subject to the specific predictions that could affect the P300 to antonyms. So, its elicitation does not depend on when orthographic and semantic information become available during reading. Rather, semantic information alone determines the evaluation process for non-antonyms, so that the

ERP effects triggered by the first fixation on these words should mirror previous results, as regards the P300.

In light of these considerations, the biphasic pattern found for the antonym paradigm should be distributed across two fixation positions in natural reading. The last fixation before the critical target word might show a monophasic P300 for antonyms. If visual acuity restrictions exert a heavy influence at this point, the parafoveal P300 should either be reduced in amplitude or entirely absent. Given that the P300 for antonyms does result from pre-lexical information, there should be no foveal P300 because pre-lexical information will already have been processed – the foveal fixation occurs after the decision about the antonym was made. If, however, the concurrent evaluation of prelexical orthographic and lexical-semantic information is an essential prerequisite for the P300 to antonyms, there should be no P300 at all in Experiment 2 because neither fixation provides all the necessary information. Furthermore, if the automatic spread of activation facilitates the recognition of orthographic forms which are pre-activated, parafoveal N400 amplitude may be reduced for antonyms (also in the absence of the P300) and perhaps for related non-antonyms as well. Regarding the first fixation on the target word, the two non-antonym conditions are predicted to engender an N400/P300 biphasic pattern. In each case the P300 indicates the final stage of a binary decision-making process in which the target word is categorized as the expected antonym or an unexpected non-antonym, respectively; the N400 indicates linguistic processing proper. Only the P300 to antonyms should be susceptible to the parafoveal preview in reading, as the P300 to non-antonyms depends on semantic information alone that is provided by the first fixation.

The hypotheses put forward in the preceding paragraphs rely wholly on the notion of decision-making and the variability as to the linguistic layer that allows for a final decision in stimulus categorization. One open question concerns whether the emphasis on decision-making is the most parsimonious way to account for the existing data and to derive hypotheses for Experiment 2. The bulk of experiments investigating the role of attention in the elicitation of the P300 have revealed that selective attention plays a crucial role in psychophysiological and psycholinguistic experiments. In fact, it is possible to derive hypotheses without reference to decision-making experiments or the distinction between lexical and pre-lexical information. These hypotheses will be subsumed under the attention-based account.

Predictions from the attention-based account center around two main points crucial to the P300-eliciting event. One of these is again stimulus quality, as reflected in the drop in visual acuity from fovea to parafovea. The second factor concerns the allocation of attention in reading, which can be regarded as either focused or divided, depending on one's theoretical premises. This is closely intertwined with the ongoing debate about whether information processing in reading is serial or parallel (see, e.g., Reichle, Liversedge, et al., 2009, for recent discussion); clearly, if more than one word undergoes lexical processing, attention must be divided across these words. Chapter 2.3.2 presented two computational models of eye movements in reading, the E-Z Reader model and the SWIFT model, which make very different hypotheses concerning the allocation of attention.

The E-Z Reader model assumes that attention is focused on the foveal word, whereas parafoveal words are processed pre-attentively, allowing for parallel processing of prelexical features only (cf. Reichle et al., 2006). Attention is supposed to be fully focused on the foveal word and to shift to the parafoveal saccade target only shortly before saccade execution. This predicts no (or only tiny) effects for the ERPs collected from the last fixation prior to the critical word. If attention alone drives P300 effects, the P300 for antonyms should emerge when the word is first fixated. Under these circumstances, the existence of a parafoveal P300 is very unlikely, due to the deprivation in attention. Similarly, one would not expect to find significant parafoveal N400 priming effects induced by spreading activation, because the deprivation of attention has been shown to inhibit priming effects. Again, only foveal N400 priming effects with graded amplitudes for the two non-antonym conditions are expected when attention is allocated in a serial manner in reading. In fact, if parafoveal information is explicitly denied to affect processing strategies, ERPs collected during natural reading should exactly mirror those collected in RSVP settings.

In contrast to this strictly serial model, the SWIFT model assumes an attention gradient so that several words can undergo lexical access – hence full identification – at once. This implies that attention is not only divided in parafoveal processing, but also during the identification of the foveal word. The model also assumes specific interactions between lexical processing and saccade-target selection; these interactions are supposed to account for the empirical finding that highly predictable words are skipped more often than unpredictable words. One corollary of these interactions is that lexical processing of highly predictable words (such as antonyms) must be well advanced in a large proportion of trials before these words are directly fixated. On this account, a parafoveal P300 to expected antonyms is far more likely to occur than in the serial E-Z Reader model. Processing of the unexpected non-antonyms remains unaffected by these specific interactions. On the other hand, the divided allocation of attention predicts that parafoveal N400 priming effects could occur, leading to reduced N400 amplitudes to non-antonyms related to the expexted antonym. Foveal N400 priming effects are also conceivable if spreading activation and contextual integration should turn out to be successive steps, each of which converges with another fixation position. In these cases, a P300 should follow the N400 indexing lexical access for non-antonyms, because semantic processing is necessary to evaluate these targets.

From this perspective, Experiment 2 may be able to provide further insight into whether lexical processing is limited to foveally inspected words or whether it extends to parafoveal words as well, i.e. whether it is serial or parallel. Experiment 2 may also supplement previous studies on the neuronal correlates of parafoveal preprocessing at the word level by providing data from the sentence level, and further elucidate the extent to which these neuronal effects are passed on to the oculomotor system. This is particularly important in view of the few previous endeavors to record ERPs and eye movements concurrently (Baccino & Manunta, 2005; Simola et al., 2009). These studies investigated the influence of parafoveal preview by presenting two adjacent single words, a prime and a target, and instructing their participants to fixate on the prime first, before initiating a saccade to the target. The task was to perform a lexical decision, plus a lexical relatedness evaluation when the target was a real word. Although the results of both studies indicate that the lexical status of the target (which could be either a word or a non-word) as well as the lexical association between prime and target exerted very early effects – within the first 200 – 300 ms of the first fixation on the prime – the typical P300 and N400 time windows were not examined. Therefore, it is not clear whether lexical effects in parafoveal processing may also affect late components, especially when the critical words are embedded in a sentence.

With regard to the eye movement data collected in Experiment 2, chapter 2.2.2 suggests that the high-constraint context (*X is the opposite of Y*) should facilitate the processing of antonyms (*black – white*) as opposed to related (*black – yellow*) and unrelated non-antonyms (*black – nice*). Since the sentence context counteracts any benefit of semantic relatedness (i.e., any non-antonym is contextually incongruent regardless of its relatedness to the correct antonym), eye movement measures should not show processing facilitation for related targets relative to unrelated targets. It is difficult to derive specific hypotheses based on the existing data on parafoveal-on-foveal effects in reading. It is conceivable that the fixation durations on the pre-critical region (*the opposite of*) vary as a function of the degree to which the parafoveally perceived word matches the predicted word. It is less certain whether these effects are related to parafoveal orthographic processing or parafoveal semantic processing. One can infer that the antonym context should elicit parafoveal orthographic effects, which can emerge under serial and parallel processing conditions. Antonyms should differ from related and unrelated non-antonyms, as the unexpected word forms of all non-antonyms should generally lead to increased fixation durations. The existence of parafoveal semantic processing depends on the assumed allocation of attention, so that related and unrelated targets should only differ in terms of parafoveal-on-foveal effects if attention is distributed across several words. Only then can semantic relatedness exert a significant influence during parafoveal processing.

Experiment 3: the P300 in RSVP

The combination of EEG and eye-tracking methods rendered it necessary to change some aspects of the typical experimental setting, such as the number of trials per condition (see chapter 4.3). Thus, some of the ERP results obtained in Experiment 2 may be confounded by such technical changes to the general experimental setting, which could make the comparison with Experiment 1 and the Roehm et al. (2007) study somewhat less reliable. To exclude this possibility, Experiment 3 was carried out with the same stimulus set as in Experiment 2, the only exception being that the RSVP method was used. If the results from Experiment 2 are due to deviations from typical EEG stimulus sets, Experiment 3 should reveal a pattern similar to what Experiment 2 revealed. If, however, the findings from Experiment 2 are in fact related to the specific characteristics of natural reading, as opposed to RSVP and listening, Experiment 3 should pattern with Experiment 1.

IV Experiments

4 Looking for the P300 in natural input modalities of language comprehension

4.1 Experiment 1: Antonymy comprehension in the auditory modality

Previous experimental results concerning the comprehension of antonymy relations have been obtained either with a cross-modal (Kutas & Iragui, 1998) or a rapid serial visual (RSVP) presentation (Roehm et al., 2007). Critically, the target word was presented visually in both studies. The results are largely comparable. However, there could be slightly different results using purely auditory sentence presentation with respect to the temporal domain. For instance, Holcomb & Neville (1990) found an earlier and longer-lasting priming effect in the auditory modality, as compared to the visual modality. The first experiment is a replication of Roehm et al.'s first experiment, differing only in the mode of presentation by using auditory stimuli. When antonyms are compared with related and unrelated non-antonyms in a sentence context, one should find a threefold distinction between conditions, with antonyms exhibiting a P300 component and the other two conditions evidencing a graded N400 priming effect followed by a P300.

4.1.1 Method

Participants

Thirty students from the Max-Planck Institute subject pool in Leipzig participated in the experiment (15 female; mean age: 24, range: 19-29 years). All participants were right-handed (as assessed by a modified German version of the Edinburgh Handedness Inventory; Oldfield, 1971) and monolingual native speakers of German. They had no auditory disorders, and none of them reported neurological or psychiatric disorders. The experiment began after participants gave their informed consent. One further participant had to be excluded from all analyses due to excessive artifacts in the EEG.

Materials

The sentence sets were identical to those reported in Roehm et al. (2007). Stimulus construction and allocation to experimental lists was the same as in the Roehm et al. study. Eighty triplets of sentences were created, resulting in a total of 240 stimuli. The stimuli were allocated across four lists using a Latin Square design in order to balance the number of 'yes' and 'no' responses. Each list contained 160 sentences, which consisted of 80 antonym pairs [ANT] and 40 pairs for the related [REL] and unrelated [NON] conditions each. The lists were presented in two versions with different pseudo-randomized orders. All sentences were simple main clauses as given in examples (5) through (7).

(5) Das Gegenteil von schwarz ist weiß. [ANT]

'The opposite of black is white.'

(6) Das Gegenteil von schwarz ist blau. [REL]

'The opposite of black is blue.'

(7) Das Gegenteil von schwarz ist nett. [NON]

'The opposite of black is nice.'

Auditory stimuli were spoken by a male voice and stored on a hard disk. Prior to the experiment, each stimulus was edited for later synchronization with ERP and reaction time data and was normalized digitally when necessary. The spoken onset of the target word was taken to be the critical time event for averaging ERPs to the target.

Apparatus and Procedure

The EEG was recorded from 60 Ag/AgCl scalp electrodes that were positioned according to the international 10-20 system; the ground electrode was placed at electrode position C2. All EEG channels were referenced to the left mastoid, but re-referenced to linked mastoids offline. The EOG was recorded via two electrodes placed at the outer canthi of the eyes and two electrodes placed above and below the right eye. All EEG and EOG channels were amplified with a BrainAmp amplifier and recorded with a digitization rate of 500 Hz. The EEG was filtered online using a 250-Hz high-cutoff filter and a DC low-cutoff filter. To avoid slow signal drifts, the EEG was filtered offline with a 0.3- to 20-Hz bandpass filter. The plots for the grand average ERPs were smoothed with a 8-Hz lowpass filter, but all analyses were calculated over unfiltered data. The ERPs were averaged per participant and per condition from 200 ms before word onset to 1000 ms post word onset. No baseline correction was used.

Each trial started with a fixation cross that was displayed on the screen for 500 ms before the sentence was played to the participant via loudspeakers. The fixation cross remained on the screen during the entire sentence and participants were asked to refrain from blinking throughout this time. A sentence verification task was used in this experiment. After the appearance of a '?' indicating that the subject was to make a judgment on whether the sentence just heard was a true statement or not, the subject was given 3 s to respond. The intertrial interval (ITI) was 3 s.

Data preparation and analyses

The ERP and behavioral reaction time data were analyzed using a repeated-measures analyses of variance (ANOVAs) design with the factor condition [COND] (ANT vs. REL vs. NON). To keep the ERP data comparable to the other two experiments reported in this thesis, only a subset of electrodes was analyzed. Thus, the topographical factor region of interest (ROI) was defined for lateral ROIs (left-frontal (F3, F7), right-frontal (F4, F8), left-central (C3, T7), right-central (C4-T8), left-posterior (P3, P7), and right-posterior (P4, P8)) and for midline electrodes (frontal (Fz), central (Cz) and posterior (Pz)). To avoid type I errors resulting from violations of sphericity, alpha was corrected as proposed by (Huynh and Feldt, 1970) when effects with more than one degree of freedom in the numerator were evaluated. In these cases, the original degrees of freedom are reported with the corrected alpha. Whenever statistical tests justified pair-wise comparisons involving the factor COND, these were adjusted according to a modified Bonferroni procedure (Keppel, 1991). Alpha was set to $p < .033$ in these cases.

4.1.2 Results

Behavioral data

The mean error rates and reaction times are given in Table 4.

	Error rates (% correct)	Reaction time (ms)
Antonyms (ANT)	99.1 (1)	401 (139)
Related (REL)	94.5 (6)	434 (175)
Unrelated (NON)	99.6 (1)	391 (136)

Table 4. Mean participant error rates and reaction time. Standard deviation is given in parentheses.

Although the error rates across all conditions were at ceiling, the main effect of COND was reliable across participants and items ($F_1(2,58) = 21.51$, $p < .001$; $F_2(2,158) = 14.44$, $p < .001$). Pair-wise comparisons indicated that related target words differed significantly from both antonyms and unrelated non-antonyms, and that antonyms and unrelated non-antonyms did not differ from one another (ANT vs. REL: $F_1(1,29) = 19.45$, $p < .001$; $F_2(1,79) = 12.46$, $p = .001$; ANT vs. NON: $F_1(1,29) = 2.47$, $p < .2$; $F_2(1,79) = 3.45$, $p < .07$; REL vs. NON: $F_1(1,29) = 26.06$, $p < .001$; $F_2(1,79) = 17.34$, $p < .001$). Thus, the related non-antonym conditions yielded more incorrect answers than the other two conditions.

The disadvantage for related non-antonyms is also visible in the reaction time data. The main effect of COND was highly reliable across participants and items ($F_1(2,58) = 9.94$, $p < .001$; $F_2(2,158) = 15.43$, $p < .001$). Pair-wise comparisons revealed that the reaction time for related non-antonyms was significantly longer than for either antonyms or unrelated non-antonyms, and that the 10-ms difference between antonyms and unrelated non-antonyms was not significant (ANT vs. REL: $F_1(1,29) = 7.28$, $p < .02$; $F_2(1,79) = 16.11$, $p < .001$; ANT vs. NON: $F_1(1,29) = 2.53$, $p < .2$; $F_2(1,79) = 2.69$, $p < .2$; REL vs. NON: $F_1(1,29) = 16.00$, $p < .001$; $F_2(1,79) = 19.01$, $p < .001$). In sum, related non-antonyms impose visible processing costs on the language comprehension system, as reflected in a higher percentage of errors and longer reaction time.

ERP data

Grand-average ERPs are shown in Figure 1. Visual inspection of the ERP waves suggests that the ERPs in response to the three conditions start to diverge at around 200 ms post word onset. The antonym condition appeared to be more positive than the other two conditions, and the positivity exhibited a clear centro-parietal maximum peaking at around 300 ms. The related and unrelated non-antonym conditions, by contrast, showed a pronounced negative deflection lasting until approximately 550 ms after onset of the target word. The negativity also exhibited a central-parietal maximum. Furthermore, the related condition showed a peak latency for the negativity similar to the antonym condition at central sites, whereas there was no clear peak for the non-antonym conditions across posterior scalp sites. Finally, both non-antonym conditions appeared to be more positive going between 700 and 950 ms post word onset as opposed to the antonym condition. The scalp distribution of the late positivity appeared to be restricted to the centro-parietal ROIs with an additional small left-lateralized tendency. The time windows for the statistical tests on mean amplitudes were chosen in accordance with the visual inspection results: 200 ms to 550 ms for the N400/P300 time window and a 710 ms to 950 ms time window for the late positivity.

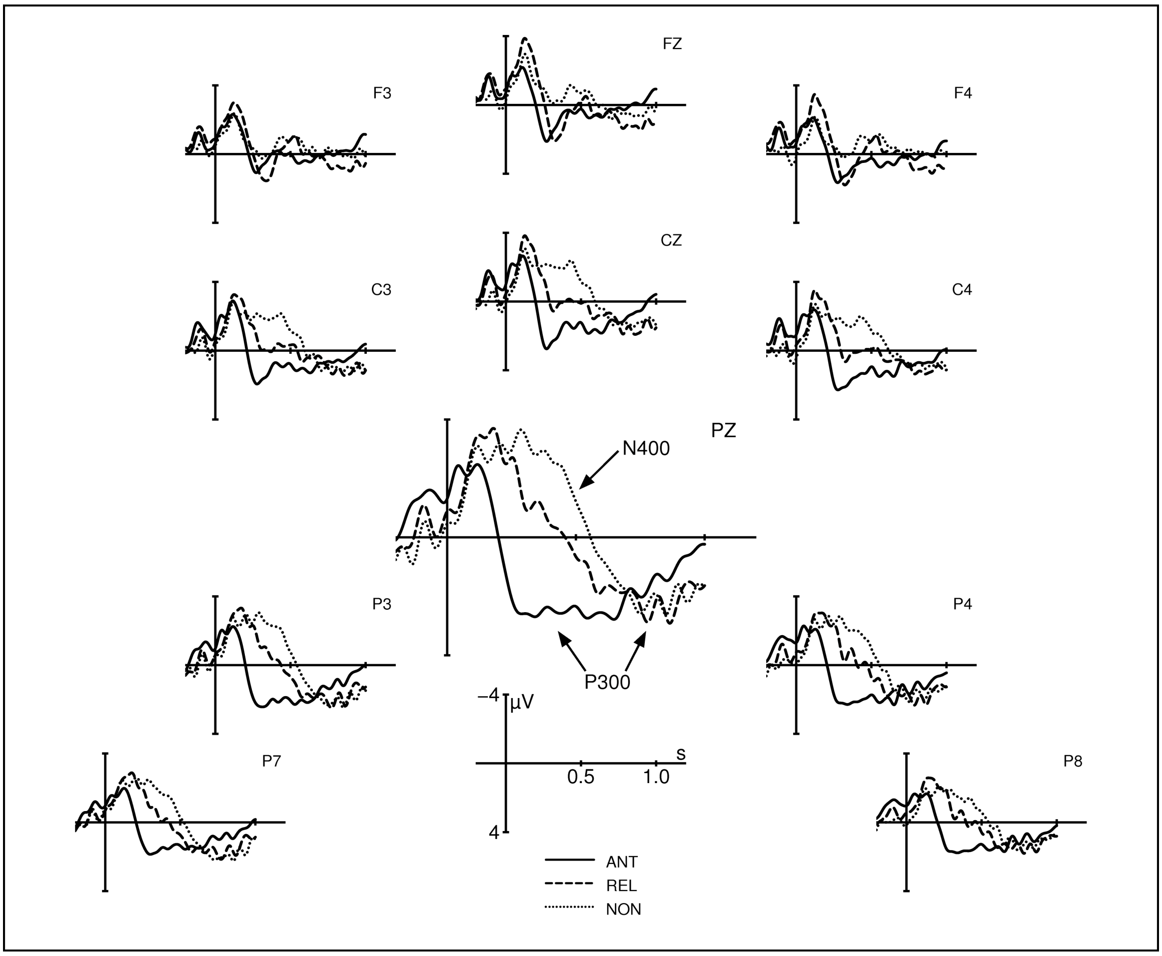


Figure 1. Grand-average ERPs (N = 30) time-locked to the onset of the target word (onset at the vertical bar). Negativity is plotted upwards in this and the following ERP figures.

P300/N400 time window: 200 ms – 550 ms

ANOVAs for the midline electrode sites revealed a reliable main effect of COND ($F(2,58) = 110.47$, $p < .001$) as well as an interaction between ROI and COND ($F(4,116) = 64.75$, $p < .001$). After resolving the interaction, the effect of COND was significant at each ROI and largest at parietal sites (frontal: $F(2,58) = 19.26$, $p < .001$); central: $F(2,58) = 108.16$, $p < .0001$; parietal: $F(2,58) = 168.29$, $p < .001$). Further single comparisons revealed that, for all electrode sites, the contrast between antonyms and related words was significant (frontal: $F(1,29) = 5.07$, $p < .033$; central: $F(1,29) = 85.61$, $p < .001$; parietal: $F(1,29) = 130.33$, $p < .001$), as were the contrasts between antonyms and unrelated words (frontal: $F(1,29) = 28.79$, $p < .001$; central: $F(1,29) = 183.11$, $p < .001$; parietal: $F(1,29) = 353.51$, $p < .001$) and between related and unrelated words (frontal: $F(1,29) = 16.24$, $p < .0001$; central: $F(1,29) =$

40.93, $p < .001$; parietal: $F(1,29) = 38.22$, $p < .001$). In each case, the difference between the conditions was larger at central and parietal sites than at frontal ROIs.

The statistical tests for the lateral ROIs largely confirmed these findings, with the exception of frontal ROIs, in that they registered two main effects of COND ($F(2,58) = 75.60$, $p < .001$) and ROI ($F(5,145) = 3.19$, $p < .04$) and an interaction between the two factors ($F(10, 290) = 43.36$, $p < .001$). Resolving the interaction by ROI revealed a significant effect of COND for all lateral ROIs except the left-frontal ROI (left-frontal: $F(2,58) = 2.00$, $p < .2$; right-frontal: $F(2,58) = 6.17$, $p < .01$; left-central: $F(2,58) = 58.12$, $p < .001$; right-central: $F(2,58) = 62.00$, $p < .001$; left-parietal: $F(2,58) = 144.19$, $p < .001$; right-parietal: $F(2,58) = 124.99$, $p < .001$). Again, the effects became larger at central and parietal ROIs. Pair-wise comparisons for the factor COND showed that antonyms differed from related targets over central and parietal scalp regions, but not over frontal sites (left-frontal: $F(1,29) = 1.18$, $p < .3$; right-frontal: $F(1,29) = 1.75$, $p < .2$; left-central: $F(1,29) = 42.02$, $p < .001$; right-central: $F(1,29) = 41.51$, $p < .001$; left-parietal: $F(1,29) = 132.60$, $p < .001$; right-parietal: $F(1,29) = 109.41$, $p < .001$). Antonyms differed significantly from unrelated words at all ROIs except the left-frontal ROI (left-frontal: $F(1,29) = 1.05$, $p < .4$; right-frontal: $F(1,29) = 9.53$, $p < .01$; left-central: $F(1,29) = 105.85$, $p < .001$; right-central: $F(1,29) = 108.18$, $p < .001$; left-parietal: $F(1,29) = 263.33$, $p < .001$; right-parietal: $F(1,29) = 209.15$, $p < .001$). The same pattern emerged between related and unrelated words (left-frontal: $F(1,29) = 3.42$, $p < .08$; right-frontal: $F(1,29) = 4.94$, $p < .04$; left-central: $F(1,29) = 25.25$, $p < .001$; right-central: $F(1,29) = 25.91$, $p < .001$; left-parietal: $F(1,29) = 35.34$, $p < .001$; right-parietal: $F(1,29) = 29.99$, $p < .001$).

Overall, the statistical tests support the conclusions drawn from the visual inspection of the ERP waves. There was a pronounced positivity, a P300, for the expected antonyms, whereas both unexpected non-antonyms exhibited a centro-parietal negativity, the N400. This component further distinguished the non-antonym conditions by revealing a larger amplitude for unrelated non-antonyms.

Late positivity time window: 710 ms – 950 ms

The ANOVAs for the midline sites registered two main effects of ROI ($F(2,58) = 35.37$, $p < .001$) and COND ($F(2,58) = 7.06$, $p < .01$), but the interaction failed to reach significance ($F(4,116) = 1.56$, $p < .3$). Single comparisons revealed that antonyms differed significantly from both related ($F(1,29) = 11.86$, $p < .01$) and unrelated non-antonyms ($F(1,29) = 8.17$, $p < .01$), whereas the latter two did not differ from each other ($F(1,29) = 2.44$, $p < .2$).

With respect to lateral ROIs, the statistical tests registered two main effects of ROI ($F(5,145) = 31.95$, $p < .001$) and COND ($F(2,58) = 4.12$, $p < .03$) and a reliable interaction ($F(10, 290) = 9.54$, $p < .001$). The resolution of the interaction revealed significant effects of COND for all ROIs, except for the right-central ROI (left-frontal: $F(2,58) = 4.08$, $p < .03$; right-frontal: $F(2,58) = 4.42$, $p < .02$; left-central: $F(2,58) = 4.11$, $p < .03$; right-central: $F(2,58) = 1.19$, $p < .4$; left-parietal: $F(2,58) = 21.11$, $p < .001$; right-parietal: $F(2,58) = 5.34$, $p < .01$). Further pair-wise comparisons revealed that antonyms differed significantly from related words bilaterally over parietal sites and at the left-frontal ROI (left-frontal: ($F(1,29) = 5.91$, $p < .03$; right-frontal: ($F(1,29) = 1.62$, $p < .3$; left-central: ($F(1,29) = 4.55$, $p < .05$; right-central: $F < 1$; left-parietal: ($F(1,29) = 18.35$, $p < .0001$; right-parietal: ($F(1,29) = 11.68$, $p < .01$). Contrary to this, antonyms differed from unrelated targets over left and marginally over right parietal electrode sites and at the left-central ROI (left-frontal: $F < 1$; right-frontal: ($F(1,29) = 3.37$, $p < .08$; left-central: ($F(1,29) = 10.57$, $p < .01$; right-central: $F < 1$; left-parietal: ($F(1,29) = 61.15$, $p < .001$; right-parietal: ($F(1,29) = 4.81$, $p < .04$). Finally, related and unrelated words differed bilaterally at frontal electrodes, with the related condition being more positive (left-frontal: ($F(1,29) = 6.03$, $p < .03$; right-frontal: ($F(1,29) = 7.37$, $p < .02$;

left-central: $F < 1$; right-central: ($F(1,29) = 2.02$, $p < .2$; left-parietal: ($F(1,29) = 2.07$, $p < .2$; right-parietal: $F < 1$).

In summary, the two non-antonym conditions elicited a late positivity at central and parietal electrodes, whereas the antonym condition was more negative going. In addition, the two non-antonym conditions differed from each other only at frontal electrode sites, a result of the stronger positive deflection for related non-antonyms in these areas.

4.1.3 Discussion

The present data fully replicate the findings from the first of a series of experiments in the visual domain by Roehm and colleagues (2007). The authors found a P300 (P3b) for antonyms embedded in a sentence context, occurring concurrently with the N400 for non-antonym targets. These unexpected words elicited an N400, the amplitude of which differed according to semantic relatedness. If prime and target were semantically related, N400 amplitude was slightly reduced. The N400 for non-antonyms was followed by a late positivity. In the present auditory experiment, antonyms elicited a P300 peaking at about 250 ms post word onset, while the two non-antonym conditions engendered an N400, starting at about 250 ms and lasting until 550 ms post onset. The N400 amplitude for related non-antonyms was reduced, compared to the one for unrelated non-antonyms. After the N400, both non-antonym conditions engendered a P300. The positivity for related non-antonyms was more pronounced frontally than for unrelated non-antonyms, which is similar to the finding reported by Roehm and colleagues. The major difference from the predecessor study lies in the latency shifts, indicating enlarged onset and offset latencies of the components under investigation.

Chapter III put forward two main hypotheses for the experiment. The first one stated that the overall pattern of results obtained with the auditory modality should mirror previous findings in the visual domain. Specifically, the P300/N400 pattern should vary according to the expectedness of the target word completing the sentence. Clearly, the present data confirm this hypothesis. The P300 to the expected antonym emerged in the same time interval as the N400 for the unexpected non-antonyms, which themselves only produced a P300 *after* the N400 peak. This finding lends further support to the central claim of the thesis: the decision-making process for categorizing the target word depends on the ease with which the word can be identified. As only antonyms fit in the predicted word form and the predicted meaning, the P300, reflecting accomplished categorization, emerges in a different time window than for non-antonyms. For these unexpected target words, lexical access is necessary to decide whether the perceived word is the expected antonym of the prime, as prelexical information already signals that non-antonyms are deviant. When the meaning of the unexpected non-antonyms is retrieved, they are also rejected with regard to the semantic prediction. After the negative evidence from both levels of prediction (prelexical and lexical) is collected, they can be securely be categorized as non-antonyms. Accordingly, the P300 here peaks after the N400, the component that reflects lexical processing. The second hypothesis dealt with modality-specific differences. As expected, the present results show earlier onset latencies for the antonym-P300 and the N400 as well as enlarged offset latencies for all components (cf. Holcomb & Neville, 1990).

As a whole, these data provide further converging evidence that the antonym N400 is virtually undetectable because of the temporal overlap between the P300 and N400 (Bentin, 1987; Kutas & Iragui, 1998; Roehm et al., 2007), and that this overlap applies to the visual and the auditory domain. The open question thus is why there isn't even any residual N400 activity for antonyms. Two hypotheses were proposed to account for this. One the one hand, the pattern is, in fact, reminiscent of the findings for non-words. Holcomb & Neville (1990) found that non-words elicited a pronounced P300 without any N400 activity, which they interpreted as reflecting the deviant orthographic and phonological – hence prelexical –

features of non-words. Non-words are judged as such by means of these prelexical features. Thus, antonyms may also be categorized based on prelexical features, which might not trigger an N400 component. On the other hand, the simultaneous availability of prelexical and semantic information, both fulfilling the predictions about word form and meaning, may be responsible for the P300 overlapping with the N400. As antonyms are the predicted sentences completion, they can be successfully categorized as antonyms. The P300 therefore co-occurs with the N400, whereas categorizing non-antonyms is more demanding (see above).

Another question is to which extent the semantic relatedness between prime and target reduces N400 amplitudes when participants are instructed to make an antonymy judgment. Considering that antonyms exhibited a reduced N400 amplitude relative to related non-antonyms when a lexical decision task was used (Roehm et al., 2007, Experiment 2), the N400 for antonyms should be more reduced than for related non-antonyms using the antonymy judgment. This facilitative effect, however, cannot be evaluated when the N400 for the former is entirely overridden by the P300. It was argued in chapter 2.2 that the availability of a parafoveal preview in natural reading might provide data that could help to answer these questions. If the P300 and N400 are triggered by events that respond to different processes (of which only mechanisms generating the P300 are susceptible to the parafoveal preview), one may be able to dissociate the P300/N400 co-occurrence in natural reading because the components would then re-distribute to different fixation positions. This was tested in Experiment 2.

4.2 Experiment 2: Antonymy comprehension in natural reading

4.2.1 Method

Participants

48 students (24 female; mean age: 23.7 years, range: 19-31) from the University of Mainz were paid to participate in the experiment after giving informed consent. Two further participants had to be excluded due to low calibration accuracy or recording problems during the experiment. All participants were right-handed native speakers of German, as assessed by the German version of the Edinburgh handedness test (Oldfield, 1971). All participants reported normal or corrected-to-normal vision and no history of any neurological or psychiatric disorders. They were naïve to the purpose of the experiment. None of them participated in either Experiment 1 or 3.

Materials

The sentences were adapted from the original stimuli used by Roehm and colleagues. 40 lexical sets of sentences of the types presented in examples (8) through (10) were constructed. All critical word pairs consisted of adjectives that varied in their semantic relatedness from being true antonyms, semantically related pairs, or semantically unrelated pairs. Thus, the critical word (underlined in the examples 8 through 10 below) was either the predicted antonym [ANT], an unpredicted, yet semantically related word [REL], or an unpredicted and semantically unrelated word [NON]. The critical words did not differ from each other in word length or frequency.[26]

[26] We computed repeated-measures ANOVAs for the mean frequency and length for the three conditions to ensure that the chosen sample of stimuli did not differ in length or frequency, which are known to affect readers' eye movements (Rayner, 1998). Mean word length per condition: ANT: 5.9; REL: 5.7; NON: 6.3. ($F(2,78) = 1.30$, $p < .3$). Mean raw frequency per

(8) Dass schwarz das Gegenteil von weiß ist, hat Gertrud gestern behauptet. [ANT]

that black the opposite of white is, has Gertrud yesterday claimed

"Gertrud claimed yesterday that black is the opposite of white."

(9) Dass schwarz das Gegenteil von blau ist, hat Gertrud gestern behauptet. [REL]

"Gertrud claimed yesterday that black is the opposite of blue."

(10) Dass schwarz das Gegenteil von nett ist, hat Gertrud gestern behauptet. [NON]

"Gertrud claimed yesterday that black is the opposite of nice."

The critical items were allocated to 4 experimental lists according to a Latin Square design, so that each participant saw 10 sentences per condition and no lexical item twice. The critical items were interspersed with 150 filler sentences of different structures. The stimuli were presented in a pseudo-randomized order. A comprehension question followed each sentence. These yes/no questions probed various parts of the critical sentence, and in about 25% of all cases they required an antonymy judgment ("Is this true?"). The number of trials per condition and the ratio of items and fillers were chosen in accordance with experimental designs in eye movement research. This deviation from typical stimulus arrangements in EEG experiments was necessary to keep participants from adopting specific strategies when reading highly similar sentences.

Apparatus and Procedure

Eye movements were recorded with an *EyeLink 1000* eyetracker with an angular resolution of 10-30 min of arc. The sampling rate was 1000 Hz. Viewing was binocular, but only the right eye was recorded. Stimuli were displayed on a 21-inch monitor and participants were seated 60 cm from the computer screen. At this distance, 3.2 characters equalled 1° of visual angle. The EEG was recorded from 17 Ag/AgCl scalp electrodes that were fixed to the scalp by

condition: ANT: 10237; REL: 11651; NON: 4338 ($F(2,78) = 2.10$, $p < .2$). Frequency values were derived from the Leipziger Wortschatz corpus (wortschatz.uni-leipzig.de).

means of an elastic cap (Easy Cap). Electrode positions followed the international 10-20 system; the ground electrode was placed at AFz. All EEG channels were referenced to the left mastoid, but re-referenced to linked mastoids offline. The EOG was recorded via 3 bipolar pairs of peri-ocular electrodes placed at the outer canthi of and above and below each eye. Electrode impedances were kept below 5 kΩ. All EEG and EOG channels were amplified with a BrainAmp amplifier and recorded with a digitization rate of 250 Hz. The EEG was filtered using a 0.3- to 20-Hz bandpass filter to avoid slow signal drifts in the EEG. The grand average ERPs displayed in the figures were smoothed with a 8-Hz lowpass filter, but all analyses were calculated over unfiltered data.

Participants were tested individually and 12 participants were randomly assigned to one of the experimental lists. Upon arrival at the lab, they were informed about the experimental procedure. After the preparation of the electrodes, participants were seated in the eye-tracking system. A calibration routine was performed prior to the experiment, which served to assure a high resolution of the eye-movement record. 10 practice trials preceded each experiment, so that participants could familiarize themselves with the experimental procedure. At the beginning of each trial, participants were instructed to fixate a rectangle, located at the left of the sentence. After the eye-tracking system registered a valid fixation in that region, the sentence was displayed with a fixed delay of 500 ms. Calibration was checked prior to the start of each trial and repeated in case of poor calibration performance. The participants were told to read the sentences for comprehension and press a button once they had finished. A sentence was never displayed for longer than 10 s. The question appeared immediately after the participants pressed a button to terminate the sentence display. The participants had to respond to the questions both as accurately and as quickly as possible. A question never appeared for longer than 15 s. Thus, there was no interstimulus interval (ISI) between words in the critical sentences and the intertrial interval (ITI) varied according to calibration accuracy. A short break followed every 120 trials. The entire experimental session lasted about 2.5 hours.

Data preparation and analyses

The EEG and eye movement (EM) recordings were synchronized via a TTL pulse that was sent from the presentation PC to both recording systems at the beginning and end of a trial. The identification of fixation-related potentials was carried out offline. First, the eye movement record of each participant was screened for erroneous trials. Less than 1% of all trials was excluded due to track losses near/at the critical region or due to technical errors. Blinks and fixations below 80 ms or above 800 ms were deleted when they were located in the critical regions; short fixations do not reflect cognitive processing and unusually long fixations are likely to reflect track losses (cf. Rayner & Pollatsek, 1989). If this resulted in a no-fixation event for one of the critical regions, these trials did not enter the computation of fixation-related potentials. No trials were excluded due to this criterion. All fixation-related potentials were triggered relative to fixation onset. For the eye-movement analysis short fixations with less than 80 ms in duration were incorporated into an adjacent fixation if the distance between these two did not exceed one character, and fixations less than 40 ms in duration were treated similarly if they were within one character of an adjacent fixation. Short fixations that did not meet these criteria were excluded from further analysis. Fixations longer than 800 ms were also excluded. In total, less than 1% of fixations were eliminated based on these criteria.

The ERP waves were time-locked and computed relative to two fixation points: the first fixation on the critical word and the last fixation before readers launched a saccade to the critical word. To define the fixation position relative to the important parts of the sentence, the experimental sentences were divided into the regions of analysis given in example (11).

(11) Dass/ schwarz/ das Gegenteil von/ weiß/ ist,/ hat Gertrud/ gestern behauptet.
pretarget region target

The average ERP waves were calculated per subject and per condition from 200 ms before the onset of the critical fixation to 1200 ms post-onset. Grand averages were then computed over all participants. No trials were excluded from averaging based on ocular artefacts. Note that the ERPs reported here were not computed relative to a baseline interval. The main assumption of a baseline correction is that there are no principled differences between the critical conditions in the pre-critical time interval. However, this assumption cannot be upheld in a natural reading situation. Specifically, the fact that reading patterns across and within subjects are largely heterogeneous, together with the availability of a parafoveal preview, which could influence the brain's reaction on the (parafoveally visible) critical word, render a baseline correction relatively problematic for the present experiment.

The ERP and EM data were analyzed using a repeated-measures analyses of variance (ANOVAs) design with the factor condition [COND] (ANT vs. REL vs. NON). For the ERP data, the topographical factor region of interest (ROI) was defined for lateral ROIs (left-central (C3, T7), right-central (C4-T8), left-posterior (P3, P7), and right posterior (P4,P8)) and for midline electrodes (central (Cz) and posterior (Pz)). As the ERP components of interest (N400, P300) have their maximum at central-parietal electrode sites, frontal electrodes were from further analyses excluded to circumvent possible contamination from eye movement artifacts.[27] To avoid type I errors resulting from violations of sphericity, alpha was corrected as proposed by Huynh and Feldt (1970) when effects with more than one degree of freedom in the numerator were evaluated. In these cases, the original degrees of freedom are reported with the adjusted alpha value. Whenever statistical tests justified pair-wise comparisons involving the factor COND, these were adjusted according to a modified Bonferroni procedure (Keppel, 1991). For the eye movement data, ANOVAs involving the

[27] Note that analyses involving ROIs for frontal electrodes led to results that were identical to those reported here.

factor COND were calculated treating participants and items as random factors F_1 and F_2, and the modified Bonferroni procedure was also applied to adjust alpha for pair-wise comparisons. In these cases, alpha was set to $p < .033$.

Note finally that the modifications in the stimulus set render the analysis of behavioral data (question accuracy and reaction time) difficult. Only about one fourth of the comprehension questions required an antonymy judgment, whereas the other questions probed the lexical content of the sentences. Therefore, a false answer in the latter does not unambiguously point to comprehension difficulty due to semantic relatedness between prime and target. In consideration of this, the report of the behavioral results concerning question answering follows standards in psycholinguistic studies using eye movements as a tool to investigate language comprehension, by reporting only the average accuracy across all participants.

A second and more important deviation from common analysis standards with regard to the ERP data concerns the choice of time windows. The recording of ERP waves during normal reading is a fairly new technique, so there is no a priori defined time window in which one would expect to find effects resulting from the experimental manipulation. The literature is as yet inconclusive regarding possible temporal shifts of ERPs in natural reading. While some report parafoveal semantic relatedness effects well before the N400 (Baccino & Manunta, 2005; Simola et al., 2009), others have reported foveal predictability effects within the typical N400 time window (Dimigen et al., 2006). The apparent difference between these two groups of studies is that the former used single words, whereas the latter used sentences. Thus, one might expect to find a temporal distribution of effects that corresponds to those established in previous EEG work (see Hutzler, Braun, Vo, Engl, Hofmann, Dambacher, Leder, & Jacobs, 2007, for further evidence of correspondences between ERPs and fixation-related potentials).

However, there is still the problem of overlapping ERPs. In order to determine whether the present manipulation of lexical relatedness and/or predictability had a significant influence on ERPs triggered by a fixation, the following strategy was employed. Visual inspection of the ERP waves is relational. Not only are ERPs on the critical word inspected, but so are ERPs to pre-critical and, when appropriate, also to post-critical regions in a sentence (see Experiment 5). The aim of this procedure is to avoid relating an ERP effect to the wrong fixation position. For example, a seemingly early effect time-locked to fixation *n* may in fact be a late effect time-locked to fixation *n*-1, which overlaps in time with the onset of the next fixation and hence with early exogenous components. Reversely, an unusually late ERP onset during fixation *n*-1 may actually be a component that is triggered by fixation *n*. Thus, the interpretation of any effect in the ERP reported here will necessarily be relative in nature: an ERP effect is taken to be functionally relevant if and only if there are no effects either in a preceding fixation or in a subsequent fixation that approach or even confirm traditional time windows. For this and the following co-registration experiments (Experiments 4 and 5), the last fixation prior to fixating the critical word or region of text and the first fixation on that word or region of text are chosen as triggering events for the ERPs. Note that the last fixation before readers launch a saccade to the target word can be located anywhere before the target. As an approximate of where it is most likely located, the launch site of the saccade leaving the pretarget region is reported. There is, however, no one-to-one correspondence between the two fixation measures.

4.2.2 Results

Behavioral results

The average accuracy was 87%. Accuracy rates for the antonymy judgment showed that there were few false answers for the antonym and the unrelated non-antonym condition, while participants made more mistakes for related non-antonyms.[28]

ERP results

Figure 2 shows the ERPs time-locked to the first fixation on the target word. Here, both non-antonym conditions are more negative in the time window between 250 and 400 ms and the unrelated non-antonym condition also exhibits a late positivity with an onset latency of about 450 ms post word onset, while the related non-antonym condition shows a weak positivity with an onset latency of about 620 ms. Figure 3 shows the ERPs time-locked to the last fixation in the pretarget region. As can be seen, the unrelated non-antonym condition is more negative going between 250 and 400 ms after fixation onset. The related non-antonym and the antonym conditions appear not to differ from one another. Additionally, both non-antonym conditions exhibit a late negativity between 450 and 600 ms and each of them also engenders a late positivity, the onset latency of which varies somewhat. The positivity for the unrelated non-antonym condition has an onset latency of about 600 ms, whereas the one for the related non-antonym condition is considerably smaller and delayed with an onset latency of about 850 ms post word onset.

The late negativity and the late positivity found for the last fixation are more accurately shaped and approach traditional N400 and late-positivity time windows with the first fixation.

[28] The main motivation behind the comprehension questions was to ascertain that participants attentively read the sentences without directing their attention to the critical parts of the experimental stimuli. Therefore, it was necessary to probe various parts of the sentence. This made it impossible to fully balance those questions requiring the antonymy judgment across items and participants. Thus, each participant judged the antonymy relation in between 7 and 8 out of all items presented. The small number of items makes statistical tests impossible. Yet, the values summed for all participants are remarkably similar to those reported in the Roehm et al. study. Across all participants and items, there were only 7 false answers in the antonym condition and 6 false answers in the unrelated non-antonym condition, whereas there were 50 incorrect answers in the related non-antonym condition.

It appears that for these components, the first fixation is a better indicator of the time when the information triggering them becomes available.[29] Visual inspection then suggests that the last fixation in the pretarget region gives rise to an enhanced negative ERP wave (N400) for unrelated non-antonyms, whereas the later negativity and late positivity are linked to the first fixation on the actual target word at which the respective ERP waves appear less smeared. The first fixation thus shows a bi-phasic pattern consisting of an enlarged N400 and a late positivity for the two non-antonym conditions. Statistical tests were therefore carried out between 250 and 400 ms for the last fixation, while two time windows were selected for the first fixation, one between 250 and 400 ms, and one between 450 and 860 ms.

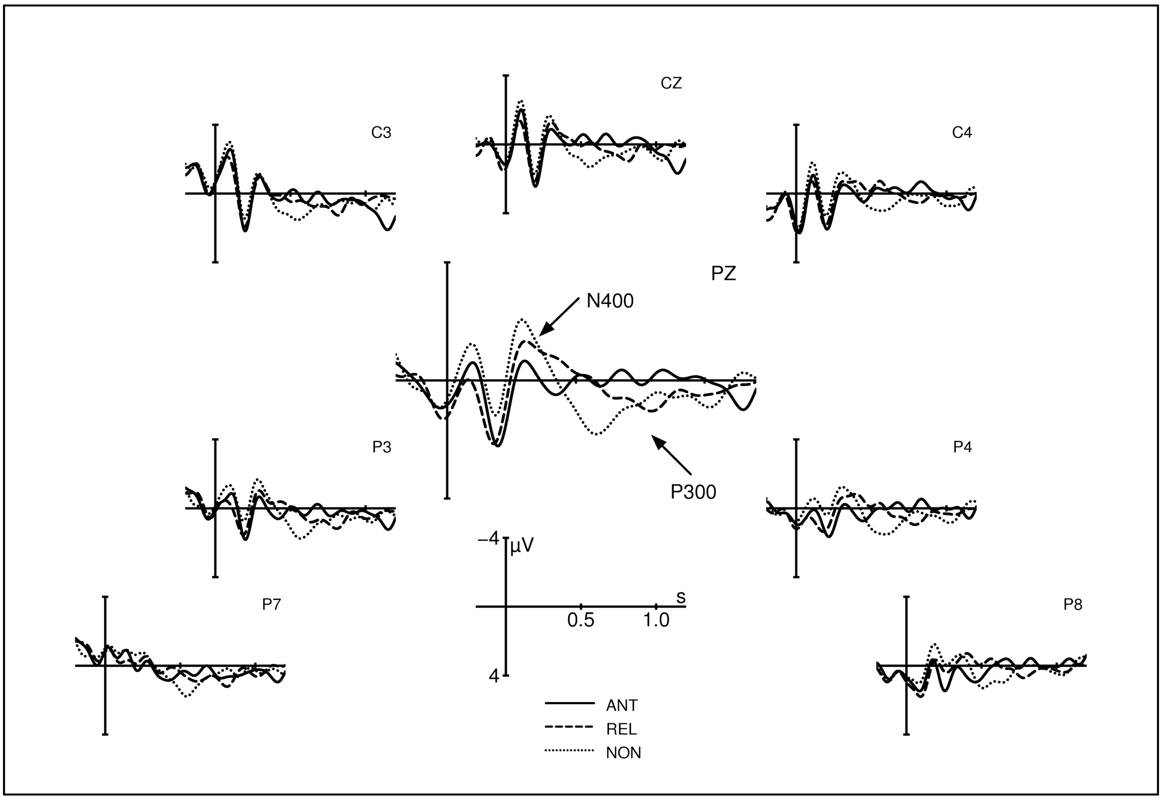


Figure 2. Grand-average ERPs (N = 48) time-locked to the onset of the first fixation on the target word (onset at the vertical bar).

[29] A similar correlation between triggering event and overall shape of ERP waves has also been found for the auditory domain. For example, when the critical manipulation involves the recognition of the word category, the ERPs in response to the violation condition appear less smeared when the word category recognition point is used for averaging. ERPs relative to word onset, by contrast, are less stable (e.g., Van der Brink & Hagoort, 2004; Wolff, Schlesewsky, Hirotani, & Bornkessel-Schlesewsky, 2008).

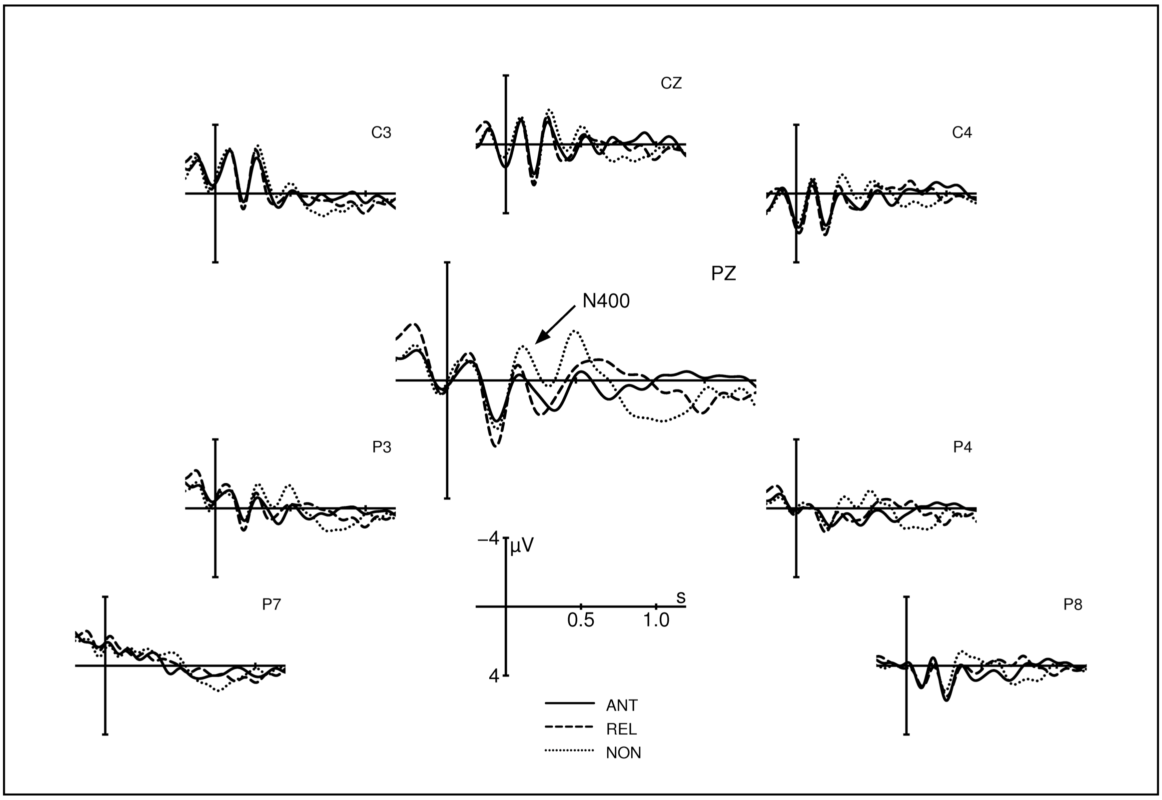


Figure 3. Grand-average ERPs (N = 48) time-locked to the onset of the last fixation before the target word (onset at the vertical bar).

First fixation analyses

N400 time window: 250 ms - 400 ms

For the midline electrodes, there was a main effect of COND ($F(2,94) = 3.16$, $p < .05$) and an interaction between the factors ROI and COND ($F(2,94) = 4.86$, $p < .02$). Resolving the latter for ROI revealed an effect of COND for the parietal ROI (Pz: $F(2,94) = 6.12$, $p < .004$), but not for the central ROI (Cz: $F < 1.03$, n.s.). For the parietal midline ROI, pairwise comparisons of the factor COND showed a marginal contrast between antonyms and related non-antonyms ($F(1,47) = 4.32$, $p < .05$), a highly significant contrast between antonyms and unrelated non-antonyms ($F(1,47) = 13.05$, $p < .0001$), but no difference between related and unrelated non-antonyms ($F < 1.5$, n.s.).

The analysis across lateral ROIs exhibited a largely similar pattern: a very marginal main effect of COND ($F(2,94) = 2.69$, $p < .08$) and a highly reliable interaction between ROI and COND ($F(6,282) = 7.81$, $p < .001$). The interaction resulted from differences in the

COND factor across the right lateralized ROIs (right-central: $F(2,94) = 4.85$, $p < .01$; right-parietal: $F(2,94) = 8.14$, $p < .0001$), whereas left-lateralized ROIs showed no significant effect at all (left-central: $F < 1$; left-parietal: $F(2,94) = 2.45$, $p < .1$). Further pairwise comparisons revealed that related non-antonyms differed from antonyms only in the right-parietal ROI (right-central: $F(1,47) = 2.07$, $p < .2$; right-parietal: $F(1,47) = 7.76$, $p < .008$). Unrelated non-antonyms, by contrast, were significantly different from antonyms across both right-lateral ROIs (right-central: $F(1,47) = 10.44$, $p < .003$; right-parietal: $F(1,47) = 15.54$, $p < .0001$). Finally, there was no difference in any ROI between the two non-antonym conditions (right-central: $F(1,47) = 2.57$, $p < .2$; right-parietal: $F < 1.44$, n.s.).

To sum up, the statistical tests confirm the visual impression that both non-antonym conditions give rise to an enhanced N400 relative to the antonym condition when the critical word is first fixated. In contrast to previous studies, the antonym condition did not elicit a P300, and the amplitude of the N400 did not differ between related and unrelated non-antonym conditions.

Late positivity time window: 450 ms - 860 ms

The analyses for the midline ROIs registered two main effects of ROI ($F(1,47) = 4.99$, $p < .04$) and COND ($F(2,94) = 14.11$, $p < .001$) and no reliable interaction ($F(2,94) = 1.74$, $p < .2$). Pair-wise comparisons showed a significant difference between antonyms and related non-antonyms ($F(1,47) = 7.14$, $p < .02$) and between antonyms and unrelated non-antonyms ($F(1,47) = 26.25$, $p < .001$), with antonyms being less positive. Unrelated non-antonyms words differed from the related non-antonyms ($F(1,47) = 7.54$, $p < .01$) in being more positive.

A similar pattern emerged with the lateral ROIs, where the main effects of ROI ($F(3,141) = 23.42$, $p < .001$) and COND ($F(2,94) = 11.39$, $p < .001$) were reliable. In addition, the interaction between them was significant ($F(6,282) = 3.48$, $p < .01$). Resolving the interaction by ROI revealed significant effects of COND in every ROI (left-central: $F(2,92) = 6.06$, $p < .01$; right-central: $F(2,94) = 7.19$, $p < .01$; left-parietal: $F(2,94) = 9.71$, $p < .0001$; right-parietal: $F(2,94) = 12.19$, $p < .001$). When the levels of COND were compared separately, antonyms differed from related non-antonyms only in the left-lateral ROIs (left-central: $F(1,47) = 9.13$, $p < .01$; right-central: $F(1,47) < .1$; left-parietal: $F(1,47) = 6.58$, $p < .02$; right-parietal: $F(1,47) = 2.04$, $p < .2$). Unrelated non-antonyms differed bi-laterally from the antonyms and right-laterally from the related non-antonyms (NON vs. ANT: left-central: $F(1,47) = 8.56$, $p < .01$; right-central: $F(1,47) = 8.63$, $p < .01$; left-parietal: $F(1,47) = 18.46$, $p < .0001$; $F(1,47) = 24.66$, $p < .001$; NON vs. REL: left-central: $F(1,47) < 1$; right-central: $F(1,47) = 11.91$, $p < .01$, left-parietal: $F(1,47) = 3.62$, $p < .07$; right-parietal: $F(1,47) = 10.30$, $p < .01$).

Taken together, both non-antonym conditions engendered a P300 following the N400. These P300 components varied topographically, with related non-antonyms more positive over left-lateral ROIs, and unrelated non-antonyms more positive over right-lateral ROIs. Figure 2 shows that the positivity for related non-antonyms is smaller in amplitude and exhibits a delay in peak latency when compared to the positivity for unrelated non-antonyms. Therefore, the peak amplitude and latency were computed in addition to the mean amplitude measure reported above.

Late positivity time window: peak amplitude 450 ms – 860 ms

Table 5 gives the mean peak amplitude and peak latency for each of the three conditions, collapsed across participants.

	Peak amplitude (µV)	**Peak latency (ms)**
Antonym (ANT)	4.12 (1.35)	637 (83)
Related (REL)	4.59 (1.75)	662 (75)
Unrelated (NON)	5.10 (1.81)	622 (75)

Table 5. P300 peak amplitude and latency. Standard deviation is given in parentheses.

The statistical test for the peak amplitude at midline ROIs only revealed a reliable main effect of COND (ROI: $F(1,47) = 1.87$, $p < .2$; COND: $F(2,94) = 8.12$, $p < .0001$; ROI x COND: $F < 1$). Single comparisons showed that the related conditions and the antonym condition did not differ from each other, while the related and the unrelated condition differed marginally (ANT vs. REL: $F(1,47) = 3.26$, $p < .08$; REL vs. NON: $F(1,47) = 4.62$, $p < .04$). There was a clear difference between antonyms and unrelated target words, with the latter exhibiting a significantly more positive ERP wave ($F(1,47) = 17.27$, $p < .0001$).

The analysis for the lateral ROIs registered two main effects of ROI and COND, but the interaction failed to reach significance (ROI: $F(3,141) = 7.00$, $p < .01$; COND: $F(2,94) = 4.71$, $p < .02$; ROI x COND: $F(6,282) = 1.09$, $p < .4$). The main effect of COND was driven by a significant contrast between the antonym and the unrelated non-antonym condition; the contrasts in the other pair-wise comparisons fell short of significance (ANT vs. REL: $F(1,47) = 3.11$, $p < .09$; ANT vs. NON: $F(1,47) = 10.05$, $p < .01$; REL vs. NON: $F(1,47) = 1.80$, $p < .2$).

There were no effects for midline ROIs in peak latency analyses (ROI: $F(1,47) = 1.73$, $p < .2$; COND: $F(2,94) = 1.84$, $p < .2$; ROI x COND: $F < 1$). The main effect of COND, however, was reliable for the lateral ROIs (ROI: $F < 1$; COND: $F(2,94) = 3.78$, $p < .03$; ROI x. COND: $F < 1$). Crucially, there was a significant difference in peak latency between related and unrelated non-antonyms, while the other single comparisons did not approach significance (Ant vs. REL: $F(1,47) = 2.66$, $p < .2$; ANT vs. NON: $F(1,47) = 1.02$, $p < .4$; REL vs. NON: $F(1,47) = 8.45$, $p < .01$).

Altogether, the analyses for the peak amplitude and latency confirmed visual inspection: the P300 for related non-antonyms peaked reliably later than the P300 for unrelated non-antonyms at lateral electrode sites. Related non-antonyms tended to have smaller amplitudes at midline electrode sites. Finally, unrelated non-antonyms were significantly more positive than antonyms the chosen time window. It is not surprising that these results are somewhat weaker than those presented for the mean amplitude values. The mean amplitude reductions in the case of unrelated non-antonyms may either be a result of real amplitude modulations or a larger variability in peak latency. This variability is more detrimental to peak measures because it necessarily renders the data more heterogeneous, and significance levels will thus be smaller.

Last Fixation

N400 time window: 250 ms - 400 ms

Statistical tests for the midline electrode sites registered a main effect of ROI ($F(1,47) = 23.19$, $p < .001$) and a main effect of COND ($F(2,94) = 4.14$, $p < .02$), but no interaction between these two factors ($F < 1$). Pair-wise comparisons involving the factor COND revealed the following pattern: while there was no difference between antonyms and related non-antonyms ($F < 1$), unrelated non-antonyms differed significantly from antonyms ($F(1,47) = 5.07$, $p < .03$) as well as from related non-antonyms ($F(1,47) = 6.57$, $p < .02$).

This pattern was replicated in the analyses for the lateral electrode sites. Here, the omnibus ANOVA registered a main effect of ROI ($F(3,141) = 43.89$, $p < .001$), a main effect of COND ($F(2,94) = 6.22$, $p < .01$) , and no interaction ($F < 1$). Pair-wise comparisons of the three levels of the factor COND revealed that antonyms and related non-antonyms did not differ from each other ($F < 1$), whereas unrelated non-antonyms led to significantly different ERP waves than both the antonyms ($F(1,47) = 8.87$, $p < .01$) and the related non-antonyms ($F(1,47) = 8.90$, $p < .01$).

In sum, the statistical tests were able to confirm the enhanced negativity for the unrelated non-antonyms, as opposed to the other two conditions. It appears that both the expected antonyms and semantically related non-antonyms show an N400 with reduced amplitude (relative to unrelated non-antonyms), and that this effect is based on parafoveal information. Critically, there was no hint of a P300 for the antonym condition.

Eye movement results

Table 6 gives the mean fixation durations for the pretarget and the target region. The statistical tests for the fixations corresponding to the ERP data reported above are described in the following paragraphs. Additional analyses that are thought to rule out some confounds with regard to the last fixation measure are reported in Appendix B.

Target word

Antonyms clearly led to shorter *first fixation* durations as opposed to the two non-antonym conditions, which only differed by 3 ms. The main effect of COND was highly significant ($F_1(2,94) = 12.02$, $p < .001$; $F_2(2,78) = 10.54$, $p < .001$). Pair-wise comparisons revealed that the antonym condition differed significantly from both the related and unrelated non-antonym conditions (ANT vs. REL; $F_1(1,47) = 22.45$, $p < .001$; $F_2(1,39) = 16.72$, $p < .001$; ANT vs. NON: $F_1(1,47) = 12.83$, $p = .001$; $F_2(1,39) = 14.10$, $p = .001$), which did not differ from one another (REL vs. NON: all Fs < 1). Antonyms led to *gaze durations* that were about 40 ms shorter than gaze durations for the two non-antonym conditions. This was reflected in the highly significant main effect of COND ($F_1(2,94) = 25.50$, $p < .001$; $F_2(2,78) = 13.32$, $p < .001$). Since comparisons revealed that antonyms differed from non-antonyms, which did not differ from each other (ANT vs. REL; $F_1(1,47) = 40.71$, $p < .001$; $F_2(1,39) = 22.89$, $p < .001$; ANT vs. NON: $F_1(1,47) = 29.89$, $p < .001$; $F_2(1,39) = 16.03$, $p < .001$; REL vs. NON: $F_1(1,47) = 1.49$, $p < .2$; $F_2 < 1$).

	Pretarget region *the opposite of*			**Target word (word *n*)** *white/yellow/nice*		
	Landing position[30]	Launch site[31]	Last fixation	First fixation	Gaze duration	Landing position
Antonym (ANT)	3.6 (1)	7.7 (3)	184 (29)	213 (36)	248 (48)	2.2 (1)
Related (REL)	3.5 (1)	7.6 (3)	190 (27)	232 (33)	292 (54)	2.3 (1)
Unrelated (NON)	3.3 (1)	7.6 (3)	186 (24)	229 (36)	285 (61)	2.1 (1)

Table 6. Participant mean fixation durations (in ms) and positions (in characters) by condition and region. Standard deviation is given in parentheses.

Clearly, these effects mirror the ERP results elicited by the first fixation on the target word. The two non-antonym conditions, which gave rise to enhanced N400 amplitudes, also produced increased fixation durations, as revealed in the first fixation and gaze duration measures. It is also notable that the eye movement record mirrors language-related ERPs exclusively, while the P300 results don’t appear to have a correlate here. For example, one would expect increased fixation durations for the related non-antonym condition relative to the unrelated non-antonym condition. This was only the case for eye movement measures indicating later processing stages, e.g. total time (see Appendix C for statistical analyses).

Last fixation in the pretarget region

There were no differences in the duration of the last fixation, as differences between the conditions did not exceed 6 ms. The statistical tests for the last fixation prior to the target

[30] Landing position is measured relative to the beginning of the region.
[31] Launch site is measured relative to the end of the region.

word revealed no reliable effects ($F_1(2,94) = 1.31$, $p < .3$; $F_2(2,78) = 1.23$, $p < .3$). The parafoveal N400 effect found for the unrelated condition therefore did not lead to increased fixation duration. In other words, there was no parafoveal-on-foveal effect in the eye-tracking record.

Landing position in the pretarget region and the target region

There were no reliable effects for the position of the first fixation in the pretarget region ($F_1(2,94) = 2.50$, $p < .1$; $F_2(2,78) = 1.27$, $p < .3$). Neither were any differences in the target region found ($F_1(2,94) = 1.45$, $p < .3$; $F_2(2,78) = 1.98$, $p < .2$).

Launch site in the pretarget region

There were no significant effects for the position of the last fixation in the pretarget region (all Fs < 1). Although the last fixation does not necessarily coincide with the position of the launch site across all trials, its position must have been sufficiently remot from the target word to avoid foveal processing. So, the parafoveal N400 effect cannot result from differences in fixation position.

4.2.3 Discussion

The present experiment, in which ERPs and eye movements were concurrently recorded while participants read sentences for comprehension, has led to several intriguing findings. These findings bear important implications regarding the interpretation of both ERP and eye movement data in language comprehension studies. Here, the antonym paradigm was used to assure maximum comparability with results from prior studies. Participants read sentences such as *dass schwarz das Gegenteil von weiß ist, hat Gertrud gestern behauptet* ('Gertrud claimed yesterday that black is the opposite of white'), in which the second antonym is expected given the preceding sentential context. All other words occurring at the position of the expected antonym share a cloze probability of zero. In addition to this cloze probability distinction this, it is possible to distinguish the unexpected targets based on whether they are semantically related to the prime, i.e. the first antonym. Previous ERP results in the visual domain (i.e., using RSVP) and the findings from the auditory experiment (see chapter 4.1) revealed distinct brain responses for antonyms as opposed to unrelated or semantically related non-antonyms. Antonyms engendered a P300 that peaked in the same interval as the N400 elicited by the two non-antonym conditions. These also evidenced a graded N400 effect, as the N400 amplitude for unexpected and semantically unrelated words was more pronounced than for unexpected and related words. Following the N400, both non-antonyms engendered a P300, the amplitude of which was somewhat reduced for semantically related target words.

The predictions for the experiment came from two general accounts of how the processing of related and unrelated words could differ; these accounts have been developed in chapter III to derive predictions for the experiment. In the decision-making account, it was hypothesized that there may be little or no P300 activity for antonyms, whereas the P300 for each of the non-antonyms would follow an N400. The presence or absence of the antonym-P300 was assumed to depend on the type of linguistic information that enables the decision-making process. If prelexical orthographic information alone is sufficient to elicit the P300, there should be a small parafoveal P300 to antonyms. The amplitude of this P300 should also be reduced, due to the lower perceptual distinctiveness of parafoveal stimuli. If, however, the P300 is elicited only if sufficient information from both the prelexical orthographic and the semantic level is collected, then antonyms shouldn't elicit a P300 at all. This is because the language processing system makes a prediction about a specific word form and a specific meaning corresponding to the expected antonym. At either fixation, however, information from one level in the word recognition process is missing, so that the decision on the antonym is characterized by insufficient evidence. Neither of the hypotheses predicts the occurrence of a foveal P300, but only the former predicts a parafoveal P300. If it should be the case that antonyms do not elicit a P300, one would be able to detect a parafoveal N400 priming effect, since there would be no component-overlapping to obscure it. The amplitudes for antonyms and related non-antonyms would then be reduced. These amplitudes might also differ from each other, provided antonyms receive higher pre-activation. In the attention-based account, attention alone was assumed to drive the P300 genesis. If attention in reading is strictly serial, very little or no attention would be allocated to parafoveal words and there wouldn't be any parafoveal ERP effects. As a consequence, there wouldn't be any differences to prior ERP results. If, by contrast, attention is allocated to several words in parallel, a parafoveal P300 is very likely because lexical processing of highly predictable words begins in the parafovea. Parafoveal N400 effects would also be possible under these circumstances.

When ERPs were time-locked to two successive fixations in natural reading the following pattern emerges: on the last fixation before the eyes move to the critical word (i.e. *white* in the example above), there are reduced N400 amplitudes for antonyms and related non-antonyms (e.g., *blue*). When the eyes first fixate the critical word, the two non-antonym conditions give rise to enhanced N400 amplitudes, indistinguishable from each other. In addition, the non-antonym conditions engender a P300, with its amplitude and latency varying according to the degree of semantic relatedness. Crucially, the ERP waves suggest that antonyms do *not* trigger a P300 in the present experiment. Note that probability of occurrence cannot account for this. All critical sentences were embedded in 150 filler sentences and thus, their probability of occurrence is quite low. Furthermore, if probability had influenced the P300, non-antonyms wouldn't elicit a P300 either because the three conditions occurred with equal probability. The fixation results show a somewhat deviant pattern that differs from the ERP results in showing no effect corresponding to the parafoveal N400. The foveal N400 effect, however, is clearly mirrored by the first fixation and gaze duration on the critical word. Both non-antonym conditions caused longer fixation durations than the antonym condition.

The results from Experiment 2 suggest that the purely attention-based account is less suited to explain the data pattern as a whole. The serial-attention hypothesis can account for the foveal N400 priming effect and the lack of a parafoveal P300. However, the emergence of a parafoveal N400 priming effect and the missing foveal P300 are not compatible with this account. This finding suggests that, contrary to the assumptions of serial models of eye movements in reading, more attention is allocated to parafoveal information and that parafoveal information is not merely processed in a pre-attentive manner (see chapter 2.3.2 above for explanation). The present results are also only partly compatible with the parallel attention hypothesis, which states that attention is divided over several words and that parafoveal words can be lexically processed, especially when they are highly predicted by the processor. If highly predictable words had already been processed in the parafovea, antonyms should have given rise to a parafoveal P300. This was not the case. The account does correctly predict parafoveal N400 effects. However, these effects do not necessarily reflect lexical processing when they are induced by unpredictable words (see chapter 2.3.2 for details). This prediction was in fact correct, as related non-antonyms did evidence a parafoveal N400 priming effect, which was caused by the spread of activation between orthographic forms.

The interaction between the decision-making account for the P300, specific predictions for upcoming words, and the spreading activation theory, on the other hand, seems to be able to account for the present data as a whole. Consider first the parafoveal N400 effect, which can be interpreted as reflecting priming due to spreading activation between related orthographic forms. In the spreading-activation theory proposed by Collins & Loftus (1975), semantic memory is conceived of as a network of semantic concepts ordered along the lines of semantic relatedness and semantic association. Semantic concepts are represented by distinct nodes. Each node is connected to other nodes based on inherently shared semantic features or co-occurrence probability. For example, the color term *red* is strongly connected to other color terms such as *orange* or *yellow*. It is also connected to the semantic node *fire*, but this connection is weaker than the relation to color terms. The *fire* node, in turn, is connected to other concepts such as *fire engine* or *house* and so forth (cf. Collins & Loftus, 1975, pp. 411). The crucial point is that concepts are more closely connected when they share more properties. The strength of this interconnection between two nodes determines the degree of priming facilitation. It is assumed that activation spreads out automatically from a concept (e.g., the prime in a semantic-priming context) when it is processed. The fact that activation decreases gradually over time and with intervening activity suggests that a higher number of activated (i.e., primed) concepts leads to a smaller amount of activation for each concept. Another assumption is that the semantic matching process relies on the collection of positive and negative evidence. Collins & Loftus (1975) claim that in order to decide whether two concepts match each other, enough positive or negative evidence has to be gathered (i.e., until the evidence reaches a critical threshold). The weighting of positive and negative evidence is asymmetrical so that a recognized mismatch on one property between the concepts can lead to a negative decision, whereas more than one property of the concepts must match for a positive decision to be taken.

The following process is conceivable in the generation of the parafoveal N400 priming effect. When the prime is encountered, activation automatically spreads to all related nodes (cf. Figure 4a). However, the pretarget region *the opposite of* constitutes negative evidence for the related non-antonyms. Their activation level drops because the pretarget region narrows down the possible set of expected sentence completions, so that only the proper antonym is expected (Figure 4b).

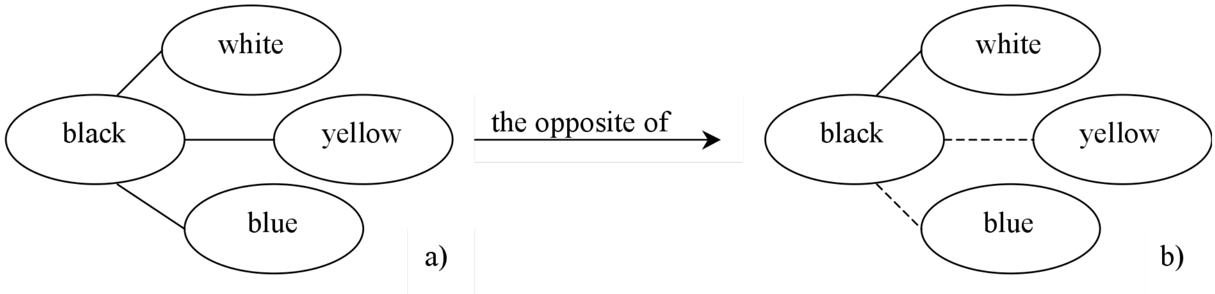


Figure 4. Scheme for the proposed spread of activation during reading sentences like *black is the opposite of white*. 4a) shows activation levels when the first antonym (*black*) is fixated, 4b) shows decrease in activation levels for non-antonyms that are semantically related (dashed lines), when the pretarget region is read.

Although the sentence context immediately before the target word impedes further activation for any non-antonym concept, there is still a distinction between related non-antonyms and unrelated non-antonyms. It appears that the residual activation for unexpected related words is the crucial determinant. Despite their unexpectedness, unrelated non-antonyms are primed in the same way when their orthographic form is perceived in the parafovea, as both antonymic word forms and word forms of semantically related words show a reduced parafoveal N400 amplitude of similar size.[32] This may be surprising given that antonyms should be more activated because they share more features with the prime than related words. However, the parafoveal priming effect suggests that the automatic spread of activation enhances activation for all related non-antonyms, but in doing so, it is restricted to the orthographic domain. The parafoveal priming effect shows that, while the processing of pre-activated strings is facilitated, lexical-semantics plays no important role at this stage in word recognition, i.e.

[32] Collins & Loftus (1975) argue for a separate lexical network that is associated with the semantic network. In the lexical network all phonological and orthographic information is stored, and importantly, lexical nodes may be associated to one or more corresponding nodes in the semantic network. Moreover, the authors propose that priming across both networks can operate in at least three different ways. For example, when perceiving the prime *black*, priming may be limited to the lexical network, producing facilitation for words that are phonologically or orthographically similar to the prime. It may, however, also prime exclusively concepts in the semantic network, which are related to the semantic concept of the prime. A third possibility is that priming may operate in both networks together. The present parafoveal N400 priming effect suggests that pre-activation in the semantic network is transferred to the lexical network, where the orthographic form associated with the preactivated semantic concept is also activated. Thus, the parafoveal N400 is not a result of orthographic neighborhood effects, but of two networks sharing activation levels.

there may be no, or not enough, negative feedback from the superordinate semantic level to distinguish related word forms from antonyms. This suggests further that there was no parafoveal semantic processing involved in producing the parafoveal N400 priming effect. Semantically unrelated non-antonyms, by contrast, cannot benefit from a pre-activation of orthographic features, so their parafoveal N400 can be considered an index of normal processing effort induced by a lack of lexical associates in the preceding context or sentence (see Neely, 1977, who argues that spreading activation cannot produce inhibition effects). Similar findings have been reported recently (Laszlo & Federmeier, 2009) in a study in which orthographic neighborhood effects between expected sentence completions and unexpected words, pseudowords, or non-words were investigated. The data clearly show that the N400 is reduced to strings that are orthographically related to the expected word. A critical point is that this effect did not interact with the type of string presented (i.e., words and the two types of non-words were not distinguishable with the N400 modulation). So, the preactivation of predicted words also extends to orthographic neighbors.

Foveal N400 effects provide further evidence for the distinction of word form and lexical-semantics. Both related and unrelated non-antonyms enhanced the foveal N400 amplitude to an equal extent, which can be taken as reflecting their shared cloze probability of zero. This finding adds to the existing body of research which shows that the amplitude of the N400 is an inverse function of cloze probability (cf. Kutas & Hillyard, 1984b). At the same time, the foveal N400 effects are a striking contrast to other previous findings, i.e., reports that semantically related but unexpected sentence completions show an N400 amplitude that is reduced, as compared to semantically unrelated and unexpected completions (cf. Federmeier & Kutas, 1999; Kutas & Federmeier, 2000; Roehm et al., 2007). This decrement in amplitude has been explained mostly in terms of overlapping features in semantic entries, i.e. since related words share semantic features with the expected completions, their lexical access is facilitated (Federmeier & Kutas, 1999). This facilitation was enforced in high-constraint contexts because predictions about upcoming words were probably more specific and therefore the relatedness between concepts was more helpful in lexical access.

As related target words no longer show a processing advantage over unrelated targets in the present experiment, some mechanism different from spreading activation or semantic-feature overlap must be responsible for this result. The contrast between prior findings and the present foveal N400 may be reconciled as follows. Foveal N400 effects purely mirror the cloze probability of the items. While reading the stimuli, readers develop very specific expectations for only one word, involving a specific antonymic meaning and a specific orthographic form. This top-down mechanism is grounded on previously read sentence parts and focuses solely on the meaning of the target word when it is fixated first. Any sentence completion other than the antonym is a marked deviation from this expectation, and therefore incompatible with the preceding sentence context, regardless of semantic relatedness. While the orthographic prediction results in the parafoveal N400 priming effect, the semantic prediction leads to the foveal N400 priming effect for the expected antonym. This assertion has two important consequences for the interpretation of this and previous results within the antonym paradigm. Specifically, the mechanism held responsible for the foveal N400 effect is de-coupled from the mechanism underlying the parafoveal N400 priming effect. This implies that part of the N400 amplitude reduction reported previously (cf. Experiment 1; Roehm et al., 2007) may result from the temporal co-occurrence of these two mechanisms. RSVP and auditory presentation do not allow for a temporal dissociation between facilitation through pre-activated word forms and facilitation due to predictions about the meaning of upcoming words. From this perspective, the facilitation for related non-antonyms (or, in general, within-category violations) may not represent feature overlap, but correlates with word form pre-activation that is shifted to an earlier point in reading, the parafoveal N400. Natural reading permits this remarkable temporal shift because parts of the word’s visual and phonological codes are processed before the word is fixated initially and its meaning retrieved. In other words, two successively processed stages (prelexical form and lexical-semantics) in word recognition and the independent mechanisms determining the ease of processing for each stage coincide in RSVP and the auditory domain, since these two types of presentation do not have an equivalent to the parafoveal preview in natural reading. Therefore, the parafoveal N400 priming effect for related words and the indistinguishable foveal N400s for related and unrelated target words constitute the lexical-relatedness benefit that has been found in previous ERP experiments in the antonym paradigm. This pattern of results strongly supports the hypothesis that the processing of orthography and semantics is separated temporally in word recognition processes in the present design. As discussed below, the P300 data provides converging evidence for this hypothesis, while disconfirming the claim that orthography alone prompted the P300 response found in previous experiments.[33]

It was argued above that the presence or absence of a parafoveal P300 is decisive as to whether prelexical input alone, or the combination of prelexical and lexical information is responsible for the P300 to antonyms in previous experiments. The present P300 results support the hypothesis that the antonym-P300 is only elicited when prelexical information and lexical information temporally coincide in processing. If it was induced by prelexical orthographic information alone, a parafoveal P300 should have emerged. Although it was predicted that, due to the detrimental effects of parafoveal stimulus degradation, such a parafoveal P300 should exhibit a smaller amplitude compared to previous reports, the absence of the component was not predicted. The parafoveal N400 priming effect demonstrates that parafoveal orthographic information is used for linguistic processes. However, it also shows that all pre-activated word forms are treated alike, regardless of whether their meaning is in fact the expected antonym. Therefore, it is unlikely that orthography alone can trigger antonym recognition – and hence, the P300 – because it is insufficient to provide a sound

[33] Another explanation for the P300 could lie in the lack of experimental power induced by stimulus alterations. This possibility will be examined further in Experiment 3.

basis for full target categorization (i.e., including semantic processing). Semantic information alone also seems inadequate to elicit the P300 response to antonyms. There was no foveal P300 elicited by antonyms in the present experiment. The absence of any significant P300 activity here suggests that the P300 to antonyms (or highly related target words) in RSVP and auditory settings only arises if prelexical input and lexical information become available for processing at the same time. If only one of these domains is available, i.e. if they are temporally dissociated, there is no P300 to antonyms.

The critical question is why both information types have to coincide to elicit a P300. One straightforward answer could be that the orthographic input perceived parafoveally is too degraded to provide strong evidence for target categorization (see chapter 1.2.3). Uncertainty about stimulus identity through stimulus degradation may explain the absence of a parafoveal P300. However, stimulus degradation is less of an issue in foveal vision, so this explanation alone is somewhat implausible to account for the absence of the foveal P300. Nonetheless, the notion of uncertainty in decision-making is worth more detailed consideration. Results from signal-detection experiments show that P300 amplitude correlates positively, and its latency negatively, with the confidence with which an event is categorized (cf. Hillyard et al., 1971; Parasuraman & Beatty, 1980; see also chapter 1.2.2). In these experiments, participants were required to detect simple sensory stimuli despite background noise. Considering that these stimuli were, for example, simple tone bursts which could be identified based on their frequency, it becomes evident that matters are more difficult with language comprehension. Words are multi-dimensional complex objects carrying a wealth of different information types (i.e., phonology, orthography, semantics), and it is not clear if or to what extent they interact in word recognition. A mere summation of evidence from each domain is implausible, as this would predict a foveal P300, since by the time the semantic information is processed, orthographic information would already be available. It is noteworthy that linguistic processing does not stopp when a saccade begins, but the visual information is continuously processed (Irwin, 1998). Orthographic processing might already be finished before the onset of the first fixation on the antonym. This implies that, when an input modality such as reading imposes a processing manner, in which word recognition is dissociated across several distinct time points (as with discrete fixations), there could not be any cascaded processing involving the different domains (for explanation of cascaded processing, see McClelland, 1973; for discussion of serial vs. cascaded processing architectures, see McElree & Griffith, 1995). Although evidence for target identification is enriched for each processed domain, these pieces of evidence may not be summed to exceed the threshold needed to finally categorize that word as the expected target stimulus. Therefore, any single-domain decision about the word is too uncertain to yield a P300 in reading, because evidence from the other domain is missing to meet both predictions. In other words, a decision about the perceived word is made for each fixation position in natural reading, i.e., the predictions about the meaning and orthographic from of an upcoming word are met at different time points. Parafoveal orthographic information is used to meet the prediction about the upcoming word form, while foveal semantic information is used to meet the prediction about the expected meaning. This dissociation of prediction-related decision-making inhibits the occurrence of the P300 at either fixation, because only one prediction is met at a time. Only if both predictions are met within one fixation (as with RSVP) or within one discrete auditory event, the P300 can emerge as an index of successful target categorization.

The foveal P300 effects for the non-antonym conditions offer further insight into the processes underlying the comprehension of target words in the antonym paradigm. The unrelated non-antonyms engendered an N400/P300 pattern that is noticeably similar to prior findings. As argued above, they are the least affected by the pre-activation processes that determine processing facilitation. N400 amplitude to these target words increases as a function of cloze probability, and the P300 follows the N400 component, because lexical access is necessary to categorize the target word as a non-antonym. Interestingly, the P300 peak amplitude was smaller and its peak latency increased in the related non-antonym condition. These effects may eventually help resolve the often observed discrepancy between ERP and behavioral results. Although related non-antonyms showed a processing advantage in terms of reduced N400 amplitude, they caused a great processing disruption when accuracy rates and reaction times were considered (Roehm et al., 2007). Here, the reduced P300 in response to related non-antonyms may index the problems in decision-making. That is, semantic relatedness impedes the categorization of related words as being either antonymic or not, producing an inhibitory effect of semantic neighborhood. The P300 amplitude reduction may therefore be due to two (not mutually exclusive) factors. First, participants could be less confident in their decisions, which is known to reduce P300 amplitudes (Hillyard et al., 1971; Parasuraman & Beatty, 1980). Second, the P300 latency could vary considerably across trials, i.e. in some trials, the decision may occur earlier than in others. This could also account for reduced mean amplitudes and may operate independently of uncertainty. Interestingly, these data suggest that the previously reported amplitude reduction for related targets may not wholly result from component overlap between the P300 and the preceding N400 component (see chapter 1.3.1). With the present results in mind, they may also support the assumption of reduced confidence in decision-making or higher variability in evaluation time.

The first fixation and the first pass time measures revealed longer reading times for the two non-antonym conditions, reflecting their cloze probability of zero. Thus, the eye movement pattern fully replicates the foveal N400 effects. As predicted, fixation times did not differ between semantically related and unrelated non-antonyms. This result supports previous findings on the relatively weak influence of semantic relatedness, which only speeds up reading time if it is additionally supported by a congruent context (Morris, 1994; Calvo & Meseguer, 2002). This was not the case in the present design in which non-antonyms were always incongruent with the preceding sentence context. However, semantic relatedness did affect reading time measures such as total time or go past time, which are assumed to reflect relatively late processing stages (see Appendix C). This suggests that the interference between lexical relatedness and target categorization, as reflected in the P300, affects comprehensive eye movement measures, whereas it does not significantly influence eye movement measures that reflect earlier lexical processing stages. This intriguing finding bears important implications as to the question of which eye movement measures capture linguistic processing proper and which also encompass domain-general processes such as stimulus evaluation (cf. chapter 5.3).

To summarize, the present results highlight the importance of spreading activation in the comprehension of stimuli in the antonym paradigm. Processing facilitation due to pre-activation only impacts those stimuli that are susceptible to spreading activation between related concepts. The parafoveal N400 priming effects caused by spreading activation are thus exclusively facilitative in nature; there is no inhibition for unrelated concepts because they are not pre-activated at any stage of processing. This may explain why the unrelated condition is the only one which exactly mirrors results obtained with RSVP or auditory presentation. Moreover, this experiment is the first to show the *linguistic* processing efforts (i.e., detectable N400 amplitude reduction) for highly expected antonyms, without using a secondary task (e.g., a lexical decision task as in the study of Roehm et al. 2007) to avoid confounding effects due to other cognitive processing correlates. In foveal vision, the N400 to the expected antonym is reduced in amplitude relative to both non-antonym conditions. In a similar vein, the data also suggests that the latency shift of the P300 in the previously reported N400/P300 pattern seems to be entirely dependent on input modality. The P300 to antonyms arises only if information relevant for all predictions are processed in a cascaded manner. This is the case with RSVP and auditory presentation, whereas there seems to be no lexical processing of parafoveal words in natural reading (see the General Discussion for more explanation).

The lack of an antonym-P300 and the parafoveal N400 are the most intriguing findings here, and as such, deserve special scrutiny in terms of possible confounds. One of these could be related to the mixture of critical stimuli and filler sentences, which followed standards of eye movement studies, rather than EEG standards. This issue will be addressed in Experiment 3, in which the identical set of stimuli was presented using the RSVP technique.

4.3 Experiment 3: no P300 in natural reading – a validation with rapid serial visual presentation (RSVP)

The above findings – that natural reading inhibits the occurrence of a parafoveal P300, and that there is no difference in the foveal N400s in the non-antonym conditions – are intriguing. However, these findings need to be elaborated further, as Experiment 2 differed from Experiment 1 and previous ERP experiments on antonym processing, in two important ways: (i) there were fewer items per condition and (ii) the ratio between critical sentences and fillers was significantly different. Specifically, each participant was presented with a total of 10 target stimuli in each condition. There were furthermore 150 filler sentences involving various structures. This is a significant deviation from the typical psycholinguistics EEG design that may have affected experimental power. In fact, EEG experiments investigating language processing present participants with usually no less than 40 items per condition to accommodate for the low signal-to-noise ratio (see chapter 1.1.1). Experiment 1, for example, presented each participant with 80 target stimuli, 80 non-target stimuli, and without any fillers. Thus, the design used in Experiment 2 may have suffered from too few critical items and the resulting low signal-to-noise ratio.

Experiment 3 was therefore conducted with the same stimulus set used in Experiment 2 (including fillers). The predictions are straightforward. If the low signal-to-noise ratio (or the small number of critical stimuli) is decisive, Experiment 3 should pattern with the foveal ERP results of Experiment 2. That is, antonyms should show no P300 and there should also be no difference in the amplitude of the N400 between the non-antonym conditions. If, however, the results from Experiment 2 are due to the uniqueness of the parafoveal preview in reading and its peculiar interaction with task-relevant processing in word recognition, Experiment 3 should differ from Experiment 2. In this case, Experiment 3 is assumed to fully replicate Experiment 1 because neither the auditory modality nor the RSVP method provides a preview of upcoming words.

4.3.1 Method

Participants

48 students from the University of Mainz (25 female; mean age: 24, range: 20-27) were paid to participate in this experiment after giving informed consent. None of them had participated in Experiments 1 or 2. All participants were right-handed native speakers of German, as assessed by the German version of the Edinburgh handedness test (Oldfield, 1971). One further participant had to be excluded due to low performance in the comprehension task. All participants reported normal or corrected-to-normal vision and no history of any neurological or psychiatric disorders. They were naïve to the purpose of the experiment.

Materials

The stimuli, filler sentences, and allocation to experimental lists were identical to Experiment 2.

Apparatus and Procedure

The EEG system, EEG and EOG recordings, and electrode configuration and preparation are identical to those in Experiment 2. Participants were seated in a dimly lit room. At the beginning of each trial, a fixation point was displayed in the centre of the screen for 400 ms and followed by a 200 ms blank screen, before the first word of a sentence was presented. The critical stimuli were displayed word by word. Words were presented with a duration of 300 ms, followed by an interstimulus interval (ISI) of 200 ms. Noun phrases in the filler sentences were presented for 400 ms. A blank screen was presented for 500 ms after the sentence-final word. The entire comprehension question was then presented for 4000 ms. The intertrial interval (ITI) was 1000 ms. Participants were instructed to respond yes or no to the comprehension question by pressing a button on a game pad. The distribution of yes and no answers to the buttons was counterbalanced across all participants. The participants were randomly assigned to one of the 4 experimental lists. The experimental session lasted about 2 hours.

Data preparation and analyses

The average ERP waves were calculated per subject per condition 200 ms before the onset of the critical word to 1200 ms post-onset. Grand averages were then computed over all participants. The EEG was filtered using a 0.3- to 20-Hz bandpass filter to avoid slow signal drifts in the EEG. The grand average ERPs displayed in the figures were smoothed with a 8-Hz lowpass filter, but all analyses were calculated with unfiltered data. Trials for which the comprehension question was answered incorrectly were excluded from the averaging, as were trials contaminated by ocular or amplifier saturation artifacts (the EOG rejection threshold was 40 μV).

Repeated-measures ANOVAs were calculated for the mean amplitudes per time window and per condition involving the factor COND (ANT vs. REL vs. NON). Regions of interest (ROIs) for the factor ROI were defined as follows: there were 3 ROIs for the midline electrode sites (Fz, Cz, Pz), and 6 ROIs for the lateral electrode sites (left-frontal: F7, F3; right-frontal: F8, F4; left-central: T7, C3; right-central: T8, C4; left-posterior: P7, P3; right-posterior: P8, P4). To avoid type I errors resulting from violations of sphericity, alpha was corrected as proposed by Huynh and Feldt (1970) when effects with more than one degree of freedom in the numerator were evaluated. In these cases, the original degrees of freedom are reported with the corrected alpha. Whenever statistical tests justified pair-wise comparisons involving the factor COND, these were adjusted according to a modified Bonferroni procedure (Keppel, 1991) to avoid overestimating the comparisons. Alpha was set to $p < .033$ in these cases.

For the behavioral data, one-factorial repeated-measures ANOVAs with the factor COND (ANT vs. REL vs. NON) were calculated for error rates and reaction times in each condition, treating subjects and items as random factors F_1 and F_2. Pair-wise comparisons were also adjusted according to the modified Bonferroni procedure.

4.3.2 Results

Behavioral data

The error rates and reaction times of the comprehension task are given in Table 7.

	Error rates (% correct)	Reaction time (ms)
Antonyms (ANT)	87.3 (9)	1648 (289)
Related (REL)	79.8 (14)	1729 (298)
Unrelated (NON)	86.9 (10)	1679 (303)

Table 7. Mean participant error rates and reaction time. Standard deviation is given in parentheses.

Participants made more correct answers in the antonym and the unrelated non-antonym conditions than in the related non-antonym condition. The main effect of COND was reliable across participants and marginal across items ($F_1(2,94) = 8.56$, $p < .001$; $F_2(2,78) = 2.86$, $p < .07$). Pair-wise comparisons showed that the related non-antonym condition was associated with marginally higher error rates than each of the other conditions, which in turn did not differ from each other (ANT vs. REL: $F_1(1,47) = 10.84$, $p < .01$; $F_2(1,39) = 3.82$, $p < .06$; ANT vs. NON: all Fs < 1; REL vs. NON: $F_1(1, 47) = 13.48$, $p = .001$; $F_2(1, 39) = 3.76$, $p = .06$).

The related non-antonym condition also exhibited longer reaction times than the other two conditions, resulting in a marginal main effect of COND ($F_1(2,94) = 3.05$, $p < .06$; $F_2 (2,78) = 3.16$, $p < .05$). Pair-wise comparisons indicated that only the contrast between the related non-antonym condition and the antonym condition was statistically reliable (ANT vs. REL: $F_1(1,47) = 5.61$, $p < .03$; $F_2(1,39) = 5.36$, $p < .03$; ANT vs. NON: $F_1(1, 47) = 1.03$, $p < .4$; $F_2 < 1$; REL vs. NON: $F_1(1, 47) = 2.11$, $p < .2$; $F_2(1, 39) = 2.61$, $p < .2$). Hence, the related non-antonym condition leads participants to make more errors and it also increases the time participants take to answer a question. The reduced significance of the results may be due to the low number of items used and to the fact that the questions probed various parts of the sentence. Nevertheless, these behavioral results mirror those reported for Experiment 1.

ERP data

Visual inspection of the ERP waves presented in Figure 5 suggests the emergence of three main components in response to the critical conditions. The ERP waves begin to diverge at about 250 ms post word onset, with the antonym condition exhibiting a pronounced positivity that peaks at about 300 ms. Contrary to this, both non-antonym conditions engender a large negative deflection peaking at about 380 ms. The distribution of these two components revealed a clear anterior-posterior distinction, such that they are more pronounced at central-parietal sites. In addition, the related and unrelated non-antonym conditions are more positive going than the antonym condition between 460 ms and 650 ms post word onset. The time windows for the statistical tests on mean amplitudes were chosen in accordance with the visual inspection results: 250 ms to 400 ms for the P300/N400 time window and a 460 ms to 650 ms time window for the late positivity.

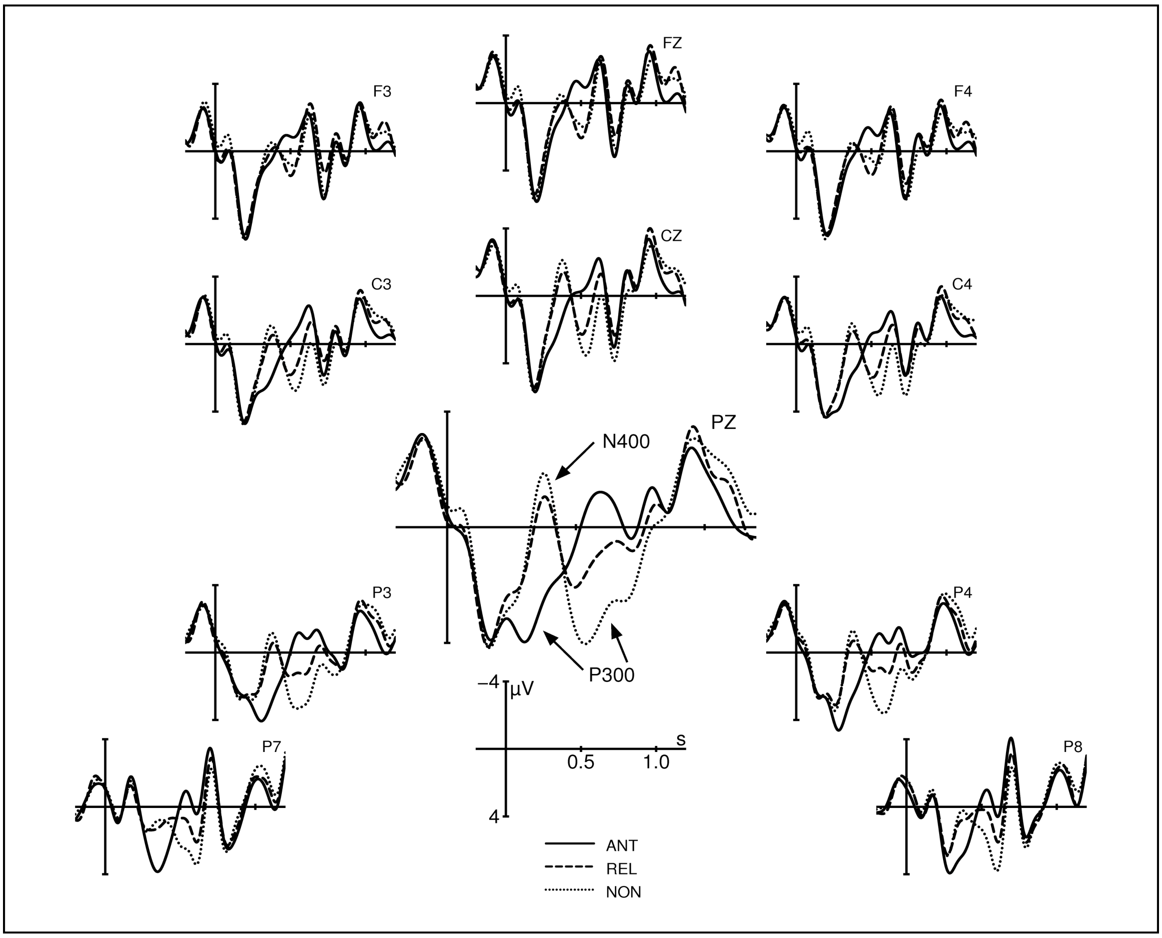


Figure 5. Grand-average ERPs (N = 48) time-locked to the onset of the target word (onset at the vertical bar).

P300/N400 time window: 250 ms – 400 ms

The ANOVAs for the midline electrode sites registered a main effect of COND ($F(2,94) = 15.60$, $p < .001$) and an interaction between factors ROI and COND ($F(4,188) = 14.15$, $p < .001$). When the interaction was resolved for the factor ROI, there were significant effects of COND at each electrode site (frontal: $F(2,94) = 3.80$, $p < .03$; central: $F(2,94) = 16.96$, $p < .001$; parietal: $F(2,94) = 21.34$, $p < .001$). Furthermore, single comparisons involving the factor COND revealed that antonyms differed from related words and from unrelated non-antonyms at all three electrodes (ANT vs. REL: frontal: $F(1,47) = 5.01$, $p < .03$; central: $F(1,47) = 20.17$, $p < .001$; parietal: $F(1,47) = 26.34$, $p < .001$; ANT vs. NON: frontal: $F(1,47) = 5.27$, $p < .03$; central: $F(1,47) = 22.96$, $p < .001$; parietal: $F(1,47) = 28.83$, $p < .001$), while the two non-antonym conditions did not significantly differ from each other (REL vs. NON:

all Fs < 1). It was also evident that these effects were more pronounced at central-parietal electrodes.

The analyses for the lateral ROIs largely mirrored the pattern for midline sites. The main effects of ROI ($F(5,235) = 11.81$, $p < .001$) and COND ($F(2,94) = 13.44$, $p < .001$) were reliable, as was the interaction between the two factors ($F(10,470) = 15.96$, $p < .001$). Resolving the interaction by ROI revealed that COND yielded significant effects at all electrode sites except the left and right frontal ROIs (left-frontal: $F < 1$; right-frontal $F < 1$; left-central: $F(2,94) = 12.53$, $p < .001$; right-central: $F(2,94) = 14.02$, $p < .001$; left-parietal: $F(2,94) = 23.89$, $p < .001$; right-parietal: $F(2,94) = 24.45$, $p < .001$). Single comparisons involving the factor COND yielded significant results for the comparison of antonyms vs. related non-antonyms (left-central: $F(1,47) = 12.25$, $p = .001$; right-central: $F(1,47) = 16.97$, $p < .0001$; left-parietal: $F(1,47) = 27.97$, $p < .001$; right-parietal: $F(1,47) = 29.06$, $p < .001$) as well as for antonyms vs. unrelated non-antonyms (left-central: $F(1,47) =19.51$, $p < .0001$; right-central: $F(1,47) = 20.42$, $p < .001$; left-parietal: $F(1,47) = 32.59$, $p < .001$; right-parietal: $F(1,47) = 30.87$, $p < .001$). Related and unrelated non-antonyms, however, did not differ statistically from each other (left-central: $F(1,47) = 1.45$, $p < .3$; right-central: $F < 1$; left-parietal: $F(1,47) = 1.07$, $p < .4$; right-parietal: $F(1,47) = 1.01$, $p < .4$). Again, the reliable effects between antonyms and non-antonyms appeared to be stronger over parietal sites.

In light of the converging evidence that a high-constraint context bears facilitative effects on unexpected but semantically related non-antonyms, the apparent similarity in N400 amplitude between related and unrelated non-antonyms is surprising. Although visual inspection of the respective amplitudes reveals a very small difference between 330 and 380 ms, this visual impression was not confirmed by statistical tests for either the mean amplitude or the peak amplitude (see Appendix D for statistical reports). To summarize, the antonym condition engendered a P300 in this time window, whereas both non-antonym conditions gave rise to indistinguishable N400 effects.

Late positivity time window: 460 ms – 650 ms

Over midline electrodes, both main effects of ROI ($F(2,94) = 47.98$, $p < .001$) and COND ($F(2,94) = 27.20$, $p < .001$) as well as the interaction ($F(4,188) = 22.02$, $p < .001$) yielded significant results. When the interaction was resolved for ROI, there were reliable effects of COND at each electrode (frontal: $F(2,94) = 10.72$, $p < .0001$; central: $F(2,94) = 28.85$, $p < .001$; parietal: $F(2,94) = 38.71$, $p < .001$). Single comparisons for the factor COND at each electrode further showed that the antonym condition was significantly different from the related non-antonym condition (frontal: $F(1,47) = 18.28$, $p < .0001$; central: $F(1,47) = 20.27$, $p < .001$; parietal: $F(1,47) = 18.26$, $p < .0001$) and from the unrelated non-antonym condition (frontal: $F(1,47) = 17.10$, $p < .0001$; central: $F(1,47) = 55.99$, $p < .001$; parietal: $F(1,47) = 74.35$, $p < .001$). The single comparison of the related and unrelated non-antonyms also revealed a significant difference at central and parietal sites (frontal: $F < 1$; central: $F(1,47) = 9.85$, $p < .01$; parietal: $F(1,47) = 21.16$, $p < .001$).

This pattern was replicated at the lateral electrode sites. The main effects of ROI ($F(5,235) = 41.97$, $p < .001$) and COND ($F(2,94) = 22.86$, $p < .001$) were reliable. There was also a significant interaction ($F(10,470) = 23.88$, $p < .001$), which yielded reliable effects of COND at all electrodes locations except the left-frontal ROI (left-frontal: $F(2,94) = 1.25$, $p < .3$; right-frontal: $F(2,94) = 3.30$, $p < .05$; left-central: $F(2,94) = 25.17$, $p < .001$; right-central: $F(2,94) = 27.16$, $p < .001$; left-parietal: $F(2,94) = 37.45$, $p < .001$; right-parietal: $F(2,94) = 42.17$, $p < .001$). The single comparisons involving the factor COND further registered a reliable contrast between antonyms and related non-antonyms (right-frontal: $F(1,47) = 8.36$, $p < .01$; left-central: $F(1,47) = 25.62$, $p < .001$; right-central: $F(1,47) = 28.71$, $p < .001$; left-parietal: $F(1,47) = 22.71$, $p < .001$; right-parietal: $F(1,47) = 19.23$, $p < .0001$) and a reliable difference between antonyms and unrelated non-antonyms (right-frontal: $F(1,47) = 1.48$, $p < .3$; left-central: $F(1,47) = 47.84$, $p < .001$; right-central: $F(1,47) = 47.20$, $p < .001$; left-parietal: $F(1,47) = 72.48$, $p < .001$; right-parietal: $F(1,47) = 77.21$, $p < .001$). Related non-antonyms also differed reliably from unrelated non-antonyms at central and parietal sites (right-frontal: $F(1,47) = 1.51$, $p < .3$; left-central: $F(1,47) = 5.39$, $p < .03$; right-central: $F(1,47) = 5.04$, $p < .03$; left-parietal: $F(1,47) = 16.41$, $p < .0001$; right-parietal: $F(1,47) = 24.65$, $p < .001$).

In sum, the unrelated non-antonym condition registered the most pronounced P300, whereas the amplitude of the related non-antonym condition was somewhat smaller. There was no P300 to antonyms in this time window.

4.3.3 Discussion

The results from Experiment 3 are easily summarized. The results replicate the major pattern found in Experiment 1: antonyms elicited a P300 that peaked slightly earlier than the N400 elicited by each of the non-antonym conditions, and following the N400, both non-antonym conditions revealed a P300 with a reduced amplitude for the related condition. There was, however, one exception: the present experiment patterned with Experiment 2 in that it failed to replicate the established N400 amplitude reduction for related targets. This may be the result of the low number of items per condition, which could lower experimental power significantly. The relatedness facilitation did not produce a reliable effect, even though visual inspection of the ERP waves suggests a very small trend in the expected direction.

Nevertheless, the main result stemming from this experiment is that the majority of the findings in Experiment 2 do not result from differences in materials and experimental design, because Experiment 3 should have produced results identical to the foveal ERP effects in Experiment 2. Experiment 3 clearly mirrored Experiment 1 in every important aspect regarding the P300 component. One can now conclude that the parafoveal effects in Experiment 2 are caused by the availability of parafoveal preview in reading, which is not present in either RSVP or the auditory modality, and that the lack of the P300 in Experiment 2 does not stem from differences in experimental design.

4.4 Summary

This series of experiments set out to investigate the main hypothesis of the present thesis; the P300 in priming studies reflects the completion of a decision-making process through which the target word is categorized according to task instructions (e.g., lexical decision, antonymy judgment). It was hypothesized that the pronounced P300 for targets highly related to their prime peaks concurrently with the N400 to unrelated targets because categorization is easier in these cases due to semantic priming. When targets were encountered in a sentence context, the categorization of words followed along the lines of predictions made by the language comprehension system. With the antonym paradigm, both a specific word form and a specific meaning are predicted. Since only antonyms fit in these two predictions, there is a P300 showing that they can be easily identified. Any other sentence completion, by contrast, shows an N400 effect due to the unexpectedness, and categorization is delayed because of the increased processing demands in lexical processing. Thus, the P300 reflecting the decision on unexpected or semantically unrelated words peaks after the N400. Alternatively, it was proposed that the P300 to related target words is brought about by the use of prelexical information alone, whereas the decision on unrelated targets requires lexical information. A further hypothesis stated that the current presentation methods in ERP experiments are inadequate to distinguish between these two alternatives concerning the type of information that is used in the decision-making process. It was also claimed that natural reading offers the possibility to distinguish between the use of prelexical and lexical information in the genesis of the P300 because these different information types are linked to different fixation positions. That is, while prelexical information is processed in the parafovea, lexical-semantic information is processed in foveal vision. This temporal dissociation distinguishes reading from the auditory modality and rapid serial visual presentation (RSVP).

Experiment 1 was designed to replicate visually evoked ERP findings which show that the antonym paradigm gives rise to a particular data pattern (cf. Roehm et al., 2007). Apart from some modality-specific differences concerning onset and offset latencies, this experiment fully replicated the results from Roehm et al. (2007) who used identical stimuli. Antonyms elicited a P300 peaking in the same time epoch as the N400 elicited by the non-antonym conditions. The non-antonyms were further distinguished by their varying degrees of semantic relatedness. When targets were semantically related to the prime word, the N400 amplitude was reduced compared to its amplitude with semantically unrelated words. Non-antonyms also showed a P300 following the N400 time window, reduced, again, for related words.

The second experiment used a novel presentation method which allowed ERPs and eye movements to be recorded concurrently. This experiment revealed new ERP processing correlates because the natural reading situation made it possible to measure brain responses for parafoveal information, i.e. before the target word was first fixated. The identification of orthographic input led to a parafoveal N400 priming effect, which could involve underlying processes different from the underlying processes in foveal vision. Specifically, as the parafoveal N400s to antonyms and non-antonyms were similarly reduced in amplitude, the parafoveal N400 effect was presubably caused by the spread of activation between related concepts in the lexicon. There was no parafoveal P300 to antonyms. During the first fixation on the target word, the non-antonym conditions elicited indistinguishable foveal N400 amplitudes. The ensuing P300 to both non-antonym conditions was reduced in amplitude and with increased peak latency when the non-antonym was semantically related to the prime. Again, there was no P300 activity for antonyms. Thus, it was concluded that the concurrent availability of prelexical and lexical information appears to be the critical determinant in P300 generation. When the decision on the word-form prediction and the meaning prediction are made at different time points (i.e., at separate fixation positions), the P300 is inhibited.

The third experiment was conducted to rule out some confounds with regard to alterations in the experimental setup that were necessary in the second experiment. Experiment 3 confirmed that the results from Experiment 2 could be derived from the availability of parafoveal preview in reading, with results identical to those from the first experiment. There was, however, one exception to this conclusion. Both Experiments 2 and 3 did not show a reduction in N400 amplitude for semantically related targets. This was interpreted as reflecting the low signal-to-noise ratio.

V General Discussion

This thesis aimed, on the one hand, to investigate how endogenous ERP components such as the P300 and the N400 might vary under different input modalities; and, on the other hand, to study the relationship between two widely-used on-line methods in psycholinguistic research, event-related potentials and eye movements in reading, which offer a high temporal resolution and multi-dimensional measures of language comprehension. Despite such similarities, the empirical outcomes regarding the time course of language processing have so far eluded a plausible cross-methodological comparison, co-existing more or less independently. For instance, the language-related ERP effects such as the N400 or the P600/ late positivity do not precede readers' eye movements when identical stimulus sets are compared across methods, as one might expect. Rather, the eyes seem to be faster than the corresponding brain response (Sereno & Rayner, 2003), which is at odds with the claim that (neuro)cognitive processes drive eye movements to a large extent. It is this temporal divergence that severely limits a possible comparison of the two methods when data from independent experiments are compared with each other.

To overcome this limitation, a novel method was introduced, namely the concurrent recording of ERPs and eye movements. Reading is considered an input modality that basically differs from other input techniques such as auditory presentation or rapid serial visual presentation (RSVP) in one important aspect: how much and when information becomes available to the processor. The consequences of a feature unique to reading, the parafoveal preview of upcoming words, were examined using the antonym paradigm. In this paradigm, participants read or hear highly constraining sentences such as *black is the opposite of white* in which the language processing system is prompted to strongly anticipate one particular word (i.e. *white*), including a specific word form and a specific meaning. Several experiments have demonstrated that distinctive patterns with the N400 and the P300 are produced in this paradigm, depending on whether the predicted word or an unpredicted word occurs. The combination of the antonym paradigm with the novel co-registration technique used in this thesis revealed that the dissociation of subprocesses in visual word recognition highly affects the P300 to predicted words, leading to a full inhibition of the component. This dissociation can also separate the different mechanisms underlying the N400 priming effect into parafoveal and foveal processing steps. To further generalize these findings, two additional experiments were carried out varying word order and animacy of verbal arguments. While the results confirmed the reliability of parafoveal N400 effects (and, to some extent, also the negative effect that the parafoveal preview has on the P300), they also showed that, for the tracking of parafoveal processing, a combined ERP and eye-movement recording can only be used in those instances where the measuring fixation is located on words that do not induce any processing difficulty themselves. Thus, the data presented here contribute to psycholinguistic research in at least two ways. The reported findings represent a direct comparison of ERPs and eye movements collected in a single experiment. On the other hand, they highlight the crucial role of input modality in measuring ERP correlates of cognitive processing, particularly language comprehension. Both of these aspects lead not only to a better understanding of the connection between neuronal and behavioral data in (visual) language comprehension, but also to a more precise formulation of hypotheses about the architectural features of language comprehension models, when such features are presumed to function identically across several input modalities.

5.1 The P300 in language comprehension

A number of studies have aimed to unravel the determinants of the P300 and, after more than four decades of research, much is now known about the factors inducing the P300 in psychophysiological experiments. Studies investigating language comprehension across different structures and linguistic domains have often reported a P300, but it has not received as systematic an investigation into its emergence as, for example, the N400 has. The literature on lexical processing and semantic relatedness between words is probably one of the clearest examples of this. While the N400 is assumed to be functionally related to lexical processing, the concurrently emerging P300 is either treated as a problematic disturbance that may mask N400 effects due to component overlap, or is reported without any functional interpretation whatsoever (but see Bentin et al., 1999; Holcomb, 1988). Given its persistent occurrence in these experiments, one must assume a common source generating the P300 across all the experiments and make serious attempts to fill the explanatory gap. The account presented here hypothesized that the P300 reflects the accomplishment of target categorization according to task instructions (e.g., lexical decision, semantic relatedness judgment). Depending on the type and complexity of the stimulus, the P300 peaked either after or concurrently with the N400 elicited in response to the lexical processing of words. Thus, it is clear that the P300 in these experiments is not language-specific (i.e., it does not respond to a special linguistic domain such as semantics), but reflects the domain-general process of decision-making that makes use of linguistic information.

According to such a decision-making account, the P300 amplitude and latency should vary proportionally with the informational content that the stimulus provides (e.g., Hillyard et al., 1971; Kerkhof, 1978; Parasuraman & Beatty, 1980; cf. chapter 1.2.2). If a stimulus cannot be recognized unambiguously as the task relevant target, confidence declines accordingly, and P300 amplitude declines and latency increases. While the target stimulus in these sorts of experiments is most often a tone burst, the target stimulus in lexical processing experiments is a word, which offers far more information that must be taken into account for a decision. Furthermore, the P300 to target words highly related to their prime exhibits an earlier onset latency than the P300 to unrelated target words. These two aspects led to the basic research question regarding the P300. What type of linguistic information is critical to induce the P300 and how much evidence for target recognition must be gathered? Specifically, it was claimed that in highly constraining contexts (such as the antonym paradigm), the language comprehension system makes two separate predictions: one about an upcoming word form and one about an upcoming meaning (see Laszlo & Federmeier, 2009, for independent evidence that orthographic forms are pre-activated by predictions). Both predictions must be confirmed for the target word to be categorized as the expected antonym. As orthographic and semantic information become available to the processor simultaneously with auditory and RSVP methods, a robust P300 emerges, as evidenced by Experiments 1 and 3. Reading, however, is different, in that it temporally dissociates prelexical orthographic and semantic processing in word recognition. Prelexical information is processed in the parafovea, and continues during the execution of a saccade (Irwin, 1998). Thus, processing of the orthographic form may be accomplished well before the eye first fixates the antonym, i.e., before information from the lexical-semantic level is received. Hence, the orthographical and semantic predictions are then confirmed at different fixation positions. Consequently, there is not enough evidence at either fixation point to confidently categorize the perceived input as the expected antonym. This leads to the hypothesis that there should be no P300 to antonyms when ERPs are collected during reading, as the level of confidence in identifying/ categorizing antonyms is too low. This prediction was confirmed, in that antonyms did not elicit a P300, while unexpected non-antonyms did. The absence of the P300 to antonyms was confirmed across two fixation positions, which were chosen to index parafoveal and foveal processing of the antonym. This striking finding was not due to changes in the typical ERP experimental design (e.g., a smaller number of trials per condition), as it was shown that the same stimuli using RSVP in a follow-up experiment patterned with the auditory ERP experiment. Thus, the differing results from the reading experiment are due to properties of the input modality, not the changes in experimental design.

What can the absence of the P300 to antonyms tell us about the categorization process for these items in reading? The comparison of all three experiments indicates that, in natural reading, neither prelexical (parafoveal) nor lexical (foveal) processing steps alone can provide sufficient information to elicit the P300. The temporal coincidence of these steps in listening and RSVP does provide enough information to elicit the P300. What the data indicate is that the categorization process must not be split up into separate stages, i.e. the entire informational content of the predicted stimulus from both linguistic domains has to be present to elicit the P300.

The P300 to antonyms in previous ERP studies can thus be interpreted as resulting from the ease of target categorization. The pre-activation of antonyms is associated with top-down predictions, which are all met as soon as the target is perceived auditorily or using the RSVP method. In reading, by contrast, expectations are only partially met at each stage of processing, i.e., the parafoveal and foveal processing steps. Although parafoveal information suffices to meet expectations at the prelexical level, it is insufficient to meet any semantics-oriented predictions. In a similar vein, lexical information, when taken from the first fixation on the antonym, can fulfill the prediction for a particular meaning. The first fixation on the antonym also provides orthographic input, but the parafoveal N400 priming effect suggests that orthographic input has been processed more or less completely during the preceding fixation. So, parafoveal orthographic information is perhaps used to make a first attempt at categorizing the stimulus. The lack of the P300 may then indicate that, overall, the evidence is not compelling enough for a confident decision. Alternatively, the positive evidence from one level and the negative, i.e. unmet, evidence from the other may cancel each other out, resulting again in a low level of confidence for any decision about the perceived input. These considerations are in accordance with common interpretations of the P300 stipulating that it is elicited after the completion of an evaluation process (e.g., Kok, 2001; Donchin & Coles, 1998; Magliero et al., 1984). It also fits in with the findings from signal-detection experiments that show a decline in P300 amplitude and an increase in latency when the perceptual properties of the stimulus only provide the basis for a decision with a relatively high degree of uncertainty (e.g., Hillyard et al., 1971; Kerkhof, 1978; Parasuraman & Beatty, 1980; cf. chapter 1.2.2). Thus, there is no clear indicator of the categorization of antonyms because of the low confidence level in the decision-making, which is a result of the dissociation of information across two fixation positions. This account does not claim that there is no decision at all for antonyms. This claim would imply that processing strategies not only change substantially across input modalities, but also across experimental conditions. There is no reason why antonyms should not be categorized, when non-antonyms clearly are.

As related and unrelated non-antonyms fall outside the range of either predicted word forms or meanings, their processing is not subject to the same requirements as antonyms. The processing advantage for related non-antonyms, as reflected in the parafoveal N400 priming effect, is a consequence of the automatic spread of activation between related words, but it is not related to predictions for semantically related non-antonyms (e.g., *black – blue*). The same basic mechanism can thus be assumed for both related and unrelated non-antonyms. The language comprehension system is not prepared to encounter non-antonymic orthographic forms, so that they are considered irrelevant for decision-making during parafoveal processing. In foveal vision it becomes clear that their meaning is not the predicted antonym, leading to an increase in foveal N400 amplitudes. Only after their meaning is retrieved, can non-antonyms be classified as such, which gives rise to the P300 component. The dissociation of prelexical orthographic and lexical-semantic information is thus irrelevant to their categorization, as with the RSVP data from Roehm et al. (2007) and the findings from Experiment 3 reported here. Unexpected non-antonymic words are thus categorized as such mainly based on their lexical meaning. Consequently, the foveal P300 peaks after the N400, which reflects the well-known inverse relationship between N400 amplitude and cloze probability. Strikingly, the two non-antonyms are also categorized with different levels of subjective confidence. That is, an unexpected word which happens to be semantically related to the actual antonym cannot be unequivocally categorized, because semantic relatedness interferes with the categorization process. Semantic relatedness between words thereby lowers confidence and prolongs evaluation time. This adverse interference effect is reflected in the reduced amplitude and the increased peak latency of the P300 component to related non-antonyms in Experiments 2 and 3. Experiment 1 showed a similar pattern, although it was somewhat less statistically reliable. Interestingly, this categorization problem also affected behavioral measures such as error rates and reaction times in Experiments 1 and 3, as well as total reading time in Experiment 2 (see below, chapter 5.3).

The results on the P300 to antonyms suggest that empirical data in support of one particular model of word recognition or lexical processing must be carefully evaluated with regard to input modality. Although the P300 itself does not depend on a particular type of linguistic information, it does require a sufficient amount of linguistically defined evidence for target identification. This evidence must be gathered from the same linguistic domains that the predicted stem from (at least in high-constraint contexts) and there must be partial overlap in the processing of information from these domains (i.e., the prelexical and the lexical domains). When an input modality such as reading prevents such a cascaded processing manner, the P300 to antonyms is effectively erased. While this is the first ERP finding that confirms the assumption that lexical processing/ word recognition in reading is serial (cf. Reichle, Liversedge, et al., 2009), it also imposes a severe limitation on the formulation of modality-independent principles (in models of word recognition) stating whether prelexical and lexical information is generally processed in a serial or cascaded manner (cf. McClelland, 1973; see McElree & Griffith, 1995, for a discussion of different accounts).

The P300 to non-antonyms is also revealing with respect to the debate on the relationship between the P600/ late positivity and the P300. Despite the strong evidence suggesting that the P300 is elicited consistently by far more than merely an improbable event, the probability manipulation has remained the critical test case for the question of whether the P600 is an instance of the target-related P300 or not (see chapter 1.3.2 for references). The present experiments did not manipulate the probability of the critical stimuli. They clearly show that the problem of component confusion is much more aggravated once task environment is considered with all its dimensions. While Experiment 1 employed an antonymy judgment, Experiments 2 and 3 intermingled the antonymy judgment task with a lexical probe task, which occurred more frequently. Nonetheless, a P300 to non-antonym targets was found in all experiments. Evidently, the sentence structure *x is the opposite of y* contributes much more to the overall task environment than the instructions given to participants. Future investigations into the nature of the late positivity in language processing should therefore not focus on experimental probability – especially considering that it bears no importance to linguistic operations – and should aim instead at disentangling the factors inherent to task environment (e.g., sentence context, task instruction) to find out whether the late positivity is a member of the P300 family.

First attempts to clarify the influence of task instruction on the late-positivity complex have revealed that the acceptability judgment task could be a good way to disentangle the P300 and P600 components (cf. Hagoort, Brown, & Groothusen, 1993; Haupt et al., 2008; Osterhout et al., 1996). However, Experiment 4 revealed that the distinction between P300 and P600 is also evident without alterations in task instruction. Specifically, this experiment examined the processing of word order variations and interference effects induced by the animacy of verbal arguments. One late positivity was found at the verb position for object-initial clauses; another late positivity was related to animacy-induced interference caused by two NPs sharing the same animacy status. Interestingly, these positivities showed different topographies, the former being much more confined in its overall distribution. The late positivity to object-initial clauses was interpreted as reflecting the lower degree of well-formedness of these structures in German – this kind of late positivity has been argued to belong to the P3b family (cf. Bornkessel & Schlesewsky, 2006). The slightly different topographies of the two late positivities were taken to suggest that the P300 is probably to some extent susceptible to the availability of the parafoveal preview. In these structures the first fixation on the verb provided the relevant information, as the language comprehension system probably forms predictions only in the semantic domain, concerning specific semantic properties of the verb (accusative case in this experiment). So, the dissociation between orthographic and semantic processing is not as deleterious as in Experiment 2. Nevertheless, the distribution of the late positivity was limited when compared with the distiribution of the late positivity that reflected semantic interference in the processing of verbal arguments. One may conclude tentatively that members of the P3b family are in general prone to the parafoveal preview, in that, for example, the partially parallel processing of foveal and parafoveal information leads to a deprivation of the attentional resources necessary for the generation of the P300.

5.2 Parafoveal vs. foveal N400

In the light of the current debate over how many words can be simultaneously processed in natural reading (see Reichle, Liversedge, et al., 2009, for a recent discussion), the presence of a parafoveal N400 is a challenge to interpret, as the N400 component is usually taken to reflect lexical-semantic processes (e.g., Kutas & Federmeier, 2000). In Experiment 2, by contrast, the parafoveal N400 was interpreted to be a reflection of prelexical processes. However, the validity of this interpretation may be called into question on the grounds of two aspects. From the perspective of eye movement research, the parafoveal N400 may be the result of foveal processing in mislocated fixations. Furthermore, the issue of whether there is only one monolithic N400 component indexing lexical-semantic integration, or a poly-functional N400 family, has not yet been settled.

The parafoveal N400 is a so-called parafoveal-on-foveal effect in reading. These effects can be distinguished into two subtypes, one related to prelexical processes, the other to lexical processing (Drieghe et al., 2008). This is a pivotal distinction to make, as the debate on serial vs. parallel word encoding mainly centers around the lexical parafoveal-on-foveal effect. This is because the lexical parafoveal-on-foveal effect represents a major challenge to those computational models of eye movements in reading that assume a serial encoding of words (e.g., E-Z Reader, Reichle et al., 2003). As laid out in chapter 2.3.2 above, serial models can predict prelexical parafoveal-on-foveal effects to a certain extent, but are incompatible with lexical parafoveal-on-foveal effects. On the other hand, parallel models of reading such as SWIFT (Engbert et al., 2005) assume that attention is gradually allocated such that more than one word at a time can undergo lexical processing, and are thus capable of dealing with these lexical parafoveal-on-foveal effects.

Let us first outline the arguments for the interpretation of the parafoveal N400 as an index of prelexical processing, i.e., as a priming effect based on orthographic form. Parafoveal priming for antonyms and unexpected related non-antonyms can be reconciled with serial models such as the E-Z Reader model when it is assumed that the parafoveal preview suffices to evaluate upcoming orthographic forms. However, the priming effect for the unexpected related non-antonym is not predicted by the present formulation of the SWIFT model, which assumes that parafoveal lexical processing occurs mainly for highly predictable words. Following this assumption, only antonyms should evidence a parafoveal priming effect. Recall that the antonym paradigm provides a high-constraint context in which only antonyms are anticipated, rendering all other continuations as unexpected, regardless of the lexical relatedness manipulation. If predictability were to exert an influence so early on in processing, *all* types of unexpected words should have elicited N400s with enlarged amplitudes. On top of that, the strongest evidence for the parafoveal N400 as reflecting prelexical processes stems from the absence of a parafoveal P300 to antonyms. As argued in detail in the preceding section, the emergence of the P300 to antonyms presupposes the temporal overlap of orthographic and semantic processing, which, however, did not take place in Experiment 2. Similarly, the indistinguishable parafoveal N400 priming effects for antonyms and unexpected related targets suggest that deeper semantic analyses do not take place at this early level of word recognition.

The lack of a P300 effect to antonyms also rules out the possibility that mislocated fixations can be responsible for the parafoveal N400. The mislocated-fixation account has been proposed by proponents of serial models of eye movement control to explain some of the lexical parafoveal-on-foveal effects reported in the literature (cf. Drieghe et al., 2008; Rayner, Warren, Juhasz, & Liversedge, 2004; Rayner et al., 2003; but see Kennedy, 2008). This line of explanation assumes that lexical parafoveal-on-foveal effects are a result of inaccurate saccade targeting. That is to say, saccades intended to land on the target word fall short and land somewhere on the final letters of a pretarget word. Because the foveal area of the retina translates into approximately 3 letters, these mislocated fixations could still involve foveal processing of the target word, even though they are located on the word preceding the target. If the data from Experiment 2 resulted for the most part from such mislocated fixations, there should have been a profound P300 effect to antonyms because mislocated fixations should lead to the simultaneous availability of orthographic and semantic information. At the very least, there should have been a modulation of the parafoveal N400 priming effect with smaller amplitudes for antonyms (*black – white*) as opposed to related non-antonyms (*black – blue*). Even if only a few trials elicited a P300, the mean amplitude of the N400 to antonyms would be reduced in comparison with the mean N400 amplitude to unexpected related non-antonyms. However, neither prediction is confirmed by the data from Experiment 2.

One can thus conclude that the N400 in natural reading is elicited by both parafoveal and foveal information. While the former reflects prelexical processes, the latter is related to lexical properties of the target word in highly constrained antonym sentences. The foveal N400 further confirms the effect of cloze probability on N400 amplitude, with a low cloze probability leading to enhanced amplitude. In contrast to previous findings, the foveal N400 did not evidence an advantage for semantically related target words over unrelated targets. Such an advantage was only found for the parafoveal N400. This may be taken as evidence that the processing benefit for related non-antonyms in terms of reduced N400 amplitude does not result from specific predictions about upcoming words (cf. Kutas & Iragui, 1998; Roehm et al., 2007), but from the automatic spread of activation from antonyms to semantic neighbors. Critically, dissociating the spread of activation between words and the contextual integration of words along the lines of parafoveal and foveal N400s helps resolve the debate as to whether the N400 component reflects the automatic spread of activation or specific predictions. The emergence of two distinct N400 components, one for automatically spreading activation and the other for specific predictions, suggests that the N400 is not a monolithic component. Rather, it appears to comprise a component family, so that the functional interpretation of each instance of the N400 relies on the detailed description of the context in which it occurs. As listening and RSVP conditions cannot bring about a temporal dissociation of automatic processes and target-related predictions, both are conflated within a single component at a particular point in time.

Experiments 4 and 5 were designed to further investigate the reliability of the parafoveal N400 and to extend the range of constructions in which ERP correlates of parafoveal processing can be found. While there was no parafoveal ERP effect in Experiment 4, Experiment 5 revealed that the animacy status of the upcoming NP might already be acquired in the parafovea. This finding is clearly important, inasmuch as parafoveal N400 effects seem to be reliable across different sets of stimuli. It also suggests that the nature of the parafoveal N400 depends on the linguistic environment, such that there is no one-to-one mapping between the parafoveal N400 and either prelexical or lexical processing.

Surprisingly, the extent to which lexical-semantic information is extracted from the parafovea contradicts the common assumption that more parafoveal information is acquired from highly predictable words than for unexpected words (cf. Balota et al., 1985; Rayner & Well, 1996). In other words, high-constraint contexts are thought to enable the language comprehension system to acquire prelexical and, to some extent, also lexical information from the parafovea. Experiment 5 used unconstrained contexts in which the occurrence of an animate or inanimate NP was equally probable. So it is rather unexpected that the parafoveal N400 resulted from parafoveal semantic processing, whereas the parafoveal N400 priming effect in Experiment 2, using a high-constraint context, was due to only orthographic information from highly predictable words. One explanation for this may lie in the strategic processing involved in high-constraint contexts. The language comprehension system relied on parafoveal orthographic information in Experiment 2 because this information was sufficient to confirm whether the perceived word form was in fact the predicted antonym. Specifically, since the parafoveal orthographic information made it possible to decide whether or not the expected word would come up next, the language comprehension could prepare for the upcoming word. In unconstrained contexts such as in Experiment 5, however, there are no specific predictions regarding an upcoming word form (and meaning), so parafoveal orthographic information cannot meet any prediction. Therefore, more parafoveal preprocessing of words may be necessary for unhindered processing in foveal vision. This conclusion is speculative at present, since these parafoveal ERP results are the first of their kind. Clearly, further linguistic contexts must be examined to assess whether parafoveal ERPs indeed show the opposite effect relative to the eye movement measure, regarding the extent of parafoveal preprocessing and its interaction with contextual constraint. Finally, the comparison of Experiments 4 and 5 indicated that the range of linguistic stimuli eliciting parafoveal N400 effects might be limited. As it seems, a parafoveal N400 effect only arises when the word, on which the fixation assumed to capture parafoveal processing of the upcoming critical word is located, does not induce processing effort itself.

5.3 On the correlation between ERP and eye movement data

One of the remarkable findings of this series of experiments is that there is no parafoveal-on-foveal effect in the eye-movement record that corresponds to the parafoveal N400 effects found in Experiments 2 or 5. This null effect is especially interesting in view of the semantic and structural properties of the antonym stimuli used in Experiment 2 because both types should have caused parafoveal-on-foveal effects in the eye-tracking record. For example, many studies have found that words that are highly predictable from the preceding context are likely to induce parafoveal-on-foveal effects (e.g., Kliegl et al., 2006; Pynte et al., 2008). Some recent studies have also investigated whether parafoveal-on-foveal effects may be evoked by non-adjacent words (e.g., Angele et al., 2008; Kliegl et al., 2007; Rayner et al., 2007), meaning that the foveal word (word *n*) and the critical parafoveal word (word *n* + 2) are separated by yet another word (word *n* +1). It was found that these parafoveal-on-foveal effects from word *n* + 2 on word *n* are mostly found when word *n* + 1 is a very short word with less than 4 letters in length (see Angele et al., 2008, for discussion). Note that in the stimuli *dass schwarz das Gegenteil von weiß ist, hat Gertrud gestern behauptet* ('Gertrud claimed yesterday that black is the opposite of white.') the average launch site for the saccade targeting the critical word was on the word *Gegenteil* ('opposite'). The preposition *von* ('of') is likely to be recognized without fixation due to its short length and the fact that it is a function word. The null effect is thus unexpected, given that the antonym paradigm, with its highly predictable target words, should have led to a behavioral parafoveal-on-foveal effect. Note, however, that the behavioral null result converges with the study by Kliegl and colleagues (Kliegl et al., 2007) who found a parafoveal-on-foveal effect from word *n* + 2 only when the word *n* + 1 was a content word, not when it was a function word.

Overall, this co-registration experiment provides further evidence that not all neuronal effects necessarily translate into a corresponding effect in the eye-movement record. The parafoveal N400 effect in Experiment 2 was not accompanied by any significant changes in the eye movement pattern, which was also the case for the parafoveal N400 effect in Experiment 5. Remarkably, neither of the two previous single-word studies using the co-registration technique was able to find reliable parafoveal-on-foveal effects due to semantic relatedness in the eye movement data (Baccino & Manunta, 2005; Simola et al., 2009). This is particularly intriguing in light of the eye-mind span hypothesis that assumes "the eye-movement data can provide a good reflection of the moment-to-moment cognitive processes associated with reading" (Rayner & Sereno, 1994, p. 61). Lexical factors (e.g., lexical frequency or predictability) have been shown to influence fixation duration in a nearly endless number of studies and many researchers claim that these factors are one of the driving forces behind eye movements in reading. In other words, they are considered a major factor concerning the *when* decision of eye movements, i.e. they determine the duration of a fixation and when the eyes leave a word (cf. the reviews by Clifton et al., 2007; Liversedge, Paterson, & Pickering, 1998; Rayner, 1998; Staub & Rayner, 2007). However, by revealing a de-coupling between neuronal and ocular effects, the present data highlight that the eyes are first and foremost a means of information delivery and not a passive indicator of cognitive processing disruption. Some part of the oculomotor activity – and hence effects in the eye movement record – should therefore be regarded as being controlled by an input-driven mechanism. If this is true, there is no need for the neuronal parafoveal-on-foveal effects to translate into eye movement behavior as prolonged fixation duration. As laid out in the preceding chapter, inaccurate parafoveal information obviously does not help to decide whether the upcoming word form also carries the expected antonymic meaning (as, otherwise, antonyms and related non-antonyms should have shown differences in their parafoveal N400 priming effects). At this point, the eyes are not indexing moment-to-moment cognitive processes.

Interestingly, initial research on the resolution of syntactic ambiguities and the role of regressive saccades in syntactic reanalysis suggested that significant effects in the eye movement record are probably input-driven. Mitchell et al. (2008) found that the first regressive saccade from the point of disambiguation targeted the immediately preceding word in most cases, and was then followed by a series of regressive saccades targeting the locus of ambiguity. It was striking that the first regressive saccade appeared to be launched independently of any of the linguistic variables examined. The authors went on to argue that one purpose of regressive saccades in reading is to allow the language system to catch up with its processing lag, and not to provide new input. When their "Time-Out hypothesis" (Mitchell et al., 2008, p. 269) is extended to fixations as well, it predicts that linguistic processes require an input stop, given processing is so fundamentally disrupted (or far from a threshold indicating completion of word recognition) at some stage that newly gained information may lead to a complete collapse. On the other hand, when new information is recognized as (potentially) deviant and further evidence is needed to support this evaluation, the eyes should not dwell on their current position, but move on to parts of the text that may deliver the awaited evidence. This may explain why the fixation duration was not increased with the corresponding parafoveal N400 effect in Experiments 2 and 5. Receiving more inaccurate parafoveal information would not have helped in securely identifying the upcoming target word. In other words, the deviance of an orthographic form was noticed neuronally, but this information was not passed on to the oculomotor system.[34] In Experiment 5, where semantic

[34] Alternatively, one could argue that the parafoveal information was processed too late to substantially influence saccade programming. However, studies on parafoveal processing in different writing systems (e.g., alphabetic and logographic) suggest that parafoveal information is already extracted from the beginning of the ongoing fixation (cf. Inhoff, Eiter,

(animacy) information seemed to be acquired from the parafovea, the disadvantageous animacy status of the NP did not increase the duration of the corresponding fixation, because the language processing system might have been searching for further input to solve this problem. These considerations do not undermine the eye-mind span hypothesis, but actually extend it, inasmuch as the eyes are not simply regarded as an output organ that indicates processing difficulties, but as a tool that offers different ways to overcome them. In this sense, the primary function of eye movements is to provide new information, whenever these may be helpful in word recognition, or to prevent new visual information from being perceived, as long as the currently processed linguistic input has not exceeded a certain threshold that indicates the end of its processing.

The formulation of the eye-mind hypothesis as cited above does make some correct predictions with regard to other divergences between ERPs and eye movements. Some of the reported data highly suggest that the original definition of the eye-mind span, specifically its implicit reference to how domain-general processes can affect readers' eye movements, should be taken seriously. First, the differences between early and late measures of the eye movement results reported for Experiment 2 indicate what kind of cognitive processes may be reflected in the eye-tracking record. Specifically, total reading time shows a disadvantage for the related non-antonyms (i.e., *black – blue*) relative to the unrelated non-antonyms (i.e., *black – nice*) – beside the effect that reading times in both conditions were longer than with expected antonyms (see Appendix C). When eye movements are defined as being driven by linguistic processes, one should expect that they mirror the foveal N400 effect (antonyms < unexpected related/ unrelated). However, this only holds when early measures of eye movements such as first fixation duration are considered. The total time results are more

& Radach, 2005; Yen, Radach, Tzeng, Hung, & Tsai, 2009). Note that furthermore the absolute N400 peak is no index of when, in the course of the fixation, information became available for processing (Bornkessel-Schlesewsky & Schlesewsky, 2008; Dogil, Frese, Haider, Roehm, & Wokurek, 2004).

compatible with the P300 results, which showed that related non-antonyms were more difficult to categorize (relative to unrelated non-antonyms) and therefore elicited a P300 with reduced amplitude and increased latency. This was interpreted as reflecting that semantic relatedness interfered with the cloze probability of related words during target word evaluation. Although the meaning of these non-antonyms was clearly unexpected, it still exhibited some semantic relation to the expected antonym. This made target evaluation more difficult for related non-antonyms, and total times reflect this difficulty in categorizing the related non-antonyms as opposed to unrelated non-antonyms. Because the categorization of the latter type of stimuli is not subject to interference, their total reading time is shorter. Nonetheless, total reading time also reflects linguistic processing proper, as the unrelated non-antonyms still led to longer reading times than antonyms. This suggests that effects found in late measures of eye movements such as total reading time reflect an accumulation of different processes, which, critically, may not only include actual linguistic processing such as word recognition, but also more general cognitive processes. In sum, one should well consider whether and to what extent eye movements are a sensitive method to detect non-linguistic factors in reading. For example, the present impression of a correlation between P300 latency and total reading time could be further investigated by statistically correlating total reading time with P300 amplitude and/ or latency on a single-subject level. Based on the present findings, increased P300 latency and reduced amplitude would predict longer total reading times. Such investigations could be extended to derive hypotheses about whether certain eye movement measures correlate with specific linguistic or non-linguistic cognitive processes (if these processes are also reflected in different ERP components).

5.4 Outlook

The present experiments represent a first attempt to gain more insight into possible connections between two on-line methods used in language comprehension studies, and, more generally, into the link between brain and behavior. The major questions were: to what extent does input modality alter common ERP waves, and to what degree are neuronal processing effects passed on to the oculomotor system? So far the data suggested that the input modality has a noteworthy impact on ERP components such as the P300 and N400 and their interpretation, and that there is no one-to-one correspondence between eye movement patterns and ERP effects in reading.

Experiment 2 revealed that there is neither a parafoveal nor a foveal P300 for highly predicted antonyms. It was proposed that this is the case because reading prohibits the temporal overlap of prelexical orthographic and lexical-semantic processing, which is pivotal in eliciting the P300. That is, the language comprehension system makes predictions in two linguistic domains and both predictions must be met simultaneously if a P300 is to be elicited. In reading, the information relevant for these predictions are collected from two different fixation positions, with the result that the P300 is not elicited by the predicted word. This hypothesis can be tested further by employing a gaze-contingent display technique which can manipulate the availability or usefulness of a parafoveal preview (see Rayner, 1998, for an overview). When, for example, a boundary paradigm is used (Rayner, 1975), the parafoveal preview of the upcoming antonym is masked with a pseudoword (a non-word that follows the orthographic and phonological constraints of the language under investigation). So, readers acquire normal prelexical information from the parafovea, but crucially it is not useful with respect to the prediction about the specific word form belonging to the expected antonym. During the saccade to the target word, the parafoveal pseudoword is replaced with the correct antonym. Consequently, the correct orthographic and semantic information that are relevant for the predictions in both linguistic domains become available all at once during the first fixation on the antonym. This predicts that the P300 to antonyms should occur under these circumstances, as the processing of orthographic and semantic information overlap temporally. The parafoveal N400 priming effect should disappear, as pseudowords cannot be pre-activated by the word-form predictions. If, however, these pseudowords are orthographic neighbors of the expected antonym, they may show a reduced parafoveal N400 amplitude, as suggested by a recent study on the processing of orthographic neighbors of highly predicted words (Laszlo & Federmeier, 2009). The same outcome regarding the P300 should be brought about by another eye-contingent display technique, the so-called moving window technique. Here, only a limited sequence of words is visible at a time, while the rest of the sentence is masked with random letter strings. It is possible to make only the foveal word visible and mask parafoveal words. This should also lead to a P300 to antonyms during the first fixation, as there is no parafoveal preprocessing of their word forms. However, such a display technique is somewhat unnatural because readers may become aware of the unnatural masking of parafoveal text, possibly resulting in different reading strategies. Thus the boundary paradigm, as suggested above, probably represents the most natural circumstance to examine the dissociation of prelexical orthographic and lexical semantic processing and its impact on the P300 to highly predicted words.

The emergence of a parafoveal P300 is also conceivable for a limited set of words that can be identified in the parafovea. There is evidence that very short and high-frequency words, which in most cases happen to be function words, are usually skipped during reading because they have been fully recognized in the parafovea (Rayner & Sereno, 1994; Rayner, 1998). The subject-object ambiguities used in Experiment 4 appear to be a valid context to examine parafoveal P300 effects in reading. There is compelling evidence that the first case-ambiguous NP in the German Middlefield is preferably interpreted as the subject of the clause. This subject preference was attested across several experiments using a range of different experimental methods (see chapter 7.1.1 for references). Given the robustness of this parsing strategy, one can hypothesize that it leads to the strong anticipation of a particular verb inflection that is in accordance with a subject-initial word order. Experiment 4 had main verbs following the two case-ambiguous NPs (e.g.*, Dass Elfriede Statisten vergisst, hat viele meist verärgert* 'It annoyed many people that Elfriede forgets movie extras'). However, it is also possible to use analytic verb phrases, containing the auxiliary *have* or *be* and a past participle of the main verb (cf. Haupt et al., 2008; Schlesewsky & Bornkessel, 2006). In these cases, the auxiliary provides the critical number information disambiguating the ambiguous NPs, and the past participle provides information about which object case should be assigned. Considering that the inflected forms of the auxiliaries are relatively short (both in singular and plural inflection; see also Example 4 in chapter 1.3.2 above), and that there is evidence for syntactic prediction in sentence processing (see Staub & Clifton, 2006, for evidence from eye movements), it could be possible to find a parafoveal P300 before an auxiliary that disambiguates the sentence towards the anticipated subject-first word order. Two aspects are important in this line of thought. First, the prediction for the auxiliary is one-dimensional, in that there is only a syntactic prediction about a specific word order, and not, for example, about the verb meaning. So, the dissociation of information across parafoveal and foveal processing stages is irrelevant for categorizing the critical auxiliary. Second, the information (dis)confirming the subject preference would be orthographic, as only a distinction based on word length and the shape of the initial letters of the upcoming auxiliary (*hat* ('has') vs. *haben* ('have') or *ist* ('is') vs. *sind* ('are')) is necessary to decide whether the subject preference led to the correct parse.[35] The parafoveal P300 could in principle correlate with a higher skipping rate for auxiliaries that fit the subject-first prediction, given that highly predictable words are also more often skipped than unexpected ones (e.g., Balota et al., 1985;

[35] A more precise description of this type of information would be that the morphological properties of the auxiliary form have to be inferred from parafoveal information. There is evidence from an eye movement study conducted for Hebrew that morphological information may also be acquired from the parafovea (Deutsch, Frost, Pelleg, Pollatsek, & Rayner, 2003).

Rayner & Well, 1996). Object-initial sentences are not anticipated (because of the subject-first preference); therefore, the P300 cannot result from any fulfilled expectations, but it is more likely associated with the evaluation of the well-formedness of the clause (see also Haupt et al., 2008). From this perspective, one could expect to find a parafoveal P300 to subject-initial clauses, reflecting fulfilled predictions, and a foveal P300 on the auxiliary to object-initial clauses, reflecting well-formedness constraints.[36]

The anticipation of a subject-before-object structure could be further strengthened by providing the ambiguous NPs with semantic properties that make one particular syntactic function more likely than the other. Experiment 5 reported an N400 effect for conflicting prominence information for verbal arguments. An animate subject elicited an N400 relative to an inanimate subject in a passive clause because of the dispreferred assignment of the more patient-like undergoer role. The semantic properties of the animate subject argument (nominative case marking, animacy) qualified it as the more agent-like actor of the event, so the undergoer-role assignment is in conflict with the other prominence information. This interaction of syntactic and semantic information could be employed in the proposed endeavor to find a parafoveal P300 to syntactic predictions. If the first NP is endowed with prominence information (e.g., animacy, specificity) that additionally supports the subject preference, the stronger anticipation for the auxiliary could boost such effects.

All of these predictions rest on two implicit assumptions, namely the more or less universal use of parafoveal information across (i) readers and (ii) writing systems. Reading patterns differ not only when children are compared with adults, but also patterns within the adult population reveal differences regarding reading skill which seem to be related to

[36] Note that Experiment 4 reported a P300 to object-initial clauses which was interpreted as a well-formedness P300. If this effect had resulted from readers' anticipation of a particular word order, it would have given rise to a null effect. Since the semantic properties of the clause-initial NP provided a pivotal hint for either subject-initial or object-initial clauses, both word orders should have evidenced a P300 when readers encountered the verb confirming their prediction. Yet, only object-initial clauses engendered this positivity.

parafoveal processing (Ashby et al., 2005; Chace, Rayner, & Well, 2005; Kennison & Clifton, 1995). That is, unskilled adult readers appear to be less efficient in using parafoveal information (e.g., Chace et al., 2005). With the highly constraining antonym contexts used in Experiment 2, unrelated non-antonyms such as in *black is the opposite of nice* differed from antonyms and related non-antonyms (e.g. *blue*), as only the latter two exhibited a parafoveal N400 priming effect. If poor readers do indeed extract less parafoveal information while reading, word recognition processes are mainly confined to foveal vision for these readers. As a result, the parafoveal N400 priming effect could vanish, and perhaps even a foveal P300 to antonyms could occur, since less useful parafoveal information is acquired. If so, the foveal P300 amplitude and the presence of the parafoveal N400 could correlate with reading skill.

More recently, researchers have investigated in more detail the logographic writing system of Chinese because it is remarkably different to the alphabetic system used in languages such as English and German (e.g., Bai, Yan, Liversedge, Zang, & Rayner, 2008; Inhoff & Liu, 1998; Inhoff & Wu, 2005; Rayner, Li, Juhasz, & Yan, 2005; Yan, Tian, Bai, & Rayner, 2006; Yan et al., 2009; Tsai & McConkie, 2003; Yang et al., 2009; Yen, Radach, et al., 2009; Yen, Tsai, Tzeng, & Hung, 2008). It is commonly agreed that more linguistic information is encoded in a Chinese character in terms of spatial extension, as opposed to alphabetic script in which several letters must be assembled to deliver an equal amount of information. Although the previous data are somewhat inconclusive with respect to the distinction between parafoveal and foveal processing of Chinese script, they do suggest that parafoveal words may be processed differently than in alphabetic languages; they may even reach the level of semantic processing, according to some researchers (e.g., Yan et al., 2009). If the supposed semantic processing in the parafovea should also be shown when eye movements and ERPs are recorded concurrently, one can expect to find much more robust parafoveal effects. With the antonymy paradigm, one might find a parafoveal P300 that could be taken to index lexical processing of more than one word at the same time (i.e., during the last fixation before the antonym, semantic information from the upcoming expected word might be available). On the other hand, one could speculate whether character complexity imposes a limitation on processing. That is, since logographic script encodes more linguistic information in a single character, this complexity may not allow all of the relevant information to be decoded sufficiently, so that parafoveal information does not extend beyond basic orthographic feature extraction. In this case, one would expect the pattern of Experiment 2 to be replicated.

Finally, further concurrent recordings of ERP and eye movements in reading may help to better understand the relationship between neuronal and behavioral data. Many experiments have evidenced that behavioral and ERP data diverge in terms of effect size or direction (e.g., Holcomb, 1993; Chwilla et al., 2000; Roehm et al., 2007), and the present parafoveal N400 effects are just another example of this. The P300 appears to be extremely well suited to clarify the correlation between neuronal and behavioral data in language processing. It has been suggested that there is a positive correlation between P300 latency and reaction time measures, with longer P300 latency predicting longer reaction times (e.g., Kutas et al., 1977; Pfefferbaum et al., 1983; Ritter et al., 1972). With the co-registration technique used here, it became apparent that the processing disruption for related non-antonyms (e.g., *black is the opposite of blue*) increases P300 latency and reduces its amplitude. This is presumably because semantic relatedness interferes with stimulus categorization. Interestingly, late reading time measures such as the total time spent on the critical word (*blue*) showed longer reading times for the related condition than for any other. If the positive correlation between P300 latency and behavioral data also extends to reading times, a correlation on the single-subject level should reveal that P300 latency predicts reading time. This correlation may hold in particular for late reading time measures such as total reading time or second-pass reading time, since the P300 is related to non-linguistic evaluation processes (see the preceding chapter). A similar prediction can be upheld with respect to P300 amplitude/ latency, and their interaction with judgment tasks. As P300 amplitude is a function of the level of confidence in stimulus categorization (Hillyard et al., 1971; Paul & Sutton, 1972; Squires et al., 1973b, 1975a), further correlative tests may reveal to what extent the information from a given word enables an individual to confidently make a judgment about it. These two aspects of the P300, latency varying with evaluation time and amplitude varying with confidence level, could play an important role in answering some unresolved questions concerning antonym comprehension. For instance, Roehm and colleagues (Roehm et al., 2007) reported a group effect indicating that one group of participants did not show any P300 activity (neither for antonyms nor for non-antonyms), when the critical word pairs were presented in isolation and an antonymy judgment had to be performed. Interestingly, the authors also reported a marginal trend towards longer reaction times for this group. One may therefore speculate whether the group was characterized by (i) a remarkable P300 latency variability and/or (ii) low confidence level in their decisions. The present experiment did not allow for any statistical correlations between P300 amplitude/ latency and reading time because the number of items per condition in the ERP data reduces statistical power too much to produce reliable results. To completely understand the connection between neuronal processes and behavioral output, future correlative tests are evidently due.

This line of research would thus be enlightening with respect to one of the central questions debated in psycholinguistic research, namely the precise relationship between the various eye movement measures and ERP waves. First, does a particular ERP effect bring about a specific eye movement? For instance, it has been suggested that the N400 correlates with increased fixation duration and that the late positivity (P600) is tied to regressive saccades (cf. Dambacher & Kliegl, 2007; Dimigen, Sommer, & Kliegl, 2007). The present experiments suggest that there are no such clear-cut divisions. The foveal N400 effect in Experiment 2 led to increased first fixation durations for non-antonyms. However, the foveal N400 effect reported in Experiment 4, reflecting animacy-induced interference effects in sentence processing, did not increase the first fixation duration, but caused more refixations on the current word, as indicated by the increase in first pass time. So, there is no clear pattern indicating how the N400 relates to the first fixation or refixations on a word. The correlation between late positivities and eye movements is similarly uncertain. Experiment 4 revealed two different late positivities, one in response to object-initial word orders and one in response to animacy-induced interference effects. While the first was accompanied by increased first fixation durations and longer first pass time, the second one was correlated with an increase in first pass time only. Thus, late positivities do not correlate exclusively with regressive saccades, but may increase fixation durations as well. Second, correlative analyses may prove useful in a more general sense by determining whether a given ERP component correlates with so-called early or late measures of eye movements, which are assumed to reflect early and late processing stages, respectively (cf. Clifton et al., 2007). The preceding section gave a first indication of how different processes in language comprehension, i.e. lexical processes and domain-general evaluation, may influence different eye movement measures (see chapter 5.3).

Finally, the present co-registration experiments have raised several technical issues that should be addressed in future experiments. The comparison of Experiments 4 and 5 indicated that the likelihood of finding parafoveal ERP effects is smaller when the fixation assumed to capture parafoveal processing is located on a word that induces processing costs. This may be the case with ambiguous words (as exemplified by case ambiguity in Experiment 4) or with words belonging to a scrambled syntactic constituent (e.g. the NPs in Experiment 5). Another open question also concerns the last fixation trigger, which was used to track parafoveal processing. Actually, the last fixation before the critical word could be one of various fixation types such as the last fixation on the word immediately to the left of the target word, but also any fixation further remote. Critically, parafoveal information might change as a function of how close this last fixation is to the target word. Therefore, it may be informative to have a last fixation trigger only on the immediately preceding word and to compare these results with the last fixation trigger as used in this thesis. Due to the low number of trials per conditions, such a division was not feasible for the reported experiments. Limiting the analysis of parafoveal processing to fixations immediately preceding the target word is advantageous in that it reduces the possibility that the last fixation trigger coincides with other fixation positions that may show processing effects (e.g., the first fixation on another critical word), especially if they are unrelated to the desired parafoveal effect. This overlap is likely to occur when linguistic stimuli with several critical regions are examined, particularly when these regions are adjacent to each other (cf. Experiment 5 and particularly Appendix E). The problem of overlapping fixation triggers can also be limited by only analyzing those trials in which readers made a single fixation on every word or region of text. The present experiments do not address the issue of multiple fixations on a word or region of text, despite claiming that N400 effects elicited by the first fixation on a word increase the probability of refixations on that word. At present, it remains to be seen whether such refixations induce ERP effects or not. In cases where there seem to be such ERP effects triggered by the second (or any later) fixation on a word, the extent to which it is comparable to the first fixation effects needs to be investigated in detail. For example, one should take into account whether the fixation of interest follows the canonical reading direction and whether it is followed by a progressive or regressive saccade.

VI Supplement: Additional tests of the N400 as an index of parafoveal processing in reading

6 Determining the depth of processing in the parafovea

Experiment 2, in which ERPs were collected while participants read sentences for comprehension, revealed that orthographic information obtained from the parafovea can reduce the amplitude of the N400 component when a word form is primed by the preceding context (cf. chapter 4.2.2). This parafoveal N400 priming effect was one of the most intriguing findings from Experiment 2 and it is the first indication of an N400 effect elicited by parafoveal information acquired during reading. As such further investigations using different sentence structures and different linguistic manipulations are necessary to develop a more detailed description of the electrophysiological indices of parafoveal processing in reading. One obvious question concerns the depth of processing in the parafovea. It was argued above that the parafoveal N400 priming effect is brought about exclusively by orthographic preprocessing of the upcoming word, whereas semantic processing of the to-be-fixated word commences when it is first fixated.

This interpretation is in line with numerous studies indicating orthographic preprocessing without semantic preprocessing in the parafovea (see Rayner, White, et al., 2003, for review). However, one may still contend that the distance between the last fixation prior to the critical word and the onset of the critical word is too large to allow for semantic processing, as the drop in visual acuity may only enable the extraction of orthographic codes. Moreover, Experiment 2 employed sentence structures such as *dass schwarz das Gegenteil von weiß ist, hat Gertrud gestern behauptet* ('Gertrud claimed yesterday that black is the opposite of white'). On average, the launch site fixation was on the word *Gegenteil* ('opposite'). Thus, readers concurrently processed the fixated word and identified the short preposition *von* ('of'), which was then skipped in the majority of trials. Both of these aspects may have prevented the extraction of anything more than orthographic information from the critical word, as the processing of the sequence *Gegenteil von* induced a relatively high foveal load, which would reduce the parafoveal preview on the critical word (see Henderson & Ferreira, 1993, for evidence regarding the adverse effects of foveal load on the parafoveal preview). Consequently, Experiment 2 does not offer fully conclusive evidence for the claim that there is no semantic preprocessing in the parafovea. Parafoveal words being closer to the current fixation could still receive qualitatively more parafoveal preprocessing, including semantics. On the other hand, Experiment 2 represents a priming context in which a specific lexical item and its orthographic form are strongly anticipated. It remains unclear whether other (non-priming) linguistic contexts may give rise to parafoveal effects of a similar kind.

Two further experiments were conducted to shed more light on these issues and to test the electrophysiological reality of parafoveal processing in reading. These experiments used permuted word orders and varied the animacy status of one verbal argument. Experiment 4 employed verb-final clauses, manipulating the order of two fully case-ambiguous NPs, which could serve as subject or accusative object in the clause. The two NPs were additionally characterized by different nominal properties (e.g., definiteness/ specificity, animacy). Importantly, the critical words were adjacent to each other, so that the parafoveal word was much closer to the current fixation than was the case in Experiment 2. This design is useful to investigate whether semantic information is extracted from adjacent parafoveal words to reduce ambiguity costs in processing. To sketch the major finding of Experiment 4 with respect to parafoveal processing: the animacy variation induced visible processing costs, as revealed by increased N400 effects and late positivity effects. However, these effects do not appear to be due to parafoveal processing. To investigate the hypothesis that the case ambiguity effectively inhibits parafoveal ERP effects, a second experiment with a similar design (Experiment 5) was carried out. In it, the order of subject and object in the clause was again permuted, but the two NPs were unambiguously marked for either nominative or dative case. The animacy status of one noun phrase was varied similarly to the preceding experiment. Experiment 5 revealed the remarkable finding that semantic information (i.e., animacy information) is acquired from the parafovea when the critical words are not case-ambiguous. However, this parafoveal N400 effect in response to animacy information was only present for pre-critical parts of the sentence. This is taken to suggest that linguistic stimuli with more than one critical region are less suited to detect parafoveal ERP effects in linguistic processing.

VII Experiments

7.1 Experiment 4: Subject-object ambiguity and syntactic reanalysis

7.1.1 Introduction

The investigation of so-called garden path sentences has proven fruitful in detecting basic mechanisms in sentence parsing. Garden path sentences are ambiguous structures in which the parser initially adopts a specific structural analysis for the ambiguous input, which, however, turns out to be false at some later point, making it necessary to reanalyze the syntactic structure. Reanalysis refers to the process of revising the initial misanalyzed parse. A large body of data has been assembled by investigating garden path sentences across different garden path types in various language and much has been learned about the assumed underlying processes in initial processing and reanalysis (e.g., for English Frazier & Rayner, 1982; Kaan & Swaab, 2003; Staub, 2007a, b; Osterhout & Holomb, 1992; Osterhout, Holcomb, & Swinney, 1994; for Dutch, see, e.g., Brysbeart & Mitchell, 1996; Frazier & Flores D'Arcais, 1989; for Spanish see Meseguer, Carreiras, & Clifton, 2002; Carreiras, Sallilas, & Barber, 2004). As for German, subject-object ambiguities are probably one of the most well-studied instances of garden path sentences (e.g., Bader & Meng, 1999; Bornkessel, McElree, Schlesewsky, & Friederici, 2004; Friederici & Mecklinger, 1996; Friederici, Mecklinger, Spencer, Steinhauer, & Donchin, 2001; Haupt et al., 2008; Knoeferle, Habets, Crocker, & Münte, 2008; Kretzschmar, Bornkessel-Schlesewsky, Staub, Roehm, & Schlesewsky, in press; Mecklinger et al., 1995; Scheepers, Hemforth, & Konieczny, 2000; Schlesewsky, & Bornkessel, 2006; Schriefers, Friederici, & Kühn, 1995).

In German, the syntactic functions of subject and object are not determined by word order, but by case marking on the verb arguments. Specifically, nominative case is used to indicate the subject of the clause, which agrees in number and person with the verb, whereas accusative and dative case encode objects. Ambiguities in syntactic function assignment then arise because of the high degree of case syncretism in the German nominal system. For example, in the sentence fragment *die Prinzessin küsst* ('the princess kisses') the initial female noun phrase can be interpreted as bearing either nominative or accusative case. The fragment can be disambiguated by a post-verbal noun phrase (NP) that bears unambiguous case assignment, such as *die Prinzessin küsst den König* ('the princess kisses the king'), where the post-verbal masculine NP marked with the accusative signals a subject-verb-object (SVO) structure. If, by contrast, the fragment were completed with an argument bearing nominative case (*der König*), the structure would have to be interpreted as OVS. When both NPs are case-ambiguous as in *dass Peter Sängerinnen* the clause-final verb disambiguates the structure via the matching of number information. So, *dass Peter Sängerinnen küsst* would be interpreted as a subject-initial clause, whereas *küssen* would index an object-initial structure. The garden path effect has been explained as resulting from the parser's preference to assign the first case-ambiguous NP the syntactic subject function (cf. Bader & Meng, 1999). When the number feature of the verb does not match the number feature of the presumed subject or when the case-marking of the second NP enforces an object marking on the first NP so that the subject-first analysis cannot be upheld, the entire structure must be revised to build up the correct object-initial syntactic structure.[37] These reanalysis costs are visible in extended reading times (e.g., Scheepers et al., 2000), lower ratings in judgment tasks (e.g., Bader & Meng, 1999), or in pronounced ERP components (e.g., Bornkessel, McElree et al., 2004; Mecklinger et al., 1995).

The question as to which ERP component is sensitive to syntactic reanalysis in German has been subject to intense research that has yielded mixed results. While there is evidence that the ambiguous region itself enhances the amplitude of the P600/ late positivity (Frisch, Schlesewsky, Saddy, & Alpermann, 2002), the brain responses measured at the

[37] The person feature is irrelevant at this point because subject-object ambiguities only apply to the third person paradigm.

disambiguating position were originally found to elicit a P600 (Friederici & Mecklinger, 1996; Friederici et al., 2001; see Mecklinger et al., 1995, for discussion of latency changes due to sentence structure). However, two recent studies have shown that this may not hold. Bornkessel, McElree et al. (2004) hypothesized that there may be qualitative differences in reanalysis difficulty based on object case assignment. They compared verbs assigning accusative case (e.g., *besuchen* ‘to visit’) to verbs assigning dative case (e.g., *folgen* ‘to follow’). As in previous studies, accusative verbs engendered a pronounced P600 when the sentence had to be reanalyzed towards an object-initial reading, whereas dative verbs showed an N400. Bornkessel and colleagues carried out two additional experiments that lent converging support to their interpretation of the N400 effect as reflecting true reanalysis costs and as arguing against the view that the N400 to dative verbs is some instance of implausibility detection. They argued in particular that the N400 in response to dative verbs is due to the availability of unmarked dative-initial structures in the German grammar (see also Schlesewsky & Bornkessel, 2006, for a detailed account of how dative verbs benefit from unmarked dative-nominative structures containing dative-object experiencer verbs; see Wunderlich, 1997, for theoretical arguments for a dative-initial base order). This renders syntactic reanalysis easier, as it only involves the revision of case marking, whereas accusative-initial structures call for phrase structure revisions.

Haupt and colleagues (Haupt et al., 2008) recently elaborated this line of research with a detailed investigation of the effects of task environment on the processing of subject-object ambiguities (see also chapter 1.3.2). They had their participants listen to short stories that contained two fully case-ambiguous NPs followed by either a dative or an accusative verb that disambiguated the clause towards a subject-initial or an object-initial word order (cf. example 12) – similar to the items in the Bornkessel, McElree et al. study. The authors found a biphasic N400-late positivity pattern for both verb classes at the disambiguating auxiliary, the position where the initial subject preference was disconfirmed and reanalysis became necessary (12b, d).

(12) Example stimuli from Haupt et al. (2008)

(a) … dass Betram Surferinnen geärgert hat, …
that Betram$_{NOM/ACC/DAT.SG}$ surfers$_{NOM/ACC/DAT.PL}$ annoyed$_{ACC}$ has$_{SG}$
'… that Betram annoyed surfers…'

(b) … dass Betram Surferinnen geärgert haben, …
that Betram$_{NOM/ACC/DAT.SG}$ surfers$_{NOM/ACC/DAT.PL}$ annoyed$_{ACC}$ have$_{PL}$
'…that surfers annoyed Betram …'

(c) … dass Betram Surferinnen gratuliert hat, …
that Betram$_{NOM/ACC/DAT.SG}$ surfers$_{NOM/ACC/DAT.PL}$ congratulated$_{DAT}$ has$_{SG}$
'…that Betram congratulated surfers …'

(d) … dass Betram Surferinnen gratuliert haben, …
that Betram$_{NOM/ACC/DAT.SG}$ surfers$_{NOM/ACC/DAT.PL}$ congratulated$_{DAT}$ have$_{PL}$
'…that surfers congratulated Betram …'

In a second experiment, they were able to show that the occurrence of the biphasic pattern was not affected by the altering task instructions (even though it was clearly sensitive to changes in task), nor a consequence of contextual factors, as the same results emerged when the critical sentences were presented in isolation. Thus, the authors suggested that the N400 is the reanalysis component proper for all verb classes and that the ensuing late positivity reflects the markedness of the object-initial structure. Previous results that revealed no N400 to accusative verbs are accounted for as resulting from strategic processing, which causes the emergence of a P300 component and thus component overlap in the N400 time window, which reduced the possibility of finding clear N400 effects (cf. Haupt et al., 2008, general discussion). Some dative verbs did not elicit a late positivity, the authors explained this by pointing to differences in material (see also Schlesewsky, & Bornkesel, 2006, for an explanation of the finding that the late positivity for dative-first constructions disappears when case information and syntactic disambiguation are conveyed by the same lexical item).

Overall, there seems to be well-grounded evidence that the N400 in these studies indexes syntactic reanalysis, while the late positivity indexes the general markedness of the resulting structure. The present experiment was designed to provide independent support for the reanalysis N400 and to examine further possible sources of the late positivity.

Regarding the N400, both Haupt et al. (2008) and Bornkessel, McElree et al. (2004) have already put forward several arguments to mitigate the main criticism against their reanalysis account, namely that the N400 could be induced by subtle differences in plausibility between subject-initial (more plausible) and object-initial (less plausible) structures. However, a more stringent test is necessary to compare the comprehension of object-initial with subject-initial structures while clearly excluding the possibility of reanalysis. The present experiment used clauses that were structurally similar to the stimuli in the Bornkessel, McElree et al. (2004) and Haupt et al. (2008) studies. Participants read sentences that contained two NPs ambiguous between nominative, accusative, or dative case marking and a finite verb in the clause-final position (see example 13). The two NPs were always a proper name bearing the subject function and a bare plural noun bearing the object function. Thus, one and the same NP type was always used to encode a particular syntactic function. Word order was manipulated by changing the relative order of proper name and bare plural, resulting in subject-initial structures whenever a proper name occurred in the first NP position (13a, c) and object-initial structures whenever a bare plural occurred first (13b, d). Since the design was not balanced with respect to the allocation of syntactic function to NP types, syntactic ambiguity was restricted to case marking. Semantic information provided by the first NP served as the critical disambiguating information (i.e., the distinction between a specific proper name and a non-specific bare plural), inhibiting ambiguity in syntactic function assignment. Since it has been shown that language comprehension also proceeds in an incremental manner in verb-final structures (Bader & Lasser, 1994; Kamide & Mitchell, 1999), the language comprehension system should use NP semantics to adopt the correct subject-initial or object-initial reading at the beginning of the clause, preventing reanalysis effects at the clause-final verb. Consequently the subject preference should not be upheld for initial bare plurals so that there is no garden path effect. The emergence of a reanalysis N400 at the position of the verb is thus highly unlikely because initial bare plurals should always be immediately and correctly analyzed as objects under these conditions. This predicts that if the present experiment does not give rise to an N400 to object-initial clauses, it would provide direct support for the reanalysis N400, as its absence is expected due to the imbalanced design. On the other hand, if a clear N400 effect emerges for object-initial clauses, it would call into question the validity of the reanalysis N400, as only the alleged variation in plausibility could give rise to an N400 in the present design.

(13) Example stimuli for subject-object ambiguities disambiguated by accusative verbs[38]

(a) Dass Elfriede Statisten vergisst, hat viele meist verärgert. (SO-AA)
That Elfriede$_{\text{NOM/ACC/DAT.SG}}$ movie extras$_{\text{NOM/ACC/DAT.PL}}$ forgets$_{\text{ACC.SG}}$, has many often annoyed
'It annoyed many people that Elfriede forgets movie extras.'

(b) Dass Statisten Elfriede vergisst, … (OS-AA)

(c) Dass Elfriede Tabletten vergisst, … (SO-AI)
That Elfriede$_{\text{NOM/ACC/DAT.SG}}$ pills$_{\text{NOM/ACC/DAT.PL}}$ forgets$_{\text{ACC.SG}}$, …
'It annoyed many people that Elfriede forgets pills.'

(d) Dass Tabletten Elfriede vergisst, … (OS-IA)

To test the functional scope of the late positivity reported in the Haupt et al. study, the animacy status of the bare plurals was varied so that it was either animate (13a, b) or inanimate (13c, d). This test was derived from a recent neurocognitive model of language comprehension that posits two functionally dissociable types of late positivity (the extended Argument-Dependency Model (eADM); cf. Bornkessel & Schlesewsky, 2006; Bornkessel-Schlesewsky & Schlesewsky, 2009a,b).

[38] Abbreviations: SO – subject-before-object clause; AA – two animate arguments; AI/ IA – inanimate object and animate subject

The eADM model assumes three processing stages, each made up of functionally distinct processes. During the first stage, word category information and basic constituent structure are processed. This information is used in the initial phase of the second stage to determine whether the perceived word form is a potentially predicating form (i.e., a verb). Non-predicting items (i.e., NPs serving as verb arguments) are processed with respect to their prominence information in the "compute prominence" step. Prominence information is assumed to encompass semantic information that is relevant for the other step in stage two, the linking of verb arguments to the logical structure of the verb (the step of "argument linking"). The logical structure of the verb includes argument slots to which the previously encountered NPs are linked (see Van Valin, 2005, for theoretical motivation). These slots are abstract representations of the semantic roles that are involved in the event depicted by the verb. Only two generalized semantic roles, the actor and the undergoer of the event, are postulated within the eADM.

The computation of prominence information is highly important for the identification of the generalized semantic role an argument carries in a transitive event. An argument can either serve as the actor (the most agent-like argument) or undergoer (the most patient-like argument) of the event depicted by the verb (cf. Bornkessel-Schlesewsky & Schlesewsky, 2009b). Actors are prototypically independent of the event, and are also "prototypically sentient, cause the event described and […] (consciously) in control of it" (Bornkessel-Schlesewsky & Schlesewsky, 2009b, p. 41). Prominence information is primarily used to assign the actor role, which follows from the assumption that the undergoer role possesses no defining properties of its own. That is, the undergoer is conceptualized as the counterpart of a given actor, best defined by the absence of actor properties (see Primus, 1999, and Bornkessel-Schlesewsky & Schlesewsky, 2009a,b). Due to this asymmetry in the definition of actor and undergoer, all of the available prominence information is used to determine the most prominent argument, which is, by definition, the most likely candidate to carry the actor role. Prominence information is assumed to vary across languages and to be internally structured into primary and secondary information types (see Bornkessel-Schlesewsky & Schlesewsky, 2008, for a cross-linguistic application); for German, case marking on NPs is the primary means to identify the roles of preverbal arguments (e.g., nominative case is always associated with the actor role), whereas semantic information types such as animacy or definiteness/ specificity usually have a modulating capacity. Based on these types of prominence information, preverbal arguments can be assigned a generalized semantic role even before the verb is encountered. When different types of prominence information conflict with each another or lead to a situation in which more than one argument qualifies as the actor argument, it is assumed that N400 effects emerge (Bornkessel & Schlesewsky, 2009b).

When the verb is finally encountered, the arguments are linked to its logical structure and their prominence information is integrated with the semantic requirements of the verb. The model assumes different scenarios for how linking costs arise from the distribution of prominence features. These scenarios are derived from the core principle of the model, the *distinctness principle*. It posits that “the participants in an event should be as distinct as possible from one another in terms of all available dimensions of prominence” (Bornkessel-Schlesewsky & Schlesewsky, 2009b, p. 44). The scenario most relevant for the present experiment occurs when two arguments are similarly prominent with respect to the sum of their rankings on the prominence scales. The competition of arguments for the actor role is visible in processing costs that are reflected in the N400 component (see Bornkessel-Schlesewsky & Schlesewsky, 2009b, for more details). Finally, the third and final stage of processing comprises two different steps. The first of these is the “generalized mapping” between the information collected from the computation of prominence, the outcome of argument linking, and other information types such as plausibility (which is processed separately from, but in parallel with, argument prominence). The outcome of this mapping is then evaluated for its overall well-formedness. Both of these steps are reflected in late positivity effects. Late positivities elicited by well-formedness evaluations are explicitly assumed to be a member of the P300 family (Bornkessel-Schlesewsky & Schlesewsky, 2006).

Let us now consider how the animacy manipulation of the present model fits into this model. The distinctness principle predicts that processing costs will arise whenever the two arguments share the same prominence feature. Two animate NPs should therefore lead to processing costs because they are not distinct with regard to animacy. As this is related to the concept of interference developed in memory-based accounts of language processing (e.g., Gordon, Hendrick, & Johnson, 2001; McElree, 2006), it can be termed an animacy-induced interference effect. The animacy interference effect should elicit both an N400 and a late positivity, depending on sentence position (i.e., argument or verb; see below).

Note furthermore that the well-formedness of the object-initial structures is additionally reduced because bare-plural objects precede proper names. There is a general syntax-independent linearization principle stating that verbal arguments should best be ordered along prominence hierarchies, and a growing body of evidence suggests that a violation of this principle leads to visible processing costs. For example, arguments that rank higher on a given hierarchy (i.e., they are more prominent) should preferably come before arguments ranked lower on that scale, e.g. animate NPs should precede inanimate NPs and specific NPs should precede non-specific NPs (see Aissen, 2003, for theoretical foundations of argument hierarchies, and Bornkessel-Schlesewsky & Schlesewsky, 2009b, Bornkessel, Zysset, Friederici, von Cramon, & Schlesewsky, 2005; Grewe, Bornkessel, Zysset, Wiese, von Cramon, & Schlesewsky, 2006, for empirical support from functional neuroimaging). As for German, it has been claimed that the preferred linear order with respect to the definiteness/ specificity hierarchy is that definite/ specific NPs precede indefinite/ non-specific NPs (Lenerz, 1977). Consequently, proper names should precede bare plurals, with a deviant order engendering processing costs. This hypothesis has recently been confirmed in two studies, one using functional neuroimaging and one using eye movements in reading (Bornkessel-Schlesewsky, Schlesewsky, & Cramon, 2009; Kretzschmar et al., in press).

In sum, these considerations give rise to the following hypotheses. At the first NP position, one should not see a significant effect related to the computation of the word order.[39] One may see a lexical animacy effect reflected in a general disadvantage for inanimate over animate NPs, which would increase the N400 amplitude to inanimate bare plurals. At the second NP position, it is possible to evaluate the semantic properties of both NP types, enabling the computation of prominence information from each argument and their relation to one another. This view predicts that animacy-based interference should exert its influence at this stage of parsing (the "compute prominence" step), where two NPs compete for the actor role. Specifically, there should be animacy-based interference for the two animate NPs in examples (13a, b) above, which are less distinct than their counterparts in (13c, d) in terms of animacy. This lack of distinctness should result in processing disruption, measurable as an increase in N400 amplitude (see also Frisch & Schlesewsky, 2001, who found an N400 for two animate NPs in ungrammatical structures, and Roehm, Schlesewsky, Bornkessel, Frisch, & Haider, 2004, who found an increase in N400 amplitude when the actor argument ranked lower on the animacy scale than the preceding undergoer argument). One final prediction can be derived from a previous study that reported increased N400 amplitudes at the second NP of a transitive event when it was the subject of the clause (Bornkessel, Fiebach, & Friederici, 2004). This effect was taken to reflect the switch from an intransitive to a transitive structural template. The default preference for an intransitive template is also inferred from the distinctness principle, as the single argument of an intransitive relation is maximally distinct by virtue of its singularity. At the verb, animacy interference effects should influence the

[39] Since the semantic cue about word order eliminates ambiguity in syntactic function, it should also prohibit a P600 indexing syntactic function ambiguity at this position (cf. Frisch et al., 2002). Due to the imbalanced design of the present experiment, however, this contrast was not statistically examined (see below).

stages of argument linking and the generalized mapping of prominence/ linking information with other information types used in language comprehension. Processing costs in the stage of argument linking should lead to an increase in N400 amplitude. As generalized mapping processes are thought to elicit a late positivity, an enhanced positivity should be expected, because interference affects the retrieval of the two NP representations.

It is also possible to predict the occurrence of a completely different late positivity effect. If there is a late positivity related to the well-formedness evaluation of object-initial sentences, as argued by Haupt et al. (2008), it should remain unaffected by the interference manipulation and, as assumed in the eADM, it should be a member of the P300 family. This leads to an intriguing hypothesis. If the well-formedness positivity is indeed a member of the P300 family and if, as Haupt and colleagues propose, accusative verbs are especially expected in the processing of subject-object ambiguities, the present stimuli provide another context in which the occurrence of the P300 can be examined. Recall from Experiment 2 that a highly constraining context leads to the anticipation of both a specific word form and a specific meaning and that the P300 is effectively inhibited if evidence from both domains is not collected simultaneously. The present stimuli differ from these high-constraint contexts in that one semantic feature (i.e., the assignment of accusative case and generalized semantic roles associated with it) is predicted, while a particular word form is not, as this would amount to predicting a specific verb. Thus, only a semantic prediction needs to be met, which means a P300-like late positivity might emerge when the verb is first fixated and its semantic information retrieved. In other words, the temporal dissociation of orthographic and semantic processing in natural reading should not exert an equally bad influence as compared to Experiment 2, while effects of parallel information processing across words could still influence P300 amplitude. Hence, if initial bare-plural objects induce processing costs (unrelated to plausibility), they are likely to engender an increased late positivity, which is associated with the well-formedness evaluation. The robustness of this positivity and of the one engendered by animacy interference will help resolve whether a late positivity indexes the generalized mapping or the well-formedness evaluation stage in processing (i.e., if it is a member of the P300 family or not).

The predictions regarding the effects of parafoveal preview on these processing correlates are easily summarized in view of the results from Experiment 2. If substantially more than just prelexical orthographic information is extracted from parafoveal vision, both the animacy-induced and the template switch N400 effects are likely to emerge as parafoveal N400 effects. It is conceivable that the N400 elicited by the template switch could be parafoveal in nature in any case, as orthography is a productive means to distinguish nominal from non-nominal expressions (i.e., capitalization of the initial letter is indicative of nominal expressions). Thus, the word category information of the upcoming second NP can be recognized in the parafovea. Furthermore, if the parafoveal preview per se is responsible for the lack of the P300 in Experiment 2, there should be no late positivity for the well-formedness evaluation of the object-initial clauses (i.e., bare plural-initial clauses). If, on the other hand, the parafoveal preview interacts with the extent to which the linguistic context allows the prediction of a specific lexical meaning and a specific orthographic form, so that the temporal dissociation of orthographic and semantic information is only crucial in high-constraint contexts, a well-formedness positivity could come into existence when the verb position is first fixated. Only semantic information from the verb, but not the specific word form, is predictable in the present stimuli and the dissociation of orthographic processing in the parafovea and semantic processing in the fovea should not matter.

7.1.2 Method

Participants

48 students (24 female; mean age: 23.8 years; range 19-31) from the University of Mainz were paid to participate in the experiment after giving informed consent. One participant was

replaced because of very low accuracy in the comprehension task and two further participants were replaced because of low calibration accuracy or recording problems during the experiment. All participants were right-handed native speakers of German, as assessed by the German version of the Edinburgh handedness test (Oldfield, 1971). All participants reported normal or corrected-to-normal vision and no history of any neurological or psychiatric disorders. They were naïve to the purpose of the experiment.

Materials

40 sets of sentences were constructed by combining 20 different verbs with two sets of noun pairs each. The critical clauses always contained a fully case-ambiguous proper name and a fully case-ambiguous bare plural noun phrase that varied in animacy (cf. example 13). The ambiguous NPs and the disambiguating verb were combined to form a plausible subordinate clause, which was followed by a main clause. This was done to avoid sentence-wrap effects at the subordinate verb position, which likely to affect eye movement patterns. Two factors were manipulated. First, the word order between proper name and bare plural was varied, with proper names always serving as the subject of the subordinate clause. The type of the initial NP thus unambiguously indicated the syntactic word order, i.e. initial proper names resulted in a subject-initial reading, whereas initial bare plurals led to an object-initial reading. Second, the animacy status of the bare plural was changed so that the object was either animate or inanimate.

The nouns and verbs were matched for frequency and length. Proper names were on average 8.6 letters long, inanimate NPs were on average 8.8 letters long, and animate NPs 8.9. Frequency was matched by choosing nouns that exhibited similar frequency classes, as assessed by the Wortschatz corpus from the University of Leipzig (frequency classes 11 or 12; www.wortschatz.uni-leipzig.de). The verbs had a mean length of 7.5 letters and a similar frequency to the nouns. The items were distributed to 4 lists using a Latin Square design so that each participant saw 10 sentences per condition, but no item twice within a list. The critical sentences were interspersed with 140 filler sentences of different structures. The stimuli were presented in a pseudo-randomized order. A comprehension question (requiring a yes or no answer) followed each trial, probing that trial's lexical contents. Syntactic relations were probed on roughly 25% of all trials. Participants had an average accuracy rate of 82%.

Apparatus and Procedure

Apparatus and experimental procedure were identical to Experiment 2, reported above.

Data preparation and analyses

The ERP and eye movement data were prepared for analysis as described for Experiment 2. The eye movement record of each participant was screened for erroneous trials. Less than 2% of all trials were excluded due to track losses near/at the critical region or due to technical errors. Blinks and fixations below 80 ms or above 800 ms were deleted when they were located in the critical regions; short fixations do not reflect cognitive processing and unusually long fixations are likely to reflect track losses (cf. Rayner & Pollatsek, 1989). If this resulted in a no-fixation event for one of the critical regions, these trials did not enter the computation of fixation-related potentials. 3 trials were excluded due to this criterion. All fixation-related potentials were triggered relative to fixation onset. For the eye-movement analysis short fixations with less than 80 ms in duration were incorporated into an adjacent fixation if the distance between these two did not exceed one character, and fixations less than 40 ms in duration were treated similarly if they were within one character of an adjacent fixation. Short fixations that did not meet these criteria were excluded from further analysis. Fixations longer than 800 ms were also excluded. In total, less than 1% of fixations was eliminated based on these criteria.

The regions of interests (ROI) were defined identically to Experiment 2. ERP waves were averaged and analyzed separately for each of the three critical regions (first case-ambiguous NP, second case-ambiguous NP, verb). The ERPs were averaged per participant and condition from 200 ms before fixation onset to 800 ms post fixation onset. Grand averages were then computed over all participants. This time window was chosen because any components in response to the manipulations were expected to emerge before 800 ms and the chance was high to find very late effects that are probably caused by fixations downstream the sentence. Since the most reliable results from Experiment 2 were obtained when the first fixation on a word and the last fixation before that word were used as averaging points, the following analyses were restricted to these two fixation points. The length of the words used in the present experiment increased the probability of refixations on a word (cf. Kliegl et al. 2004), leading to generally fewer single fixations, so that the ERP measures probably lack statistical power because fewer subjects contribute data. The critical subordinate clause was divided into the regions of analyses as given in example (14). The same regions were used for the ERP and eye movement analyses.

(14) Dass/ Elfriede/ Statisten/ vergisst,/ …

It was mentioned above that the materials were not constructed with a fully balanced 2x2 design of the factors word order and animacy. The syntactic function of proper names and bare plurals was not counterbalanced across sentences so that the relative order between the two NP types was confounded with an experimentally induced cue for word order. This prohibited statistical tests for NP1 regarding word order effects. However, tests for the factor animacy were possible for the bare plural NP1s. Thus, only one pair-wise comparison for animacy effects was performed for NP1, whereas the effects of both word order and animacy were examined for NP2 and verb position. Alpha was adjusted for NP1 by using a modified Bonferroni procedure (Keppel, 1991) to avoid overestimating of effects obtained with pair-wise comparisons. Alpha was set to $p < .05$ for NP1. A 2x2 design involving the factors ORDER (subject-initial vs. object-initial) and ANIM (animate bare plural vs. inanimate bare plural) was computed for NP2 and verb positions. To avoid type I errors resulting from violations of sphericity, alpha was corrected as proposed by (Huynh and Feldt, 1970) when effects with more than one degree of freedom in the numerator were evaluated. In these cases, the original degrees of freedom are reported with the corrected alpha. The eye movement results were analyzed using the same statistical models as indicated above. Participants and items were treated as random factors F_1 and F_2, respectively.

7.1.3 Results

ERP results

Figure 6 gives the ERP waves collected for the first fixation on NP1. Visual inspection of the ERP waves triggered by the first fixation suggests no particularly enhanced activity for any of the experimental conditions in earlier time windows. There was, however, a small negativity for the object-initial condition involving two animate NPs, ranging between about 400 and 540 ms post fixation-onset.

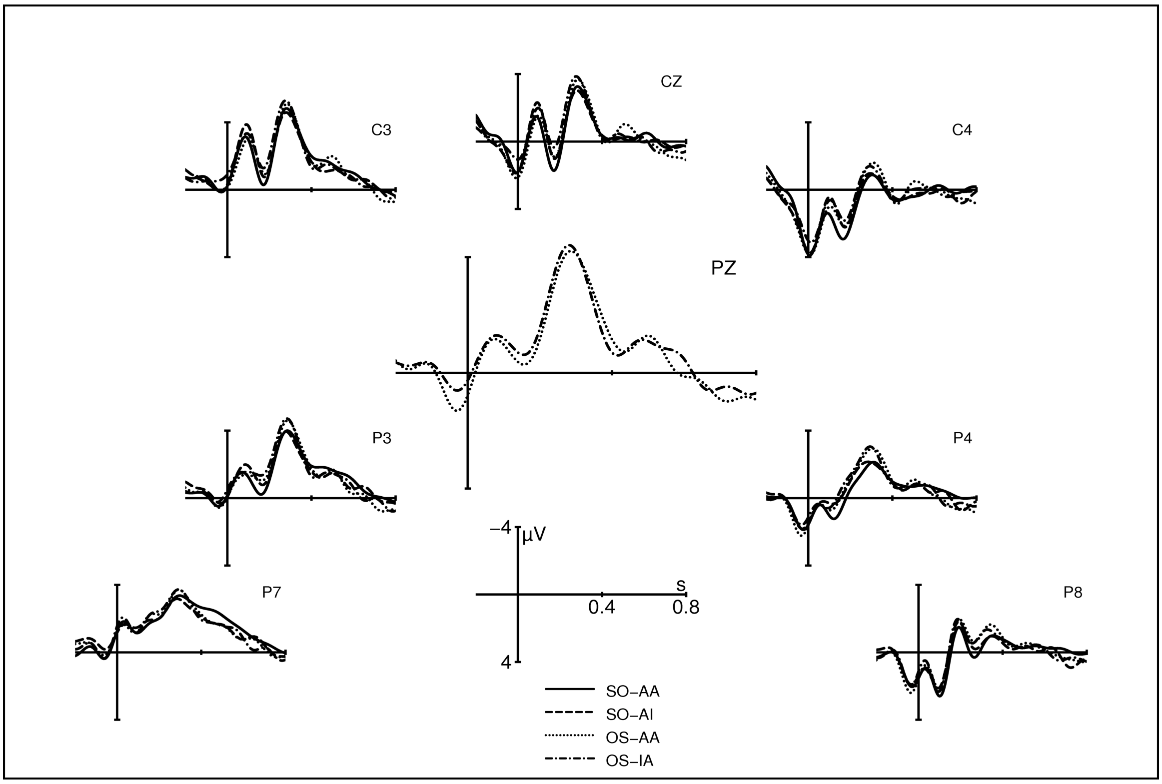


Figure 6. Grand-average ERPs (N = 48) time-locked to the onset of the first fixation on NP1 (onset at the vertical bar).

The last fixation before readers made a saccade to NP1 showed that the object-initial condition with an inanimate bare plural elicited a pronounced negativity between 200 and 300 ms post fixation-onset (cf. Figure 7). There was a very late effect between about 650 and 800 ms post fixation-onset, with an increased negative-going ERP wave to object-initial clauses with two animate NPs. However, visual inspection of the first fixation on NP2 suggests that the processing of the second NP triggered this late effect (see Figure 8).

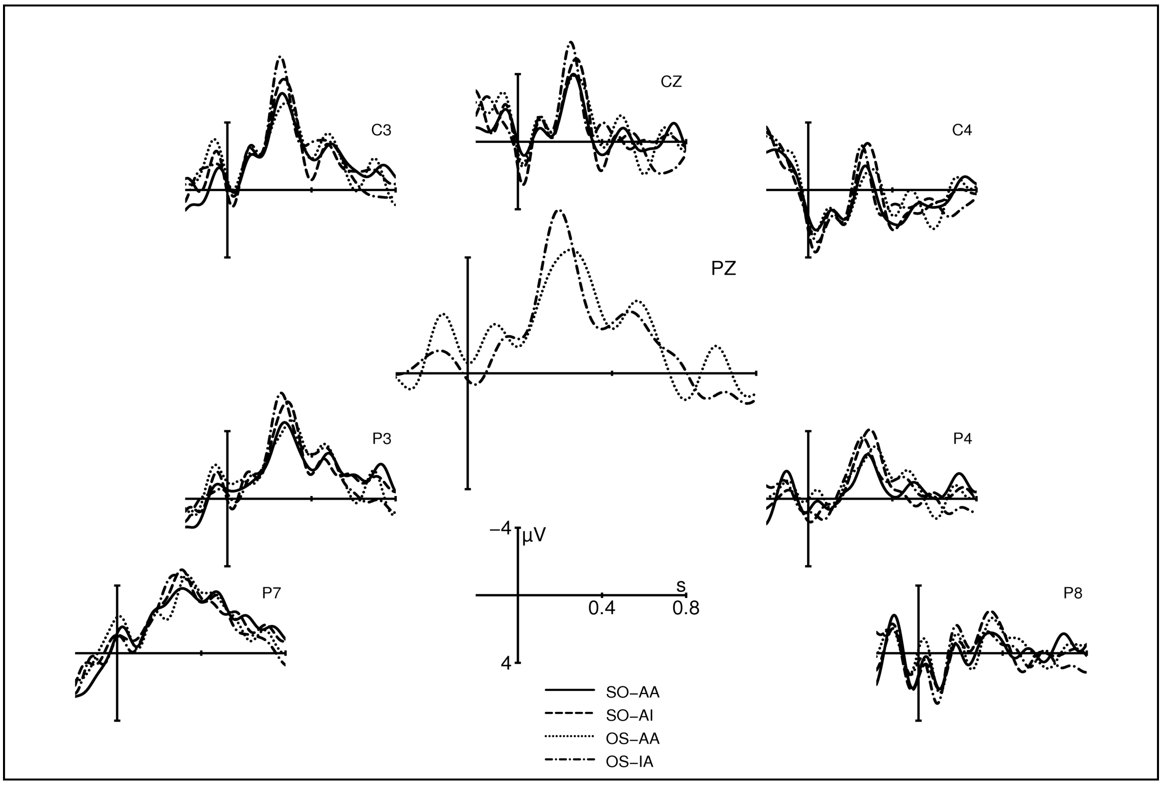


Figure 7. Grand-average ERPs (N = 30) time-locked to the onset of the last fixation before NP1 (onset at the vertical bar).

Specifically, when the second NP was first fixated, the subject-initial conditions appeared to elicit a negativity between 250 and 420 ms post fixation-onset (see also Figure 9). Similarly, when both NPs in the clause were animate, this enhanced the negative component between 250 and 400 ms (see also Figure 10). Both of these effects were present at all centro-parietal electrodes and symmetrically distributed across both hemispheres.

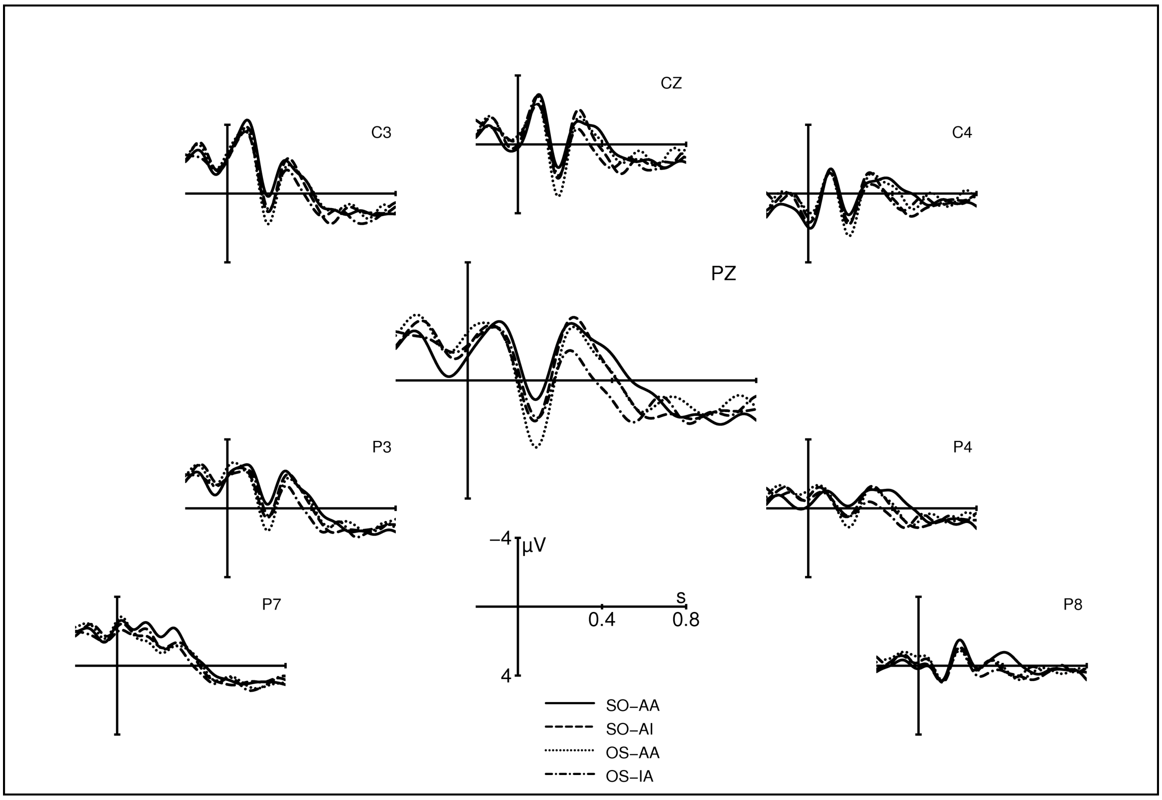


Figure 8. Grand-average ERPs (N = 48) time-locked to the onset of the first fixation on NP2 (onset at the vertical bar).

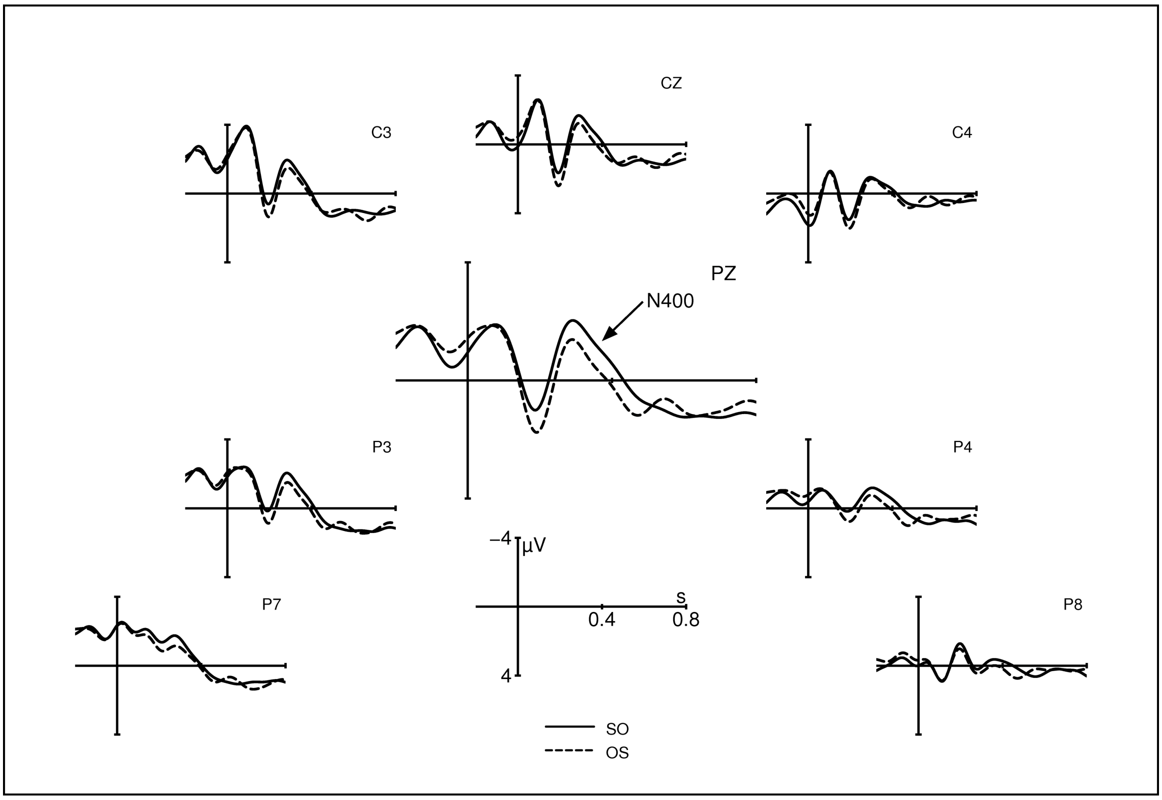


Figure 9. Grand-average ERPs (N = 48) time-locked to the onset of the first fixation on NP2 (onset at the vertical bar). Grand-average ERPs are collapsed across the animacy factor to show the influence of the word-order manipulation.

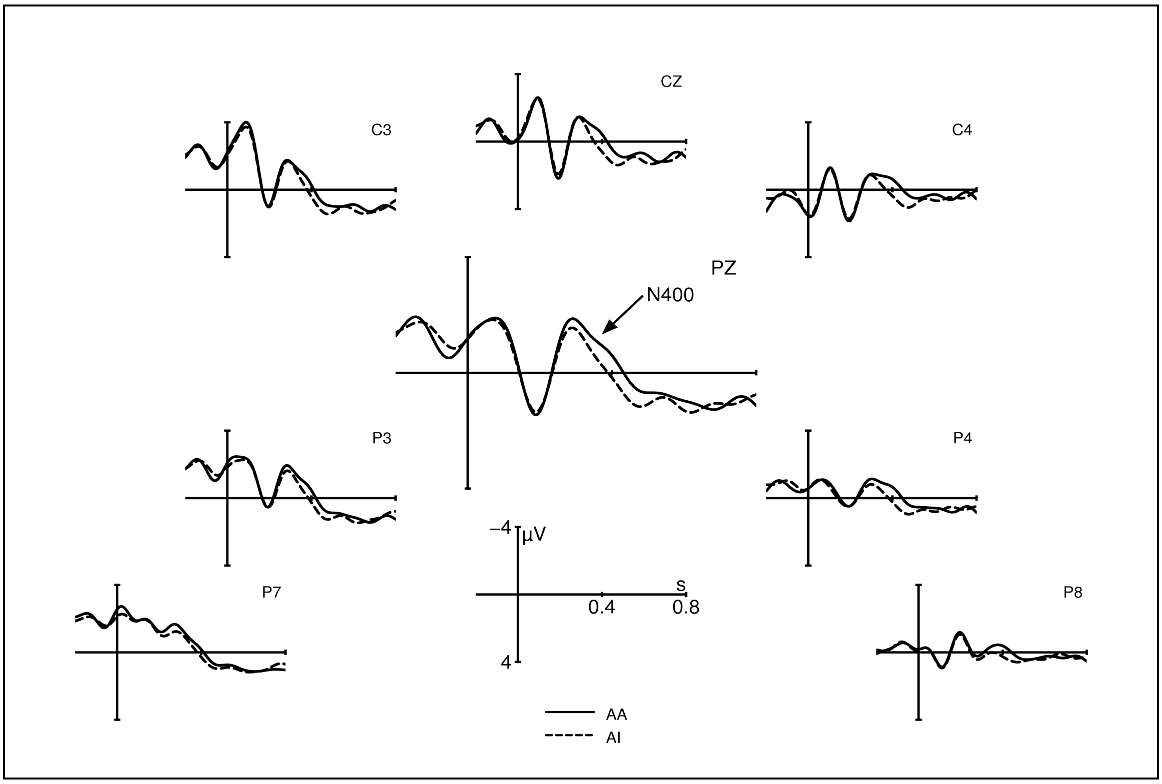


Figure 10. Grand-average ERPs (N = 48) time-locked to the onset of the first fixation on NP2 (onset at the vertical bar). Grand-average ERPs are collapsed across the word order factor to show the influence of the animacy manipulation.

Figure 11 shows ERP waves triggered by the last fixation before NP2. From about 540 ms post fixation-onset through the end of the epoch, these averages mirror the effects reported for the first fixation. Thus, these late effects are most likely the result of overlapping waves from the last fixation and first fixation measures. There was no consistent pattern in the N400 time window, with the exception that the subject-initial condition with two animate NPs seemed to be more negative going than the remaining conditions.

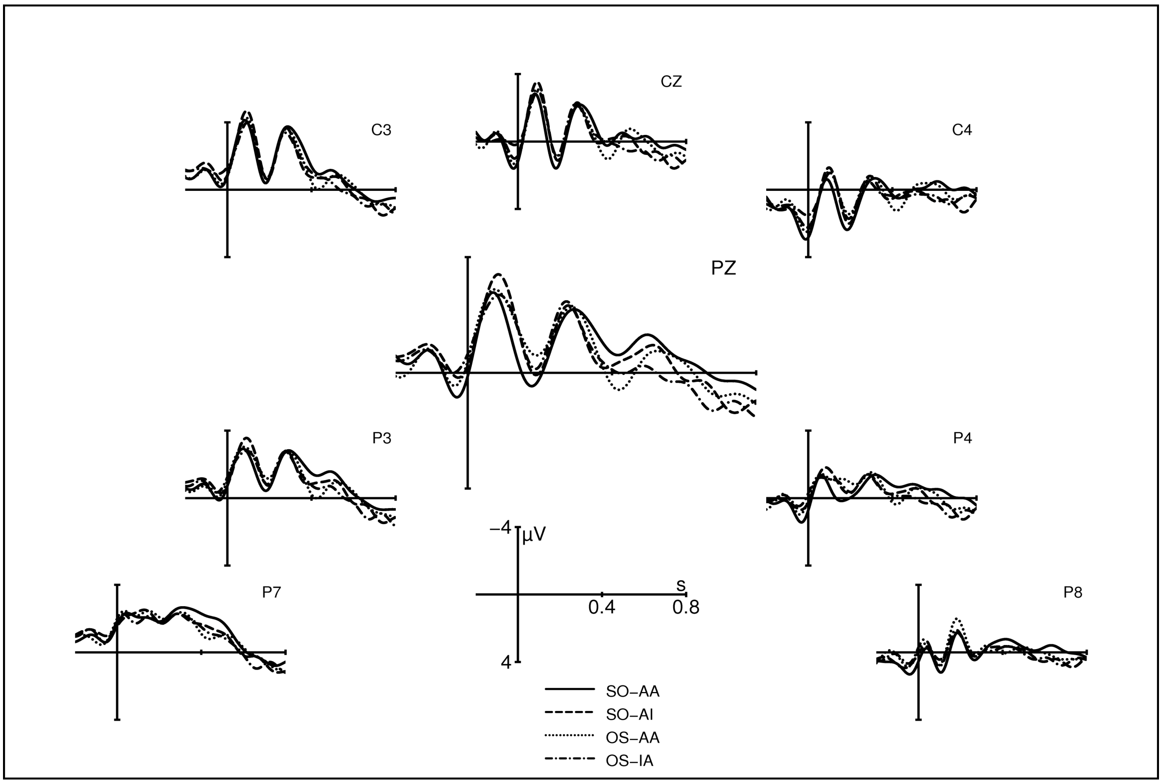


Figure 11. Grand-average ERPs (N = 48) time-locked to the onset of the last fixation before NP2 (onset at the vertical bar).

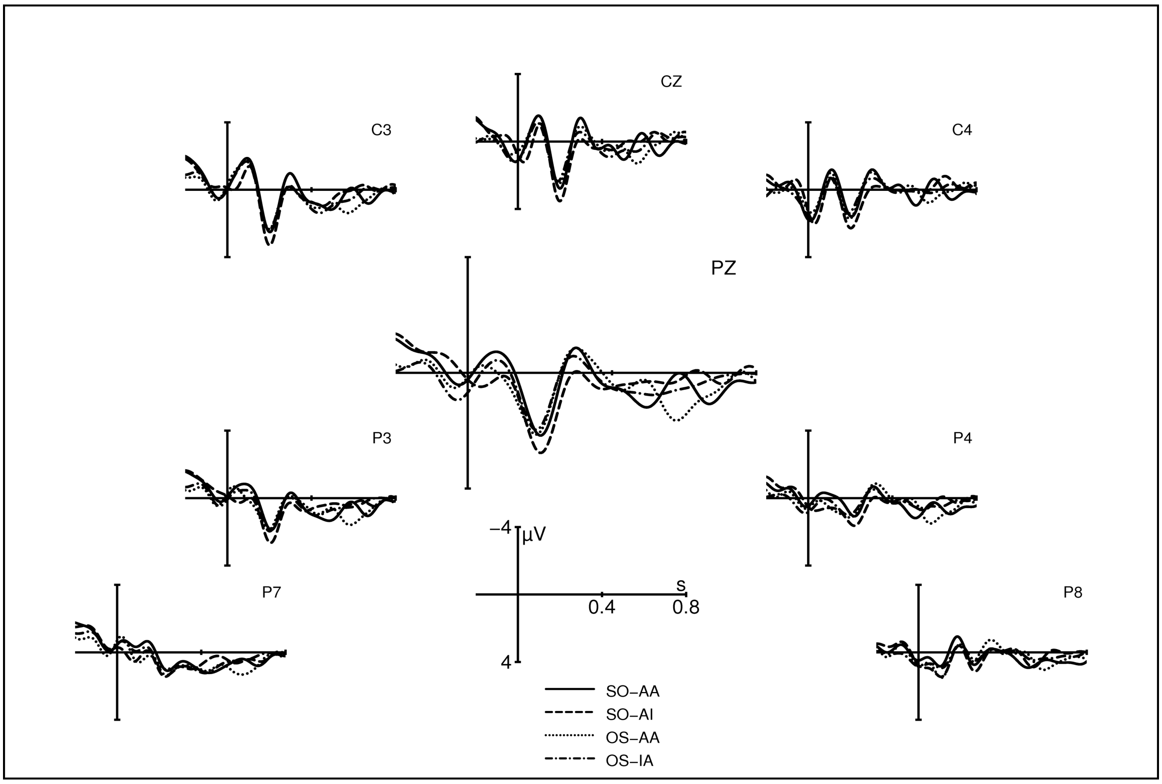


Figure 12. Grand-average ERPs (N = 48) time-locked to the onset of the first fixation on the accusative verb (onset at the vertical bar).

The ERPs elicited by the verb are given in Figures 12 and 15 for the first and last fixations, respectively. When the verb was first fixated, the object-initial conditions appeared to be more positive going than the subject-initial conditions between 500 and 700 ms post-fixation onset, and this effect was most visible across the two midline electrodes (see also Figure 13). The conditions with two animate NPs seemed to modestly enhance the negative component in the N400 time window, especially near midline electrodes (see also Figure 14). Between 500 and 700 ms post fixation-onset, the same conditions engendered a pronounced positivity compared to the two conditions containing NPs with an inanimate NP. This positivity was visible across all electrodes.

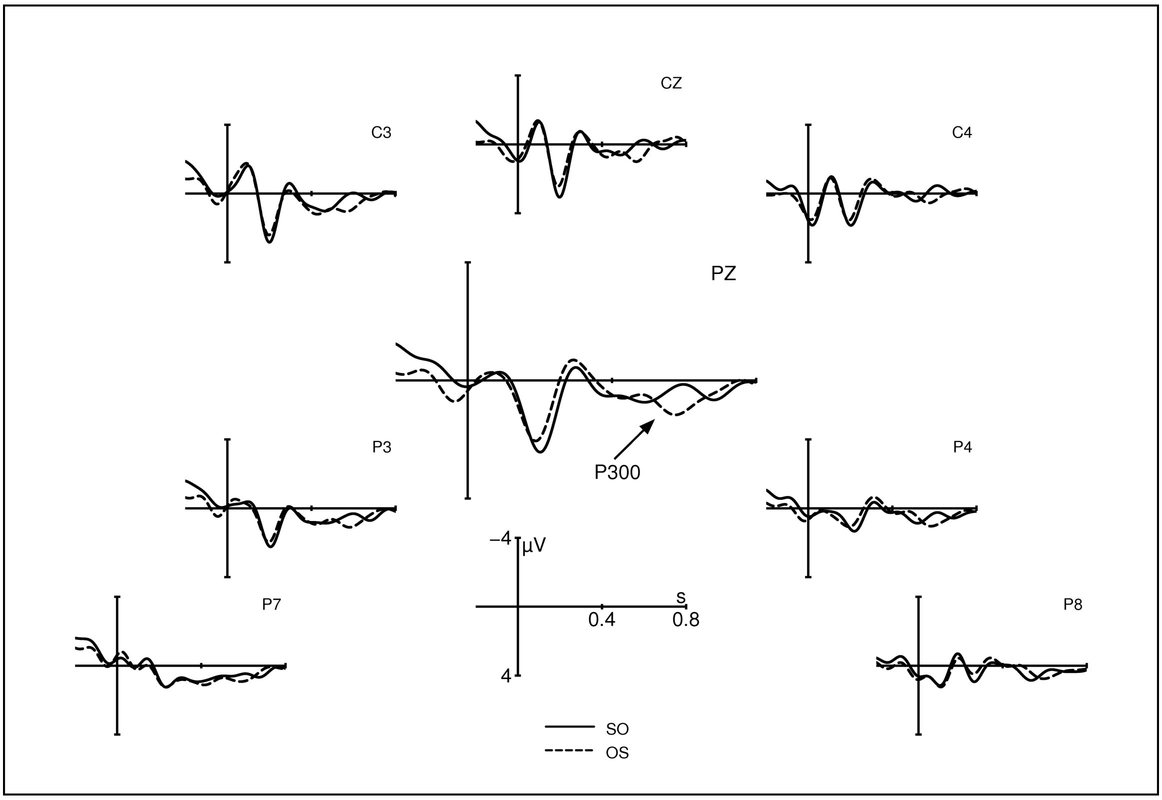


Figure 13. Grand-average ERPs (N = 48) time-locked to the onset of the first fixation on the accusative verb (onset at the vertical bar). Grand-average ERPs are collapsed across the animacy factor to show the influence of the word order manipulation.

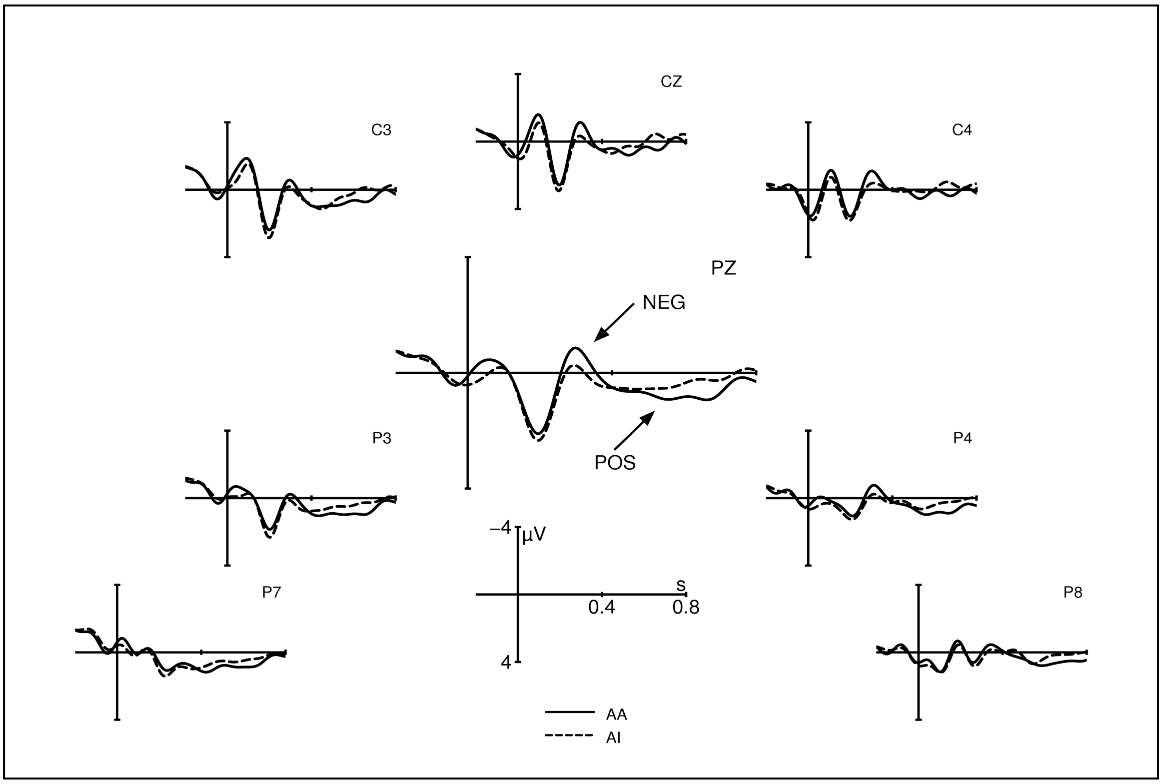


Figure 14. Grand-average ERPs (N = 48) time-locked to the onset of the first fixation on the accusative verb (onset at the vertical bar). Grand-average ERPs are collapsed across the word order factor to show the influence of the animacy manipulation.

The last fixation before an eye movement was made to the verb showed no differences between conditions in the N400 time window (Figure 15). There were later effects suggesting an increased negativity followed by a late positivity to the conditions with two animate NPs. However, given that the last fixation was on average about 250 ms in duration (see Table 8 below), these effects most likely stemmed from the first fixation reported above. Therefore, in order to avoid overestimating effects that are triggered by a later fixation position, the late effects found in the last fixation before the verb and the second NP were left out of the statistical analyses. The following time windows were chosen: 250 – 400 ms and 400 – 540 ms for NP1, 250 – 400/ 420 ms for NP2 and 250 – 400 ms and 500 – 700 ms for the verb.

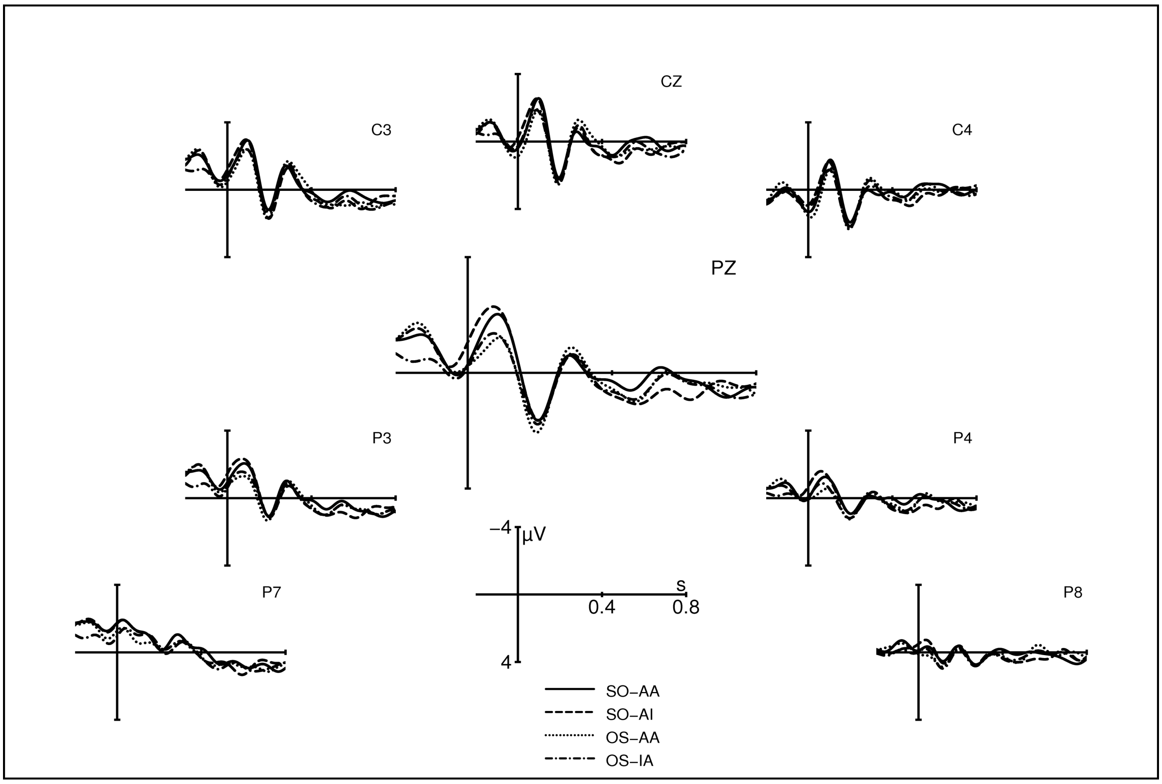


Figure 15. Grand-average ERPs (N = 48) time-locked to the onset of the last fixation before the accusative verb (onset at the vertical bar).

NP1 position: OS-AA vs. OS-IA

First fixation analyses

N400 time window: 250 ms – 400 ms

The analysis over midline electrodes revealed no significant effects involving the factor ANIM (ROI: $F(1,47) = 17.82$, $p < .0001$; ANIM: $F < 1$; ROI x ANIM: $F < 1$). Similarly, the lateral electrode sites did not reveal any significant effect for the experimental manipulation (ROI: $F(3,141) = 128.01$, $p < .001$; ANIM: $F < 1$; ROI x ANIM: $F < 1$).

Later time window: 400 ms – 540 ms

Visual inspection suggested a somewhat larger negativity for NP1 (especially at central electrode sites) when it was animate. Statistical tests, however, did not show any significant effect involving the factor ANIM for either midline sites (ROI: $F(1,47) = 32.60$, $p < .001$;

ANIM: $F < 1$; ROI x ANIM: $F < 1$) or lateral sites (ROI: $F(1,47) = 66.42$, $p < .001$; ANIM: $F < 1$; ROI x ANIM: $F < 1$).

Last fixation analyses[40]

N400 time window: 250 – 400 ms

The analysis for the midline ROIs revealed no reliable effects involving the factor ANIM (ROI: $F(1,28) = 11.36$, $p < .01$; ANIM: $F < 1$; ROI x ANIM: $F < 1$). The analysis involving lateral electrode sites confirmed these findings (ROI: $F(3,84) = 86.84$, $p < .001$; ANIM: $F < 1$; ROI x ANIM: $F(3,84) = 1.52$, $p < .3$). Visual inspection, however, suggested that between approximately 200 and 300 ms post fixation onset, inanimate bare plurals appeared to be more negative going than animate bare plurals at left-parietal electrodes. Statistical tests revealed that the animacy of the bare plural had no impact on midline electrode sites (ANIM: $F(1,29) = 2.30$, $p < .2$; all other Fs < 1.34, n.s.), whereas a reliable main effect across lateral ROIs was found (ANIM: $F(1,29) = 7.48$, $p < .02$; ROI: $F(3,87) = 94.74$, $p < .001$). The interaction with ROI was not significant ($F(3,87) = 1.76$, $p < .02$).

NP2 position

First fixation analyses

N400 time window: 250 ms – 420 ms[41]

The analysis for the midline ROIs revealed that all three main effects were significant (ROI: $F(1,47) = 10.19$, $p < .01$; ORDER: $F(1,47) = 8.38$, $p < .01$; ANIM: $F(1,47) = 6.49$, $p < .02$). The interaction between the factors ROI and ORDER fell short of significance ($F(1,47) =$

[40] Note that the complementizer *dass* 'that' preceding NP1 is a very short function word that is very likely to be skipped in reading. Therefore, only 30 out of 48 subjects contributed fixation data to the last fixation before NP1.

[41] Visual inspection of the ERPs at NP2 position suggests that ERP waves diverged slightly beyond 400 ms. The N400 time window was therefore extended to 420 ms post word-onset. The same pattern of results was obtained when the window was set to 400 ms post word-onset.

3.16, $p < .09$). All other interactions were clearly not significant (all Fs < 2.2, n.s.). The same pattern of results emerged when lateral ROIs entered the statistical analysis, except that the main effect of ORDER was only marginally significant (ROI: $F(3,141) = 10.04$, $p < .0001$; ORDER: $F(1,47) = 3.5$, $p < .06$; ANIM: $F(1,47) = 7.09$, $p < .02$). No interaction approached significance for the lateral electrode sites (all Fs < 2.64, n.s.). Apparently, the main effects of ORDER and ANIM were due to enhanced negativities for subject-initial clauses and for clauses containing two animate NPs, respectively.

Last fixation analyses

N400 time window: 250 ms – 400 ms

The analyses for this fixation position registered no reliable effects at midline electrodes (all Fs < 1.37, n.s.) and at lateral ROIs (all Fs < 2.55, n.s.).

Verb position

First fixation analyses

N400 time window: 250 ms – 400 ms

The analyses across midline electrode sites registered only a marginal main effect of animacy (ANIM: $F(1,47) = 3.99$, $p < .06$), while all other main effects and interactions were not reliable (all Fs < 2.08, n.s.). The animacy effect was due to an increase in the negativity elicited by clauses that contained two animate NPs. The analyses for the lateral electrode sites revealed a slightly different pattern. The main effect of animacy did not approach significance (ANIM: $F(1,47) = 2.86$, $p < .1$), but the interaction between ROI and ORDER was highly reliable ($F(3,141) = 6.95$, $p < .01$). However, when the interaction was resolved for ROI, there were no significant effects at either electrode site (left-central: $F(1,47) = 3.00$, $p < .09$; right-central: $F(1,47) = 1.31$, $p < .3$; left-parietal: $F < 1$; right-parietal: $F(1,47) = 2.94$, $p < .1$). No further effects approached significance here (all Fs < 1.54, n.s.).

Late positivity time window: 500 ms – 700 ms

Statistical tests revealed reliable main effects of ROI and ANIM and a marginal main effect of ORDER across midline electrodes (ROI: $F(1,47) = 16.11$, $p < .0001$; ANIM: $F(1,47) = 5.88$, $p < .02$; ORDER: $F(1,47) = 3.69$, $p < .07$). The interactions were not reliable (all Fs < 1). Statistical tests for the lateral ROIs revealed fully reliable main effects of ROI and ANIM (ROI: $F(3,141) = 23.67$, $p < .001$; ANIM: $F(1,47) = 6.05$, $p < .02$). The main effect of ORDER did not yield significance ($F(1,47) = 2.92$, $p < .1$), and no other main effect or interaction yielded significant results (all Fs < 1.83, n.s.). In sum, the statistical analyses confirm the visual impression that sentences containing two NPs with identical animacy status (i.e., two animate NPs) elicited a late positivity at the verb position, whereas sentences with NPs differing in animacy did not. There was also a locally confined positivity at midline electrodes for those clauses that began with a bare plural NP (i.e., object-first clauses).

Last fixation analyses

N400 time window: 250 ms – 400 ms

Statistical tests confirmed the visual impression that there was no activity specifically related to one of the experimental manipulations in this time window. At midline electrode sites, only the 3-way interaction between the factors ROI, ORDER and ANIM approached significance ($F(1,47) = 3.98$, $p < .06$; all other Fs < 1.22, n.s.). Resolving this interaction for ROI, however, yielded no significant results at either midline electrode (central: $F(1,47) = 2.04$, $p < .2$; parietal: $F < 1$). With the exception of the factor ROI, no effects approached significance at lateral electrode sites (ROI: $F(3,141) = 28.57$, $p < .001$; all other Fs < 1.93, n.s.).

Eye movement results

Table 8 gives the man fixation times and fixation positions across participants for each critical position. Statistical tests for the two fixation positions to which ERPs were collected are reported first, followed by supplementary analyses.

	NP1	NP2	Verb
First fixation			
SO-AA	211 (34)	255 (45)	270 (45)
SO-AI	211 (42)	249 (44)	271 (50)
OS-AA	210 (38)	254 (54)	290 (56)
OS-IA	212 (44)	240 (36)	281 (58)
Last fixation			
SO-AA	186 (68)	215 (36)	242 (42)
SO-AI	180 (51)	211 (43)	235 (42)
OS-AA	188 (60)	215 (32)	248 (48)
OS-IA	181 (52)	225 (37)	232 (37)
First pass time			
SO-AA	330 (99)	354 (100)	354 (102)
SO-AI	331 (84)	338 (86)	347 (105)
OS-AA	356 (110)	343 (102)	410 (139)
OS-IA	373 (123)	324 (80)	375 (95)
Total time			
SO-AA	446 (161)	446 (119)	433 (121)
SO-AI	438 (142)	430 (130)	421 (136)
OS-AA	622 (221)	538 (199)	577 (224)
OS-IA	598 (225)	495 (172)	488 (172)
Launch site			
SO-AA	5.1 (1)	5.6 (2)	5.4 (1)
SO-AI	5.2 (1)	5.4 (1)	5.5 (1)
OS-AA	5.1 (1)	5.2 (1)	5.6 (2)
OS-IA	5.1 (1)	5.1 (1)	5.7 (2)
Landing position			
SO-AA	2.9 (1)	3.4 (1)	3.4 (1)
SO-AI	3.0 (1)	3.3 (1)	3.4 (1)
OS-AA	2.9 (1)	3.5 (1)	3.4 (1)
OS-IA	2.8 (1)	3.4 (1)	3.3 (1)

Table 8. Participant mean fixation durations (in ms) and positions (in characters) by condition and region. Standard deviation is given in parentheses. The launch site measure gives the position (in letters relative to the end of the region) from which a saccade leaving the region was initiated.

NP1 position: OS-AA vs. OS-IA

The statistical tests for the first fixation on the first NP revealed no significant effect of animacy (all *F*s < 1). There was also no effect of animacy for the duration of the last fixation preceding the first NP (all Fs < 1). The 17-ms difference in first pass time also failed to reach significance across participants and items ($F_1(1,47) = 3.29$, $p < .08$; $F_2(1,39) = 1.79$, $p < .2$), as did the 24-ms difference in total time (F_1 and $F_2 < 1$). The two analyses for the fixation positions revealed no reliable effects (all Fs < 1).

NP2 position

The duration of the *first fixation* increased marginally if both NPs were animate (255 ms vs. 245 ms). This was reflected in a main effect of ANIM that was, however, only significant by participants ($F_1(1,47) = 5.37$, $p < .03$; $F_2(1,39) = 2.74$, $p < .2$). The effect of ORDER and the interaction between both factors did not yield significance (ORDER: Fs < 1.49, n.s.; ORDER x ANIM: Fs < 1.21, n.s.). The duration of the *last fixation* before readers entered this region revealed no reliable effects. The main effect of ORDER only approached significance by items ($F_1(1,47) = 1.87$, $p < .2$; $F_2(1,39) = 3.76$, $p < .06$) and the interaction between word order and animacy was significant by participants only ($F_1(1,47) = 4.45$, $p < .04$; $F_2(1,39) = 2.76$, $p < .2$). The main effect of ANIM was not reliable in either analysis (Fs < 1.12, n.s.).

In *first pass time*, there was a marginal main effect of ORDER ($F_1(1,47) = 3.90$, $p < .06$; $F_2(1,39) = 4.27$, $p < .05$), indicating longer reading time for subject-initial clauses (346 ms vs. 334 ms). The main effect of ANIM was only reliable in the by-participants analysis ($F_1(1,47) = 4.16$, $p < .05$; $F_2(1,39) = 2.13$, $p < .2$), suggesting a small disadvantage for the condition with two animate NPs (349 ms vs. 331 ms). The interaction was not significant (Fs < 1). A very similar pattern arose in *total time*, with a highly significant main effect of ORDER ($F_1(1,47) = 18.25$, $p < .001$; $F_2(1,39) = 22.98$, $p < .001$). This, however, was due to longer reading times for object-initial structures (517 ms vs. 438 ms). The marginal main effect of ANIM ($F_1(1,47) = 4.61$, $p < .04$; $F_2(1,39) = 2.59$, $p < .2$) resulted from increased reading times for clauses containing two animate NPS (492 ms vs. 463 ms). The interaction was again not reliable ($F_1(1,47) = 1.23$, $p < .3$; $F_2(1,39) = 1.65$, $p < .3$). There were no effects at all for the *landing position* on the second NP (ORDER: Fs < 1; ANIM: $F_1(1,47) = 2.37$, $p < .2$; $F_2(1,39) = 1.61$, $p < .3$; ORDER x ANIM: Fs < 1). The analysis for the *launch site* of the saccade leaving the word registered a reliable main effect of ORDER ($F_1(1,47) = 9.98$, $p < .01$; $F_2(1,39) = 6.56$, $p < .02$). This reflected the fact that the average launch site for object-initial clauses shifted to the end of the word (on average 5.2 letters before the beginning of the next region) as opposed to subject-initial clauses (5.5). All other effects were statistically not reliable (Fs < 1.64, n.s.).

Verb position

Only the word order manipulation had an impact on the duration of the *first fixation* on the verb, in that object-initial clauses showed increased fixation durations (286 ms vs. 271 ms). This was reflected in a fully significant main effect of ORDER ($F_1(1,47) = 6.87$, $p < .02$; $F_2(1,39) = 12.65$, $p < .001$). The main effect of ANIM (all Fs < 1) and the interaction between the two factors (all Fs < 1.54, ns.s) were not significant. The duration of the *last fixation* preceding the verb revealed a marginal main effect of ANIM ($F_1(1,47) = 6.37$, $p < .02$; $F_2(1,39) = 3.77$, $p < .06$), induced by longer fixation durations for clauses with two animate NPs (245 ms vs. 234 ms). The main effect of ORDER and the interaction were not statistically reliable (ORDER: all Fs < 1; ORDER x ANIM: $F_1(1,47) = 1.30$, $p < .3$; $F_2(1,39) = 1.83$, $p < .2$).

The analysis for *first pass time* revealed a fully significant main effect of ORDER ($F_1(1,47) = 20.25$, $p < .001$; $F_2(1,39) = 24.02$, $p < .001$) and a marginal main effect of ANIM ($F_1(1,47) = 6.11$, $p < .02$; $F_2(1,39) = 3.91$, $p < .06$). The interaction between the two factors was far from significant ($F_1(1,47) = 3.00$, $p = .09$; $F_2(1,39) = 2.27$, $p < .2$). The two main effects were due to longer reading times for object-first clauses (393 ms vs. 351 ms) and for those containing NPs with identical animacy status (382 ms vs. 361 ms), respectively. In *total time*, the main effects of ORDER ($F_1(1,47) = 41.96$, $p < .001$; $F_2(1,39) = 51.34$, $p < .001$) and

ANIM ($F_1(1,47) = 13.14$, $p = .001$; $F_2(1,39) = 8.93$, $p < .01$) were fully reliable, as was the interaction between the two ($F_1(1,47) = 13.29$, $p = .001$; $F_2(1,39) = 7.03$, $p < .02$). Resolving this interaction for ORDER revealed that the animacy manipulation had no visible influence on subject-initial clauses (all Fs < 1), while the effect of ANIM was reliable for object-initial sentences ($F_1(1,47) = 19.70$, $p < .001$; $F_2(1,39) = 13.94$, $p = .001$). This effect was mainly driven by a remarkable increase in reading times for object-initial clauses containing two animate NPs. There were no effects at all for either the *landing position* on the verb (all *F*s < 1) or the *launch site* of the saccade leaving it (all *F*s < 1.62, n.s.).

7.1.4 Discussion

The present experiment was conducted to examine the comprehension of object-initial and subject-initial subordinate clauses in German, in which both NPs were fully case ambiguous. Syntactic word order was always given by the order of the two NP types used, namely proper names and bare plurals. The NP in the clause-initial position unambiguously indicated the order of subject and object in the clause, thereby counteracting the ambiguous case marking on both NPs that allowed both subject and object readings. So, a clause-initial bare plural NP always represented the object-initial word order, whereas a proper name represented the subject-initial word order. The animacy status of the bare plural was also varied such that it was animate or inanimate, yielding interference effects when both the proper name and the bare plural in the clause were animate. It was hypothesized that the semantically induced cue about word order should inhibit any reanalysis costs at the verb. Furthermore, two arguments sharing the same animacy status were presumed to interfere with one another in the course of processing. This interference effect was predicted to occur at two positions, the second NP position where the prominence of both NPs is related to each other and the verb position where the two NPs are mapped onto the logical structure of the verb. The findings from this experiment are summarized according to the three critical sentence positions.

No effects related to any of the experimental factors were found at the first NP in both the ERP and eye movement data. Although only the animacy manipulation could be statistically evaluated, visual inspection of the ERP waves did not indicate any processing disruptions when the first NP was fixated for the first time. This suggests that the semantically induced cue about the implemented word order was effective, so that participants were aware that an initial bare plural NP indicated an object-initial word order. This suggests that the semantic cue eliminated ambiguity costs arising from the lack of informative case marking (cf. Frisch et al., 2002). The animacy status of the bare plural did not seem to play a role in processing. It is difficult to give a sound interpretation of the increased negativity to inanimate bare plurals (between 200 – 300 ms) that was found for the last fixation before the first NP. First, the status of this negativity is uncertain. N400 effects with such a short duration have not yet been reported and, importantly, none of the N400 effects reported in this thesis evidences similar N400 latencies. Second, significantly fewer participants contributed data to this effect (30 out of 48), which is problematic in light of the fact that the ERP data appear to be strongly affected by a low signal-to-noise ratio in the present experiment (see below for further details). Consequently, this negativity must be treated with caution and a well-grounded interpretation would require data replication in further experiments.

The effects at the second NP both confirm that the word order cue eliminated the ambiguity in syntactic function assignment and suggest that N400 effects do not necessarily result in increased fixation durations, but may trigger refixations instead. Consider the N400 elicited by subject-initial structures. This effect can be accounted for when the predictive capacity of the language comprehension system is taken into consideration. Previous studies have shown that an unambiguously case-marked nominative argument/ subject in sentence-initial position is preferably interpreted as the single argument of an intransitive event (cf.

Bornkessel et al., 2004). When, at a later point during sentence comprehension, an accusative argument/ object is encountered, processing costs go up because it has not been predicted. This processing effort is thus associated with the switch from an intransitive to a transitive structure and occurs for subject-before-object structures. A sentence-initial accusative argument, by contrast, leads straightforwardly to the prediction of an upcoming nominative argument, so there is no enhanced N400 for object-initial clauses at the second NP. The system is prepared to encounter a nominative argument during some stage of parsing and this prediction facilitates processing when the nominative argument is eventually perceived. This finding suggests that prediction in language comprehension is not confined to the lexical-semantic domain, but is also used at the syntax-semantics interface.

There are theoretical accounts arguing for an asymmetric relationship between undergoer and actor roles, such that the presence of an undergoer leads to the prediction of an actor, but not vice versa (Primus, 1999; see Bornkessel-Schlesewsky & Schlesewsky, 2009b, for further empirical investigation). As actor and undergoer are encoded with nominative and accusative case, respectively, the enhanced N400 for subject-initials in the present experiment is straightforward. Participants used the unequivocal association of NP type with syntactic function to anticipate the word order – as under unambiguous case-marking conditions. The interpretation as subject appeared to automatically activate an intransitive template, resulting in extra processing effort when a second argument was perceived. If this pattern holds in further co-registration experiments, it would mean that the intransitive reading is adopted even in the presence of a parafoveal preview indicating another upcoming NP that is a likely verb argument.[42] This could indicate that parafoveal information was not used effectively

[42] Capital letters are used in German orthography to indicate nominals. It is therefore likely that the initial capital letter of the second NP is perceived parafoveally. If this orthographic information is used to anticipate an upcoming noun phrase, it is conceivable that the noun phrase will be interpreted as an argument and not as an adjunct, since there is no preposition indicating a probable adjunct phrase. So, the orthographic parafoveal information should only activate the expectation for an argument to follow.

(because of enhanced foveal load, see below) or that, in general, parafoveal orthographic information is not used for initial decisions about the adopted structural template (see Clifton, Traxler, Mohamed, Williams, Morris, & Rayner, 2003, for evidence from eye movements suggesting that parafoveal preview information is not used for initial structural decisions). Alternatively, the distinctness principle – according to which arguments should be as distinct as possible (best achieved with an intransitive structure with only one argument; cf. Bornkessel-Schlesewsky & Schlesewsky, 2009b) – could be a general principle in processing that is not prone to previews of upcoming sentence material. The distinctness principle may thus be independent of input modality.

It is also notable that the extra processing cost described above does not translate into increased first-fixation duration. Instead it leads to refixations on the second noun phrase, as first pass time and total time showed a corresponding effect. Recall that increased first fixation durations accompanied the N400 effect elicited by a lexically unpredicted word (e.g., *black is the opposite of nice*; cf. Experiment 2). This supports the claim that the N400 on the critical word does not affect the duration of the fixation triggering it, but of the subsequent fixation (Dambacher & Kliegl, 2007). However, the comparison of the present findings with those from Experiment 2 demonstrates that (i) this correlation is not the default scenario, as the current fixation may be extended in accordance with an N400 effect, and (ii) an N400 effect may not simply increase the duration of the subsequent fixation, but actually trigger it. Further research is clearly needed to show whether the increase in current fixation duration and the probability of refixating are functionally separated by linguistic domain (e.g., lexical-semantic processing vs. syntactic processing). In any case, the current findings suggest that N400 effects resulting from predictive processing in the lexical-semantic domain and at the syntax-semantics interface have clear correspondents in readers' eye movements.

The N400 effect due to animacy-based interference further confirms the distinctness principle. According to this principle, processing costs occur whenever two arguments do not differ in all relevant prominence features. It was found that an animate NP2 elicits an N400 when it follows an animate NP1. This is expected on the grounds that prominence information is used to identify the actor argument. If more than one argument bears a prominence property typical of the actor role, competition between the two arguments arises. As predicted, this interference effect manifested itself in the present experiment with an enhanced N400 amplitude. The eye movement record revealed only inconsistent results. The duration of the first fixation on NP2 as well as cumulative measures of reading time (such as first pass time or total reading times) registered a modest increase in fixation time across participants. For example, two animate NPs increased first pass time by about 18 ms and total reading time by about 29 ms. The lack of significant results in the by-item analysis could reflect a lack in experimental power.

There was no N400 effect for object-initial structures at the verb, although an increase in the duration of the first fixation as well as in first-pass and total reading time was found. This supports Haupt et al.'s arguments against a plausibility-induced N400. If object-initial sentences were per se less plausible than subject-initial structures, especially when presented in isolation, then a pronounced plausibility N400 should have been found in the present experiment. As this was evidently not the case, this explanation seems to be ruled out. There was, however, a small N400 caused by animacy-induced interference, which had, however, no correlate in the eye movement pattern. Animacy also affected the late positivity that was enhanced for two animate NPs. These effects of animacy-induced interference can be accounted for straightforwardly. When the verb is first fixated, its logical structure is retrieved and the semantic representations of the two arguments are mapped onto the corresponding argument positions. Argument competition of two animate NPs causes disruption when the NPs are linked to the argument slots. When, moreover, information from all available domains is integrated, interference due to the similarity in animacy impedes the generalized mapping process. This applies to subject-initial and object-initial structures alike.

Interestingly, a small late positivity for bare plural-first clauses was found, and it was accompanied by robust effects in the eye movement record. It was argued that, in German, definite/ specific NPs preferably precede indefinite/ non-specific NPs (Lenerz, 1977). A recent eye-tracking study has supported this claim by demonstrating that clause-initial bare plurals impose extra processing costs on the language comprehension system (Kretzschmar et al., in press). Since object-initial clauses in the present experiment are inevitably associated with bare plural-first order, the late positivity and the fixation pattern may be a reflection of the dispreferred NP order.[43]

There were several divergences between the eye movement and ERP data. For example, animacy had an impact on the last fixation before the verb, but there is no electrophysiological activity associated with this fixation. Although the difference amounts to no more than 10 ms, the statistical results indicate that the effect is much more stable than the numerically larger effect in first pass time on the second NP. This pattern can be explained by assuming that the last fixation before the verb coincides with a refixation on NP2 (or NP1, if readers regressed to the beginning of the clause), but not (or only marginally) with the first fixation on that NP. If, as argued above, the N400 effect for the first fixation triggers refixations on the NP (i.e., leads to increased first pass time), the refixations may not be able to reveal anything with regard to neuronal processing. They simply follow the moment in processing (i.e., the first fixation on NP2) that initiated the sequence of refixations. If the parafoveal preview on the next word does not provide useful information to overcome the comprehension problem, the N400 effect triggered by the first fixation on a word may induce multiple fixations which are inadequate trigger positions for possible ERP components. So, one observes effects in reading time measures in the absence of clear electrophysiological

[43] In contrast to the Haupt et al. (2008) study, there was no visible P300 to the subject/ proper name-initial clauses in the present experiment, which should reflect the well-formedness of these structures. The reason might be that the small animacy-induced N400-like effect overlapped with the positivity, thereby masking it.

effects. Overall, one can infer from the current data that N400 interference effects may correlate with both increased first fixation durations and a higher probability of refixations on the critical word. However, this generalization needs support from further investigations of interference effects. Another eye movement effect that was not accompanied by a corresponding ERP effect was the significant interaction between word order and animacy in total reading times for the verb. This showed that animacy-based interference affected the comprehension of bare plural-first clauses, whereas there was no detectable influence on clauses beginning with a proper name. The main effect of animacy was not significant in first-pass or total reading time. Thus, eye movements appear to be sensitive to prominence information and to the interactive effects of different prominence hierarchies. The fact that the same interaction was not statistically reliable in the ERP data suggests that the low number of experimental items per condition used here may have a detrimental impact on ERP waves when fine-grained semantic information is processed. Alternatively, the interaction found in total reading times could reflect that animacy interference alone cannot significantly influence eye movements at this stage of processing. On the other hand, the combined effects of animacy interference and the dispreferred word order of a bare plural in clause-initial position (i.e., an object-first clause) could cause a disruption in processing that impedes the eye from progressing to new material.

Note that, in this respect, the late positivity for initial bare plurals was restricted to midline electrodes, whereas the late positivity induced by animacy interference was broadly distributed across midline and lateral electrode sites. This topographical difference may justify the functional distinction of late positivity ERP effects, as envisaged in the eADM model.[44] Specifically, the late positivity indexing animacy interference belongs to the processing stage of generalized mapping, while the late positivity for initial bare plurals is an

[44] It is also possible that the small late positivity to initial bare plurals is a consequence of component overlap or of statistically unreliable effects at lateral electrodes sites that yield significance only when they overlap additively at midline electrode sites.

indicator for reduced well-formedness. As argued in the introduction, the late positivity indexing the poor well-formedness of object-initial sentences is thought to belong to the P300 family. Consequently, the P300-like late positivity to bare plural-initial clauses should be more susceptible to environmental factors. If so, then the dissociation of orthographic and lexical processing of the predicted item is still detrimental to, but not strong enough to completely prohibit, the emergence of a P300-like component. Alternatively, the simultaneous processing of foveal and parafoveal words may cause parallel information processing that affects the P300 (cf. chapter 1.2.2). From this perspective, the interaction of word order and animacy-induced interference, as exhibited by the total reading time measure, should not be detectable in the ERP data due to the low signal-to-noise ratio in the present experiment.

Strikingly, none of the ERP effects described above are induced by parafoveal information. Animacy information and prominence computation for NP2 were associated with the first fixation on it, even though both NPs were adjacent to each other, allowing parafoveal proessing. However, it could be possible that orthographic information from NP2 evokes a parafoveal effect, signaling the need to switch from an intransitive to a transitive structure. As this was clearly not the case, it seems that not much was acquired from the parafovea. It is conceivable that, despite the semantically induced word order cue, case ambiguity still increased processing effort, so that the resulting foveal load for both NPs prohibited any parafoveal N400 effects, as less attentional resources were available to gain benefit from the parafoveal preview (cf. Henderson & Ferreira, 1993). Since the last fixation before the second NP (or the verb) was always located at another ambiguous region of text, ensuing processing costs could have blurred the critical parafoveal processing effects. To test this hypothesis, Experiment 5 examined dative-object scrambling with unambiguously case-marked arguments.

7.2 Experiment 5: Dative-object scrambling in unambiguously case-marked structures

7.2.1 Introduction

Experiment 4 further confirmed the claim that the comprehension of preverbal arguments proceeds in an incremental manner (cf. Bader & Lasser, 1994; Kamide & Mitchell, 1999). The language system processes all available information provided by the arguments immediately rather than to wait until the verb. However, none of the information seemed to be acquired from the parafovea and it was argued that case ambiguity might have reduced the parafoveal preview of upcoming words. To test this possibility, the processing of unambiguously case marked arguments was investigated when word order was either canonical or scrambled. There is compelling evidence that the comprehension of preverbal arguments entails different processing correlates. Whether these correlates arise depends on whether morphosyntactic information allows for the unequivocal assignment of actor or undergoer roles. It has been consistently shown that when case marking on the NP enables the immediate identification the semantic role, deviations from canonical word order elicit a specific ERP response, the scrambling negativity. This negative deflection has a fronto-central maximum and arises approximately 300-450 ms post-onset of the word indicating the word order deviation (Rösler, Pechmann, Streb, Röder, & Hennighausen, 1998; Schlesewsky, Bornkessel, & Frisch, 2003).

Interestingly, the scrambling negativity interacts with sentential context and verb semantics such that it emerges consistently in response to scrambled accusative objects, but inconsistently with scrambled dative objects. That is, a scrambled dative object elicits the scrambling negativity only in active sentence frames, in which they can never be interpreted as the single argument of an intransitive structure. This was the major finding of a study that examined to what extent sentential context licenses the interpretation of dative arguments

(Bornkessel, Schlesewsky, & Friederici, 2002). Bornkessel and colleagues set out to investigate two competing accounts that could explain the scrambling negativity. On one account, the subject preference would not only hold for ambiguous arguments (cf. the introduction to Experiment 4), but could in fact result from a general structural preference to interpret the first argument as the subject of the clause, which, in German, carries the higher-ranking generalized semantic role.[45] Whenever the first argument does not bear the highest semantic role, a non-canonical word order is assumed. The human parser then predicts that the second argument will bear the higher-ranking generalized semantic role, and processing costs arise. A frequency-based account, on the other hand, would assume that object-initial structures enhance processing costs because they are less frequent than subject-initial structures, as measured in language corpora. To distinguish between these two accounts, Bornkessel et al. (2002) used sentences structures for which each account made different predictions. Consider therefore the examples in (15). In (15a, b), the determiner following the finite verb elicits a scrambling negativity by signaling either dative (*dem*) or accusative case (*den*). The dative or accusative object following the finite auxiliary *hat* signals that a nominative argument will come up later in the sentence. Since the auxiliary *hat* indicates active voice for the constructions in (15a, b), the nominative argument is associated with the highest-ranking actor role. Both the frequency-based and the grammar-based accounts can explain the resulting scrambling negativity equally well. Either the occurrence of an initial non-nominative argument in these active sentences is structurally infrequent or it goes against the subject preference in German.

---

[45] This preference is assumed to apply to clause-medial positions only. The German *Vorfeld*, the structural position immediately preceding the finite verb in verb-second sentences (e.g., *Peter schlägt den Hund* ‘Peter beats the dog’, where *Peter* occupies the *Vorfeld* position), can host several syntactic constituents depending on contextual constraints.

(15) Word order scrambling in German (examples 15b-d from Bornkessel et al., 2002)

(a) Gestern hat dem Jäger der Gärtner geholfen.

yesterday has [the hunter]$_{DAT}$ [the gardener]$_{NOM}$ helped

‘Yesterday, the gardener helped the hunter.’

(b) Gestern hat den Jäger der Gärtner beruhigt.

Yesterday has [the hunter]$_{ACC}$ [the gardener]$_{NOM}$ calmed

‘Yesterday, the gardener calmed the hunter.’

(c) … dass dem Jäger der Gärtner half.

… that [the hunter]$_{DAT}$ [the gardener]$_{NOM}$ helped

‘… that the gardener helped the hunter.’

(d) … dass dem Jäger geholfen wurde.

… that [the hunter]$_{DAT}$ helped was

‘… that the hunter was helped.’

However, the two accounts predict different processing costs for dative objects following the complementizer *dass* ‘that’ in (15c). While the frequency-based account makes the same predictions as in the case of examples (15a, b), the grammar-based account predicts no principled processing costs for (15c). Specifically, Bornkessel and colleagues argued that this is due to the availability of passive constructions as in (15d) in which the initial dative argument appears in canonical order (see Bornkessel et al., 2002, for references from the theoretical literature). Remember from Experiment 4 that the parser attempts to adopt the simplest analysis during each step of processing (cf. the distinctness principle as described in chapter 7.1.1). This implies that for clauses beginning with the complementizer *dass* followed by a dative argument, the language comprehension system would first interpret the sentence fragment as a passive construction. The reasoning behind this involves what is called an intransitive structural template. In intransitive templates, only one generalized semantic role is assigned, the single argument role, which can be either actor or undergoer depending on verb information. This is essentially simpler than a transitive template, in which two roles must be assigned. Thus, it is likely that the dative argument in structures such as (15c, d) is interpreted by default as the single argument of an intransitive construction. This contrasts with accusative arguments, which cannot occur in passive voice, and thus always signal an upcoming actor argument for an active transitive clause regardless of the sentential context they are embedded in.

Bornkessel et al. compared the processing of structures such as (15c), in which the first NP was unambiguously case marked for dative case, to similar structures containing an initial accusative or nominative argument. They found that only accusative objects gave rise to a scrambling negativity relative to nominative arguments, whereas dative arguments clearly did not. This finding was in favor of the grammar-based explanation as sketched above, and incompatible with the frequency-based account. They found further support for the claim that dative arguments in clauses such as (15c) are initially interpreted as the single argument of an intransitive clause at the second NP. Specifically, when the second nominative NP is encountered, an early positivity with a parietal maximum emerges, which was taken to reflect the need to thematically reanalyze the semantic role of the clause-initial dative, since the nominative argument in active sentences receives the higher-ranking generalized semantic role, and the initial interpretation of the dative object thus turned out to be incorrect (for further evidence for thematic reanalysis effects, see Bornkessel, Schlesewsky, & Friederici, 2003).

In sum, when the linguistic context allows the parser to interpret the dative argument in the first clause-medial position as conveying the highest generalized semantic role (i.e., it is assumed to be the single argument of the clause), there is no scrambling negativity. Dative arguments engender a scrambling negativity only if their occurrence unequivocally signals a later occurring nominative argument associated with the higher semantic role. In this case, they behave like accusative objects, which are automatically associated with the lower ranking generalized semantic role (undergoer) and therefore always indicate that an actor argument (encoded with nominative case) will occur later in the sentence. This prediction concerning an upcoming actor argument is the strategy underlying the processing of structures like (15a). In accordance with this claim, studies using similar structures (i.e., main clauses in active voice that allow no intransitive reading) consistently reported a neuronal response to object scrambling independent of object case (Rösler et al., 1998; Schlesewsky et al., 2003}.

The present experiment used sentences that were similar to the stimuli in the Bornkessel et al. (2002) study. The dative object in the first clause-medial position can be assigned the role of the single argument (see example 16). Here, the auxiliary *wurde* signals a passive structure and the sentence fragment could be completed without a nominative argument (e.g. *Gestern wurde dem Reiter geholfen* ‘Yesterday, the equestrian was helped.’). Similar to the Bornkessel et al. study, these initial datives (16c, d) should thus not give rise to a scrambling negativity relative to nominative arguments in the same position (16a, b). Since the intransitive passive construction can be securely inferred from the auxiliary (i.e., in contrast to the complementizer *dass*, it excludes an active-voice reading), the first NP is most likely always interpreted as the single argument of the construction and assigned the generalized semantic role of undergoer. From this perspective, costs for template switching are unlikely when the second NP is encountered, simply because the parser makes the same predictions for both clause-initial nominative and dative arguments (i.e., an initial dative object does not lead to predictions regarding a possibly upcoming subject marked with nominative case). So, in contrast to what was found in Experiment 4, template switching should produce no reliable results at the second NP when it is the object of the clause, as the switch is necessary in all conditions. However, the switch from an intransitive to a transitive template requires the re-assignment of the semantic role from the initial dative argument to the following nominative NP and this thematic reanalysis should give rise to an early positivity to the second nominative argument for dative-initial clauses only.

(16) Example stimuli for Experiment 5[46]

(a) Gestern wurde das Pferd dem Reiter verkauft, erklärte Anna. (NOM-DAT – AA)

yesterday was [the horse]$_{\text{NOM.ANIM}}$ [the equestrian]$_{\text{DAT}}$ sold, explained Anna

'Anna explained that the horse was sold to the equestrian yesterday.'

(b) Gestern wurde der Sattel dem Reiter verkauft, erklärte Anna. (NOM-DAT – IA)

yesterday was [the saddle]$_{\text{NOM.INANIM}}$ [the equestrian]$_{\text{DAT}}$ sold, explained Anna

'Anna explained that the saddle was sold to the equestrian yesterday.'

(c) Gestern wurde dem Reiter das Pferd verkauft, … (DAT-NOM – AA)

(d) Gestern wurde dem Reiter der Sattel verkauft, … (DAT-NOM – AI)

There is one important limitation to the prediction about the early positivity indexing thematic reanalysis. Recently, Bornkessel and Schlesewsky (2006) have argued on the grounds of extensive theoretical literature on dative case that dative arguments differ from accusative arguments in their processing because only the latter are able to receive the generalized semantic undergoer role (Bornkessel & Schlesewsky, 2006, p. 792). More specifically, dative case is widely assumed to deviate from either the ideal actor or the ideal undergoer of an event and therefore doesn't qualify for a generalized semantic role. Transitive structures with a dative object are probably treated like intransitive structures on the semantic level, in which only the nominative argument possesses a generalized semantic role. If this conclusion generally holds, it will have important implications for thematic reanalysis. The initial dative in the present stimuli should still be treated as the single argument of the passive

[46] Abbreviations: NOM – nominative case; DAT – dative case; ANIM – animate noun phrase; INANIM – inanimate noun phrase; Condition labels: NOM-DAT – nominative-before-dative clause; DAT-NOM – dative-before-nominative clause; AA – animate dative object and animate subject; AI/ IA – animate dative object and inanimate subject

clause, but without the semantic undergoer role. Under these circumstances there should be no early parietal positivity, as there is no generalized semantic role that needs to be reanalyzed. However, one can expect to find a similar effect for the second nominative argument by assuming that, on a semantic level, the structure must be revised to contain one generalized semantic role (as the initial dative was associated with a zero-role structure).

In addition to the word order manipulation detailed above, the animacy of the nominative argument was varied. Since dative arguments in the two-argument structures employed in the present experiment are necessarily animate, only animacy of the nominative could be altered. The results could thus show whether animacy-induced interference effects can be found in these constructions as well. If there are such effects, they should emerge for dative-nominative structures, with an animate nominative argument following a dative argument. An alternative hypothesis can be derived from the fact that, in the passive clauses used here, the different types of prominence information could contradict one another. Both nominative case marking and animacy (as opposed to inanimacy) are preferably linked to the actor role rather than the undergoer role in German (cf. Bornkessel-Schlesewsky & Schlesewsky, 2009a). So, an animate nominative argument should give rise to a mismatch effect in any position (probably reflected in an increased N400 amplitude), because an animate undergoer is highly dispreferred.

The present experiment is also very interesting with respect to the relationship between ERP and eye movement correlates of linguistic processing. Although initial dative arguments in structures that allow a passive reading do not give rise to a scrambling negativity, they do lead to increased reading times, as measured in eye movement and self-paced reading experiments (Scheepers, 2005). This discrepancy has been used to support the claim that the RSVP presentation induces ERP waves with latencies that do not reflect the actual time course of linguistic processes in natural reading (Scheepers, 2005). If this is correct, one should either find an alignment between ERP and eye movement measures or significantly altered ERP effects in the present co-registration experiment. If, however, the RSVP measure does not significantly change the nature of ERP waves, the present data should replicate the divergence between ERPs and eye movements in natural reading. Finally, both experimental manipulations can help to investigate the extent to which parafoveal information is used in sentences that are fully unambiguous with respect to case marking and syntactic function. The findings from Experiment 4 suggested that there is no parafoveal semantic processing. However, the case ambiguity of the critical parts of the sentence may have blurred parafoveal effects. It has been argued that morpho-syntactic ambiguity induces neuronal processing costs (Frisch et al., 2002), which could result in an enhanced foveal processing load. When the currently fixated word is difficult to comprehend, little or no information is acquired from the parafovea (cf. Henderson & Ferreira, 1993). The unambiguous structures in the present experiment should therefore exert less foveal processing load on the comprehension system, facilitating parafoveal processing to a larger degree. It is thus conceivable that information about the animacy of the nominative argument is acquired in the parafovea, i.e. before the noun is first fixated.

7.2.2 Method

Participants

48 students (24 female; mean age: 23.7 years, range: 19-31) from the University of Mainz were paid to participate in the experiment after giving informed consent. Two further participants had to be excluded due to low calibration accuracy or recording problems during the experiment. All participants were right-handed native speakers of German, as assessed by the German version of the Edinburgh handedness test (Oldfield, 1971). All participants reported normal or corrected-to-normal vision and no history of any neurological or

psychiatric disorders. They were naïve to the purpose of the experiment. None of them participated in either Experiment 1 or 3.

Materials

40 sets of sentences were constructed by combining one set of 20 different verbs with two sets of noun pairs each. The NPs were matched for frequency and length. Animate nominative NPs were 6.8 letters long and inanimate nominative NPs were 6.9 letters long on average. Dative NPs were on average 7.3 letters long. Nouns were matched for frequency so that the three groups of NPs exhibited the same average log frequency class, as assessed by the Leipzig Wortschatz corpus (frequency class 11; www.wortschatz.uni-leipzig.de). All nouns were combined with the definite article, which is always three letters long. Verbs were chosen as to complement the two NPs in the most plausible way. They varied between 7 and 10 letters in length and were as frequent as the NPs (average frequency class 12). The items were distributed to 4 lists using a Latin Square design so that each participant saw 10 sentences per condition, but no item twice within a list. The critical sentences were interspersed with 140 filler sentences of different structures. The stimuli were presented in a pseudo-randomized order. A comprehension question (requiring a yes or no answer) followed each trial, probing that trial's lexical contents. Participants had an average accuracy rate of 92%.

Apparatus and Procedure

Apparatus and experimental procedure were identical to Experiment 2 reported above.

Data preparation and analyses

The ERP and eye movement data were prepared for analysis as described for Experiment 2, with the exception that the ERP data were filtered for blinks and amplifier saturation artifacts (rejection criterion was 40 μV). The overall pattern of effects did not change, but the statistical results were more stable after using the artifact correction.[47] The eye movement record of each participant was screened for erroneous trials. Less than 1% of all trials was excluded due to track losses near/at the critical region or due to technical errors. Blinks and fixations below 80 ms or above 800 ms were deleted when they were located in the critical regions; short fixations do not reflect cognitive processing and unusually long fixations are likely to reflect track losses (cf. Rayner & Pollatsek, 1989). If this resulted in a no-fixation event for one of the critical regions, these trials did not enter the computation of fixation-related potentials. No trials were excluded due to this criterion. All fixation-related potentials were triggered relative to fixation onset. For the eye-movement analysis short fixations with less than 80 ms in duration were incorporated into an adjacent fixation if the distance between these two did not exceed one character, and fixations less than 40 ms in duration were treated similarly if they were within one character of an adjacent fixation. Short fixations that did not meet these criteria were excluded from further analysis. Fixations longer than 800 ms were also excluded. In total, less than 1% of fixations were eliminated based on these criteria.

The regions of interests (ROI) for the ERP analysis were defined as in Experiments 2 and 4. ERP waves were averaged and analyzed separately for each of the two critical NP positions (see Appendix E for the verb). The ERPs were averaged per participant and per condition from 200 ms before fixation onset through 800 ms post fixation onset for the same reasons as in Experiment 4. Grand averages were then computed over all participants. As in Experiments 2 and 3, ERP waves were averaged to the first fixation on a word and the last

[47] This was not the case for both Experiment 2 and 4, where unfiltered and filtered data yielded results that were very similar with respect to statistical effect size.

fixation before that word. The sentences were divided into regions of text as given in example (17). Regions of analysis were identical for the ERP and eye movement analyses. Note that in natural reading, articles are usually skipped during first-pass reading because of their short length. The present ERP analyses therefore deviate from some previous reports (Rösler et al., 1998; Schlesewsky et al., 2003), in that ERPs were not collected from the article and the ensuing noun separately. Instead, the NP as a whole was used to identify the ERP-triggering fixations.

(17) Gestern wurde/ dem Reiter/ der Sattel/ verkauft, …

As inanimate dative arguments do not occur in transitive constructions, the design was not fully balanced with respect to the animacy manipulation. Therefore, pair-wise comparisons were carried out at each NP position separately, involving either the factor ORDER (dative-nominative [DAT-NOM] vs. nominative-dative [NOM-DAT]) or the factor ANIM (animate [AA] vs. inanimate nominative argument [AI]). Alpha was adjusted for these comparisons with the modified Bonferroni procedure (Keppel, 1991) to avoid overestimating effects obtained with pair-wise comparisons. There were two pair-wise comparisons at the first NP position: word order variation was examined by contrasting subject-initial and object-initial clauses that contained two animate NPs (conditions NOM-DAT – AA vs. DAT-NOM – AA). The impact of animacy on the processing of the subject phrase (i.e., the nominative argument) was investigated by comparing the two subject-initial conditions (NOM-DAT – AA vs. NOM-DAT – IA). Alpha was set to $p < .025$ in these cases. As for the second NP, four different pair-wise comparisons were possible. The animacy manipulation was analyzed within each word order (DAT-NOM – AA vs. DAT-NOM – AI and NOM-DAT – AA vs. NOM-DAT – IA). Possible interference effects were also examined by collapsing averages across word order (AA vs. AI). Finally, word order variation was investigated by contrasting conditions with two animate NPs (NOM-DAT – AA vs. DAT-NOM – AA). Alpha was set to

$p < .012$ for all of these comparisons. To avoid type I errors in the ERP results resulting from violations of sphericity, alpha was corrected as proposed by (Huynh and Feldt, 1970) when effects with more than one degree of freedom in the numerator were evaluated. In these cases, the original degrees of freedom are reported with the corrected alpha. The eye movement results were statistically analyzed by using the same factors and pair-wise comparisons as described above. Participants and items were treated as random factors F_1 and F_2, respectively.

7.2.3 Results

ERP results

Figure 16 gives the grand-average ERPs triggered by the first fixation on the first NP. Between approximately 250 ms and 400 ms nominative-initial clauses appear to engender a small negative deflection at central-parietal sites, which was somewhat more pronounced across the left hemisphere. As this effect also appeared to be more pronounced for the parietal electrode sites, it is most likely not related to the scrambling negativity that has its maximum at frontal-central electrode sites. Between 400 and 720 ms post fixation onset, the nominative-dative condition was also more positive going than the dative-initial condition at parietal electrodes. With respect to the two nominative-dative conditions, there is little difference in the two ERP waves until about 400 ms post fixation onset. However, between about 400 ms until 720 ms, the ERP wave is more positive going when both the nominative and the dative argument are animate.

Figure 17 shows the grand-average ERPs triggered by the last fixation before readers fixated the first NP. There is no difference in ERP waves in early time windows with respect to the word order manipulation, whereas a late effect for nominative-dative clauses with two animate NPs is visible. This effect is remarkably similar to the earlier effect found for the first fixation on NP1, which suggests that it is induced by the first fixation rather than the current last fixation.

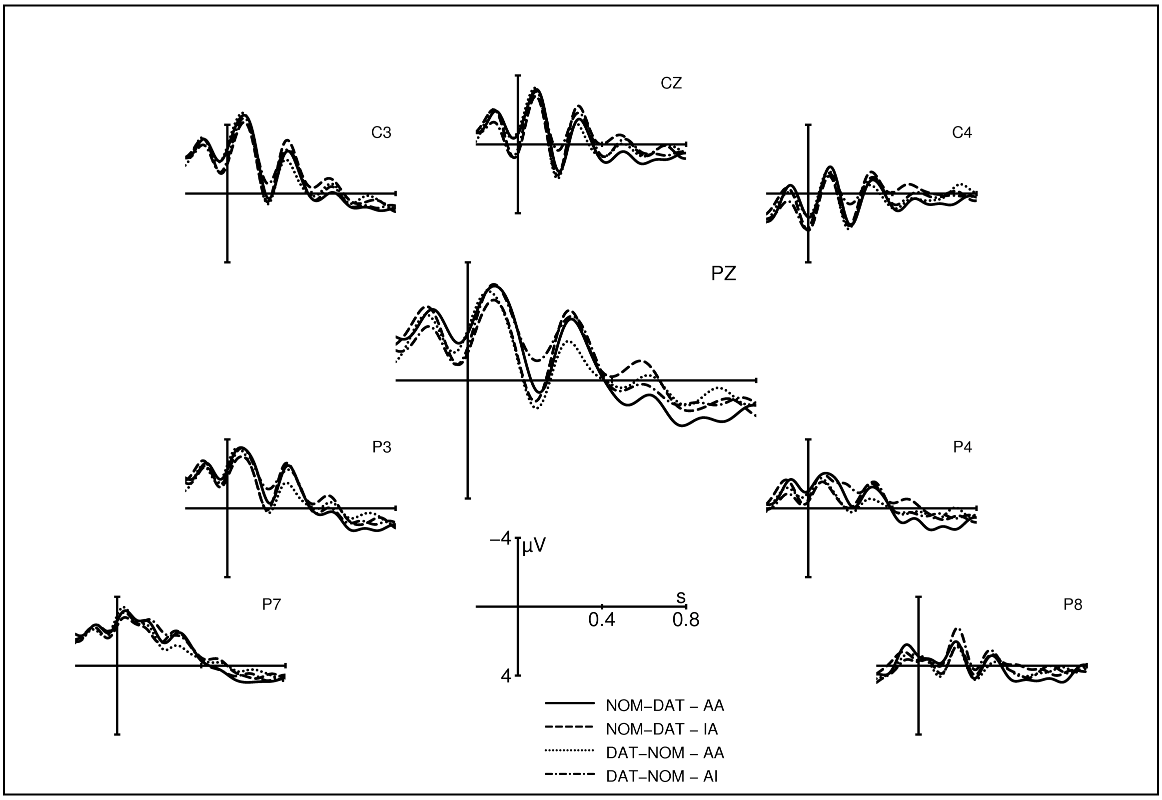


Figure 16. Grand-average ERPs (N = 48) time-locked to the onset of the first fixation on NP1 (onset at the vertical bar).

Furthermore, visual inspection suggests that the conditions with two animate NPs engendered a small negativity in the traditional N400 time window, which is most visible at parietal electrodes. Again, there is a very late effect (between 600 and 800 ms post fixation onset) in which the two animate NPs engendered a positivity attributable to the first fixation on NP1.

Figure 18 presents the grand average ERPs collected from the first fixation on the second NP. For nominative-dative orders, there appeared to be a small increased negativity for inanimate subjects at left-parietal electrodes between 250 and 400 ms post fixation onset. Animacy appeared to induce a negative deflection for dative-initial clauses with two animate NPs in the same time range, relative to the dative-initial clauses with an inanimate nominative argument. The late negative deflection for the latter is most likely related to activity at the following verb position.

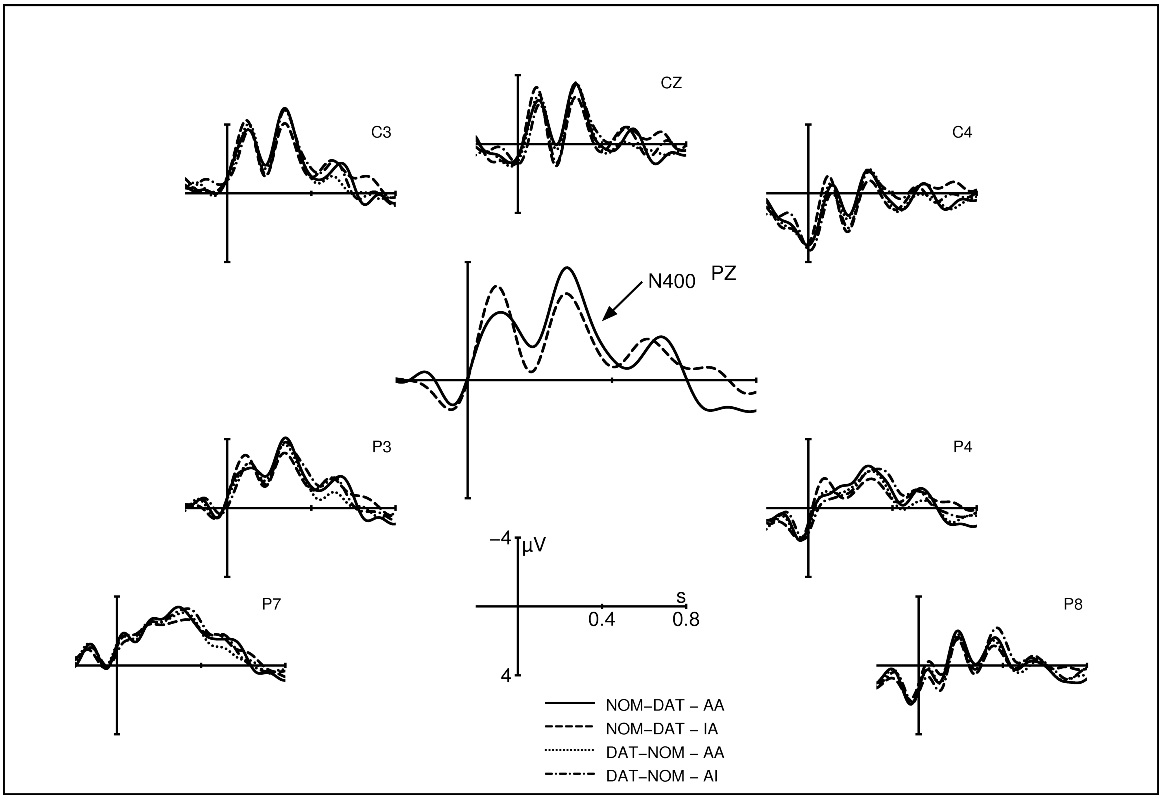


Figure 17. Grand-average ERPs (N = 48) time-locked to the onset of the last fixation before NP1 (onset at the vertical bar).

Figure 19 shows the ERPs averaged for the last fixation before NP2 was first fixated. This is virtually no difference in the ERP waves of the two nominative-initial conditions. As for the dative-initial structures, there is again a negative deflection between 400 and 600 ms post fixation onset for constructions with two animate NPs. This, however, is most likely related to the N400-like deflection triggered by the first fixation on NP2.

The N400 time window chosen for Experiments 2 and 4, based on visual inspection, is applied for all fixation positions here. For NP1, the subject-initial condition with two animate NPs was more positive going than its object-initial counterpart or the subject-first clauses with an inanimate subject. In both comparisons, the crucial difference seemed to fall within 400 ms and 720 ms post fixation onset, which was thus taken to be the late positivity window. The onset of the late effects for NP2 appears to be related to the processing of the verb, and they were therefore not statistically analyzed for this word position.

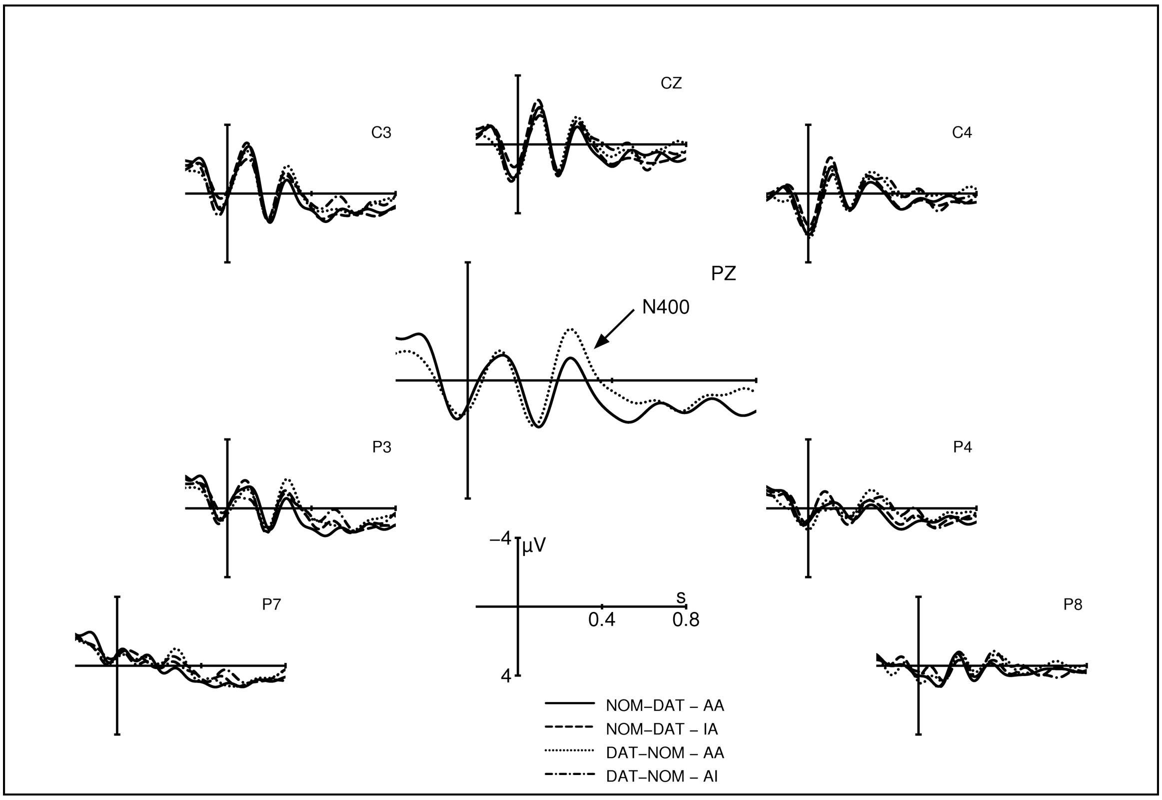


Figure 18. Grand-average ERPs (N = 48) time-locked to the onset of the first fixation on NP2 (onset at the vertical bar).

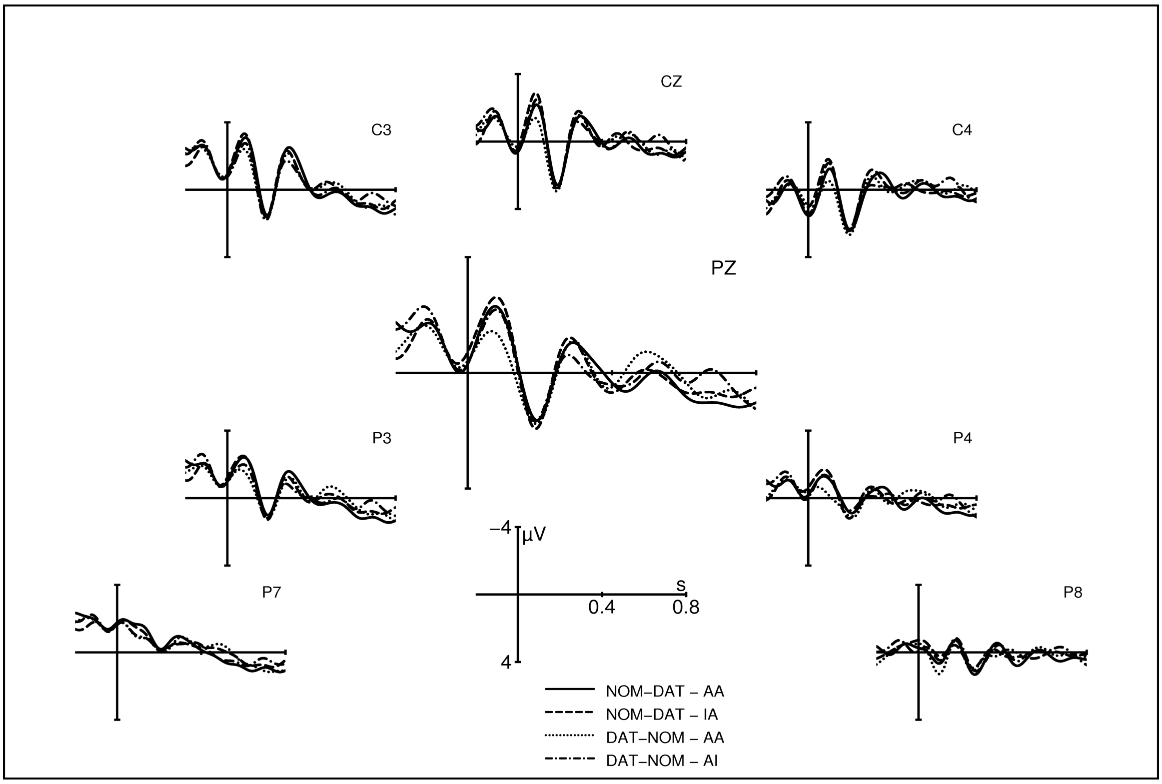


Figure 19. Grand-average ERPs (N = 48) time-locked to the onset of the last fixation before NP2 (onset at the vertical bar).

NP1

First fixation analyses

N400 time window: 250 ms – 400 ms

NOM-DAT – AA vs. NOM-DAT – IA

The analysis across midline ROIs registered a main effect of ROI and a reliable interaction between ROI and the factor ANIM, whereas the main effect of ANIM was not significant (ROI: $F(1,47) = 13.40$, $p < .0001$; ANIM: $F(1,47) = 1.46$, $p < .3$; ROI x ANIM: $F(1,47) = 10.37$, $p < .01$). Resolving the interaction for ROI yielded no significant results (central: $F(1,47) = 3.33$, $p < .08$; parietal: $F < 1$).[48] Across lateral ROIs only the main effect of ROI was significant ($F(3,141) = 41.55$, $p < .001$; all other Fs < 1).

NOM-DAT – AA vs. DAT-NOM – AA

Only the main effect of ROI was significant for both midline ROIs and lateral ROIs (midline: ROI: $F(1,47) = 18.17$, $p < .0001$; all other Fs < 2.05, n.s.; lateral: ROI: $F(3,141) = 34.93$, $p < .001$; all other Fs < 2.46, n.s.). Thus, there was no difference between nominative-initial and dative-initial clauses.

Late time window: 400 ms – 720 ms

NOM-DAT – AA vs. NOM-DAT – IA

The main effect of ANIM was fully significant across midline electrodes ($F(1,47) = 10.40$, $p < .01$), whereas the main effect of ROI only approached significance ($F(1,47) = 3.57$, $p < .07$) and the interaction was not reliable ($F < 1$). Both main effects were reliable across lateral

[48] The data that were not corrected for blinks and amplifier saturation effects revealed a reliable effect. The interaction of ROI and ANIM was highly reliable ($F(1,47) = 12.90$, $p < .0001$) and resolving it for ROI revealed a significant effect of ANIM for electrode Cz (central: $F(1,47) = 7.41$, $p < .01$; parietal: $F(1,47) = 2.02$, $p < .2$). The N400 amplitude to the inanimate nominative argument was increased at Cz.

electrode sites (ROI: $F(3,141) = 4.53$, $p < .02$; ANIM: $F(1,47) = 15.78$, $p < .0001$, but the interaction was not ($F(3,141) = 2.72$, $p < .08$). This effect confirms the visual inspection that the nominative-initial condition with two animate NPs was reliably more positive going than its counterpart with an inanimate subject phrase.

NOM-DAT – AA vs. DAT-NOM – AA

For midline electrodes, only the main effect of ORDER reached full significance ($F(1,47) = 9.45$, $p < .01$; all other *F*s < 2.51, n.s.). The tests for lateral ROIs revealed two reliable main effects (ROI: $F(3,141) = 4.33$, $p < .021$; ORDER: $F(1,47) = 10.26$, $p < .01$), whereas the interaction was not significant ($F < 1.79$, n.s.). This effect confirms the visual inspection that the subject-initial condition was reliably more positive-going than the object-initial condition.

Last fixation analyses

N400 time window: 250 ms – 400 ms

NOM-DAT – AA vs. NOM-DAT – IA

Only the main effect of ROI was reliable for the analysis across midline electrode sites (ROI: $F(1,47) = 20.10$, $p < .001$; ANIM: $F(1,47) = 4.38$, $p < .05$; ROI x ANIM: $F < 1$). The analysis for lateral ROIs revealed two significant main effects of ROI and ANIM (ROI: $F(1,47) =$ 3,141) = 117.30, $p < .001$; ANIM: $F(1,47) = 6.02$, $p < .02$), but no interaction ($F < 1$). The main effect of ANIM confirmed the visual impression of the enhanced negativity in response to the condition with an animate subject.

NOM-DAT – AA vs. DAT-NOM – AA

Statistical tests involving midline electrode sites registered a main effect of ROI ($F(1,47) = 11.69$, $p < .01$) and a marginal interaction between ROI and ORDER ($F(1,47) = 5.02$, $p =$

.029). The main effect of ORDER was not significant ($F < 1$). Resolving the interaction by ROI revealed no significant effects (all Fs < 1). Across lateral ROIs, only the main effect of ROI was significant (ROI: $F(3,141) = 113.65$, $p < .001$; all other Fs < 1.4, n.s.).

NP2

First fixation analyses

N400 time window: 250 ms – 400 ms

NOM-DAT – AA vs. NOM-DAT – IA

No effects reached significance across midline electrodes (all Fs < 2.83, n.s.). At lateral electrode sites, only the main effect of ROI was reliable (ROI: $F(3,141) = 4.86$, $p < .02$; ANIM: $F(1,47) = 3.20$, $p < .08$; ROI x ANIM: $F(3,141) = 1.66$, $p < .2$).

DAT-NOM – AA vs. DAT-NOM – AI

There were no effects for midline electrode sites (all Fs < 1.40, n.s.). At lateral ROIs, the main effect of ROI was the only reliable finding (ROI: $F(3,141) = 7.17$, $p < .01$; all other Fs < 1.24, n.s.).

AA vs. AI

No effects were reliable for midline ROIs (all Fs < 2.29, n.s.). Only the main effect of ROI was significant for the lateral ROIs (ROI: $F(3,141) = 7.19$, $p <.01$; all other Fs < 1).

In sum, there were no animacy-induced interference effects at the position of the second NP.

NOM-DAT – AA vs. *DAT-NOM – AA*

At midline electrode sites, no effects yielded significant results. The main effect of ORDER was significant only without the Bonferroni correction (ORDER: $F(1,47) = 4.39$, $p < .05$; all other *F*s < 1.96, n.s.). The main effect of ORDER was fully reliable across parietal ROIs, as was the main effect of ROI. The interaction was not significant (ROI: $F(3,141) = 6.32$, $p < .01$; ORDER: $F(1,47) = 9.59$, $p < .01$; ROI x ORDER: $F(3,141) = 1.03$, $p < .4$). This effect confirmed the visual impression that dative-initial clauses were associated with an enlarged amplitude for the N400 component, as opposed to the nominative-initial clauses.

Last fixation analyses

N400 time window: 250 ms – 400 ms

NOM-DAT – AA vs. NOM-DAT – IA

The analysis for midline electrode sites registered only a main effect of ROI ($F(1,47) = 11.03$, $p < .01$; all other *F*s < 1). A similar result was obtained in the analysis for lateral ROIs (ROI: $F(3,141) = 29.38$, $p < .001$; all other *F*s < 2.36, n.s.).

DAT-NOM – AA vs. DAT-NOM – AI

At midline electrode sites, the main effect of ROI was reliable ($F(1,47) = 14.39$, $p < .0001$), whereas all other effects were clearly not significant (*F*s < 1). The same held for lateral ROIs (ROI: $F(3,141) = 30.55$, $p < .001$; all other *F*s < 1).

AA vs. AI

The main effect of ROI was reliable in the analyses for midline and lateral ROIs, but animacy had no reliable influence (midline: ROI: $F(1,47) = 23.32$, $p < .001$; all other *F*s < 1; lateral: *R*OI: $F(3,141) = 36.38$, $p < .001$; all other *F*s < 1.33, n.s.).

In sum, animacy did not exert a noticeable influence on the processing of the second NP on either the first fixation on the NP or the last fixation preceding it.

NOM-DAT – AA vs. DAT-NOM – AA

Only the main effect of ROI registered significant results for the midline electrodes (ROI: $F(1,47) = 10.21$, $p < .01$; all other Fs < 1). The same pattern was obtained for the lateral electrode sites (ROI: $F(3,141) = 33.47$, $p < .001$; all other Fs < 1.08, n.s.).

Eye movement results

Table 9 gives the mean fixation durations and positions per region. The statistical tests below are reported separately for each region of text. The fixation positions for which ERPs were analyzed are reported first.

NP1 position

NOM-DAT – AA vs. NOM-DAT – IA

The effect of animacy did not influence the duration of the first fixation (all Fs < 1) or the last fixation preceding it ($F_1(1,47) = 1.75$, $p < .2$; $F_2(1,39) = 1.57$, $p < .3$). First pass time was not affected by the animacy of the subject (Fs < 1). In the total time measure, the effect of ANIM tended to approach significance by participants, but not by items ($F_1(1,47) = 3.19$, $p = .08$; $F_2(1,39) = 1.84$, $p < .2$). The landing position of the first fixation and the launch site of the last fixation leaving the region were not influenced by manipulating animacy (landing position: all Fs < 1; launch site: $F_1(1,47) = 2.64$, $p < .2$; $F_2(1,39) = 1.71$, $p < .2$).

	NP1	NP2
First fixation		
NOM-DAT – AA	209 (41)	229 (42)
NOM-DAT – IA	205 (37)	232 (40)
DAT-NOM – AA	215 (42)	227 (42)
DAT-NOM – AI	213 (41)	226 (42)
AA	NA	228 (40)
AI	NA	229 (38)
Last fixation		
NOM-DAT – AA	195 (27)	203 (27)
NOM-DAT – IA	190 (26)	197 (25)
DAT-NOM – AA	194 (31)	204 (31)
DAT-NOM – AI	194 (29)	200 (28)
AA	NA	203 (25)
AI	NA	198 (23)
First pass time		
NOM-DAT – AA	363 (123)	371 (94)
NOM-DAT – IA	372 (113)	390 (118)
DAT-NOM – AA	386 (119)	378 (104)
DAT-NOM – AI	393 (123)	367 (96)
AA	NA	375 (92)
AI	NA	378 (103)
Total time		
NOM-DAT – AA	495 (176)	498 (149)
NOM-DAT – IA	461 (157)	491 (162)
DAT-NOM – AA	515 (171)	504 (142)
DAT-NOM – AI	492 (147)	468 (144)
AA	NA	501 (132)
AI	NA	480 (144)
Launch site		
NOM-DAT – AA	5.4 (1)	6.6 (2)
NOM-DAT – IA	5.2 (1)	6.6 (2)
DAT-NOM – AA	5.7 (2)	6.2 (1)
DAT-NOM – AI	5.6 (1)	6.2 (2)
AA	NA	6.4 (1)
AI	NA	6.4 (2)
Landing position		
NOM-DAT – AA	3.2 (1)	3.7 (2)
NOM-DAT – IA	3.3 (1)	3.6 (2)
DAT-NOM – AA	3.2 (1)	3.7 (2)
DAT-NOM – AI	3.0 (1)	3.7 (1)
AA	NA	3.7 (2)
AI	NA	3.7 (1)
Launch site of the last fixation preceding NP1		
NOM-DAT – AA	3.6 (1)	
NOM-DAT – IA	3.5 (1)	
DAT-NOM – AA	3.6 (1)	
DAT-NOM – AI	3.8 (1)	

Table 9. Participant mean fixation durations (in ms) and positions (in characters) by condition and region. Standard deviation is given in parentheses.

NOM-DAT – AA vs. DAT-NOM – AA

Word order did not have an impact on the duration of the first fixation ($F_1(1,47) = 1.44$, $p < .3$; $F_2(1,39) = 2.44$, $p < .2$) or the duration of the last fixation before the first NP (all *F*s < 1). However, the main effect of ORDER was reliable by participants in first pass time ($F_1(1,47) = 6.82$, $p < .02$; $F_2(1,39) = 3.36$, $p < .08$), evidencing marginally longer reading times for dative-initial clauses (386 ms vs. 363 ms). It was not reliable for the total reading time measure ($F_1(1,47) = 1.49$, $p < .3$; $F_2 < 1$). Neither the landing position of the first fixation nor the launch site of the last fixation leaving the region was influenced by word order variation (landing position: all *F*s < 1; launch site: $F_1(1,47) = 1.43$, $p < .3$; $F_2(1,39) = 1.68$, $p < .3$).

NP2 position

NOM-DAT – AA vs. NOM-DAT – IA

The effect of animacy did not influence the duration of the first fixation (all *F*s < 1) or the last fixation before first fixating NP2 ($F_1(1,47) = 1.85$, $p < .2$; $F_2(1,39) = 2.16$, $p < .2$). As for cumulative reading times, neither first-pass nor total reading time revealed a reliable effect of ANIM (first pass time: $F_1(1,47) = 2.47$, $p < .2$; $F_2(1,39) = 2.34$, $p < .2$; total time: all *F*s < 1). The landing position of the first fixation and the launch site of the last fixation leaving the region were not influenced by manipulating animacy (all *F*s < 1).

DAT-NOM – AA vs. DAT-NOM – AI

Animacy had no influence on the duration of the first fixation on the second NP or the fixation preceding it (all *F*s < 1). There was also no effect of ANIM in first pass time ($F_1(1,47) = 1.38$, $p < .3$; $F_2(1,39) = 1.00$, $p < .4$), but it approached significance by participants in total time ($F_1(1,47) = 3.76$, $p < .06$; $F_2(1,39) = 2.17$, $p < .2$), manifesting slightly increased reading times when both NPs were animate (504 ms vs. 468 ms). The landing position of the first fixation and the launch site of the last fixation leaving the region were not influenced by the variation in animacy (all Fs < 1).

AA vs. AI

The interference effect was not reliable in the first fixation measure (all Fs < 1), the last fixation measure ($F_1(1,47) = 2.22$, $p < .2$; $F_2(1,39) = 2.56$, $p < .2$), in first pass time (all Fs < 1), or in total time ($F_1(1,47) = 2.61$, $p < .2$; $F_2(1,39) = 1.13$, $p < .3$). The landing position of the first fixation and the launch site of the last fixation leaving the region were not influenced by animacy-induced interference (all Fs < 1).

NOM-DAT – AA vs. DAT-NOM – AA

The fixation time measure revealed no significant results (all Fs < 1). The fixation positions also failed to reveal reliable effects (landing position: all Fs < 1; launch site: $F_1(1,47) = 4.36$, $p < .05$; $F_2(1,39) = 3.11$, $p < .09$).

7.2.4 Discussion

The present experiment aimed to show whether parafoveal semantic processing is detectable in terms of ERP correlates of processing when case marking or syntactic function assignment of the linguistic stimuli is unambiguous. The stimuli consisted of passive clauses in which the word order between an unambiguously case-marked dative argument and a nominative argument was varied. The animacy of the nominative argument was also manipulated, so that it was either animate or inanimate. This design focused on processing costs for scrambled word orders and conflicts between prominence information and linking information. While it was hypothesized that there should be no scrambling negativity for clause-initial dative arguments, the animacy manipulation was expected to influence the processing ease of the nominative argument.

The findings from this experiment can be summarized as follows. As expected, there was no hint of a scrambling negativity for dative arguments when they occupied the first clause-medial position. While there was no electrophysiological effect, the eye movement data evidenced the expected (though small) behavioral disadvantage for initial dative arguments. This suggests that the eye movement method may be sensitive to the structural frequency of dative-nominative word orders (cf. Scheepers, 2005), even when no neuronal processing costs are visible. Interestingly, this pattern replicates previous ERP and eye movement results obtained in separate studies. While Bornkessel et al. (2002) did not find a scrambling negativity for initial dative arguments, Scheepers (2005) reported increased reading times for these arguments (and the lack of any effects at the second NP position). The present experiment supports the view that this discrepancy between ERP and eye movement measures is not a result of differences in presentation methods. Rather, information concerning structural frequency and grammatical preferences are treated differently in the course of linguistic processing, with only the former passed on to the oculomotor system. The reason for this divergence is as yet unclear and remains to be tested in future experiments. In contrast to this, the variation in animacy of the nominative argument did have an impact on the processing of the first NP – interestingly, the ERP effects found for the last fixation preceding the first NP suggest that animacy information is already acquired in the parafovea. The parafoveal N400 amplitude was enhanced when the upcoming nominative argument was animate (condition NOM-DAT – AA), but not when it was inanimate (condition NOM-DAT – IA). This effect cannot be related to animacy-induced interference (cf. Experiment 4), as this would mean that semantic information from both NPs is already extracted even before the first NP is fixated. This appears somewhat implausible given the distance between the last fixation preceding NP1 and the second NP, which is too remote to deliver any useful information. Rather, it is conceivable that semantic information from the initial nominative is acquired parafoveally, and then evaluated with respect to the argument's prominence status. According to the basic assumptions of the eADM model introduced in chapter 7.1.1, the N400 effect is attributable to the mismatch of linking information and prominence information. According to Table 9 above, the last fixation before the first NP was located on the immediately preceding auxiliary *wurde* in a significant number of trials. The passive auxiliary *wurde* signals that the generalized semantic role of undergoer should be attributed to the nominative argument. However, nominative case usually encodes the argument that carries the higher generalized semantic role, namely the actor role. When the animacy information provides further evidence for the actorhood of the argument in question (i.e., it is animate), the resulting accumulation of prominence information (case marking and animacy) conflicts with the assigned undergoer role. It is this mismatch between the prominence information from the argument and the linking information from the auxiliary that probably induces this N400 effect to animate nominative arguments, as found in the present experiment. Inanimate nominative arguments, by contrast, are better undergoer arguments because their inanimacy makes them less optimal for actorhood. Note that the late positivity engendered by (the first fixation on the) initial nominative arguments is difficult to account for, as it partly overlaps with the N400 elicited by the second NP (see below). Specifically, when comparing the ERP waves triggered by the first fixation on NP1 and NP2, it is conceivable that the late positivity for nominative-initial structures with two animate NPs is at least partially due to an overlap with the N400 effect for dative-initial clauses at NP2 (e.g., the increased N400 to NP2 in dative-initial structures containing an animate nominative argument is already visible approximately 500-600 ms post-onset of NP1). Therefore no functional interpretation appears possible here, and future experiments are needed to decide whether the late positivity is also visible when there is no overlap between ERPs elicited by two adjacent NPs with similar properties. Finally, as in Experiment 2, this parafoveal effect was not mirrored in the corresponding fixation duration. In fact, no eye movement measure showed effects for the first NP that would indicate processing disruption due to the described mismatch between case and animacy information.

Importantly, based on the last row of Table 9 above, one might contend that the position of this final fixation was very close to the first NP in a majority of trials (approximately 4 letters preceding the space before the NP), thereby allowing for the possibility of foveal processing. However, this conjecture misses the fact that the determiner preceding the noun increases the distance to the noun itself, which carries the critical animacy information. Thus, the distance separating the last fixation before the first NP and the noun itself was almost double the distance suggested by the launch site measure. Foveal processing is therefore less likely to play a role in the observed parafoveal N400 effect.

This parafoveal animacy effect is important in two respects. First, it was hypothesized in the discussion of Experiment 4 that parafoveal N400 effects may not occur when foveal processing load is high, as this reduces the amount of resources that can be spared for parafoveal processing (cf. Henderson & Ferreira, 1993). In Experiment 4, foveal processing load was probably enhanced through the ambiguity in case marking for each NP. Here, by contrast, the last fixation preceding the first NP was located in a non-critical region of the sentence, so foveal processing load should be relatively low, resulting in a benefit for parafoveal processing. Second, this has important methodological implications for future experiments that aim to concurrently record ERPs and eye movements during reading. Parafoveal N400 effects are less likely to emerge if the last fixation before the critical word is located on ambiguous material, or put more generally, if the region of text preceding the critical target possesses properties that induce a high processing load. This somewhat limits the range of possible experimental designs that can be used. If, particularly, the linguistic manipulation under investigation brings about multiple adjacent regions being examined, the chance to find clear parafoveal ERP effects is weakened. The last fixation before the critical target word should therefore best be located on parts of the sentence that impose no processing effort, i.e. they should not be part of the experimental manipulation. This limitation regarding multiple-region analyses can explain the lack of parafoveal N400 effects for ambiguous NPs in Experiment 4 and for the second NP in the present experiment (considering the dative-nominative structures). The parafoveal N400 effect suggests moreover that semantic preprocessing can take place in the parafovea, in the absence of corresponding effects in fixation durations. This supports the claim that not all kinds of cognitive processing disruption are passed on to the oculomotor system (cf. chapter 5.3). Furthermore, the co-registration of ERPs and eye movements in reading may be particularly fruitful for investigations of parafoveal-on-foveal effects. Chapter 2.3.1 showed that the eye movement data on lexical parafoveal-on-foveal effects are highly heterogeneous and that most of these effects were found when language corpora were used as stimuli. The ERP method is probably less well suited when stimuli from language corpora are used for the reading task, as it is impossible to control for all external factors in corpus-based research on eye movements in reading, i.e. factors to which the ERP method is sensitive, but which are not of relevance to psycholinguistic research. However, the present experiment suggests that the combined measurement of ERPs and eye movements in a controlled experimental setting can provide new insights into parafoveal processing.

However, one open question remains. Why should a relatively non-constraining context such as in the present experiment bring about semantic processing in the parafovea when high-constraint contexts such as those in the antonym paradigm do not? This appears contradictory to the assumption that highly constraining contexts should facilitate parafoveal processing (cf. chapter 2.2.2). In order to account for this paradox, one has to remember that the high-constraint antonym context employed in Experiment 2 activates one and only one lexical entry, including its orthographical form. Given this very specific prediction, orthographic preprocessing in the parafovea is sufficient to ensure unimpeded processing, as it can already deliver critical information about the upcoming word. In the low-constraint context used here, there is no particular prediction with respect to upcoming word forms, so the language comprehension system does not gain an extra benefit from the orthographic information acquired in the parafovea (beyond saccadic programming to the next saccade target). Instead, more parafoveal information is necessary to prepare the system for upcoming words and thus the language comprehension system may engage to a larger degree in semantic preprocessing in the parafovea.

At the position of the second NP, dative-nominative word orders with two animate NPs (condition DAT-NOM – AA) enhanced the amplitude of the N400 relative to their nominative-dative counterparts (condition NOM-DAT – AA). This effect clearly contradicts the prediction that an early positivity reflecting thematic reanalysis should have emerged. One could posit that the thematic reanalysis involving the undergoer role might elicit processing correlates that are different from the thematic reanalysis effect involving the actor role, as in Bornkessel et al. (2003) study. However, this explanation lacks independent motivation for why the re-assignment of the undergoer role should differ from the reassignment of the actor role. It is more likely that the specific interaction between case and noun semantics is responsible for the N400 effect. In following the widely accepted claim that dative arguments do not receive a generalized semantic role (see the introduction), the N400 effect on the second NP in dative-nominative word orders cannot reflect thematic reanalysis, because the dative argument did not receive a generalized semantic role in the first place. The N400 effect can then be explained with two partly different accounts. One attributes the effect to the allocation of a generalized role. The N400 arises because the nominative argument is assigned the undergoer role, whereas the dative argument does not receive a generalized semantic role. The N400 might then reflect subtle differences in semantic complexity. Interestingly, a similar effect has been reported for accusative verb participles relative to dative verb participles, with the former showing an increase in N400 amplitude. This was explained to be a complexity effect that was related to the transitivity of verbs. Specifically, it was proposed that more effort has to be made to process accusative verbs because they assign two generalized semantic roles, whereas dative verbs behave like intransitive verbs in the assignment of generalized semantic roles (Schlesewsky & Bornkessel, 2006). The second explanation is identical to the explanation of the parafoveal N400 effect given above: conflicting information from case marking and animacy information of the nominative argument has to be reconciled with the assigned undergoer role. The conflict occurs when prominence information qualifies the animate NP as an ideal actor, while the passive construction assigns it the undergoer role.

Both of these accounts have their merits and disadvantages. For instance, although the first one could also explain the early positivity in the Bornkessel et al. study, it fails to account for the reversed polarity of the ERP effects across the Bornkessel et al. and the current study. The second explanation, on the other hand, can account for the N400 effects found here across both NP positions in a more parsimonious way than the first. Both accounts share one important caveat, namely that the comparisons of the respective counterpart conditions failed to yield significant results. There was no effect on the first NP when initial nominative arguments were compared to initial dative arguments (conditions NOM-DAT – AA vs. DAT-NOM – AA for NP1). Visual inspection of the first NP suggested that the N400 amplitude to nominative arguments was slightly increased relative to dative arguments, but this effect fell short of significance (cf. Figure 16). This is unexpected by either account. If, moreover, it had been only the mismatch between prominence information and role assignment that induced the N400 to animate nominative arguments at the second NP, the same effect should have been found when the animate and inanimate nominative arguments in dative-initial conditions were compared (conditions DAT-NOM – AA vs. DAT-NOM – AI). However, in spite of the visual impression that the animate nominatives were more negative than inanimate nominatives (cf. Figure 18), this effect was not statistically reliable.[49] Thus, both of these explanations can account for the statistically reliable N400 effects, but the overall pattern of results is not conclusive enough to favor one explanation over the other. It appears premature to give a definite conclusion about the exact nature of this N400 effect found for the second NP. Most of the present ERP effects were rather small and it could be possible that the generally weak experimental power is responsible for the null effects in some of the pair-wise comparisons carried out here. Further experiments are called for to clarify whether the present N400 effects at both NP positions are robust when ERPs are collected during natural reading and how they might relate the positivity elicited by thematic reanalysis (i.e., to what extent the availability of the parafoveal preview may be a decisive cause of the diverging results).

Finally, the eye movement record showed no effect corresponding to the N400 to animate nominative arguments. There was, however, a small increase in total reading time for an animate nominative argument in the dative-nominative structures. Altogether, the animacy effect is much weaker in the present experiment as opposed to Experiment 4. One possibility

[49] The latter finding could be due to the fact that for the dative-initial structures, the prominence information delivered by linear position and case marking of the NPs do not converge. Specifically, the first NP in the clause can be considered more prominent than the second NP (cf. Bornkessel-Schlesewsky & Schlesewsky, 2009b) and information from other prominence hierarchies, such as the case hierarchy, should ideally lead to nominative case-marking for the first NP, as nominative is the more prominent case in German. Since both dative-initial structures violate the preferred order of cases in general, there is an additional clash between the linear position and the case hierarchy. This might have reduced the N400 effect to animate nominative arguments relative to inanimate nominative arguments.

is that animacy only played a minor role here because of the unambiguous case marking. As described in chapter 7.1.1 prominence information such as animacy plays only a modulating role in identifying the generalized semantic role of an argument, whereas case marking is a primary means to allocate the actor and undergoer roles. Thus, animacy was presumably more important in Experiment 4 because of the case ambiguity – which was noticed despite the semantic cue about word order – whereas it was of little relevance in the role identification of the unambiguously case-marked arguments used in this experiment. While the language comprehension system registered the mismatch between the assigned undergoer role and the prominence information favoring the actor role (for the animate nominative argument), this mismatch was not powerful enough to disrupt processing in a way to prolong fixation durations or cause refixations.

7.3 Summary

Experiment 4 and 5 aimed to further study the nature of the parafoveal N400 effect found in Experiment 2. It was argued that this parafoveal N400 was solely due to orthographic processing. This supports the view that word recognition in reading is a serial process, with prelexical information processed in parafoveal vision and semantic information processed in foveal vision. With the use of stimuli that differed in their structural properties, parafoveal processing could be investigated in linguistic contexts that differed in how specific the predictions about upcoming words are and in the processing load induced by morphosyntactic ambiguity.

Experiment 4 investigated the processing of verb-final clauses with two case-ambiguous NPs. This case ambiguity rendered both NPs ambiguous as to their syntactic function. However, the type of each NP (proper name or bare plural) provided a semantic cue about its syntactic function, because a certain NP type was associated with one and the same syntactic function throughout the experiment. The results showed robust N400 and late positivity effects due to animacy-induced interference. The low wellformedness of object-initial clauses gave rise to a small late positivity. Importantly, all of these effects were wholly foveal in nature. This experiment thus provided further converging evidence for the claim that parafoveal N400 effects are not related to semantic processing. One limitation to this conclusion is that case marking ambiguity could have reduced the parafoveal preview of upcoming words, thereby reducing the likelihood of parafoveal N400 effects occurring.

To test this hypothesis, Experiment 5 was carried out with unambiguous structures involving word order scrambling and animacy variation. The data revealed a parafoveal N400 effect that resulted from semantic processing. Specifically, when the upcoming nominative argument was animate, the amplitude of the parafoveal N400 was increased (relative to an inanimate nominative argument). This parafoveal N400 was interpreted as reflecting the mismatch between different types of prominence information. This divergence between Experiments 4 and 5 points to a methodological issue, which remains to be resolved in future research. Experiment 4 revealed no parafoveal N400 effect, even though the critical words were adjacent to one another. Experiment 5, by contrast, showed such an effect between words that were further remote from each other. It was proposed that the last fixation trigger must be located on pre-critical words that do not induce processing effort themselves if parafoveal N400s are to be detected. This represents a limitation for those experimental designs in which the sentences include multiple adjacent regions of analysis.

Both experiments revealed that N400 and late positivity effects can result in increased fixation durations. Experiment 5 replicated moreover the intriguing finding that some behavioral effects are not accompanied by neuronal effects, which supports previous results from separate ERP and reading time studies. Eye movements were sensitive to scrambled dative objects, as visible in the somewhat increased reading times, but there was no specific ERP wave associated with this effect.

References

Aissen, J. (2003). Differential object marking: Iconicity vs. economy. *Natural Language & Linguistic Theory*, *21*, 435-483.

Allefeld, C., Frisch, S., & Schlesewsky, M. (2005). Detection of early cognitive processing by event-related phase synchronization analysis. *NeuroReport*, *16*(1), 13-16.

Altenmüller, E. O., & Gerloff, C. (1999). Psychophysiology and the EEG. In E. Niedermeyer & F. Lopes da Solva (Eds.), *Electroencephalography: Basic principles, clinical applications, and related fields* (4th Edition ed., pp. 637-655). Baltimore: Williams & Wilkins.

Anderson, J. E., & Holcomb, P. J. (1995). Auditory and visual semantic priming using different stimulus onset asynchronies: an event-related brain potential study. *Psychophysiology*, *32*, 177-190.

Angele, B., Slattery, T. J., Yang, J., Kliegl, R., & Rayner, K. (2008). Parafoveal processing in reading: Manipulating n + 1 and n + 2 previews simultaneously. *Visual Cognition*, *16*, 697-707.

Ashby, J., Rayner, K., & Clifton, C. (2005). Eye movements of highly skilled and average readers: differential effects of frequency and predictability. *Quarterly Journal of Experimental Psychology*, *58A*, 1065-1086.

Ashby, J., Treiman, R., Kessler, B., & Rayner, K. (2006). Vowel processing during silent reading: Evidence from eye movements. *Journal of Experimental Psychology: LearninG, Memory, and Cognition*, *32*, 416-424.

Baccino, T., & Manunta, Y. (2005). Eye-fixation-related potentials: insight into parafoveal processing. *Journal of Psychophysiology*, *19*, 204-215.

Bader, M., & Lasser, I. (1994). German verb-final clauses and sentence processing: Evidence for immediate attachment. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 225-242). Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Bader, M., & Meng, M. (1999). Subject-object ambiguities in German embedded clauses: An across-the-board comparison. *Journal of Psycholinguistic Research*, *28*, 121-143.

Bai, X., Yan, G., Liversedge, S. P., Zang, C., & Rayner, K. (2008). Reading spaced and unspaced Chinese text: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1277-1287.

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*, 364-390.

Bentin, S. (1987). Event-related potentials, semantic processes, and expectancy factors in word recognition. *Brain and Language*, *31*, 308-327.

Bentin, S., Kutas, M., & Hillyard, S.A. (1993). Electrophysiological evidence for task effects on semantic priming in auditory word processing. *Psychophysiology, 30,* 161-169

Bentin, S., Kutas, M., & Hillyard, S.A. (1995). Semantic processing and memory for attended and unattended words in dichotic listening: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 54-67.

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinicial Neurophysiology*, *60*, 343-355

Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier, J. F., & Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. *Journal of Cognitive Neuroscience*, *11* 235-260.

Berger, H. (1929). Über das Elektroenkephalogramm des Menschen [On the human electroencephalogram]. *Archiv für Psychiatrie und Nervenkrankheiten*, *87*, 527-570.

Bornkessel, I. D., Fiebach, C. J., & Friederici, A. D. (2004). On the cost of syntactic ambiguity in human language comprehension: an individual differences approach. *Cognitive Brain Research*, *21*, 11-21.

Bornkessel, I., McElree, B., Schlesewsky, M., & Friederici, A. D. (2004). Multi-dimensional contributions to garden path strength: Dissociating phrase structure from case marking. *Journal of Memory and Language*, *51*, 495-522.

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychological Review*, *113*, 787-821.

Bornkessel, I., Schlesewsky, M., & Friederici, A. D. (2002). Grammar overrides frequency: evidence from the online processing of flexible word order. *Cognition*, *85*, B21-30.

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

Bornkessel, I., Zysset, S., Friederici, A.D., von Cramon, Y., & Schlesewsky, M. (2005). Who did what to whom? The neural basis of argument hierarchies during language comprehension. *Neuroimage, 26*, 221-233.

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

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009a). *Processing Syntax and Morphology: A neurocognitive perspective* (6). Oxford: Oxford University Press.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009b). The Role of Prominence Information in the Real-Time Comprehension of Transitive Constructions: A Cross-Linguistic Approach. *Language and Linguistics Compass*, *3*, 19-58.

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

Brookhuis, K. A., Mulder, G., Mulder, L. J., Gloerich, A. B., van Dellen, H. J., van der Meere, J. J. et al. (1981). Late positive components and stimulus evaluation time. *Biological Psychology*, *13*, 107-123.

Brysbeart, M.,, & Mitchell, D.C. (1996). Modifier attachment in sentence parsing: Evidence from Dutch. *Quarterly Journal of Experimental Psychology, 49A*, 664-695.

Calvo, M. G., & Meseguer, E. (2002). Eye movements and processing stages in reading: relative contribution of visual, lexical, and contextual factors. *Journal of Spanish Psychology*, *5*, 66-77.

Camblin, C. C., Gordon, P. C., & Swaab, T. Y. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory and Language*, *56*, 103-128.

Carreiras, M., Salillas, E., & Barber, H. (2004). Event-related potentials elicited during parsing of ambiguous relative clauses in Spanish. *Cognitive Brain Research, 20*, 98-105.

Chace, K. H., Rayner, K., & Well, A. D. (2005). Eye movements and phonological parafoveal preview: effects of reading skill. *Canadian Journal of Experimental Psychology*, *59*, 209-217.

Chen, H.-C., Song, H., Lau, W. Y., Wong, K. F. E., & Tang, S. L. (2003). Developmental characteristics of eye movements in reading Chinese. In C. McBride-Chang & H.-C.

Chen (Eds.), *Reading development in Chinese children* (pp. 157-169). Westport, CT: Praeger.

Chen, H.-C., Tang, C.-K. (1998). The effective visual field in Chinese. *Reading and Writing, 10,* 245-254.

Christensen, C. A., Ford, J. M., & Pfefferbaum, A. (1996). The effect of stimulus-response incompatibility on P3 latency depends on the task but not on age. *Biological Psychology*, *44*, 121-141.

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

Chwilla, D. J., Kolk, H. H., & Mulder, G. (2000). Mediated priming in the lexical decision task: Evidence from event-related potentials and reaction time. *Journal of Memory and Language*, *42*, 314-341.

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R.P.G. van Gompel, M.H. Fischer, & W.S. Murray (Eds.), *Eye movements: A window on mind and brain* (pp. 341-372). Amsterdam: Elsevier.

Clifton, C., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, *49*, 317-334.

Collins, A. M., & Loftus, E. F. (1975). Spreading activation theory of semantic processing. *Psychological Review*, *82*, 407-428.

Combs, L. A., & Polich, J. (2006). P3a from auditory white noise stimuli. *Clinical Neurophysiology*, *117*, 1106-1112.

Comerchero, M. D., & Polich, J. (1999). P3a and P3b from typical auditory and visual stimuli. *Clinical Neurophysiology, 110*, 24-30.

Connolly, J.F., Phillips, N.A., Stewart, S.H., & Brake, W.G. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and Language, 43*, 1-18.

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

Coulson, S., King, J. W., & Kutas, M. (1998b). ERPs and domain specificity: Beating a straw horse. *Language and Cognitive Processes*, *13*, 653-672.

Courchesne, E., Hillyard, S. A., & Courchesne, R. Y. (1977). P3 waves to the discrimination of targets in homogeneous and heterogeneous stimulus sequences. *Psychophysiology*, *14*, 590-597.

Courchesne, E., Hillyard, S. A., & Galambos, R. (1975). Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalogr Clin Neurophysiol*, *39*, 131-143.

Crocker, M. W. (1994). On the nature of the principle-based sentence processor. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 245-266). Hillsdale, NJ: Erlbaum.

Dambacher, M., & Kliegl, R. (2007). Synchronizing timelines: relations between fixation durations and N400 amplitudes during sentence reading. *Brain Research*, *1155*, 147-162.

Deutsch, A., Frost, R., Pelleg, S., Pollatsek, A., & Rayner, K. (2003). Early morphological effects in reading: evidence from parafoveal preview benefit in Hebrew. *Psychonomic Bulletin & Review*, *10*, 415-422.

Dien, J. (2009). The neurocognitive basis of reading single words as seen through early latency ERPs: a model of converging pathways. *Biological Psychology*, *80*, 10-22.

Dien, J., Spencer, K. M., & Donchin, E. (2004). Parsing the late positive complex: mental chronometry and the ERP components that inhabit the neighborhood of the P300. *Psychophysiology*, *41*, 665-678.

Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., Engbert, R., & Kliegl, R. (2006). Concurrent recording of EEG and gaze position: Measuring effects of word predictability during left-to-right reading of normal sentences. *Journal of Cognitive Neuroscience (Supplement).*

Dimigen, O., Sommer, W., & Kliegl, R. (2007). *Long reading regressions are accompanied by a P600-like brain potential.* Paper presented at the European Conference on Eye Movements, Potsdam.

Ditman, T., Holcomb, P. J., & Kuperberg, G. R. (2007). An investigation of concurrent ERP and self-paced reading methodologies. *Psychophysiology*, *44*, 927-935.

Dogil, G., Frese, I., Haider, H., Roehm, D., & Wokurek, W. (2004). Where and how does grammatically geared processing take place--and why is Broca's area often involved. A coordinated fMRI/ERBP study of language processing. *Brain and Language*, *89*, 337-345.

Dominguez, A., de Vega., M, & Barber, H. (2004). Event-related brain potentials elicited by morphological, homographic, orthographic, and semantic priming. *Journal of Cognitive Neuroscience*, *16*, 598-608.

Donchin, E. (1981). Surprise! Surprise? *Psychophysiology*, *18*, 493-513.

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*, 357-374.

Drieghe, D. (2008). Foveal processing and word skipping during reading. *Psychonomic Bulletin & Review*, *15*, 856-860.

Drieghe, D., Rayner, K., & Pollatsek, A. (2005). Eye movements and word skipping during reading revisited. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 954-959.

Drieghe, D., Rayner, K., & Pollatsek, A. (2008). Mislocated fixations can account for parafoveal-on-foveal effects in eye movements during reading. *Quarterly Journal of Experimental Psychology*, *61*, 1239-1249.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, *27*, 429-446.

Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: the variation of event-related potentials with subjective probability. *Psychophysiology*, *14*, 456-467.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*, 641-655.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: a dynamical model of saccade generation during reading. *Psychological Review*, *112*, 777-813.

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

Ford, J. M., Pfefferbaum, A., Tinklenberg, J. R., & Kopell, B. S. (1982). Effects of perceptual and cognitive difficulty on P3 and RT in young and old adults. *Electroencephalography and Clinical Neurophysiology*, *54*, 311-321.

Frazier, L., & Flores D'Arcais, G.B. (1989). Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language, 28*, 331-344.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements and the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178-210.

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*, 157-176.

Friederici, A. D., Mecklinger, A., Spencer, K. M., Steinhauer, K., & Donchin, E. (2001). Syntactic parsing preferences and their on-line revisions: a spatio-temporal analysis of event-related brain potentials. *Cognitive Brain Research*, *11*, 305-323.

Friedman,D., Hakerem, G., Sutton, S., & Fleiss, J.L. (1973). Effect of stimulus uncertainty on the pupillary dilation response and the vrtex evoked potential. *Electroencephalography and Clinical Neurophysiology, 34*, 475-484.

Frisch, S., & Schlesewsky, M. (2001). The N400 reflects problems of thematic hierarchizing. *Neuroreport*, *12*, 3391-3394.

Frisch, S., Schlesewsky, M., Saddy, D., & Alpermann, A. (2002). The P600 as an indicator of syntactic ambiguity. *Cognition*, *85*, B83-92.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1411-1423.

Grainger, J., & Holcomb, P.J. (2009). Watching the word go by: On the time-course of component processes in visual word recognition. *Language and Linguistics Compass, 3*, 128-156.

Grewe, T., Bornkessel, I., Zysset, S., Wiese, R., von Cramon, D. Y., & Schlesewsky, M. (2006). Linguistic prominence and Broca's area: The influence of animacy as a linearization principle. *Neuroimage*, *32*, 1395-1402.

Gunter, T. C., & Friederici, A. D. (1999). Concerning the automaticity of syntactic processing. *Psychophysiology*, *36*, 126-137.

Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology*, *34*, 660-676.

Hagen, G. F., Gatherwright, J. R., Lopez, B. A., & Polich, J. (2006). P3a from visual stimuli: task difficulty effects. *International Journal of Psychophysiology*, *59*, 8-14.

Hagoort, P., Brown, C. M., & Swaab, T. Y. (1996). Lexical-semantic event-related potential effects in patients with left hemisphere lesions and aphasia, and patients with right hemisphere lesions without aphasia. *Brain*, *119*, 627-649.

Haupt, F. S., Schlesewsky, M., Roehm, D., Friederici, A. D., & Bornkessel-Schlesewsky, I. (2008). The status of subject-object reanalyses in the language comprehension architecture. *Journal of Memory and Language*, *59*, 54-96.

Henderson, J. M., & Ferreira, F. (1993). Eye movement control during reading: fixation measures reflect foveal but not parafoveal processing difficulty. *Canadian Journal of Experimental Psychology*, *47*, 201-221.

Henderson, J. M., Pollatsek, A., & Rayner, K. (1989). Covert visual attention and extrafoveal information during object identification. *Perception & Psychophysics*, *45*, 196-208.

Hill, H., Ott, F., & Weisbrod, M. (2005). SOA-dependent N400 and P300 semantic priming effects using pseudoword primes and a delayed lexical decision. *International Journal of Psychophysiology*, *56*, 209-221.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, *182*, 177-180.

Hillyard, S. A., Squires, K. C., Bauer, J. W., & Lindsay, P. H. (1971). Evoked potential correlates of auditory signal detection. *Science*, *172*, 1357-1360.

Hoffman, J. E., Houck, M. R., MacMillan, F. W. r., Simons, R. F., & Oatman, L. C. (1985). Event-related potentials elicited by automatic targets: a dual-task analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 50-61.

Hoffman, J. E., Simons, R. F., & Houck, M. R. (1983). Event-related potentials during controlled and automatic target detection. *Psychophysiology*, *20*, 625-632.

Holcomb, P. J. (1988). Automatic and attentional processing: an event-related brain potential analysis of semantic priming. *Brain and Language*, *35*, 66-85.

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: implications for the role of the N400 in language processing. *Psychophysiology*, *30*, 47-61.

Holcomb, P. J., & Anderson, J. E. (1993). Cross-modal semantic priming: A time-course analysis using event-related brain potentials. *Language and Cognitive Processes*, *8*, 379-411.

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes*, *5*, 281-312.

Horowitz, T. S., Fencsik, D. E., Fine, E. M., Yurgenson, S., & Wolfe, J. M. (2007). Microsaccades and attention. Does a weak correlation make an index? Reply to Laubrock, Engbert, Rolfs, and Kliegl (2007). *Psychological Science, 18*, 367-368.

Horowitz, T. S., Fine, E. M., Fencsik, D. E., Yurgenson, S., & Wolfe, J. M. (2007). Fixational eye movements are not an index of covert attention. *Psychological Science*, *18*, 356-363.

Hutzler, F., Braun, M., Vo, M. L., Engl, V., Hofmann, M., Dambacher, M. et al. (2007). Welcome to the real world: validating fixation-related brain potentials for ecologically valid settings. *Brain Research*, *1172*, 124-129.

Huynh, H., & Feldt, L. S. (1970). Conditions under which the mean square ratios in repeated measurement designs have exact F-distributions. *Journal of the American Statistical Association*, *65*, 1582-1589.

Inhoff, A. W., Eiter, B. M., & Radach, R. (2005). Time course of linguistic information extraction from consecutive words during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 979-995.

Inhoff, A. W., & Liu, W. (1998). The perceptual span and oculomotor activity during the reading of Chinese sentences. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(1), 20-34.

Inhoff, A. W., & Wu, C. (2005). Eye movements and the identification of spatially ambiguous words during chinese sentence reading. *Memory & Cognition*, *33*, 1345-1356.

Irwin, D. E. (1998). Lexical processing during saccadic eye movements. *Cognitive Psychology*, *36*, 1-27.

Isreal, J. B., Chesney, G. L., Wickens, C. D., & Donchin, E. (1980). P300 and tracking difficulty: evidence for multiple resources in dual-task performance. *Psychophysiology*, *17*, 259-273.

Isreal, J. B., Wickens, C. D., Chesney, G., & Donchin, E. (1980). The event-related brain potential as an index of display-monitoring workload. *Human Factors*, *22*, 211-224.

Johnson, R. J. (1986). A triarchic model of P300 amplitude. *Psychophysiology*, *23*, 367-384.

Johnson, R. J. (1993). On the neural generators of the P300 component of the event-related potential. *Psychophysiology*, *30*, 90-97.

Joos, M., Rötting, M., & Velichkovsky, B. M. (2003). Bewegungen des menschlichen Auges: Fakten, Methoden und innovative Anwendungen. In G. Rickheit, T. Herrmann, & W. Deutsch (Eds.), *Psycholinguistik/Psycholinguistics*. *Ein internationales Handbuch/An international Handbook* (pp. 142-168). Berlin & NY: de Gruyter.

Juhasz, B. J., Liversedge, S. P., White, S. J., & Rayner, K. (2006). Binocular coordination of the eyes during reading: word frequency and case alternation affect fixation duration but not fixation disparity. *Quarterly Journal of Experimental Psychology*, *59*, 1614-1625.

Juhasz, B. J., White, S. J., Liversedge, S. P., & Rayner, K. (2008). Eye movements and the use of parafoveal word length information in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1560-1579.

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

Kamide, Y., & Mitchell, D. C. (1999). Incremental Pre-head Attachment in Japanese Parsing. *Language and Cognitive Processes*, *14*, 631-662.

Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000). *Principles of neural science*. Appleton & Lange.

Katayama, J., & Polich, J. (1996). P300 from one-, two-, and three-stimulus auditory paradigms. *International Journal of Psychophysiology*, *23*, 33-40.

Kennedy, A. (2000). Parafoveal processing in word recognition. *Quarterly Journal of Experimental Psychology*, *53A*, 429-455.

Kennedy, A. (2008). Parafoveal-on-foveal effects are not an artifact of mis-located saccades. *Journal of Eye Movement Research*, *2*, 1-10.

Kennedy, A., & Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Research*, *45*, 153-168.

Kennison, S. M., & Clifton, C. J. (1995). Determinants of parafoveal preview benefit in high and low working memory capacity readers: implications for eye movement control. *Journal of Experimental Psychology: Learnnig, Memory, and Cognition*, *21*, 68-81.

Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Kerkhof, G. A. (1978). Decision latency: the P3 component in auditory signal detection. *Neuroscience Letters*, *8*, 289-294.

Kiefer, M., Marzinzik, F., Weisbrod,M., Scherg, M., & Spitzer, M. (1998). The time course of brain activations during response inhibition: evidence from event-related potentials in a go/no-go task. *Neuroreport, 9*, 765-770.

Kiefer, M., Weisbrod, M., Kern, I., Maier, S., & Spitzer, M. (1998). Right hemisphere activation during indirect semantic priming: evidence from event-related potentials. *Brain and Language*, *64*, 377-408.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*, 262-284.

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: the influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, *135*, 12-35.

Kliegl, R., Risse, S., & Laubrock, J. (2007). Preview benefit and parafoveal-on-foveal effects from word n + 2. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 1250-1255.

Knoeferle, P., Habets, B., Crocker, M.W., & Münte, T.F. (2008). Visual scenes trigger immediate syntactic reanalysis: Evidence from ERPs during situated spoken comprehension. *Cerebral Cortex, 18*, 789-795.

Koivisto, M., & Revonsuo, A. (2007). Electrophysiological correlates of visual consciousness and selective attention. *Neuroreport*, *18*, 753-756.

Koivisto, M., Revonsuo, A. (2008). The role of selective attention in visual awareness of stimulus features: Electrophysiological studies. *Cognitive, Affective, & Behavioral Neuroscience, 8*, 195-210.

Koivisto, M., Revonsuo, A. (2008). Comparison of event-related potentials in attentional blink and repetition blindness. *Brain Research, 1189*, 115-126.

Kok, A. (1986). Effects of degradation of visual stimulation on components of the event-related potential (ERP) in go/nogo reaction tasks. *Biological Psychology*, *23*, 21-38.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*, 557-577.

Kok, A., & Looren de Jong, H. (1980). Components of the event-related potential following degraded and undegraded visual stimuli. *Biological Psychology*, *11*, 117-133.

Kok, A., & Rooyakkers, J. A. (1986). ERPs to laterally presented pictures and words in a semantic categorization task. *Psychophysiology*, *23*, 672-683.

Kok, A., van de Vijver, F. R., & Rooijakkers, J. A. (1985). Effects of visual field, stimulus degradation, and level of practice on event-related potentials of the brain. *Psychophysiology*, *22*, 707-717.

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

Kramer, A. F., Sirevaag, E. J., & Hughes, P. R. (1988). Effects of foveal task load on visual-spatial attention: event-related brain potentials and performance. *Psychophysiology*, *25*, 512-531.

Kramer, A. F., Wickens, C. D., & Donchin, E. (1985). Processing of stimulus properties: evidence for dual-task integrality. *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 393-408.

Kretzschmar, F., Bornkessel-Schlesewsky, I., Staub, A., Roehm, D., & Schlesewsky, M. (in press). Prominence facilitates ambiguity resolution: On the interaction between specificity and word order in syntactic reanalysis. In M. J. A. Lamers & P. de Swart (Eds.), *Case, Word Order, and Prominence: Psycholinguistic and theoretical approaches to argument structure.* Berlin: Springer.

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

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

Kutas, M., & Hillyard, S. A. (1980b). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology, 11*, 99-116.

Kutas, M., & Hillyard, S. A. (1984a). Event-related brain potentials (ERPs) elicited by novel stimuli during sentence processing. *Annals of the New York Academy of Sciences*, *425*, 236-241.

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

Kutas, M., & Iragui, V. (1998). The N400 in a semantic categorization task across 6 decades. *Electroencephalography and Clinical Neurophysiology*, *108*, 456-471.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. *Science*, *197*, 792-795.

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

Laszlo, S., & Federmeier, K. D. (2009). A Beautiful Day in the Neighborhood: An Event-Related Potential Study of Lexical Relationships and Prediction in Context. *Journal of Memory and Language*, *61*, 326-338.

Laubrock, J., Engbert, R., Rolfs, M., & Kliegl, R. (2007). Microsaccades are an index of covert attention: commentary on Horowitz, Fine, Fencsik, Yurgenson, and Wolfe (2007). *Psychological Science*, *18*, 364-366.

Lenerz, J. (1977). *Zur Abfolge nominaler Satzglieder im Deutschen.* Tübingen: Gunter Narr Verlag.

Lima, S. D., & Inhoff, A. W. (1985). Lexical access during eye fixations in reading: Effects of word-initial letter sequence. *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 272-285.

Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading time. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 55-75). Elsevier.

Luck, S. J. (1998). Sources of dual-task interference: Evidence from human electrophysiology. *Psychological Science*, *9*, 223-227.

Luck, S. J. (2004). Ten simple rules for designing ERP experiments. In E. C. Handy (Ed.), *Event-Related Potentials: A methods book* (pp. 17-32). Cambridge, MA: MIT Press.

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

Luck, S. J., & Hillyard, S. A. (1990). Electrophysiological evidence for parallel and serial processing during visual search. *Perception & Psychophysics*, *48*, 603-617.

Magliero, A., Bashore, T. R., Coles, M. G., & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. *Psychophysiology*, *21*, 171-186.

Mangun, G. R., & Hillyard, S. A. (1990). Allocation of visual attention to spatial locations: tradeoff functions for event-related brain potentials and detection performance. *Perception & Psychophysics*, *47*, 532-550.

McCarthy, G., & Donchin, E. (1976). The effects of temporal and event uncertainty in determining the waveforms of the auditory event related potential (ERP). *Psychophysiology*, *13*, 581-590.

McCarthy, G., & Donchin, E. (1981). A metric for thought: a comparison of P300 latency and reaction time. *Science*, *211*, 77-80.

McCarthy, G., & Nobre, A. C. (1993). Modulation of semantic processing by spatial selective attention. *Electroencephalography and Clinical Neurophysiology*, *88*, 210-219.

McConkie, G. W., & Rayner, K. (1975). The span of effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*, 578-586.

McClelland, J.L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86*, 287-330.

McDonald, S. A. (2006). Parafoveal preview benefit in reading is only obtained from the saccade goal. *Vision Research*, *46*, 4416-4424.

McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), *The psychology of learning and motivation, Vol. 46.* San Diego: Academic Press.

McElree, B., & Griffith, T. (1995). Syntactic and thematic processing in sentence comprehension: Evidence for a temporal dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 134-157.

Mecklinger, A., Schriefers, H., Steinhauer, K., & Friederici, A. D. (1995). Processing relative clauses varying on syntactic and semantic dimensions: An analysis with event-related potentials. *Memory & Cognition*, *25*, 477-494.

Mecklinger, A., & Ullsperger, P. (1993). P3 varies with stimulus categorization rather than probability. *Electroencephalography and Clinical Neurophysiology*, *86*, 395-407.

Mitchell, D. C. (2004). On-Line Methods in Language Processing: Introduction and Historical Review. In M. Carreiras & C. Clifton, Jr. (Eds.), *The on-line study of sentence comprehension: eyetracking, ERPs and beyond* (pp. 15-32). New York: Psychology Press.

Mitchell, D. C., Shen, X., Green, M. J., & Hodgson, T. L. (2008). Accounting for regressive exe-movements in models of sentence processing: A reappraisal of the Selective Reanalysis hypothesis. *Journal of Memory and Language*, *59*, 266-293.

Morris, R. K. (1994). Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 92-103.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*, 226-254.

Neely, J.H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner, & G.W. Humphreys (Eds.), *Basic progresses in reading - Visual word ecognition* (pp. 264-333). Hillsdale, NJ: Erlbaum.

Neville, H. J., Kutas, M., Chesney, G., & Schmidt, A. L. (1986). Event-related brain protentials during initial encoding and recognition memory of congruous and incongruous words. *Journal of Memory and Language*, *25*, 75-92.

Newman, R. L., Connolly, J. F., Service, E., & McIvor, K. (2003). Influence of phonological expectations during a phoneme deletion task: evidence from event-related brain potentials. *Psychophysiology*, *40*, 640-647.

Nobre, A. C., & McCarthy, G. (1994). Language-related ERPs - Scalp distributions and modulation by word type and semantic priming. *Journal of Cognitive Neuroscience*, *6*, 233-255.

Nuthmann, A. (2006). The "Where" and "When" of eye fixations in reading. Doctoral Dissertation. University of Potsdam.

Okita, T., Wijers, A. A., Mulder, G., & Mulder, L. J. (1985). Memory search and visual spatial attention: an event-related brain potential analysis. *Acta Psychologica*, *60*, 263-292.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97-113.

Osterhout, L., & Hagoort, P. (1999). A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King, and Kutas (1998) in the P600/SPS-P300 debate. *Language and Cognitive Processes, 14*, 1-14.

Osterhout, L., & Holcomb, P. J. (1992). Event-Related Brain Potentials Elicited by Syntactic Anomaly. *Journal of Memory and Language*, *31*, 785-806.

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

Osterhout, L., McKinnon, R., Bersick, M., & Corey, V. (1996). On the Language Specificity of the Brain Response to Syntactic Anomalies: Is the Syntactic Positive Shif a Member of the P300 Family? *Journal of Cognitive Neuroscience*, *8*, 507-526.

Parasuraman, R., & Beatty, J. (1980). Brain events underlying detection and recognition of weak sensory signals. *Science*, *210*, 80-83.

Paul, D. D., & Sutton, S. (1972). Evoked potential correlates of response criterion in auditory signal detection. *Science*, *177*, 362-364.

Perea, M., & Acha, J. (2009). Space information is important for reading. *Vision Research*, *49*, 1994-2000.

Pfefferbaum, A., Ford, J., Johnson, R. J., Wenegrat, B., & Kopell, B. S. (1983). Manipulation of P3 latency: speed vs. accuracy instructions. *Electroencephalography and Clinical Neurophysiology*, *55*, 188-197.

Pfefferbaum, A., & Ford, J. M. (1988). ERPs to stimli requiring response production and inhibition: Effects of age, probability and visual noise. *Electroencephalography and Clinical Neurophysiology*, *71*, 55-63.

Pfefferbaum, A., Ford, J. M., Weller, B. J., & Kopell, B. S. (1985). ERPs to response production and inhibition. *Electroencephalography and Clinical Neurophysiology*, *60*, 423-434.

Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology*, *9*, 456-479.

Picton, T. W., & Hillyard, S. A. (1974). Human auditory evoked potentials. II. Effects of attention. *Electroencephalography and Clinical Neurophysiology*, *36*, 191-199.

Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society, 363*, 1071-1086.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*, 2128-2148.

Pollatsek, A., Raney, G. E., Lagasse, L., & Rayner, K. (1993). The use of information below fixation in reading and in visual search. *Canadian Journal of Experimental Psychology*, *47*, 179-200.

Pollatsek, A., & Rayner, K. (1982). Eye movement control in reading The role of word boundaries. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 817-833.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyota Symposium* (pp. 55-85). Hillsdale, NJ: Erlbaum.

Primus, B. (1999). *Case and thematic roles. Ergative, accusative and active.* Tübingen: Max Niemeyer Verlag.

Pritchard, W. S. (1981). Psychophysiology of P300. *Psychological Bulletin*, *89*, 506-540.

Proverbio, A. M., Esposito, P., & Zani, A. (2002). Early involvement of the temporal area in attentional selection of grating orientation: an ERP study. *Cognitive Brain Research*, *13,* , 139-151.

Pynte, J., New, B., & Kennedy, A. (2008). On-line contextual influences during reading normal text: a multiple-regression analysis. *Vision Research*, *48*, 2172-2183.

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, *7*, 65-81.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372-422.

Rayner, K., & Clifton, C. J. (2009). Language processing in reading and speech perception is fast and incremental: implications for event-related potential research. *Biological Psychology*, *80*, 4-9.

Rayner, K., Fischer, M. H., & Pollatsek, A. (1998). Unspaced text interferes with both word identification and eye movement control. *Vision Research*, *38*, 1129-1144.

Rayner, K., Juhasz, B. J., & Brown, S. J. (2007). Do readers obtain preview benefit from word N + 2? A test of serial attention shift versus distributed lexical processing models of eye movement control in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 230-245.

Rayner, K., Li, X., Juhasz, B.J., & Yan, G. (2005). The effect of word predictability on the eye movements of Chinese readers. *Psychonomic Bulletin & Review, 12*, 1089-1093.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading.* Englewood Cliffs, NJ: Prentice Hall.

Rayner, K., & Pollatsek, A. (1996). Reading unspaced text is not easy: comments on the implications of Epelboim et al.'s (1994) study for models of eye movement control in reading. *Vision Research*, *36*, 461-470.

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, *21*, 448-465.

Rayner, K., & Sereno, S. C. (1994). Eye movements in reading: Psycholinguistic studies. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 57-81). London: Academic Press.

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*, 1290-1301.

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, *3*, 504-509.

Rayner, K., Well, A. D., Pollatsek, A., & Bertera, J. H. (1982). The availability of useful information to the right of fixation in reading. *Perception & Psychophysics*, *31*, 537-550.

Rayner, K., White, S. J., Kambe, G., Miller, B., & Liversedge, S. P. (2003). On the processing of meaning from parafoveal vision during eye fixations in reading. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The Mind's Eye: Cognitive and applied aspects of eye movement research* (pp. 213-234). Amsterdam: Elsevier.

Reichle, E. D., Liversedge, S. P., Pollatsek, A., & Rayner, K. (2009). Encoding multiple words simultaneously in reading is implausible. *Trends in Cognitive Sciences*, *13*, 115-119.

Reichle, E.D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader odel of eye movement control in reading: Comparison to other models. *Behavioral and Brain Sciences, 26*, 445-476.

Reichle, E.D., Pollatsek, A., & Rayner, K. (2006). EZ Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research*, *7*(1), 4-22.

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review, 16*, 1-21.

Reingold, E. M., & Rayner, K. (2006). Examining the word identification stages hypothesized by the E-Z Reader model. *Psychological Science*, *17*, 742-746.

Richter, E. M., Engbert, R., & Kliegl, R. (2006). Current advances in SWIFT. *Cognitive Systems Research*, *7*, 23-33.

Ritter, W., Simson, R., & Vaughan, H. G. J. (1972). Association cortex potentials and reaction time in auditory discrimination. *Electroencephalography and Clinical Neurophysiology*, *33*, 547-555.

Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., & Schlesewsky, M. (2007). To predict or not to predict: influences of task and strategy on the processing of semantic relations. *Journal of Cognitive Neuroscience*, *19*, 1259-1274.

Roehm, D., Schlesewsky, M., Bornkessel, I., Frisch, S., & Haider, H. (2004). Fractionating language comprehension via frequency characteristics of the human EEG. *Neuroreport*, *15*, 409-412.

Rolke, B., Heil, M., Streb, J., & Hennighausen, E. (2001). Missed prime words within the attentional blink evoke an N400 semantic priming effect. *Psychophysiology*, *38*, 165-174.

Rösler, F. (1983). Endogenous ERPs and cognition: Probes, prospects, and pitfalls in matching pieces of the mind-body puzzle. In A. W. K. Gaillard & W. Ritter (Eds.), *Tutorials in ERP Research: Endogenous Components* (pp. 9-35). Amsterdam: North-Holland Publishing Company.

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*, 150-176.

Scheepers, C., Hemforth, B., & Konieczny, L. (2000). Linking syntactic functions with thematic roles: Psych verbs and the resolution of subject-object ambiguity. In B. Hemforth & L. Konieczny (Eds.), *German sentence processing* (pp. 95-136). Dordrecht: Kluwer Academic Publishers.

Scheepers, C. (2005). *Thematic revision before the verb?* Paper presented at the Annual CUNY Conference on Sentence Processing, Tucson, Arizona.

Scheffers, M. K., Johnson, R. J., & Ruchkin, D. S. (1991). P300 in patients with unilateral temporal lobectomies: the effects of reduced stimulus quality. *Psychophysiology*, *28*, 274-284.

Schlesewsky, M., & Bornkessel, I. (2006). Context-sensitive neural responses to conflict resolutin: Electrophysiological evidence from subject-object ambiguities in language comprehension. *Brain Research, 1098*, 139-152.

Schlesewsky, M., Bornkessel, I., & Frisch, S. (2003). The neurophysiological basis of word order variations in German. *Brain and Language*, *86*, 116-128.

Schriefers, H., Friederici, A.D., & Kühn, K. (1995). The processing of locally ambiguous relative clauses in German. *Journal of Memory and Language, 34*, 499-520.

Sereno, S. C., Pacht, J. M., & Rayner, K. (1992). The effect of meaning frequency on processing lexically ambiguous words: Evidence from eye fixations. *Psychological Science*, *3*, 296-300.

Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: eye movements and event-related potentials. *Trends in Cognitive Sciences*, *7*, 489-493.

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.

Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience*, *8*, 1391-1400.

Sessa, P., Luria, R., Verleger, R., & Dell'Acqua, R. (2007). P3 latency shifts in the attentional blink: further evidence for second target processing postponement. *Brain Research*, *1137*, 131-139.

Sheatz, G. C., & Chapman, R. M. (1969). Task relevance and auditory evoked responses. *Electroencephalography and Clinical Neurophysiology*, *26*, 468-475.

Simola, J., Holmqvist, K., & Lindgren, M. (2009). Right visual field advantage in parafoveal processing: evidence from eye-fixation-related potentials. *Brain and Language*, *111*, 101-113.

Singhal, A., & Fowler, B. (2004). The differential effects of Sternberg short- and long-term memory scanning on the late Nd and P300 in a dual-task paradigm. *Brain Res Cogn Brain Research*, *21*, 124-132.

Smid, H.G.O.M., Lamain, W., Hogeboom, M.M., Mulder, G., & Mulder, L.J.M. (1991). Psychophysiological evidence for continuous information transmission between visual search and response processes. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 696-714.

Smulders, F. T., Kok, A., Kenemans, J. L., & Bashore, T. R. (1995). The temporal selectivity of additive factor effects on the reaction process revealed in ERP component latencies. *Acta Psychologica*, *90*, 97-109.

Sommer, W., Leuthold, H., & Matt, J. (1998). The expectancies that govern the P300 amplitude are mostly automatic and unconscious. *Behavioral and Brain Sciences*, *21*, 149-168.

Squires, K.C., Hillyard, S.A., Lindsay, P.H. (1973a). Cortical potentials evoked by confirming and disconfirming feedback following an auditory discrimination. *Perception & Psychophysics, 13*, 25-31.

Squires, K.C., Hillyard, S.A., Lindsay, P.H. (1973b). Vertex potentials evoked during auditory signal detection: Relation to decision criteria. *Perception & Psychophysics, 14*, 265-272.

Squires, K. C., Squires, N. K., & Hillyard, S. A. (1975a). Vertex evoked potentials in a rating-scale detection task: relation to signal probability. *Behavioral Biology*, *13*, 21-34.

Squires, K. C., Squires, N. K., & Hillyard, S. A. (1975b). Two varietes of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, *38*, 387-401.

Squires, K. C., Wickens, C., Squires, N. K., & Donchin, E. (1976). The effect of stimulus sequence on the waveform of the cortical event-related potential. *Science*, *193*(4258), 1142-1146.

Stabler, E.P. (1994). The finite connectivity of linguistic structure. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 303-336). Hillsdale, NJ: Erlbaum.

Starr, M. S., & Inhoff, A. W. (2004). Attention allocation to the right and left of a fixated word: Use of orthographic information from multiple words during reading. *European Journal of Cognitive Psychology*, *16*, 203-225.

Staub, A. (2007a). The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 550-569.

Staub, A. (2007b). The return of the repressed: Abandoned parses facilitate syntactic reanalysis. *Journal of Memory and Language, 57*, 299-323.

Staub, A., & Clifton, C. J. (2006). Syntactic prediction in language comprehension: evidence from either.or. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 425-436.

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

Staub, A., Rayner, K., Pollatsek, A., Hyona, J., & Majewski, H. (2007). The time course of plausibility effects on eye movements in reading: evidence from noun-noun compounds. J*ournal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 1162-1169.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, *150*, 1187-1188.

Tsai, J. L., McConkie, G.W. (2003). Where do Chinese reades send their eyes? In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The Mind's Eye: Cognitive and applied aspects of eye movement research* (pp. 159-176). Amsterdam: Elsevier.

Vainio, S., Hyönä, J., & Pajunen, A. (2009). Lexical predictability exerts robust effects on fixation duration, but not on initial landing position during reading. *Experimental Psychology*, *56*, 66-74.

Van den Brink, D., & Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPs. *Journal of Cognitive Neuroscience, 16*, 1068-1084.

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

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

Van Valin, R. D., Jr. (2005). *Exploring the Syntax-Semantics Interface.* Cambridge, NY: Camridge University Press.

Verleger, R. (1988). Event-related potentials and cognition: A critique of the context updating hypothesis and an alternative interpretation of P3. *Behavioral and Brain Sciences*, *11*, 343-427.

Verleger, R. (1997). On the utility of P3 latency as an index of mental chronometry. *Psychophysiology*, *34*, 131-156.

Verleger, R. (1998). Toward an integration of P3 research with cognitive neuroscience. *Behavioral and brain sciences*, *21*, 149-168.

Verleger, R., Jaskowski, P., & Wauschkuhn, B. (1994). Suspense and surprise: on the relationship between expectancies and P3. *Psychophysiology*, *31*, 359-369.

Vogel, E. K., & Luck, S. J. (2002). Delayed working memory consolidation during the attentional blink. *Psychonomic Bulletin & Review*, *9*, 739-743.

Vogel, E. K., Luck, S. J., & Shapiro, K. L. (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1656-1674.

Wang, C. A., Inhoff, A. W., & Radach, R. (2009). Is attention confined to one word at a time? The spatial distribution of parafoveal preview benefits during reading. *Attention, Perception, and Psychophysics*, *71*, 1487-1494.

Weisbrod, M., Kiefer, M., Winkler, S., Maier, S., Hill, H., Roesch-Ely, D. et al. (1999). Electrophysiological correlates of direct versus indirect semantic priming in normal volunteers. *Cognitive Brain Research*, *8*, 289-298.

White, S. J. (2008). Eye movement control during reading: Effects of word frequency and orthographic familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 205-223.

Wickens, C., Kramer, A., Vanasse, L., & Donchin, E. (1983). Performance of concurrent tasks: a psychophysiological analysis of the reciprocity of information-processing resources. *Science*, *221*, 1080-1082.

Wijers, A.A:, Okita, T., Mulder, G., Mulder, L.J.M., Lorist, M.M., Poiesz, R., & Scheffers, M.K. (1987). Visual search and spatial attention: ERPs in focussed and divided attention conditions. *Biological Psychology, 25*, 33-60.

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

Wunderlich, D. (1997). Cause and the structure of verbs. *Linguistic Inquiry*, *28*, 27-68.

Yan, G., Tian, H., Bai, X., & Rayner, K. (2006). The effect of word and character frequency on the eye movements of Chinese readers. *British Journal of Psychology*, *97*, 259-268.

Yan, M., Richter, E. M., Shu, H., & Kliegl, R. (2009). Readers of Chinese extract semantic information from parafoveal words. *Psychonomic Bulletin & Review*, *16*, 561-566.

Yang, J., Wang, S., Xu, Y., & Rayner, K. (2009). Do Chinese readers obtain preview benefit from word n + 2? Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1192-1204.

Yen, M. H., Radach, R., Tzeng, O. J., Hung, D. L., & Tsai, J. L. (2009). Early parafoveal processing in reading Chinese sentences. *Acta Psychologica*, *131*, 24-33.

Yen, M. H., Tsai, J. L., Tzeng, O. J. L., & Hung, D. L. (2008). Eye movements and parafoveal word processing in reading Chinese. *Memory & Cognition*, *36*, 1033.

Appendix

Appendix A: Supplementary single-fixation analysis for Experiment 2

One might contend that the first fixation measure is not the optimal point for averaging ERPS, as the first fixation can be the first of multiple fixations on a target word. First fixation durations could thus show processing efforts specifically linked to refixations on the target word. ERPs were averaged for the subset of trials in which the target was fixated only once to investigate whether there are differences in the ERP results depending on whether the target was refixated or not.

Results

Figure 20 gives an overview of the ERPs time-locked to the single fixation on the target, with the possibility of having multiple fixations in the pretarget region. The ERP waves differ from both the last fixation and the first fixation results with respect to the N400 effects reported above. There seems to be a slightly enhanced negativity between 190 and 300 ms for unrelated words at left-parietal and mid-parietal electrodes, while by and large the waves converge at all other electrodes. Interestingly, there is again a pronounced late positivity between 450 and 650 ms for unrelated targets, which is distributed mainly across central and parietal electrodes in both hemispheres. Additionally, related target words also exhibit a late positive wave between 650 and 900 ms post word onset, which is distributed over central and parietal electrodes in the right hemisphere. Statistical tests were carried out with the three time windows defined according to the visual inspection. The negativity was investigated within 190 ms and 300 ms, and the two positivities were examined in a time window ranging from 450 to 860 ms.

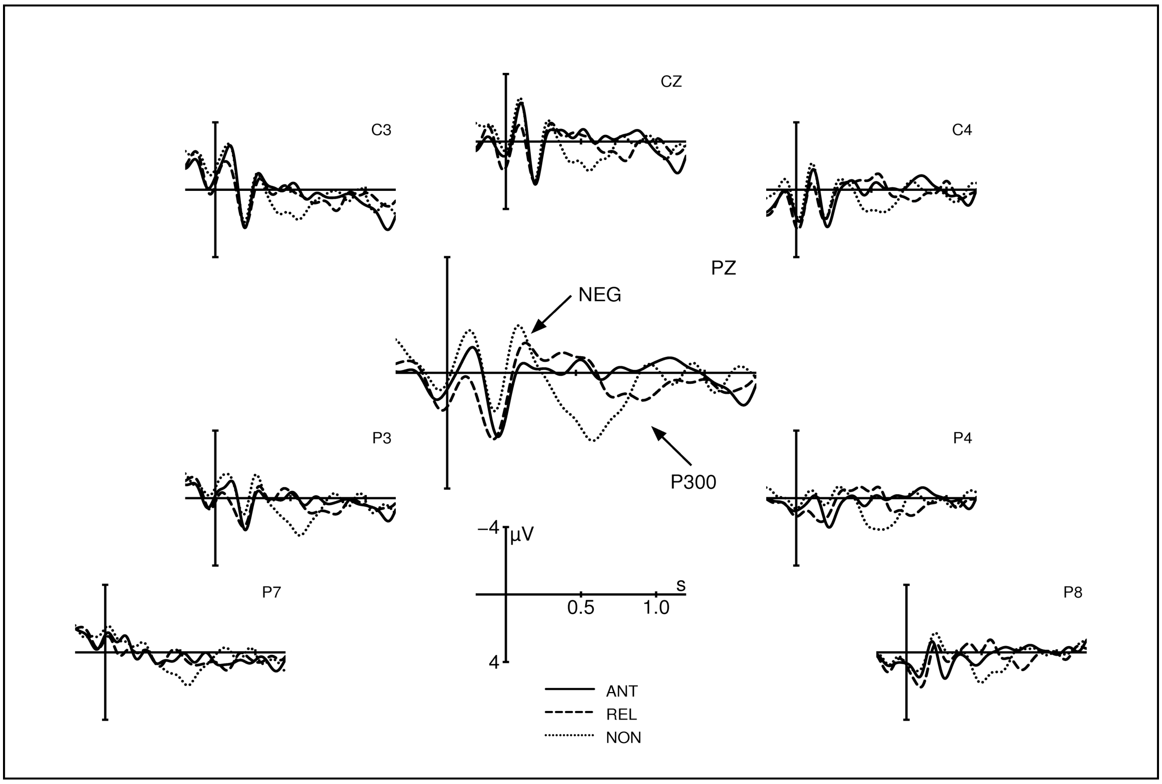


Figure 20. Grand-average ERPs (N = 48) time-locked to the onset of the single fixation on the target word (onset at the vertical bar).

Negativity time window: 190 ms – 300 ms

The analysis for midline electrodes registered no reliable effects (ROI: $F(1,47) = 2.92$, $p < .1$ COND: $F(2,94) = 2.43$, $p < .1$; interaction: $F(2,94) = 2.57$, $p < .1$). The analysis for lateral ROIs revealed a reliable main effect of ROI ($F(3,141) = 3.14$, $p < .05$) and a significant interaction between ROI and COND ($F(6,282) = 3.57$, $p < .01$). When the interaction was resolved by ROI, only electrodes at right-parietal sites showed a reliable effect of COND (left-central: F < 1; right-central: $F(2,94) = 2.34$, $p < .02$; left-parietal: $F(2,94) = 2.52$, $p < .09$; right-parietal: $F(2,94) = 6.33$, $p < .01$). Further single comparisons showed that antonyms differed from related and unrelated words at the right-parietal ROI (ANT vs. REL: $F(1,47) = 6.28$, $p < .02$; ANT vs. NON: $F(1,47) = 13.21$, $p < .0001$). Unrelated and related targets did not differ from each other ($F(1,47) = 1.42$, $p < .3$).

The single fixation analysis replicates the N400 effects found with the first fixation triggers. Both non-antonym conditions exhibited N400-like negativities that were increased in amplitude as opposed to the antonym condition. In contrast to the first fixation results, the onset latency of these effects was somewhat earlier and the effect size was less strong.

Late positivity time window: 450 ms – 860 ms

At midline sites, the ANOVAs revealed a fully significant main effect of COND ($F(2,94) = 9.55$, $p < .0001$), whereas the main effect of ROI and the interaction between both factors fell short of significance (ROI: $F(1,47) = 2.48$, $p < .2$; ROI x COND: $F(2,94) = 1.67$, $p < .2$). Further pair-wise comparisons for the factor COND revealed marginally reliable differences between antonyms and related non-antonyms ($F(1,47) = 4.57$, $p < .04$) and between related and unrelated non-antonyms (NON vs. REL: $F(1,47) = 4.59$, $p < .04$), whereas the contrast between antonyms and unrelated non-antonyms was significant (NON vs. ANT: $F(1,47) = 20.81$, $p < .001$).

At lateral electrode sites, there was a main effect of ROI ($F(3,141) = 9.33$, $p < .0001$) and a main effect of COND ($F(2,94) = 5.34$, $p <. 01$), as well as a marginal interaction between the two ($F(6,282) = 2.22$, $p < .07$). When this interaction was resolved by ROI, it revealed significant effects at all electrode sites except the left-central ROI (left-central: $F(2,94) = 1.48$, $p < .3$; right-central: $F(2,94) = 4.01$, $p < .02$; left-parietal: $F(2,94) = 4.17$, $p < .02$; right-parietal: $F(2,94) = 8.49$, $p < .0001$). Further single comparisons for the factor COND revealed no reliable contrast between antonyms and related non-antonyms (all Fs < 2.95, n.s.). Unrelated non-antonyms, by contrast, differed reliably from antonyms and, at the right-parietal ROI, marginally from related non-antonyms (NON vs. ANT: right-central: $F(1,47) = 7.12$, $p < .02$; left-parietal: $F(1,47) = 10.22$, $p < .01$; right-parietal: $F(1,47) = 20.25$, $p < .001$; NON vs. REL: right-central: $F(1,47) = 4.42$, $p < .05$; left-parietal: $F(1,47) = 2.16$, $p < .2$; $F(1,47) = 4.76$, $p < .04$). Altogether, these analyses confirm the presence of a late positive component in response to unrelated non-antonyms. Related non-antonyms also showed a late positive component, which was, however, smaller in amplitude and thus less reliable. Thus, these findings mirror the first-fixation analyses that revealed P300s for both non-antonym conditions.

Eye movement results

Single fixation on the target word

Single fixation duration confirmed the first fixation results, with antonyms eliciting the shortest fixation durations overall (ANT: 214 ms (38); REL: 240 ms (40); NON: 239 ms (47); Standard deviation in parentheses). The main effect of COND was highly significant ($F_1(2,94) = 16.18$, $p < .001$; $F_2(2,78) = 8.92$, $p < .001$) and single comparisons revealed that antonyms differed from both related and unrelated words, which, however, did not differ from each other (ANT vs. REL: $F_1(1,47) = 30.38$, $p < .001$; $F_2(1,39) = 12.42$, $p = .001$; ANT vs. NON: $F_1(1,47) = 18.31$, $p < .001$; $F_2(1,39) = 16.29$, $p < .001$; REL vs. NON: all Fs < 1). These eye movement results are in accordance with the ERP results, which showed a small N400-like negativity for the two non-antonym conditions.

Summary

Overall, the single fixation analyses showed no notable difference to the first fixation analysis. There were, however, latency shifts, especially for the N400 component. Its latency was earlier and somewhat shorter for ERPs triggered by the single fixation, as opposed to ERPs triggered by the first fixation. At present, it is difficult to give a sound functional interpretation of this shift. Although latency variability for the N400 is widely acknowledged, there are very few systematic investigations into the matter. Holcomb (1993), for example, investigated the influence of stimulus degradation on N400 latency and amplitude and reported no significant effects. The P300 effects were nearly identical across first and single fixation measures. The P300 to unrelated non-antonyms peaked earlier and was more pronounced in amplitude than the P300 to related non-antonyms. The statistical tests for the

P300 to related non-antonyms was, however, somewhat weaker in the single-fixation analysis than in the first-fixation analysis.

There also appears to be a converse relationship between ERPs and fixation results when both single and first fixations are considered. ERPs averaged to the first fixation on the target were more pronounced with respect to mean amplitude as opposed to ERPs time-locked to single fixations. In terms of fixation duration, the opposite pattern emerged in that the difference between antonyms and non-antonyms was larger in single fixation duration (with a difference of about 26 ms) than in first fixation duration (with a difference of about 18 ms)[50]. However, these findings need further elaboration, as fewer single-fixation trials entered the ERP analysis than did first-fixation trials. This may have contributed substantially to the relatively weaker effects in the single fixation analysis.

Appendix B: Supplementary eye movement analyses for the pretarget region in Experiment 2

A striking finding was that only the foveal N400 effect appeared to correlate with fixation durations, whereas the parafoveal N400 priming effect did not have a correlate in terms of reduced fixation durations for antonyms and related non-antonyms. However, this result must be viewed with some caution because the measure of last fixation averages across different types of fixations such as the single fixation, the first fixation, or the second or n^{th} fixation on a word or region of text. Since processing disruption may affect fixation positions in different ways, this measure is a less precise one. Similarly, it is not clear whether the second or n^{th} fixation on a word is driven by the linguistic manipulation under consideration or whether it is

[50] This was computed as the difference between the mean of the antonym condition and the mean collapsed across the two non-antonym conditions.

a consequence of the systematic range error of saccades (Rayner, 1998). The oculomotor system is biased towards an average saccade length of about 7 – 9 spaces, so refixations may also follow a corrective saccade resulting from a prior mislocated fixation. When the incoming saccade either undershoots or overshoots the critical word, the resulting first fixation on it is not located near the optimal viewing position and one or more corrective saccades are likely to follow (Nuthmann, 2006). To exclude the possibility that the null effect in the pretarget region is due to variance induced by fixations that are not associated with the parafoveal processing of the target word itself, the single fixation, first fixation, and gaze duration for the pretarget were computed as well.[51]

Results

Table 10 gives the additional measures for the pre-target region. As can be seen, there were no differences across conditions in the single fixation or first fixation measure for this region, which corresponds to the null effect in the last fixation (which is repeated here for convenience; for statistical reports, see chapter 4.2.2). There were, however, somewhat larger differences in gaze duration. The statistical tests are reported below.

First fixation on the pretarget region

There was no reliable effect of COND for the first fixation in the pretarget region ($F_1(2,94) = 2.07$, $p < .2$; $F_2(2,78) = 2.14$, $p < .2$).

---

[51] One may argue of course that single fixation or first fixation durations are not very informative for multi-word regions provided each word in a region is fixated at least once. Note, however, that the pretarget region (*das Gegenteil von* ‘the opposite of’) consisted of one content word and two very short function words. It is well known that function words and words with less than 4 letters are usually skipped because they can be already identified in the parafovea (cf. Rayner, 1998). Indeed, Table 6 suggest that (i) the initial landing position is close to the beginning of the content word *Gegenteil* (‘opposite’) and (ii) more importantly, the average launch site evidences that the preposition *von* (‘of’) is almost never fixated.

Single fixation on the pretarget region

Statistical tests yielded no significant main effect of COND for this region (all Fs < 1).

	Pretarget region *the opposite of*			
	Single fixation	First fixation	Gaze duration	Last fixation
Antonym (ANT)	217 (69)	211 (41)	397 (112)	184 (29)
Related (REL)	217 (63)	205 (38)	401 (135)	190 (27)
Unrelated (NON)	213 (38)	203 (29)	372 (89)	186 (24)

Table 10. Participant mean fixation durations (in ms) by condition. Standard deviation is given in parentheses.

Gaze duration on the pretarget region

The means suggest a disadvantage for related target non-antonyms, as opposed to both antonyms and unrelated non-antonyms, which, unexpectedly, exhibited the shortest gaze duration. This impression was partially supported by the statistical tests that revealed a main effect of COND ($F_1(2,94) = 3.89$, $p < .03$; $F_2(2,78) = 3.32$, $p < .05$). Pair-wise comparisons further showed that related non-antonyms differed from antonyms only in the by-subject analysis ($F_1(1,47) = 5.14$, $p = .03$; $F_2 < 1$) and reliably across subjects and items from unrelated non-antonyms ($F_1(1,47) = 5.66$, $p < .03$; $F_2(1,39) = 6.22$, $p < .02$), while there was no reliable difference between antonyms and unrelated non-antonyms ($F_1 < 1$; $F_2(1,39) = 3.82$, $p < .06$).

Summary

Overall, the single and first fixations in the pretarget region confirm the null effect found for the last fixation. The gaze duration effect, however, offers a paradox. Neither single fixation nor first fixation durations differed across conditions in the pretarget region.[52] Moreover, in a pre-test study using a subset of the present stimuli (i.e., 32 out of 40 items were used), there was no corresponding effect in gaze duration, whereas all the other effects for the pretarget region were fully replicated.[53] So, the effect in gaze duration seems to be somewhat arbitrary and hard to reconcile with the data from the other fixation measures, which are more consistent with each other.

Appendix C: Supplementary eye movement analyses for the target region in Experiment 2

The ERP waves time-locked to the first fixation on the critical word revealed that the P300 to unexpected related non-antonyms was reduced in amplitude and increased in latency. This effect was not mirrored in the duration of the first fixation. Eye movement measures that are presumed to reflect later processing stages in reading were examined to investigate whether the processes underlying the P300 might affect later eye movement measures. Other reasons for this investigation were that (i) the domain-general P300 followed the N400 reflecting linguistic processing proper and that (ii) related targets have led to increased reaction times and error rates in previous studies (see also Experiments 1 and 3). The P300 effect might then be visible in the eye-tracking record at a later time point. The eye movement measures total reading time, go past time, and first-pass regressions were examined for the target word, the word immediately following the target word, and the remainder of the sentence. Total reading time is defined as the total time that a reader spends on a given word or region of text,

[52] Note that the mean duration of the second fixation in the pretarget region showed results identical to single and first fixation. Across participants, the mean fixation duration per condition were: ANT: 178 ms, REL: 180 ms, NON: 176 ms (all *F*s < 1).
[53] Mean participant gaze duration (with standard deviation given in parentheses): ANT: 380 ms (88), REL: 370 ms (100), NON: 394 ms (104); main effect of COND: $F_1(2,78) = 1.94$, $p < .2$; $F_2(2,62) = 1.23$, $p < .3$.

including regressions. Go past time is the reading time from first encountering a word until first leaving it (following the canonical reading direction) – this includes re-reading time of earlier parts of the sentence. First-pass regressions reflect the proportion of trials in which the reader leaves the critical word with a regressive saccade. Note that while total time is seen as a general measure that reflects late processing stages, go past time and first-pass regressions can also index fairly early processes when they are measured for the critical region of text. For post-critical material, however, they can also reflect later processing stages (cf. Clifton et al., 2007). Table C.1 gives the mean fixation duration and percentage of regressions for the target word, the spillover word, and the remainder of the sentence. The regions were divided as given in example C.1 below.

(C.1) Regions of analysis

Dass/ schwarz/ das Gegenteil von/ weiß/ ist,/ hat Gertrud/ gestern gemeint.

That/ black/ the opposite of/ white/ is,/ has Gertrud/ yesterday claimed

‘Gertrud claimed yesterday that black is the opposite of white.’

The ANOVAs were carried out treating participants and items as random factors F_1 and F_2, respectively. The factor entering the ANOVA was COND (condition; ANT vs. REL vs. NON). Whenever pair-wise comparisons were justified, the respective alpha was set to $p < .033$, according to the modified Bonferroni procedure (Keppel, 1991).

Results

Table 11 shows the mean participant reading times and percent regressions per condition for the critical regions.

Target word

On the target word, both reading time measures revealed a main effect of COND (go past time: $F_1(2,94) = 19.13$, $p < .001$; $F_2(2,78) = 17.83$, $p < .001$; total time: $F_1(2,94) = 47.12$, $p <$

.001; $F_2(2,78) = 21.27$, $p < .001$). In *go past time*, this effect was due to longer fixation durations for the non-antonyms as compared to the antonyms. The contrast between related and unrelated non-antonyms was, however, not significant (ANT vs. REL: $F_1(1,47) = 36.75$, $p < .001$; F_2(1,39= 31.65, $p < .001$; ANT vs. NON: $F_1(1,47) = 25.00$, $p < .001$; $F_2(1,39) = 28.59$, $p < .001$; REL vs. NON: $F_1(1,47) = 1.00$, $p < .4$; $F_2 < 1$). As for *total reading time*, the contrasts between antonyms and all non-antonyms were also highly significant, while the 24-ms difference between related and unrelated non-antonyms only approached significance in the analysis by participants (ANT vs. REL: $F_1(1,47) = 76.92$, $p < .001$; $F_2(1,39) = 38.40$, $p < .001$; ANT vs. NON: $F_1(1,47) = 66.50$, $p < .001$; $F_2(1,39) = 22.31$, $p < .001$; REL vs. NON: $F_1(1,47) = 5.09$, $p < .03$; $F_2(1,39) = 2.52$, $p < .2$). The analysis for the *first-pass regressions* revealed no reliable main effect of COND for this region ($F_1(2,94) = 2.58$, $p < .09$; $F_2(2,78) = 3.97$, $p < .03$).

	Target ***weiß***	**Spillover** ***ist***	**Main clause 1** ***hat Gertrud***	**Main clause 2** ***gestern gemeint***
Go past time				
Antonyms (ANT)	264 (54)	204 (88)	315 (77)	887 (370)
Related (REL)	340 (98)	254 (75)	354 (87)	1054 (415)
Unrelated (NON)	326 (111)	239 (71)	326 (74)	961 (457)
First pass regressions				
Antonyms (ANT)	5 (8)	2 (4)	1 (3)	28 (18)
Related (REL)	8 (10)	6 (9)	3 (5)	41 (22)
Unrelated (NON)	9 (15)	4 (6)	2 (5)	36 (22)
Total time				
Antonyms (ANT)	264 (67)	194 (53)	377 (120)	718 (250)
Related (REL)	357 (97)	248 (71)	428 (121)	788 (292)
Unrelated (NON)	333 (94)	228 (69)	391 (107)	742 (257)

Table 11. Participant mean reading times (in ms) and percent regressions by condition and region. Standard deviation is given in parentheses

Spillover word

Go past time revealed a main effect of COND ($F_1(2,94) = 6.54$, $p < .003$; $F_2(2,78) = 7.32$, $p = .001$), resulting from increased reading times for the non-antonym conditions, as opposed to the antonyms. Again, the contrast between related and unrelated words was not significant (ANT vs. REL: $F_1(1,47) = 10.68$, $p < .01$; $F_2(1,39) = 13.92$, $p = .001$; ANT vs. NON: $F_1(1,47) = 5.35$, $p < .03$; $F_2(1,39) = 5.28$, $p < .03$; REL vs. NON: $F_1(1,47) = 1.59$, $p < .3$; $F_2(1,39) = 2.21$, $p < .2$). The same pattern was registered for the *total reading time* in this region (main effect of COND: $F_1(2,94) = 22.05$, $p < .001$; $F_2(2,78) = 7.84$, $p = .001$). However, beside a significant contrast between antonyms and both non-antonyms conditions, the comparison between related and unrelated non-antonyms showed a modest trend toward longer reading times for related non-antonyms (ANT vs. REL: $F_1(1,47) = 38.83$, $p < .001$; $F_2(1,39) = 11.86$, $p = .001$; ANT vs. NON: $F_1(1,47) = 35.21$, $p < .001$; $F_2(1,39) = 8.87$, $p < .01$; REL vs. NON: $F_1(1,47) = 4.45$, $p = .04$; $F_2(1,39) = 2.15$, $p < .2$). The *first-pass regressions* evidenced a main effect of COND ($F_1(2,94) = 5.90$, $p < .01$; $F_2(2,78) = 6.22$, $p < .01$) that resulted from a significantly higher number of regressions out of this region, for the related non-antonym condition as compared to the antonym condition. The remaining pair-wise comparisons revealed only marginally reliable effects (ANT vs. REL: $F_1(1,47) = 9.12$, $p < .01$; $F_2(1,39) = 12.48$, $p = .001$; ANT vs. NON: $F_1(1,47) = 5.27$, $p < .03$; $F_2(1,39) = 2.96$, $p < .1$; REL vs. NON: $F_1(1,47) = 2.51$, $p < .2$; $F_2(1,39) = 3.26$, $p < .08$).

Main clause 1

The main effect of COND was highly significant for *go past time* ($F_1(2,94) = 10.46$, $p < .001$; $F_2(2,78) = 6.42$, $p < .01$). This was due to significantly longer reading times for related non-antonyms, compared to unrelated non-antonyms and antonyms, which did not differ from each other (ANT vs. REL: $F_1(1,47) = 15.93$, $p < .001$; $F_2(1,39) = 10.46$, $p < .01$; ANT vs. NON: $F_1(1,47) = 1.89$, $p < .2$; $F_2(1,39) = 1.00$, $p < .4$; REL vs. NON: $\mathrm{F}_1(1,47) = 10.17$, $p < .01$; $F_2(1,39) = 6.46$, $p < .02$). *Total reading times* confirmed this pattern, in that the main effect was highly significant ($F_1(2,94) = 6.30$, $p < .01$; $F_2(2,78) = 4.64$, $p < .02$). The contrast between antonyms and related non-antonyms was also reliable, whereas the comparison of antonyms with unrelated non-antonyms was not. The critical comparison between related and unrelated non-antonyms was significant by participants, but only without the Bonferroni adjustment was it significant in the by-item analysis (ANT vs. REL: $F_1(1,47) = 9.04$, $p < .01$; $F_2(1,39) = 7.06$, $p < .02$; ANT vs. NON: $F_1(1,47) = 1.13$, $p < .3$; $F_2(1,39) = 1.05$, $p < .4$; REL vs. NON: $F_1(1,47) = 7.58$, $p < .01$; $F_2(1,39) = 4.38$, $p < .05$). Finally, *first-pass regressions* also revealed a significant main effect of COND ($F_1(2,94) = 5.21$, $p < .01$; $F_2(2,78) = 3.35$, $p < .05$). Pair-wise comparisons, however, registered only one reliable contrast between antonyms and related non-antonyms (ANT vs. REL: $F_1(1,47) = 11.30$, $p < .01$; $F_2(1,39) = 5.96$, $p < .02$; ANT vs. NON: $F_1(1,47) = 3.77$, $p < .06$; $F_2(1,39) = 2.34$, $p < .2$; REL vs. NON: $F_1(1,47) = 2.04$, $p < .2$; $F_2(1,39) = 1.55$, $p < .3$).

Main clause 2

In the final part of the sentence, *go past time* evidenced a reliable main effect of COND ($F_1(2,94) = 7.25$, $p < .002$; $F_2(2,78) = 3.71$, $p < .03$). Pair-wise comparisons revealed further that antonyms differed reliably from related non-antonyms, whereas the other comparisons were not statistically significant – although the by-participant analyses showed very modest tendencies for unrelated non-antonyms to differ from antonyms, and for related non-antonyms to induce longer reading times than unrelated non-antonyms (ANT vs. REL: $F_1(1,47) = 15.46$, $p < .001$; $F_2(1,39) = 7.07$, $p < .02$; ANT vs. NON: $F_1(1,47) = 4.01$, $p < .06$; $F_2(1,39) = 2.08$, $p < .157$; REL vs. NON: $F_1(1,47) = 3.29$, $p < .08$; $F_2(1,39) = 1.77$, $p < .2$). In *total reading time*, the main effect of COND did not yield any fully significant results, even though the reading times mirrored the numerical trends found in go past time ($F_1(2,94) = 3.71$, $p < .03$; $F_2(2,78) = 2.11$, $p < .2$). *First pass regressions*, by contrast, revealed a highly significant main effect of COND ($F_1(2,94) = 9.87$, $p < .001$; $F_2(2,78) = 5.94$, $p < .01$) that was due to the fact that antonyms elicited fewer regressions than either related or unrelated non-antonyms. The two non-antonym conditions did not reliably differ from one another (ANT vs. REL: $F_1(1,47) =$

18.51, $p < .001$; $F_2(1,39) = 9.98$, $p < .01$; ANT vs. NON: $F_1(1,47) = 10.51$, $p < .01$; $F_2(1,39) = 8.43$, $p < .01$; REL vs. NON: $F_1(1,47) = 2.15$, $p < .2$; $F_2(1,39) = 1.30$, $p < .3$).

Summary

In sum, these data show a two-fold processing pattern. On the target word and the immediately following word, antonyms are read faster relative to all non-antonyms. This mirrors the single and first fixation data. There are moreover small trends towards longer total reading times in these regions for related non-antonyms as opposed to unrelated non-antonyms. This disadvantage for related non-antonyms becomes fully visible in the first part of the main clause, where they exhibit the longest reading times and the highest proportion of first-pass regressions. The effect for unexpected related non-antonyms could thus be connected to the difficulty in judging whether these words are in fact antonyms – and hence to the P300 effect triggered by the first fixation on the related non-antonym. The fact that total time shows this effect slightly earlier than go past time is attributable to the special nature of the go-past time measure, which encompasses first pass time of the region under investigation and re-reading time in earlier parts of the sentence, leading to a more heterogeneous data pattern than in total time. Although the effects in total time appear to be smeared across the regions of text, the overall pattern suggests that readers were puzzled by the unexpected related non-antonyms, and consequently re-read the critical word and later parts of the sentence more often than in the other conditions. This pattern is remarkably similar to the increase in reaction times and error rates for unexpected related non-antonyms that has been reported previously (Roehm et al., 2007). In light of these findings, the P300 amplitude and latency may very well be predictors of a participant's eye-movement record involving late measures of eye movements in reading.

Appendix D: Supplementary ERP analyses for Experiment 3

Mean N400 amplitude in the time window between 330 and 380 ms

The time window was chosen based on visual inspection of Figure 5, which suggested a small amplitude difference between the related and unrelated non-antonym conditions. Factors entering the statistical analyses are described in chapter 4.3.1. The midline analyses registered two significant main effects and a significant interaction (ROI: $F(2,94) = 4.41$, $p < .04$; COND: $F(2,94) = 19.38$, $p < .001$; ROI x COND: $F(4,188) = 16.21$, $p < .01$). After resolving the interaction for ROI, single comparisons showed reliable effects at central and parietal electrode sites and a marginal effect at the frontal electrode site (frontal: $F(2,94) = 3.08$, $p < .06$; central: $F(2,94) = 23.63$, $p < .001$; parietal: $F(2,94) = 26.66$, $p < .001$). Pair-wise comparisons involving the factor COND revealed that antonyms differed reliably from both related and unrelated targets across central and parietal electrode sites, where the latter did not differ from each other (ANT vs. REL: frontal: F(1,47) = 2.82, p < .1; central: $F(1,47) = 26.21$, $p < .001$; parietal: $F(1,47) = 30.19$, $p < .001$; ANT vs. NON: frontal: $F(1,47) = 4.93$, $p < .032$; central: $F(1,47) = 33.01$, $p < .001$; parietal: $F(1,47) = 37.76$, $p < .001$; REL vs. NON: frontal: $F < 1$; central: $F(1,47) = 2.68$, $p < .2$; parietal: $F(1,47) = 2.85$, $p < .1$).

Identical effects were observed for the lateral ROIs. The main effects and the interaction were highly reliable (ROI: $F(5,235) = 13.73$, $p < .001$; COND: $F(2,94) = 18.05$, $p < .001$; ROI x COND: $F(10,470) = 16.35$, $p < .001$). The interaction was due to significant contrasts at central and parietal ROIs, whereas frontal ROIs showed no significant effects (left-frontal: $F < 1$; right-frontal: $F < 1$; left-central: $F(2,94) = 20.94$, $p < .001$; right-central: $F(2,94) = 17.88$, $p < .001$; left-parietal: $F(2,94) = 29.36$, $p < .001$; right-parietal: $F(2,94) = 26.29$, $p < .001$). Pair-wise comparisons revealed that antonyms again differed significantly from related and unrelated targets, which did not differ from each other (ANT vs. REL: left-central: $F(1,47) = 18.46$, $p < .0001$; right-central: $F(1,47) = 21.07$, $p < .001$; left-parietal: $F(1,47) =$

30.02, $p < .001$; right-parietal: $F(1,47) = 28.73$, $p < .001$; ANT vs. NON: left-central: $F(1,47) = 33.78$, $p < .001$; right-central: $F(1,47) = 25.83$, $p < .001$; left-parietal: $F(1,47) = 42.52$, $p < .001$; right-parietal: $F(1,47) = 35.94$, $p < .001$; REL vs. NON.: left-central: $F(1,47) = 3.55$, $p < .07$; right-central: $F < 1$; left-parietal: $F(1,47) = 3.40$, $p < .08$; right-parietal: $F(1,47) = 2.24$, $p < .2$).

The statistical tests showed no reliable difference in N400 amplitude between related and unrelated non-antonyms. As expected, both non-antonym conditions differed reliably from antonyms, which elicited a P300 in this epoch.

Peak amplitude analysis for the N400 time window between 240 and 430 ms

A larger time window was chosen for the peak amplitude analysis to assure that a peak could be found per participant and condition. Factors entering the statistical analyses are described in chapter 4.3.1.The mean peak amplitudes per condition are given in Table 12 below.

	Mean peak amplitude (µV)
Antonyms (ANT)	-1.24 (2.53)
Related (REL)	-3.12 (2.40)
Unrelated (NON)	-3.55 (2.22)

Table 12. Mean peak amplitudes for the N400 in Experiment 3. Standard deviation is given in parentheses.

The analyses for peak amplitudes revealed a pattern similar to the mean amplitudes. At midline electrode sites, the main effect of COND and the interaction with ROI yielded significance (ROI: $F(2,94) = 2.67$, $p < .1$; COND: $F(2,94) = 11.23$, $p < .0001$; $F(4,188) = 10.49$, $p < .001$). After resolving the interaction by ROI, only central and parietal electrodes registered significant effects of COND (frontal: $F(2,94) = 1.03$, $p < .4$; central: $F(2,94) = 10.65$, $p < .0001$; parietal: $F(2,94) = 21.24$, $p < .001$). Pair-wise comparisons involving the factor COND revealed reliable contrasts between antonyms and each of the non-antonym conditions, whereas the latter did not differ in their peak amplitude (ANT vs. REL: central: $F(1,47) = 9.60$, $p < .01$; parietal: $F(1,47) = 22.53$, $p < .001$; ANT vs. NON: central: $F(1,47) = 16.96$, $p < .0001$; parietal: $F(1,47) = 35.11$, $p < .001$; REL vs. NON: central: $F(1,47) = 2.45$, $p < .2$; parietal: $F(1,47) = 1.74$, $p < .2$).

At lateral electrode site, all main effects as well as their interaction were significant (ROI: $F(5,235) = 6.00$, $p < .0001$; COND: $F(2,94) = 12.12$, $p < .001$; ROI x COND: $F(10,470) = 9.48$, $p < .001$). The resolution of the interaction revealed significant contrasts involving the factor COND for all ROIs except the two frontal ROIs (left-frontal: $F < 1$; right-frontal: $F < 1$; left-central: F(2,94) = 10.55, p < .0001; right-central: $F(2,94) = 9.96$, $p < .0001$; left-parietal: $F(2,94) = 23.16$, $p < .001$; right-parietal: $F(2,94) = 18.78$, $p < .001$). Further single comparisons showed that antonyms differed from the related and unrelated conditions, while the related and unrelated non-antonym conditions did not differ significantly from each other (ANT vs. REL: left-central: $F(1,47) = 8.18$, $p < .01$; right-central: $F(1,47) = 13.06$, $p < .0001$; left-parietal: $F(1,47) = 26.46$, $p < .001$; right-parietal: $F(1,47) = 20.43$, $p < .001$; ANT vs. NON: left-central: $F(1,47) = 15.13$, $p < .0001$; right-central: $F(1,47) =14.71$, $p < .0001$; left-parietal: $F(1,47) = 31.96$, $p < .001$; right-parietal: $F(1,47) = 28.89$, $p < .001$; REL vs. NON: left-central: $F(1,47) = 4.13$, $p < .05$; right-central: $F < 1$; left-parietal: $F(1,47) = 1.44$, $p < .3$; right-parietal: $F(1,47) = 1.17$, $p < .3$).

As is apparent, the analysis for the N400 peak amplitudes also failed to find a reliable difference between the related and unrelated non-antonym conditions, while both of these differed significantly from the antonym condition.

Appendix E: Supplementary analyses for the post-critical region in Experiment 5

ERP results

Figure 21 gives the grand-average ERPs triggered by the first fixation on the past participle of the dative verb. Both dative-initial conditions seem to be more negative going from fixation onset until about 200 ms. This effect, however, appears to stem from the last fixation before the verb, particularly in the dative-initial condition with two animate NPs. There was no difference between the conditions in the N400 time window. The dative-initial conditions differed from the nominative-initial conditions between 400 ms and 520 ms post fixation-onset in being more positive. Between about 580 and 720 ms post fixation-onset the conditions with two animate NPs appear to be more positive-going than the other conditions (with a more pronounced positivity for the object-initial condition with two animate NPs).

Figure 22 shows the grand-average ERPs time-locked to the last fixation before the past participle of the dative verb. Dative-initial clauses with two animate NPs begin to be more negative by about 250 ms post fixation onset and this negativity lasts until about 600 ms. This effect was most visible at midline electrode sites and the following right-lateral electrodes. The nominative-initial condition with two animate NPs is more negative between about 580 and 720 ms, especially at midline electrode sites.

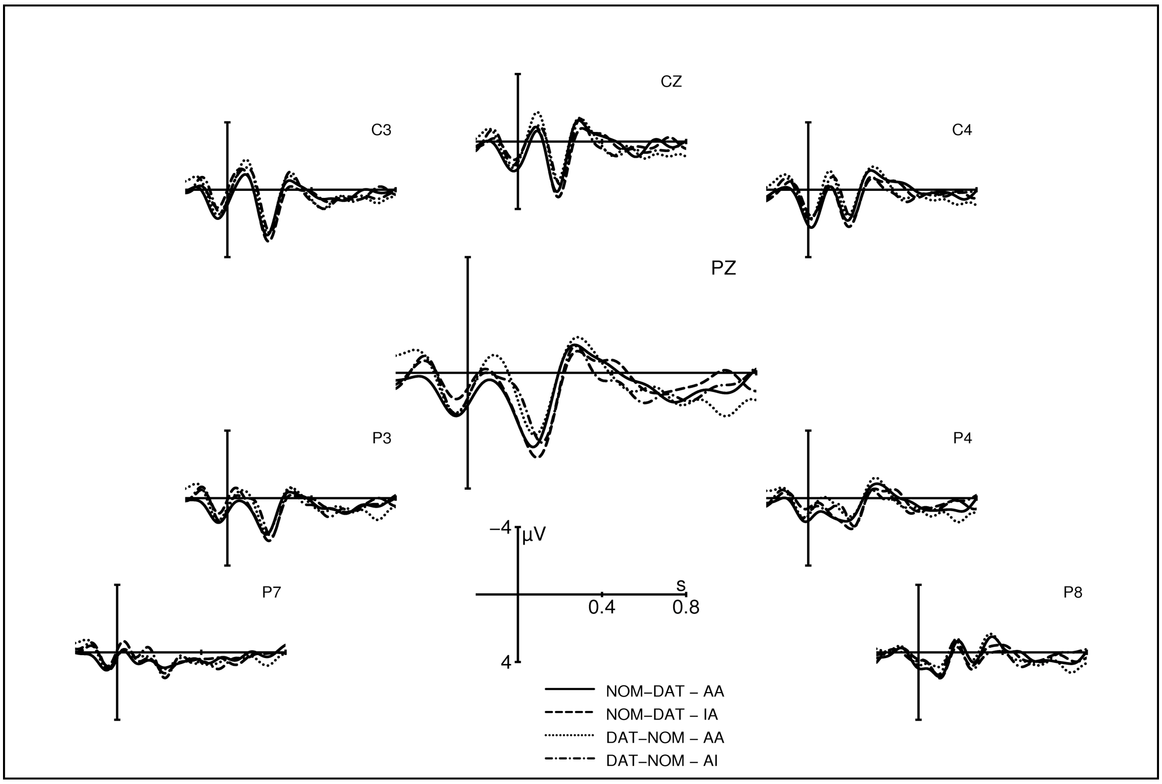


Figure 21. Grand-average ERPs (N = 48) time-locked to the onset of the first fixation on the dative verb (onset at the vertical bar).

Two aspects of the last fixation trigger are noteworthy. First, the increased negativity for the dative-initial clauses with two animate arguments is remarkably similar to the foveal N400 effect found for NP2. Since the second NP immediately precedes the past participle, one cannot exclude the possibility that ERPs averaged for the last fixation before the past participle coincides with ERPs averaged for the first fixation on NP2. Another argument against the reliability of these last-fixation effects is that, at a functional level, they would entail that the logical structure of the verb must be inferred from parafoveal information. Although Experiment 5 exhibited a clear effect of parafoveal semantic processing, this effect was found for a word that does not induce processing effort itself. This scenario, however, does not apply for the last fixation before the verb, which is located on NPs inducing processing difficulty. The foveal N400 effect for the second NP shows that the second NP leads to processing costs when it is an animate nominative argument. Thus, based on the finding that increased foveal load seems to inhibit parafoveal ERP effects, it is rather unlikely that the last fixation before the verb truly indicates semantic processing.

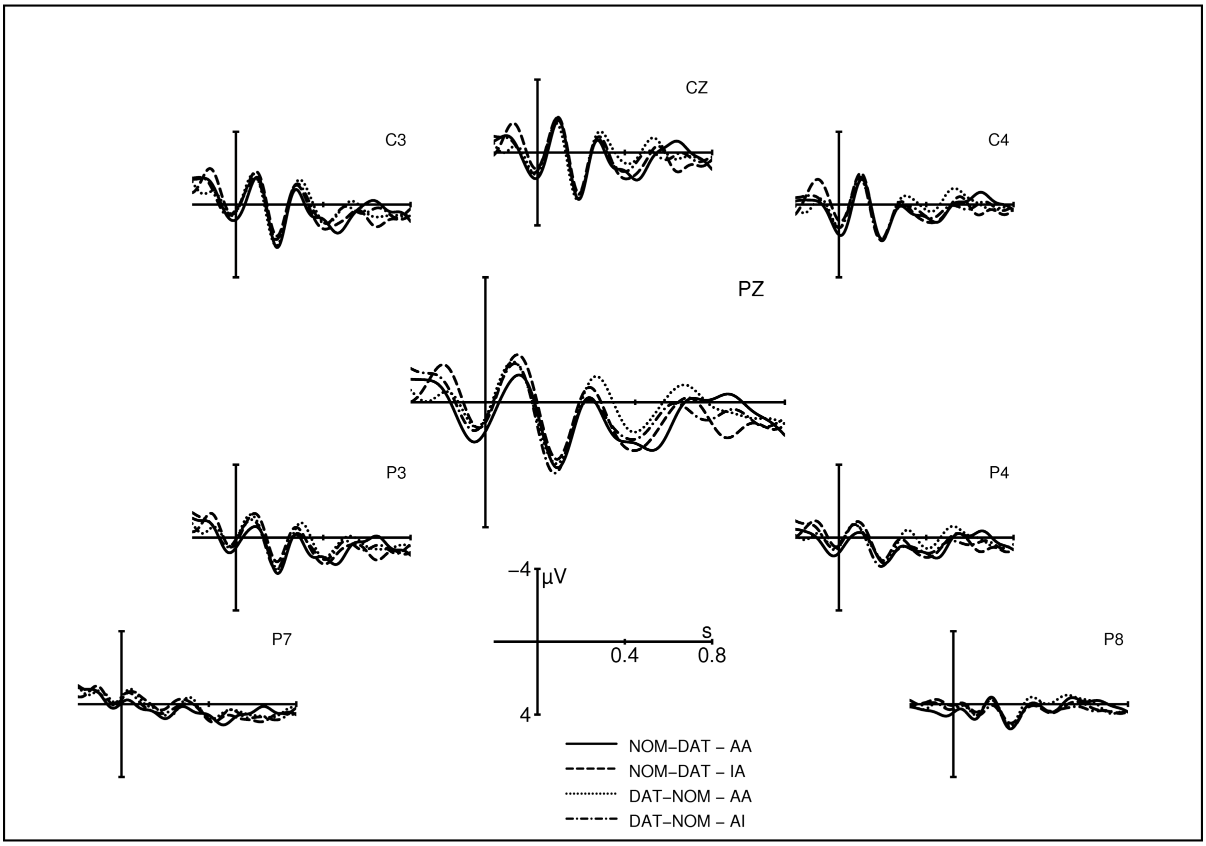


Figure 22. Grand-average ERPs (N = 48) time-locked to the last fixation before the dative verb (onset at the vertical bar).

The second noteworthy aspect refers to the divergence between parafoveal and foveal effects for the past participle. The ERPs triggered by the last fixation before the verb participle show a late negativity, which was, however, not found for the first fixation on the verb participle. All the preceding analyses revealed that such late ERP effects are most often linked to the subsequent fixation, not to the current one (see especially Experiment 2). This can be taken to suggest that the last fixation trigger, as defined in this thesis, may capture electrophysiological activity that cannot be unequivocally related to a specific fixation position. Specifically, the last fixation can be located on any word preceding the region under investigation, so that the late negativity might also be related to the foveal N400 effect found for NP2. In consideration of these aspects, a sound interpretation of the present parafoveal ERP effects appears impossible. The following statistical report is therefore limited to the ERPs averaged for the first fixation on the past participle of the verb.

The pair-wise comparisons were adjusted according to the modified Bonferroni procedure (Keppel, 1991). Since the same comparisons were made as for NP2, alpha was set to $p < .012$. To avoid type I errors in the ERP results resulting from violations of sphericity, alpha was corrected as proposed by Huynh and Feldt (1970) when effects with more than one degree of freedom in the numerator were evaluated. In these cases, the original degrees of freedom are reported with the corrected alpha. The eye movement results were statistically analyzed by using the same factors and pair-wise comparisons as described above. Participants and items were treated as random factors F_1 and F_2, respectively.

First fixation analyses

N400 time window: 250 ms – 400 ms

NOM-DAT – AA vs. NOM-DAT – IA

There were no reliable effects for the midline ROIs (all *F*s < 1). At lateral ROIs, the main effect of ROI was reliable, while the main effect of ANIM and the interaction between the two factors were not reliable (ROI: $F(3,141) = 19.25$, $p < .001$; all other *F*s < 1.85, n.s.).

DAT-NOM – AA vs. DAT-NOM – AI

There were no reliable effects for the midline ROIs (all *F*s < 1.72, n.s.). At lateral electrode sites, the main effect of ROI and the interaction between ROI and ANIM were significant (ROI: $F(3,141) = 8.64$, $p < .0001$; ANIM: $F(1,47) = 1.97$, $p < .2$; ROI x ANIM: $F(3,141) = 6.60$, $p < .01$). Resolving this interaction showed a significant effect of ANIM for the right-central ROI (left-central: $F < 1$; right-central: $F(1,47) = 13.21$, $p < .0001$; left-parietal: $F < 1$; right-parietal: $F(1,47) = 4.96$, $p < .04$). This effect, however, appeared to come from eye movement artifacts in the EEG (see Appendix F for ERP figures including EOG electrodes).

AA vs. AI

There were no reliable effects for the midline ROIs (all *F*s < 1.11, n.s.). At lateral ROIs, the main effect of ROI was fully reliable, but the main effect of ANIM only approached significance (ROI: $F(3,141) = 19.05$, $p < .001$; ANIM: $F(11,47) = 3.03$, $p < .09$). The interaction was reliable (ROI: $F(3,141) = 7.32$, $p < .0001$). Resolving the interaction by ROI showed significant effects of ANIM in the right-lateral ROIs (left-central: $F < 1$; right-central: $F(1,47) = 18.23$, $p < .0001$; left-parietal: $F < 1$; right-parietal: $F(1,47) = 6.80$, $p < .02$). These effects, however, appeared to come from eye movement artifacts in the EEG (see Appendix F for ERP figures including EOG electrodes).

NOM-DAT – AA vs. DAT-NOM – AA

There were no reliable effects for the midline ROIs (all *F*s < 1). At lateral ROIs, only the main effect of ROI was significant (ROI: $F(3,141) = 21.19$, $p < .001$; all other *F*s < 1).

Late positivity time window I: 400 ms – 520 ms

NOM-DAT – AA vs. NOM-DAT – IA

There were no reliable effects for the midline ROIs (all *F*s < 1.12, n.s.). Only the main effect of ROI was reliable across lateral electrode sites (ROI: $F(3,141) = 10.79$, $p < .001$; all other *F*s < 1.30, n.s.).

DAT-NOM – AA vs. DAT-NOM – AI

There were no reliable effects for the midline ROIs (all *F*s < 1). Only the main effect of ROI approached significance across lateral electrode sites (ROI: $F(3,141) = 4.35$, $p < .02$; all other *F*s < 1).

AA vs. AI

The analysis for the midline ROIs registered no reliable effects (all *F*s < 1). Across lateral electrode sites, only the main effect of ROI was significant (ROI: $F(3,141) = 9.74$, $p < .0001$; all other *F*s < 1).

NOM-DAT – AA vs. DAT-NOM – AA

There were no reliable effects at midline ROIs (all *F*s < 2.49, n.s.). At lateral ROIs, only the main effect of ROI was significant (ROI: $F(3,141) = 9.16$, $p < .0001$; all other *F*s < 1.75, n.s.).

Late positivity time window II: 580 ms – 720 ms

NOM-DAT – AA vs. NOM-DAT – IA

The main effect of ROI and the interaction between ROI and ANIM were reliable for the midline ROIs, while the main effect of ANIM was not (ROI: $F(1,47) = 7.84$, $p < .01$; ANIM: $F < 1$; ROI x ANIM: $F(1,47) = 4.11$, $p < .05$). There were no significant effects after the interaction was resolved by ROI (all *F*s < 2.04, n.s.). There were no reliable effects for the lateral ROIs (all *F*s < 2.09, n.s.).

DAT-NOM – AA vs. DAT-NOM – AI

There were no reliable effects for either the midline ROIs (ROI: $F(1,47) = 3.32$, $p < .08$; all other *F*s < 1.28) or the lateral ROIs (all *F*s < 2.33, n.s.) .

AA vs. AI

The main effect of ROI and the interaction between ROI and ANIM were significant across midline electrode sites (ROI: $F(1,47) = 6.77$, $p < .02$; COND: $F(1,47) = 1.18$, $p < .3$; ROI x ANIM: $F(1,47) = 4.81$, $p < .04$). Resolving the interaction by ROI revealed no reliable effects (all *F*s < 2.79, n.s.). At lateral ROIs, only the interaction between ROI and ANIM approached significance. When this interaction was resolved by ROI, only the right-parietal ROI showed an effect of ANIM, which was due to a small positive peak in the conditions with two animate NPs (right-parietal: $F(1,47) = 4.17$, $p < .05$; all other Fs < 1).

NOM-DAT – AA vs. DAT-NOM – AA

Only the main effect of ROI was significant across midline electrode sites (ROI: $F(1,47) = 12.18$, $p < .01$; all other Fs < 2.79, n.s.). At lateral ROIs, only the main effect of ROI was marginally significant (ROI: $F(3,141) = 4.73$, $p < .013$; all other Fs < 2.35, n.s.)

Eye movement results

Table 13 gives the mean fixation durations and positions for the past participle of the verb. The fixation positions for which ERPs were analyzed are reported first.

Past participle of the dative verb		
	First Fixation	Total time
NOM-DAT – AA	275 (54)	458 (146)
NOM-DAT – IA	265 (51)	438 (142)
DAT-NOM – AA	269 (44)	449 (119)
DAT-NOM – AI	256 (46)	412 (133)
AA	272 (44)	453 (122)
AI	261 (45)	425 (132)
	Last fixation	Launch site
NOM-DAT – AA	223 (46)	6.3 (2)
NOM-DAT – IA	220 (33)	6.3 (2)
DAT-NOM – AA	216 (39)	6.5 (2)
DAT-NOM – AI	207 (36)	6.4 (2)
AA	219 (39)	6.4 (1)
AI	214 (30)	6.3 (2)
	First pass time	Landing position
NOM-DAT – AA	364 (103)	3.7 (1)
NOM-DAT – IA	353 (94)	3.5 (1)
DAT-NOM – AA	363 (104)	3.6 (1)
DAT-NOM – AI	344 (104)	3.7 (1)
AA	364 (96)	3.6 (1)
AI	349 (95)	3.6 (1)

Table 13. Participant mean fixation durations (in ms) and positions (in characters) by condition. Standard deviation is given in parentheses.

NOM-DAT – AA vs. NOM-DAT – IA

The main effect of ANIM failed to reach significance for the duration of the *first fixation* ($F_1(1,47) = 2.14$, $p < .2$; $F_2(1,39) = 2.27$, $p < .2$) and the *last fixation* (Fs < 1). *First pass time* ($F_1(1,47) = 1.88$, $p < .2$; $F_2(1,39) = 1.43$, $p < .3$) and *total time* ($F_1(1,47) = 2.86$, $p < .1$; $F_2(1,39) = 1.92$, $p < .2$) were unaffected by the animacy manipulation. There ere no significant differences across *landing positions* ($F_1(1,47) = 1.28$, p < .3; $F_2 < 1$) or *launch sites* (all Fs < 1).

DAT-NOM – AA vs. DAT-NOM – AI

In the *first fixation* measure, there was a 13-ms difference between the two dative-initial conditions, which resulted in a significant main effect of ANIM ($F_1(1,47) = 6.16$, $p < .02$; $F_2(1,39) = 5.71$, $p = .022$). Clauses with two animate NPs showed increased fixation durations on the participle. The duration of the *last fixation* before the verb did not reveal a reliable effect ($F_1(1,47) = 2.27$, $p < .2$; $F_2(1,39) = 1.94$, $p < .2$). *First pass time* showed a small increase for the condition with two animate NPs, which was marginally significant across participants ($F_1(1,47) = 5.14$, $p = .028$; $F_2(1,39) = 3.49$, $p < .07$). In *total time*, two animate NPs led to longer total reading times in this region (449 ms vs. 412 ms; $F_1(1,47) = 10.26$, $p < .01$; $F_2(1,39) = 5.16$, $p = .029$). The animacy manipulation did not influence the *landing position* or *launch site* measure (all Fs < 1).

AA vs. AI

The duration of the *first fixation* was longer for two animate NPs relative two NPs differing in animacy, as reflected in a significant main effect of ANIM ($F_1(1,47) = 6.28$, $p < .02$; $F_2(1,39) = 7.85$, $p < .01$). The animacy manipulation did not influence duration of the *last fixation* before the verb ($F_1(1,47) = 1.81$, $p < .2$; $F_2(1,39) = 1.21$, $p < .3$). *First pass time* on the participle was increased by 15 ms for clauses with two animate NPs ($F_1(1,47) = 7.23$, $p = .01$;

$F_2(1,39) = 5.61$, $p = .023$). The 28-ms disadvantage for two animate NPs also reached significance in *total time* ($F_1(1,47) = 12.21$, $p = .001$; $F_2(1,39) = 6.73$, $p < .02$). The *landing position* and *launch site* measures showed no significant result (all *F*s < 1).

NOM-DAT – AA vs. DAT-NOM – AA

The main effect of ORDER failed to reach significance in *first fixation* (all *F*s < 1), *last fixation* ($F_1(1,47) = 1.75$, $p < .2$; $F_2(1,39) = 3.07$, $p < .09$), first pass time (all *F*s < 1), and in *total time* (all *F*s < 1). There were no differences for the *landing position* or the *launch site* (all *F*s < 1).

Summary

The present analyses reveal that there were no reliable ERP effects for the past participle of the verb following the critical NPs. All reading time measures (except the last fixation) revealed a systematic disadvantage for two animate NPs in the dative-initial conditions; this finding extends the marginal influence of animacy on reading times found for NP2 (see chapter 7.2.3). Overall, these data strongly suggest that tracking ERP effects for post-critical parts of the sentence is difficult, because (i) parafoveal processing effects are possibly contaminated by foveal effects from preceding regions of text, and because (ii) fixation positions further downstream the sentence appear to give rise to stronger smearing effects, which make it difficult to detect ERP effects clearly.

Appendix F: Stimulus materials for Experiments 1 – 5

Experiment 1

Antonym

1. Das Gegenteil von groß ist klein
2. Das Gegenteil von krank ist gesund
3. Das Gegenteil von schwarz ist weiss
4. Das Gegenteil von lang ist kurz
5. Das Gegenteil von ungenau ist genau
6. Das Gegenteil von schön ist hässlich
7. Das Gegenteil von leer ist voll
8. Das Gegenteil von stark ist schwach
9. Das Gegenteil von arm ist reich
10. Das Gegenteil von leicht ist schwer
11. Das Gegenteil von lebendig ist tot
12. Das Gegenteil von billig ist teuer
13. Das Gegenteil von faul ist fleißig
14. Das Gegenteil von geschlossen ist offen
15. Das Gegenteil von freundlich ist unfreundlich
16. Das Gegenteil von links ist rechts
17. Das Gegenteil von oben ist unten
18. Das Gegenteil von hinten ist vorne
19. Das Gegenteil von hier ist dort
20. Das Gegenteil von hell ist dunkel
21. Das Gegenteil von glänzend ist matt
22. Das Gegenteil von waagerecht ist senkrecht
23. Das Gegenteil von schmutzig ist sauber
24. Das Gegenteil von kalt ist warm
25. Das Gegenteil von gebraucht ist neu
26. Das Gegenteil von lieben ist hassen
27. Das Gegenteil von schnell ist langsam
28. Das Gegenteil von nass ist trocken
29. Das Gegenteil von nackt ist bekleidet
30. Das Gegenteil von süss ist sauer
31. Das Gegenteil von weich ist hart
32. Das Gegenteil von betrunken ist nüchtern
33. Das Gegenteil von krumm ist gerade
34. Das Gegenteil von schlecht ist gut
35. Das Gegenteil von Feuer ist Wasser
36. Das Gegenteil von tolerant ist intolerant
37. Das Gegenteil von klug ist dumm
38. Das Gegenteil von Nacht ist Tag
39. Das Gegenteil von Himmel ist Erde
40. Das Gegenteil von Kind ist Erwachsener
41. Das Gegenteil von Ankunft ist Abfahrt
42. Das Gegenteil von geben ist nehmen
43. Das Gegenteil von anziehen ist ausziehen
44. Das Gegenteil von öffnen ist schließen

45. Das Gegenteil von einschalten ist ausschalten
46. Das Gegenteil von zuknöpfen ist aufknöpfen
47. Das Gegenteil von eingraben ist ausgraben
48. Das Gegenteil von Abend ist Morgen
49. Das Gegenteil von Ausgang ist Eingang
50. Das Gegenteil von Abrüstung ist Aufrüstung
51. Das Gegenteil von passiv ist aktiv
52. Das Gegenteil von jung ist alt
53. Das Gegenteil von Amateur ist Profi
54. Das Gegenteil von Anfang ist Ende
55. Das Gegenteil von vertraut ist fremd
56. Das Gegenteil von Berg ist Tal
57. Das Gegenteil von bergauf ist bergab
58. Das Gegenteil von dick ist dünn
59. Das Gegenteil von aufmachen ist zumachen
60. Das Gegenteil von Krieg ist Frieden
61. Das Gegenteil von Freund ist Feind
62. Das Gegenteil von Sommer ist Winter
63. Das Gegenteil von getrennt ist zusammen
64. Das Gegenteil von weinen ist lachen
65. Das Gegenteil von schlafend ist wach
66. Das Gegenteil von immer ist nie
67. Das Gegenteil von Mann ist Frau
68. Das Gegenteil von vorher ist nachher
69. Das Gegenteil von alles ist nichts
70. Das Gegenteil von Osten ist Westen
71. Das Gegenteil von ja ist nein
72. Das Gegenteil von trinken ist essen
73. Das Gegenteil von Vorspeise ist Nachspeise
74. Das Gegenteil von Fußboden ist Decke
75. Das Gegenteil von dein ist mein
76. Das Gegenteil von ledig ist verheiratet
77. Das Gegenteil von schreiben ist lesen
78. Das Gegenteil von minus ist plus
79. Das Gegenteil von rund ist eckig
80. Das Gegenteil von Regen ist Sonne

Related Non-Antonym

1. Das Gegenteil von groß ist dick
2. Das Gegenteil von krank ist kränklich
3. Das Gegenteil von schwarz ist gelb
4. Das Gegenteil von lang ist hoch
5. Das Gegenteil von ungenau ist vage
6. Das Gegenteil von schön ist normal
7. Das Gegenteil von leer ist halbvoll
8. Das Gegenteil von stark ist halbstark
9. Das Gegenteil von arm ist zufrieden
10. Das Gegenteil von leicht ist mittelschwer
11. Das Gegenteil von lebendig ist halbtot

12. Das Gegenteil von billig ist preiswert
13. Das Gegenteil von faul ist bemüht
14. Das Gegenteil von geschlossen ist angelehnt
15. Das Gegenteil von freundlich ist höflich
16. Das Gegenteil von links ist mittig
17. Das Gegenteil von oben ist dazwischen
18. Das Gegenteil von hinten ist zentral
19. Das Gegenteil von hier ist nebenan
20. Das Gegenteil von hell ist trüb
21. Das Gegenteil von glänzend ist streifig
22. Das Gegenteil von waagerecht ist diagonal
23. Das Gegenteil von schmutzig ist staubig
24. Das Gegenteil von kalt ist lau
25. Das Gegenteil von gebraucht ist benutzt
26. Das Gegenteil von lieben ist mögen
27. Das Gegenteil von schnell ist bedächtig
28. Das Gegenteil von nass ist feucht
29. Das Gegenteil von nackt ist halbnackt
30. Das Gegenteil von süss ist bitter
31. Das Gegenteil von weich ist elastisch
32. Das Gegenteil von betrunken ist beschwipst
33. Das Gegenteil von krumm ist eckig
34. Das Gegenteil von schlecht ist mäßig
35. Das Gegenteil von Feuer ist Wind
36. Das Gegenteil von tolerant ist ambivalent
37. Das Gegenteil von klug ist schlau
38. Das Gegenteil von Nacht ist Abend
39. Das Gegenteil von Himmel ist Luft
40. Das Gegenteil von Kind ist Jugendlicher
41. Das Gegenteil von Ankunft ist Hinfahrt
42. Das Gegenteil von geben ist behalten
43. Das Gegenteil von anziehen ist umziehen
44. Das Gegenteil von öffnen ist zuhalten
45. Das Gegenteil von einschalten ist betreiben
46. Das Gegenteil von zuknöpfen ist abreißen
47. Das Gegenteil von eingraben ist umgraben
48. Das Gegenteil von Abend ist Mittag
49. Das Gegenteil von Ausgang ist Durchgang
50. Das Gegenteil von Abrüstung ist Umrüstung
51. Das Gegenteil von passiv ist träge
52. Das Gegenteil von jung ist mittelalt
53. Das Gegenteil von Amateur ist Experte
54. Das Gegenteil von Anfang ist Mitte
55. Das Gegenteil von vertraut ist bekannt
56. Das Gegenteil von Berg ist Hügel
57. Das Gegenteil von bergauf ist voran
58. Das Gegenteil von dick ist füllig
59. Das Gegenteil von aufmachen ist verpacken
60. Das Gegenteil von Krieg ist Konflikt
61. Das Gegenteil von Freund ist Kollege
62. Das Gegenteil von Sommer ist Herbst

63. Das Gegenteil von getrennt ist angrenzend
64. Das Gegenteil von weinen ist grinsen
65. Das Gegenteil von schlafend ist dösend
66. Das Gegenteil von immer ist manchmal
67. Das Gegenteil von Mann ist Kind
68. Das Gegenteil von vorher ist jetzt
69. Das Gegenteil von alles ist manches
70. Das Gegenteil von Osten ist Süden
71. Das Gegenteil von ja ist vielleicht
72. Das Gegenteil von trinken ist schlürfen
73. Das Gegenteil von Vorspeise ist Hauptgericht
74. Das Gegenteil von Fußboden ist Wand
75. Das Gegenteil von dein ist unser
76. Das Gegenteil von ledig ist geschieden
77. Das Gegenteil von schreiben ist hören
78. Das Gegenteil von minus ist null
79. Das Gegenteil von rund ist gerade
80. Das Gegenteil von Regen ist Nebel

Unrelated Non-Antonym

1. Das Gegenteil von groß ist grün
2. Das Gegenteil von krank ist wellig
3. Das Gegenteil von schwarz ist nett
4. Das Gegenteil von lang ist flott
5. Das Gegenteil von ungenau ist lieblich
6. Das Gegenteil von schön ist müde
7. Das Gegenteil von leer ist braun
8. Das Gegenteil von stark ist stumpf
9. Das Gegenteil von arm ist breit
10. Das Gegenteil von leicht ist doof
11. Das Gegenteil von lebendig ist hohl
12. Das Gegenteil von billig ist munter
13. Das Gegenteil von faul ist rosa
14. Das Gegenteil von geschlossen ist hager
15. Das Gegenteil von freundlich ist viereckig
16. Das Gegenteil von links ist bunt
17. Das Gegenteil von oben ist oval
18. Das Gegenteil von hinten ist klebrig
19. Das Gegenteil von hier ist spitz
20. Das Gegenteil von hell ist grob
21. Das Gegenteil von glänzend ist zärtlich
22. Das Gegenteil von waagerecht ist traurig
23. Das Gegenteil von schmutzig ist salzig
24. Das Gegenteil von kalt ist sanft
25. Das Gegenteil von gebraucht ist geizig
26. Das Gegenteil von lieben ist haarig
27. Das Gegenteil von schnell ist holzig
28. Das Gegenteil von nass ist lila
29. Das Gegenteil von nackt ist blind

30. Das Gegenteil von süss ist kahl
31. Das Gegenteil von weich ist zornig
32. Das Gegenteil von betrunken ist verbrannt
33. Das Gegenteil von krumm ist ängstlich
34. Das Gegenteil von schlecht ist schlank
35. Das Gegenteil von Feuer ist Sessel
36. Das Gegenteil von tolerant ist farbenfroh
37. Das Gegenteil von klug ist kühl
38. Das Gegenteil von Nacht ist Hund
39. Das Gegenteil von Himmel ist Ofen
40. Das Gegenteil von Kind ist Ratte
41. Das Gegenteil von Ankunft ist Durchfall
42. Das Gegenteil von geben ist schnarchen
43. Das Gegenteil von anziehen ist hinfallen
44. Das Gegenteil von öffnen ist wachsen
45. Das Gegenteil von einschalten ist zuhören
46. Das Gegenteil von zuknöpfen ist abwaschen
47. Das Gegenteil von eingraben ist durchlesen
48. Das Gegenteil von Abend ist Treppe
49. Das Gegenteil von Ausgang ist Weinglas
50. Das Gegenteil von Abrüstung ist Federung
51. Das Gegenteil von passiv ist kursiv
52. Das Gegenteil von jung ist rot
53. Das Gegenteil von Amateur ist Hausschuh
54. Das Gegenteil von Anfang ist Pflanze
55. Das Gegenteil von vertraut ist kantig
56. Das Gegenteil von Berg ist Maus
57. Das Gegenteil von bergauf ist musisch
58. Das Gegenteil von dick ist morsch
59. Das Gegenteil von aufmachen ist kitzeln
60. Das Gegenteil von Krieg ist Osterei
61. Das Gegenteil von Freund ist Birne
62. Das Gegenteil von Sommer ist Plastik
63. Das Gegenteil von getrennt ist gekocht
64. Das Gegenteil von weinen ist fliegen
65. Das Gegenteil von schlafend ist schimmelnd
66. Das Gegenteil von immer ist fröhlich
67. Das Gegenteil von Mann ist Stuhl
68. Das Gegenteil von vorher ist talwärts
69. Das Gegenteil von alles ist oder
70. Das Gegenteil von Osten ist Kissen
71. Das Gegenteil von ja ist Topf
72. Das Gegenteil von trinken ist lügen
73. Das Gegenteil von Vorspeise ist Atommüll
74. Das Gegenteil von Fußboden ist Umfrage
75. Das Gegenteil von dein ist neun
76. Das Gegenteil von ledig ist chemisch
77. Das Gegenteil von schreiben ist flattern
78. Das Gegenteil von minus ist Dativ
79. Das Gegenteil von rund ist blutig
80. Das Gegenteil von Regen ist Chinese

Experiments 2 and 3

Antonym

1. Dass krank das Gegenteil von gesund ist, hat Grete heute gemeint
2. Dass lebendig das Gegenteil von tot ist, hat Detlef neuerdings behauptet
3. Dass passiv das Gegenteil von aktiv ist, hat Norbert gestern geglaubt
4. Dass schwarz das Gegenteil von weiß ist, hat Antje neulich vermutet
5. Dass gebraucht das Gegenteil von neu ist, hat Axel heute gemutmaßt
6. Dass nass das Gegenteil von trocken ist, hat Gertrud neuerdings gemeint
7. Dass krumm das Gegenteil von gerade ist, hat Karsten gestern behauptet
8. Dass geschlossen das Gegenteil von offen ist, hat Ilse neulich geglaubt
9. Dass schlafend das Gegenteil von wach ist, hat Roland heute vermutet
10. Dass männlich das Gegenteil von weiblich ist, hat Paula neuerdings gemutmaßt
11. Dass links das Gegenteil von rechts ist, hat Erwin gestern gemeint
12. Dass ungenau das Gegenteil von genau ist, hat Silke neulich behauptet
13. Dass nackt das Gegenteil von bekleidet ist, hat Peter heute geglaubt
14. Dass sehend das Gegenteil von blind ist, hat Marga neuerdings vermutet
15. Dass leer das Gegenteil von voll ist, hat David gestern gemutmaßt
16. Dass ledig das Gegenteil von verheiratet ist, hat Judith neulich gemeint
17. Dass süss das Gegenteil von sauer ist, hat Günther heute behauptet
18. Dass schnell das Gegenteil von langsam ist, hat Marlies neuerdings geglaubt
19. Dass dick das Gegenteil von dünn ist, hat Friedrich gestern vermutet
20. Dass kalt das Gegenteil von warm ist, hat Sandra neulich gemutmaßt
21. Dass groß das Gegenteil von klein ist, hat Vincent heute gemeint
22. Dass freundlich das Gegenteil von unfreundlich ist, hat Hanna neuerdings behauptet
23. Dass schön das Gegenteil von hässlich ist, hat Siegfried gestern geglaubt
24. Dass tolerant das Gegenteil von intolerant ist, hat Tanja neulich vermutet
25. Dass klug das Gegenteil von dumm ist, hat Wolfram heute gemutmaßt
26. Dass hell das Gegenteil von dunkel ist, hat Maren neuerdings gemeint
27. Dass lang das Gegenteil von kurz ist, hat Sascha gestern behauptet
28. Dass schlecht das Gegenteil von gut ist, hat Lena neulich geglaubt
29. Dass leicht das Gegenteil von schwer ist, hat Erich heute vermutet
30. Dass faul das Gegenteil von fleißig ist, hat Inge neuerdings gemutmaßt
31. Dass arm das Gegenteil von reich ist, hat Robert gestern gemeint
32. Dass alt das Gegenteil von jung ist, hat Heike neulich behauptet
33. Dass stark das Gegenteil von schwach ist, hat Stefan heute geglaubt
34. Dass billig das Gegenteil von teuer ist, hat Daniel neuerdings vermutet
35. Dass schmutzig das Gegenteil von sauber ist, hat Kristina gestern gemutmaßt
36. Dass weich das Gegenteil von hart ist, hat Andrea neulich gemeint
37. Dass rund das Gegenteil von eckig ist, hat Klaas heute behauptet
38. Dass glänzend das Gegenteil von matt ist, hat Felix neuerdings geglaubt
39. Dass getrennt das Gegenteil von zusammen ist, hat Holger gestern vermutet
40. Dass vertraut das Gegenteil von fremd ist, hat Sonja neulich gemutmaßt

Related Non-Antonym

1. Dass krank das Gegenteil von fit ist, hat Grete heute gemeint
2. Dass lebendig das Gegenteil von erschöpft ist, hat Detlef neuerdings behauptet

3. Dass passiv das Gegenteil von träge ist, hat Norbert gestern geglaubt
4. Dass schwarz das Gegenteil von gelb ist, hat Antje neulich vermutet
5. Dass gebraucht das Gegenteil von benutzt ist, hat Axel heute gemutmaßt
6. Dass nass das Gegenteil von feucht ist, hat Gertrud neuerdings gemeint
7. Dass krumm das Gegenteil von eckig ist, hat Karsten gestern behauptet
8. Dass geschlossen das Gegenteil von angelehnt ist, hat Ilse neulich geglaubt
9. Dass schlafend das Gegenteil von dösend ist, hat Roland heute vermutet
10. Dass männlich das Gegenteil von sexuell ist, hat Paula neuerdings gemutmaßt
11. Dass links das Gegenteil von mittig ist, hat Erwin gestern gemeint
12. Dass ungenau das Gegenteil von vage ist, hat Silke neulich behauptet
13. Dass nackt das Gegenteil von halbnackt ist, hat Peter heute geglaubt
14. Dass sehend das Gegenteil von kurzsichtig ist, hat Marga neuerdings vermutet
15. Dass leer das Gegenteil von halbleer ist, hat David gestern gemutmaßt
16. Dass ledig das Gegenteil von geschieden ist, hat Judith neulich gemeint
17. Dass süss das Gegenteil von bitter ist, hat Günther heute behauptet
18. Dass schnell das Gegenteil von lahm ist, hat Marlies neuerdings geglaubt
19. Dass dick das Gegenteil von breit ist, hat Friedrich gestern vermutet
20. Dass kalt das Gegenteil von lau ist, hat Sandra neulich gemutmaßt
21. Dass groß das Gegenteil von dick ist, hat Vincent heute gemeint
22. Dass freundlich das Gegenteil von höflich ist, hat Hanna neuerdings behauptet
23. Dass schön das Gegenteil von normal ist, hat Siegfried gestern geglaubt
24. Dass tolerant das Gegenteil von kritisch ist, hat Tanja neulich vermutet
25. Dass klug das Gegenteil von schlau ist, hat Wolfram heute gemutmaßt
26. Dass hell das Gegenteil von trüb ist, hat Maren neuerdings gemeint
27. Dass lang das Gegenteil von hoch ist, hat Sascha gestern behauptet
28. Dass schlecht das Gegenteil von mäßig ist, hat Lena neulich geglaubt
29. Dass leicht das Gegenteil von komplex ist, hat Erich heute vermutet
30. Dass faul das Gegenteil von bemüht ist, hat Inge neuerdings gemutmaßt
31. Dass arm das Gegenteil von wohlhabend ist, hat Robert heute gemeint
32. Dass alt das Gegenteil von erwachsen ist, hat Heike neuerdings behauptet
33. Dass stark das Gegenteil von kräftig ist, hat Stefan gestern geglaubt
34. Dass billig das Gegenteil von günstig ist, hat Daniel neulich vermutet
35. Dass schmutzig das Gegenteil von staubig ist, hat Kristina heute gemutmaßt
36. Dass weich das Gegenteil von innen ist, hat Andrea neuerdings gemeint
37. Dass rund das Gegenteil von gerade ist, hat Klaas gestern behauptet
38. Dass glänzend das Gegenteil von gestreift ist, hat Felix neulich geglaubt
39. Dass getrennt das Gegenteil von nah ist, hat Holger heute vermutet
40. Dass vertraut das Gegenteil von bekannt ist, hat Sonja neuerdings gemutmaßt

Unrelated Non-Antonym

1. Dass krank das Gegenteil von froh ist, hat Grete gestern gemeint
2. Dass lebendig das Gegenteil von hohl ist, hat Detlef neulich behauptet
3. Dass passiv das Gegenteil von kursiv ist, hat Norbert heute geglaubt
4. Dass schwarz das Gegenteil von nett ist, hat Antje neuerdings vermutet
5. Dass gebraucht das Gegenteil von komisch ist, hat Axel gestern gemutmaßt
6. Dass nass das Gegenteil von lila ist, hat Gertrud neulich gemeint
7. Dass krumm das Gegenteil von ängstlich ist, hat Karsten heute behauptet
8. Dass geschlossen das Gegenteil von kompliziert ist, hat Ilse neuerdings geglaubt
9. Dass schlafend das Gegenteil von traurig ist, hat Roland gestern vermutet

10. Dass männlich das Gegenteil von ärgerlich ist, hat Paula neulich gemutmaßt
11. Dass links das Gegenteil von bunt ist, hat Erwin heute gemeint
12. Dass ungenau das Gegenteil von lieblich ist, hat Silke neuerdings behauptet
13. Dass nackt das Gegenteil von blind ist, hat Peter gestern geglaubt
14. Dass sehend das Gegenteil von tropisch ist, hat Marga neulich vermutet
15. Dass leer das Gegenteil von braun ist, hat David heute gemutmaßt
16. Dass ledig das Gegenteil von chemisch ist, hat Judith neuerdings gemeint
17. Dass süss das Gegenteil von kahl ist, hat Günther gestern behauptet
18. Dass schnell das Gegenteil von laut ist, hat Marlies neulich geglaubt
19. Dass dick das Gegenteil von heiß ist, hat Friedrich heute vermutet
20. Dass kalt das Gegenteil von sanft ist, hat Sandra neuerdings gemutmaßt
21. Dass groß das Gegenteil von grün ist, hat Vincent gestern gemeint
22. Dass freundlich das Gegenteil von praktisch ist, hat Hanna neulich behauptet
23. Dass schön das Gegenteil von müde ist, hat Siegfried heute geglaubt
24. Dass tolerant das Gegenteil von künstlich ist, hat Tanja neuerdings vermutet
25. Dass klug das Gegenteil von kühl ist, hat Wolfram gestern gemutmaßt
26. Dass hell das Gegenteil von grob ist, hat Maren neulich gemeint
27. Dass lang das Gegenteil von nervös ist, hat Sascha heute behauptet
28. Dass schlecht das Gegenteil von schlank ist, hat Lena neuerdings geglaubt
29. Dass leicht das Gegenteil von brutal ist, hat Erich gestern vermutet
30. Dass faul das Gegenteil von rosa ist, hat Inge neulich gemutmaßt
31. Dass arm das Gegenteil von breit ist, hat Robert heute gemeint
32. Dass alt das Gegenteil von leise ist, hat Heike neuerdings behauptet
33. Dass stark das Gegenteil von wahr ist, hat Stefan gestern geglaubt
34. Dass billig das Gegenteil von munter ist, hat Daniel neulich vermutet
35. Dass schmutzig das Gegenteil von scharf ist, hat Kristina heute gemutmaßt
36. Dass weich das Gegenteil von zornig ist, hat Andrea neuerdings gemeint
37. Dass rund das Gegenteil von gut ist, hat Klaas gestern behauptet
38. Dass glänzend das Gegenteil von zärtlich ist, hat Felix neulich geglaubt
39. Dass getrennt das Gegenteil von gekocht ist, hat Holger heute vermutet
40. Dass vertraut das Gegenteil von lustig ist, hat Sonja neuerdings gemutmaßt

Experiment 4

Subject-before-object clause, two animate NPs (SO-AA)

1. Dass Heinrich Nachbarn besucht, hat alle sehr verwundert
2. Dass Johannes Touristen betreut, hat manche oft erbost
3. Dass Benjamin Fotografen bewertet, hat jeden immer überrascht
4. Dass Christoph Franzosen verfolgt, hat einige meist verärgert
5. Dass Alexandra Chinesen versteht, hat viele wirklich verwundert
6. Dass Albrecht Ingenieure bemerkt, hat alle stark erbost
7. Dass Christine Passanten anschaut, hat manche sehr überrascht
8. Dass Elisabeth Christen erkennt, hat jeden oft verärgert
9. Dass Katharina Bischöfe beobachtet, hat einige immer verwundert
10. Dass Natascha Generäle verbirgt, hat viele meist erbost
11. Dass Charlotte Genossen begehrt, hat alle wirklich überrascht
12. Dass Thorsten Komplizen bekämpft, hat manche stark verärgert
13. Dass Friedhelm Theologen besiegt, hat jeden sehr verwundert
14. Dass Kathleen Klienten erobert, hat einige oft erbost
15. Dass Ferdinand Rivalen vereint, hat viele immer überrascht

16. Dass Wilfried Dozenten begreift, hat alle meist verärgert
17. Dass Marianne Pädagogen bejubelt, hat manche wirklich verwundert
18. Dass Hildegard Regisseure bewacht, hat jeden stark erbost
19. Dass Rupprecht Monarchen verehrt, hat einige sehr überrascht
20. Dass Berthold Gesellen vergisst, hat viele oft verärgert
21. Dass Brigitte Studenten besucht, hat alle immer verwundert
22. Dass Friedrich Kandidaten betreut, hat manche meist erbost
23. Dass Sebastian Athleten bewertet, hat jeden wirklich überrascht
24. Dass Matthias Demokraten verfolgt, hat einige stark verärgert
25. Dass Maximilian Biologen versteht, hat viele sehr verwundert
26. Dass Christina Kardinäle bemerkt, hat alle oft erbost
27. Dass Alexander Katholiken anschaut, hat manche immer überrascht
28. Dass Hubertus Senioren erkennt, hat jeden meist verärgert
29. Dass Siegfried Griechen beobachtet, hat einige wirklich verwundert
30. Dass Angelika Chirurgen verbirgt, hat viele stark erbost
31. Dass Anneliese Dirigenten begehrt, hat alle sehr überrascht
32. Dass Christiane Kameraden bekämpft, hat manche oft verärgert
33. Dass Jonathan Feinde besiegt, hat jeden immer verwundert
34. Dass Eberhard Mandanten erobert, hat einige meist erbost
35. Dass Stephanie Schwestern vereint, hat viele wirklich überrascht
36. Dass Annemarie Philosophen begreift, hat alle stark verärgert
37. Dass Waltraud Redakteure bejubelt, hat manche sehr verwundert
38. Dass Magdalena Regenten bewacht, hat jeden oft erbost
39. Dass Franziska Solisten verehrt, hat einige immer überrascht
40. Dass Elfriede Statisten vergisst, hat viele meist verärgert

Subject-before-object clause, inanimate object (SO-AI)

1. Dass Heinrich Apotheken besucht, hat alle wirklich verwundert
2. Dass Johannes Umfragen betreut, hat manche stark erbost
3. Dass Benjamin Geschenke bewertet, hat jeden sehr überrascht
4. Dass Christoph Flugzeuge verfolgt, hat einige oft verärgert
5. Dass Alexandra Zeitungen versteht, hat viele immer verwundert
6. Dass Albrecht Signale bemerkt, hat alle meist erbost
7. Dass Christine Maschinen anschaut, hat manche wirklich überrascht
8. Dass Elisabeth Schranken erkennt, hat jeden stark verärgert
9. Dass Katharina Planeten beobachtet, hat einige sehr verwundert
10. Dass Natascha Pistolen verbirgt, hat viele oft erbost
11. Dass Charlotte Klamotten begehrt, hat alle immer überrascht
12. Dass Thorsten Bakterien bekämpft, hat manche meist verärgert
13. Dass Friedhelm Krankheiten besiegt, hat jeden wirklich verwundert
14. Dass Kathleen Laufstege erobert, hat einige stark erbost
15. Dass Ferdinand Substanzen vereint, hat viele sehr überrascht
16. Dass Wilfried Formeln begreift, hat alle oft verärgert
17. Dass Marianne Uniformen bejubelt, hat manche immer verwundert
18. Dass Hildegard Trophäen bewacht, hat jeden meist erbost
19. Dass Rupprecht Motorräder verehrt, hat einige wirklich überrascht
20. Dass Berthold Mikrofone vergisst, hat viele stark verärgert
21. Dass Brigitte Moscheen besucht, hat alle sehr verwundert
22. Dass Friedrich Verträge betreut, hat manche oft erbost

23. Dass Sebastian Dokumente bewertet, hat jeden immer überrascht
24. Dass Matthias Raketen verfolgt, hat einige meist verärgert
25. Dass Maximilian Magazine versteht, hat viele wirklich verwundert
26. Dass Christina Schilder bemerkt, hat alle stark erbost
27. Dass Alexander Kunstwerke anschaut, hat manche sehr überrascht
28. Dass Hubertus Parkplätze erkennt, hat jeden oft verärgert
29. Dass Siegfried Satelliten beobachtet, hat einige immer verwundert
30. Dass Angelika Gewehre verbirgt, hat viele meist erbost
31. Dass Anneliese Krawatten begehrt, hat alle wirklich überrascht
32. Dass Christiane Kalorien bekämpft, hat manche stark verärgert
33. Dass Jonathan Zigaretten besiegt, hat jeden sehr verwundert
34. Dass Eberhard Medaillen erobert, hat einige oft erbost
35. Dass Stephanie Textilien vereint, hat viele immer überrascht
36. Dass Annemarie Plaketten begreift, hat alle meist verärgert
37. Dass Waltraud Frisuren bejubelt, hat manche wirklich verwundert
38. Dass Magdalena Reaktoren bewacht, hat jeden stark erbost
39. Dass Franziska Instrumente verehrt, hat einige sehr überrascht
40. Dass Elfriede Tabletten vergisst, hat viele oft verärgert

Object-before-subject clause, two animate NPs (OS-AA)

1. Dass Nachbarn Heinrich besucht, hat alle immer verwundert
2. Dass Touristen Johannes betreut, hat manche meist erbost
3. Dass Fotografen Benjamin bewertet, hat jeden wirklich überrascht
4. Dass Franzosen Christoph verfolgt, hat einige stark verärgert
5. Dass Chinesen Alexandra versteht, hat viele sehr verwundert
6. Dass Ingenieure Albrecht bemerkt, hat alle oft erbost
7. Dass Passanten Christine anschaut, hat manche immer überrascht
8. Dass Christen Elisabeth erkennt, hat jeden meist verärgert
9. Dass Bischöfe Katharina beobachtet, hat einige wirklich verwundert
10. Dass Generäle Natascha verbirgt, hat viele stark erbost
11. Dass Genossen Charlotte begehrt, hat alle sehr überrascht
12. Dass Komplizen Thorsten bekämpft, hat manche oft verärgert
13. Dass Theologen Friedhelm besiegt, hat jeden immer verwundert
14. Dass Klienten Kathleen erobert, hat einige meist erbost
15. Dass Rivalen Ferdinand vereint, hat viele wirklich überrascht
16. Dass Dozenten Wilfried begreift, hat alle stark verärgert
17. Dass Pädagogen Marianne bejubelt, hat manche sehr verwundert
18. Dass Regisseure Hildegard bewacht, hat jeden oft erbost
19. Dass Monarchen Rupprecht verehrt, hat einige immer überrascht
20. Dass Gesellen Berthold vergisst, hat viele meist verärgert
21. Dass Studenten Brigitte besucht, hat alle wirklich verwundert
22. Dass Kandidaten Friedrich betreut, hat manche stark erbost
23. Dass Athleten Sebastian bewertet, hat jeden sehr überrascht
24. Dass Demokraten Matthias verfolgt, hat einige oft verärgert
25. Dass Biologen Maximilian versteht, hat viele immer verwundert
26. Dass Kardinäle Christina bemerkt, hat alle meist erbost
27. Dass Katholiken Alexander anschaut, hat manche wirklich überrascht
28. Dass Senioren Hubertus erkennt, hat jeden stark verärgert
29. Dass Griechen Siegfried beobachtet, hat einige sehr verwundert

30. Dass Chirurgen Angelika verbirgt, hat viele oft erbost
31. Dass Dirigenten Anneliese begehrt, hat alle immer überrascht
32. Dass Kameraden Christiane bekämpft, hat manche meist verärgert
33. Dass Feinde Jonathan besiegt, hat jeden wirklich verwundert
34. Dass Mandanten Eberhard erobert, hat einige stark erbost
35. Dass Schwestern Stephanie vereint, hat viele sehr überrascht
36. Dass Philosophen Annemarie begreift, hat alle oft verärgert
37. Dass Redakteure Waltraud bejubelt, hat manche immer verwundert
38. Dass Regenten Magdalena bewacht, hat jeden meist erbost
39. Dass Solisten Franziska verehrt, hat einige wirklich überrascht
40. Dass Statisten Elfriede vergisst, hat viele stark verärgert

Object-before-subject clause, inanimate object (OS-IA)

1. Dass Apotheken Heinrich besucht, hat alle sehr verwundert
2. Dass Umfragen Johannes betreut, hat manche oft erbost
3. Dass Geschenke Benjamin bewertet, hat jeden immer überrascht
4. Dass Flugzeuge Christoph verfolgt, hat einige meist verärgert
5. Dass Zeitungen Alexandra versteht, hat viele wirklich verwundert
6. Dass Signale Albrecht bemerkt, hat alle stark erbost
7. Dass Maschinen Christine anschaut, hat manche sehr überrascht
8. Dass Schranken Elisabeth erkennt, hat jeden oft verärgert
9. Dass Planeten Katharina beobachtet, hat einige immer verwundert
10. Dass Pistolen Natascha verbirgt, hat viele meist erbost
11. Dass Klamotten Charlotte begehrt, hat alle wirklich überrascht
12. Dass Bakterien Thorsten bekämpft, hat manche stark verärgert
13. Dass Krankheiten Friedhelm besiegt, hat jeden sehr verwundert
14. Dass Laufstege Kathleen erobert, hat einige oft erbost
15. Dass Substanzen Ferdinand vereint, hat viele immer überrascht
16. Dass Formeln Wilfried begreift, hat alle meist verärgert
17. Dass Uniformen Marianne bejubelt, hat manche wirklich verwundert
18. Dass Trophäen Hildegard bewacht, hat jeden stark erbost
19. Dass Motorräder Rupprecht verehrt, hat einige sehr überrascht
20. Dass Mikrofone Berthold vergisst, hat viele oft verärgert
21. Dass Moscheen Brigitte besucht, hat alle immer verwundert
22. Dass Verträge Friedrich betreut, hat manche meist erbost
23. Dass Dokumente Sebastian bewertet, hat jeden wirklich überrascht
24. Dass Raketen Matthias verfolgt, hat einige stark verärgert
25. Dass Magazine Maximilian versteht, hat viele sehr verwundert
26. Dass Schilder Christina bemerkt, hat alle oft erbost
27. Dass Kunstwerke Alexander anschaut, hat manche immer überrascht
28. Dass Parkplätze Hubertus erkennt, hat jeden meist verärgert
29. Dass Satelliten Siegfried beobachtet, hat einige wirklich verwundert
30. Dass Gewehre Angelika verbirgt, hat viele stark erbost
31. Dass Krawatten Anneliese begehrt, hat alle sehr überrascht
32. Dass Kalorien Christiane bekämpft, hat manche oft verärgert
33. Dass Zigaretten Jonathan besiegt, hat jeden immer verwundert
34. Dass Medaillen Eberhard erobert, hat einige meist erbost
35. Dass Textilien Stephanie vereint, hat viele wirklich überrascht
36. Dass Plaketten Annemarie begreift, hat alle stark verärgert

37. Dass Frisuren Waltraud bejubelt, hat manche sehr verwundert
38. Dass Reaktoren Magdalena bewacht, hat jeden oft erbost
39. Dass Instrumente Franziska verehrt, hat einige immer überrascht
40. Dass Tabletten Elfriede vergisst, hat viele meist verärgert

Experiment 5

Nominative-before-dative clause, two animate NPs (NOM-DAT – AA)

1. Gestern wurde die Studentin dem Dozenten gezeigt, bestätigte Bert
2. Heute wurde das Pferd dem Reiter verkauft, erklärte Anna
3. Neulich wurde der Affe dem Clown geraubt, vermutete Jutta
4. Soeben wurde die Kollegin dem Verkäufer angeboten, erwähnte Klaus
5. Jetzt wurde der Knecht dem Baron geliefert, beharrte Lisa
6. Eben wurde der Schüler dem Lehrer geschickt, bestätigte Ronny
7. Gestern wurde der Lehrling dem Fachmann übergeben, erklärte Meike
8. Heute wurde die Patientin dem Mediziner überlassen, vermutete Jörg
9. Neulich wurde der Helfer dem Forscher gesandt, erwähnte Ulla
10. Soeben wurde der Soldat dem Oberst empfohlen, beharrte Malte
11. Jetzt wurde der Hirsch dem Jäger geschenkt, bestätigte Hans
12. Eben wurde die Schlange dem Züchter gestohlen, erklärte Petra
13. Gestern wurde der Sohn dem Vater abgenommen, vermutete Nora
14. Heute wurde das Schaf dem Hirten geklaut, erwähnte Jens
15. Neulich wurde der Häftling dem Aufseher entzogen, beharrte Maria
16. Soeben wurde der Mandant dem Juristen zugeteilt, bestätigte Josef
17. Jetzt wurde der Wähler dem Minister gegönnt, erklärte Bert
18. Eben wurde der Arbeiter dem Direktor geliehen, vermutete Anna
19. Gestern wurde die Schuldige dem Zöllner entrissen, erwähnte Jutta
20. Heute wurde der Mitarbeiter dem Professor zugewiesen, beharrte Klaus
21. Gestern wurde der Betreuer dem Teenager gezeigt, bestätigte Lisa
22. Heute wurde der Hahn dem Bauern verkauft, erklärte Ronny
23. Neulich wurde der Dackel dem Wärter geraubt, vermutete Meike
24. Soeben wurde der Diener dem Herzog angeboten, erwähnte Jörg
25. Jetzt wurde der Anfänger dem Handwerker geliefert, beharrte Ulla
26. Eben wurde das Model dem Maler geschickt, bestätigte Malte
27. Gestern wurde die Gefangene dem Botschafter übergeben, erklärte Hans
28. Heute wurde der Verbrecher dem Polizisten überlassen, vermutete Petra
29. Neulich wurde der Sanitäter dem Verwundeten gesandt, erwähnte Nora
30. Soeben wurde der Trainer dem Sportler empfohlen, beharrte Jens
31. Jetzt wurde der Welpe dem Buben geschenkt, bestätigte Maria
32. Eben wurde der Vogel dem Rentner gestohlen, erklärte Josef
33. Gestern wurde das Kind dem Onkel abgenommen, vermutete Bert
34. Heute wurde der Hund dem Jungen geklaut, erwähnte Anna
35. Neulich wurde der Kunde dem Berater entzogen, beharrte Jutta
36. Soeben wurde der Gast dem Fahrer zugeteilt, bestätigte Klaus
37. Jetzt wurde der Mörder dem Verräter gegönnt, erklärte Lisa
38. Eben wurde der Dolmetscher dem Diplomaten geliehen, vermutete Ronny
39. Gestern wurde die Angeklagte dem Ermittler entrissen, erwähnte Meike
40. Heute wurde der Pfleger dem Kranken zugewiesen, beharrte Jörg

Nominative-before-dative clause, inanimate nominative NP (NOM-DAT – IA)

1. Gestern wurde das Manuskript dem Dozenten gezeigt, bestätigte Ulla
2. Heute wurde der Sattel dem Reiter verkauft, erklärte Malte
3. Neulich wurde das Kostüm dem Clown geraubt, vermutete Hans
4. Soeben wurde der Betrieb dem Verkäufer angeboten, erwähnte Petra
5. Jetzt wurde die Krone dem Baron geliefert, beharrte Nora
6. Eben wurde das Buch dem Lehrer geschickt, bestätigte Jens
7. Gestern wurde der Befund dem Fachmann übergeben, erklärte Maria
8. Heute wurde der Notfall dem Mediziner überlassen, vermutete Josef
9. Neulich wurde der Artikel dem Forscher gesandt, erwähnte Bert
10. Soeben wurde der Sektor dem Oberst empfohlen, beharrte Anna
11. Jetzt wurde der Helm dem Jäger geschenkt, bestätigte Jutta
12. Eben wurde das Futter dem Züchter gestohlen, erklärte Klaus
13. Gestern wurde der Verband dem Vater abgenommen, vermutete Lisa
14. Heute wurde der Knüppel dem Hirten geklaut, erwähnte Ronny
15. Neulich wurde das Gewehr dem Aufseher entzogen, beharrte Meike
16. Soeben wurde das Abkommen dem Juristen zugeteilt, bestätigte Jörg
17. Jetzt wurde der Erfolg dem Minister gegönnt, erklärte Ulla
18. Eben wurde das Protokoll dem Direktor geliehen, vermutete Malte
19. Gestern wurde das Haschisch dem Zöllner entrissen, erwähnte Hans
20. Heute wurde der Unterricht dem Professor zugewiesen, beharrte Petra
21. Gestern wurde das Zeugnis dem Teenager gezeigt, bestätigte Nora
22. Heute wurde der Wagen dem Bauern verkauft, erklärte Jens
23. Neulich wurde die Pistole dem Wärter geraubt, vermutete Maria
24. Soeben wurde die Krone dem Herzog angeboten, erwähnte Josef
25. Jetzt wurde das Material dem Handwerker geliefert, beharrte Bert
26. Eben wurde die Farbe dem Maler geschickt, bestätigte Anna
27. Gestern wurde die Nachricht dem Botschafter übergeben, erklärte Jutta
28. Heute wurde die Verhaftung dem Polizisten überlassen, vermutete Klaus
29. Neulich wurde das Zertifikat dem Verwundeten gesandt, erwähnte Lisa
30. Soeben wurde die Medizin dem Sportler empfohlen, beharrte Ronny
31. Jetzt wurde das Einrad dem Buben geschenkt, bestätigte Meike
32. Eben wurde die Tasche dem Rentner gestohlen, erklärte Jörg
33. Gestern wurde die Last dem Onkel abgenommen, vermutete Ulla
34. Heute wurde das Paket dem Jungen geklaut, erwähnte Malte
35. Neulich wurde das Vertrauen dem Berater entzogen, beharrte Hans
36. Soeben wurde das Auto dem Fahrer zugeteilt, bestätigte Petra
37. Jetzt wurde der Untergang dem Verräter gegönnt, erklärte Nora
38. Eben wurde das Gutachten dem Diplomaten geliehen, vermutete Jens
39. Gestern wurde das Geheimnis dem Ermittler entrissen, erwähnte Maria
40. Heute wurde das Zimmer dem Kranken zugewiesen, beharrte Josef

Dative-before-nominative clause, two animate NPs (DAT-NOM – AA)

1. Gestern wurde dem Dozenten die Studentin gezeigt, bestätigte Bert
2. Heute wurde dem Reiter das Pferd verkauft, erklärte Anna
3. Neulich wurde dem Clown der Affe geraubt, vermutete Jutta

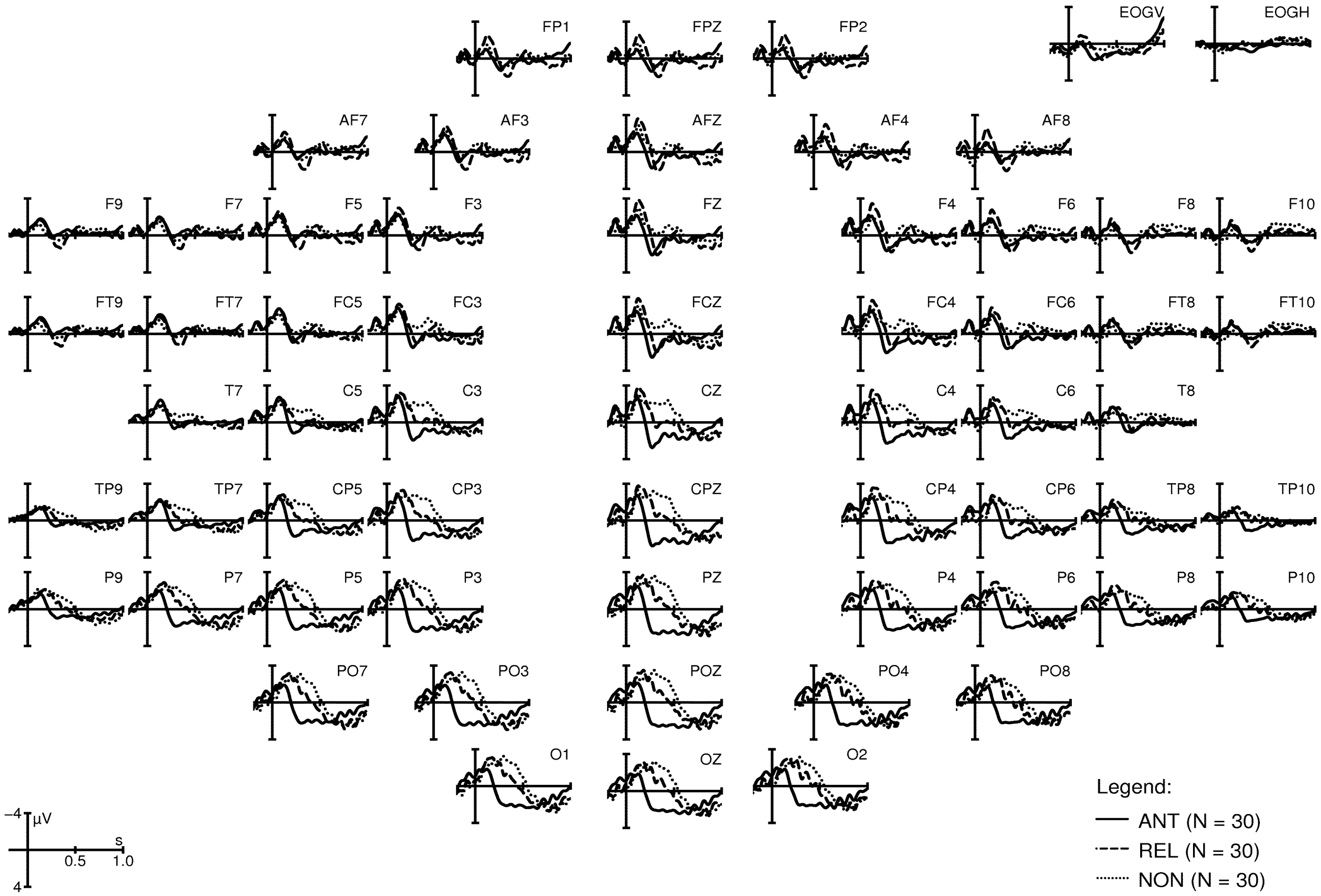
4. Soeben wurde dem Verkäufer die Kollegin angeboten, erwähnte Klaus
5. Jetzt wurde dem Baron der Knecht geliefert, beharrte Lisa
6. Eben wurde dem Lehrer der Schüler geschickt, bestätigte Ronny
7. Gestern wurde dem Fachmann der Lehrling übergeben, erklärte Meike
8. Heute wurde dem Mediziner die Patientin überlassen, vermutete Jörg
9. Neulich wurde dem Forscher der Helfer gesandt, erwähnte Ulla
10. Soeben wurde dem Oberst der Soldat empfohlen, beharrte Malte
11. Jetzt wurde dem Jäger der Hirsch geschenkt, bestätigte Hans
12. Eben wurde dem Züchter die Schlange gestohlen, erklärte Petra
13. Gestern wurde dem Vater der Sohn abgenommen, vermutete Nora
14. Heute wurde dem Hirten das Schaf geklaut, erwähnte Jens
15. Neulich wurde dem Aufseher der Häftling entzogen, beharrte Maria
16. Soeben wurde dem Juristen der Mandant zugeteilt, bestätigte Josef
17. Jetzt wurde dem Minister der Wähler gegönnt, erklärte Bert
18. Eben wurde dem Direktor der Arbeiter geliehen, vermutete Anna
19. Gestern wurde dem Zöllner die Schuldige entrissen, erwähnte Jutta
20. Heute wurde dem Professor der Mitarbeiter zugewiesen, beharrte Klaus
21. Gestern wurde dem Teenager der Betreuer gezeigt, bestätigte Lisa
22. Heute wurde dem Bauern der Hahn verkauft, erklärte Ronny
23. Neulich wurde dem Wärter der Dackel geraubt, vermutete Meike
24. Soeben wurde dem Herzog der Diener angeboten, erwähnte Jörg
25. Jetzt wurde dem Handwerker der Anfänger geliefert, beharrte Ulla
26. Eben wurde dem Maler das Model geschickt, bestätigte Malte
27. Gestern wurde dem Botschafter die Gefangene übergeben, erklärte Hans
28. Heute wurde dem Polizisten der Verbrecher überlassen, vermutete Petra
29. Neulich wurde dem Verwundeten der Sanitäter gesandt, erwähnte Nora
30. Soeben wurde dem Sportler der Trainer empfohlen, beharrte Jens
31. Jetzt wurde dem Buben der Welpe geschenkt, bestätigte Maria
32. Eben wurde dem Rentner der Vogel gestohlen, erklärte Josef
33. Gestern wurde dem Onkel das Kind abgenommen, vermutete Bert
34. Heute wurde dem Jungen der Hund geklaut, erwähnte Anna
35. Neulich wurde dem Berater der Kunde entzogen, beharrte Jutta
36. Soeben wurde dem Fahrer der Gast zugeteilt, bestätigte Klaus
37. Jetzt wurde dem Verräter der Mörder gegönnt, erklärte Lisa
38. Eben wurde dem Diplomaten der Dolmetscher geliehen, vermutete Ronny
39. Gestern wurde dem Ermittler die Angeklagte entrissen, erwähnte Meike
40. Heute wurde dem Kranken der Pfleger zugewiesen, beharrte Jörg

Dative-before-nominative clause, inanimate nominative NP (DAT-NOM – AI)

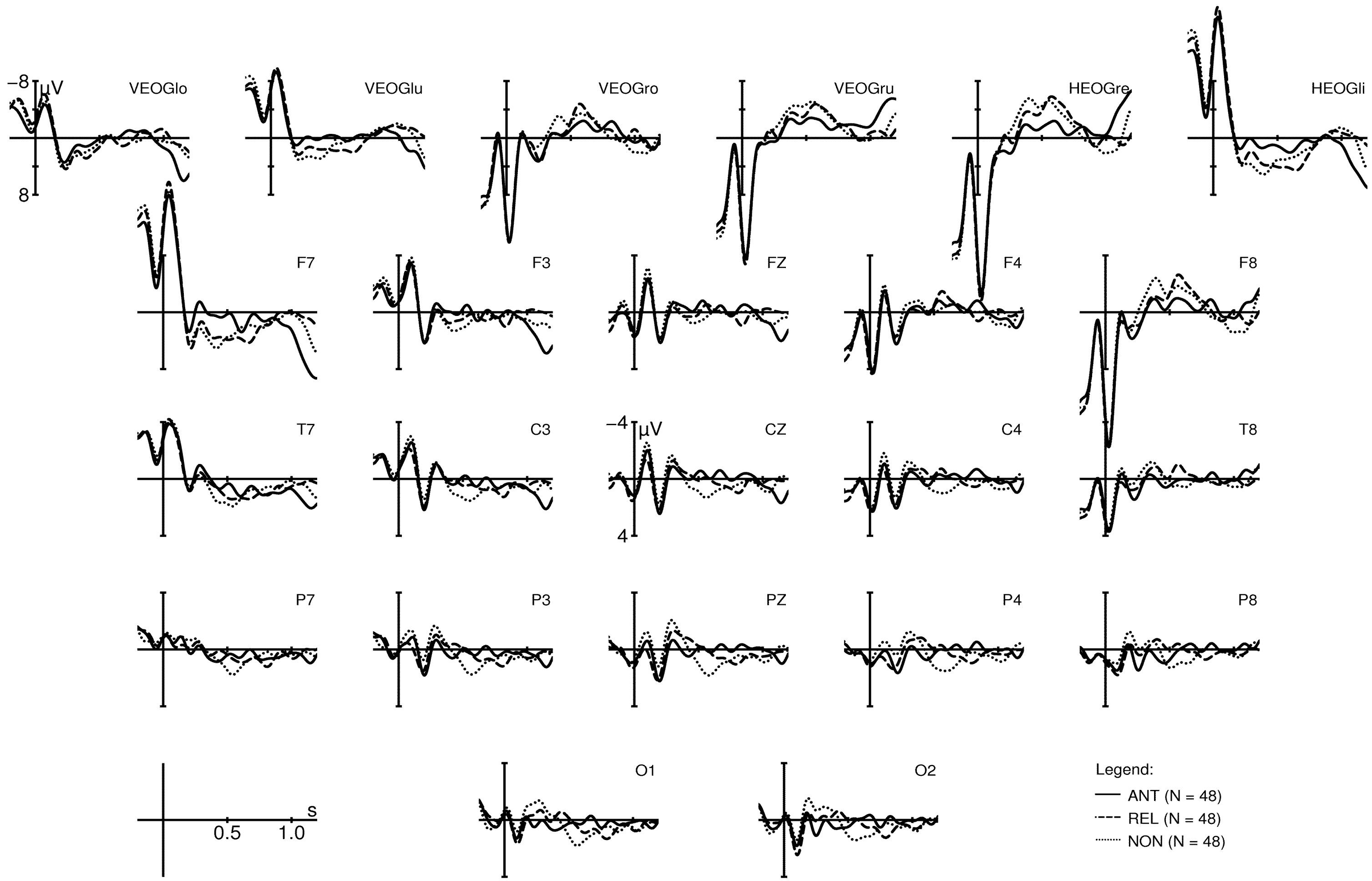
1. Gestern wurde dem Dozenten das Manuskript gezeigt, bestätigte Ulla
2. Heute wurde dem Reiter der Sattel verkauft, erklärte Malte
3. Neulich wurde dem Clown das Kostüm geraubt, vermutete Hans
4. Soeben wurde dem Verkäufer der Betrieb angeboten, erwähnte Petra
5. Jetzt wurde dem Baron die Krone geliefert, beharrte Nora
6. Eben wurde dem Lehrer das Buch geschickt, bestätigte Jens
7. Gestern wurde dem Fachmann der Befund übergeben, erklärte Maria
8. Heute wurde dem Mediziner der Notfall überlassen, vermutete Josef
9. Neulich wurde dem Forscher der Artikel gesandt, erwähnte Bert
10. Soeben wurde dem Oberst der Sektor empfohlen, beharrte Anna

11. Jetzt wurde dem Jäger der Helm geschenkt, bestätigte Jutta
12. Eben wurde dem Züchter das Futter gestohlen, erklärte Klaus
13. Gestern wurde dem Vater der Verband abgenommen, vermutete Lisa
14. Heute wurde dem Hirten der Knüppel geklaut, erwähnte Ronny
15. Neulich wurde dem Aufseher das Gewehr entzogen, beharrte Meike
16. Soeben wurde dem Juristen das Abkommen zugeteilt, bestätigte Jörg
17. Jetzt wurde dem Minister der Erfolg gegönnt, erklärte Ulla
18. Eben wurde dem Direktor das Protokoll geliehen, vermutete Malte
19. Gestern wurde dem Zöllner das Haschisch entrissen, erwähnte Hans
20. Heute wurde dem Professor der Unterricht zugewiesen, beharrte Petra
21. Gestern wurde dem Teenager das Zeugnis gezeigt, bestätigte Nora
22. Heute wurde dem Bauern der Wagen verkauft, erklärte Jens
23. Neulich wurde dem Wärter die Pistole geraubt, vermutete Maria
24. Soeben wurde dem Herzog die Krone angeboten, erwähnte Josef
25. Jetzt wurde dem Handwerker das Material geliefert, beharrte Bert
26. Eben wurde dem Maler die Farbe geschickt, bestätigte Anna
27. Gestern wurde dem Botschafter die Nachricht übergeben, erklärte Jutta
28. Heute wurde dem Polizisten die Verhaftung überlassen, vermutete Klaus
29. Neulich wurde dem Verwundeten das Zertifikat gesandt, erwähnte Lisa
30. Soeben wurde dem Sportler die Medizin empfohlen, beharrte Ronny
31. Jetzt wurde dem Buben das Einrad geschenkt, bestätigte Meike
32. Eben wurde dem Rentner die Tasche gestohlen, erklärte Jörg
33. Gestern wurde dem Onkel die Last abgenommen, vermutete Ulla
34. Heute wurde dem Jungen das Paket geklaut, erwähnte Malte
35. Neulich wurde dem Berater das Vertrauen entzogen, beharrte Hans
36. Soeben wurde dem Fahrer das Auto zugeteilt, bestätigte Petra
37. Jetzt wurde dem Verräter der Untergang gegönnt, erklärte Nora
38. Eben wurde dem Diplomaten das Gutachten geliehen, vermutete Jens
39. Gestern wurde dem Ermittler das Geheimnis entrissen, erwähnte Maria
40. Heute wurde dem Kranken das Zimmer zugewiesen, beharrte Josef

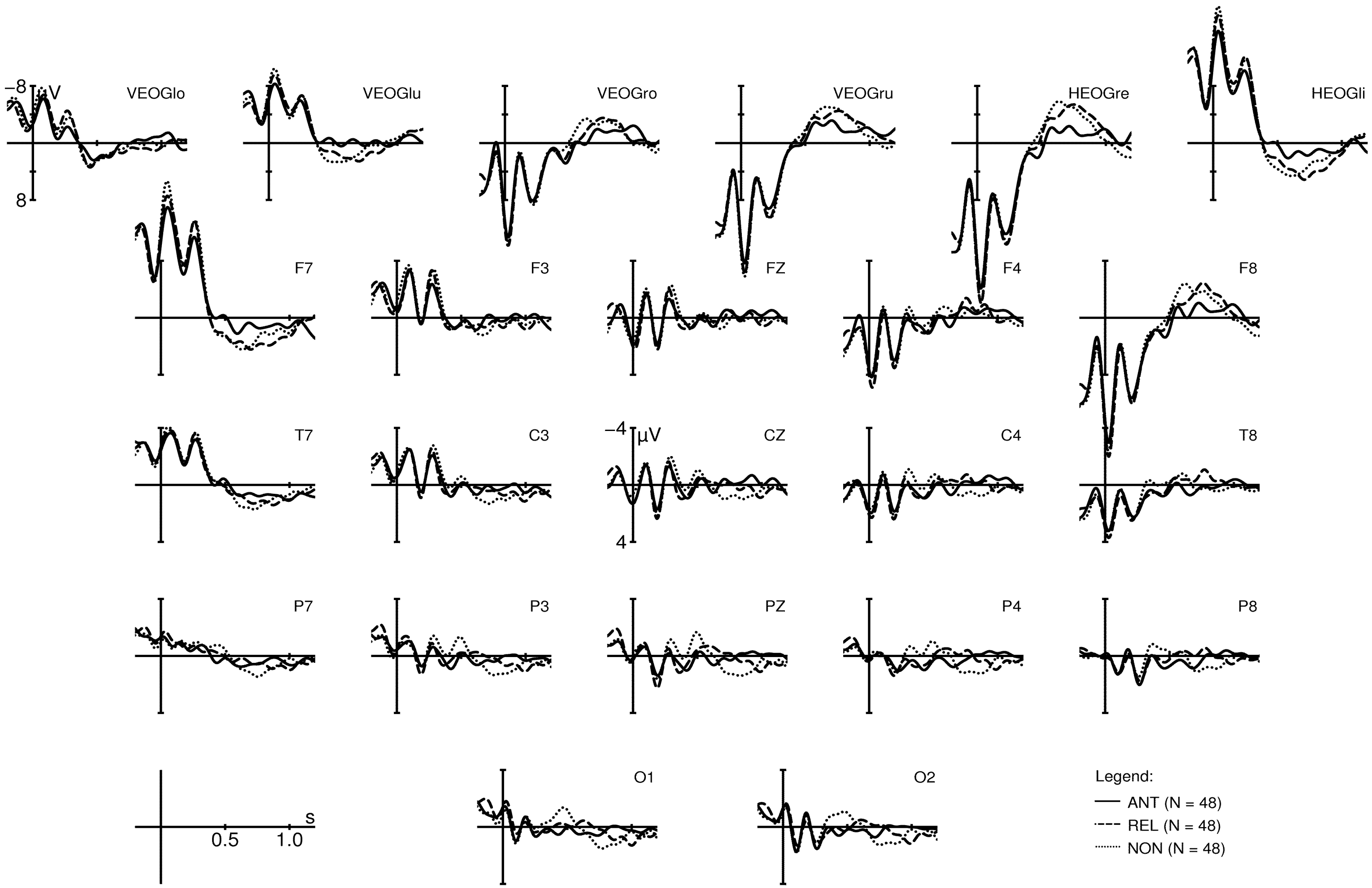
Appendix G: Supplementary figures for Experiments 1 – 5



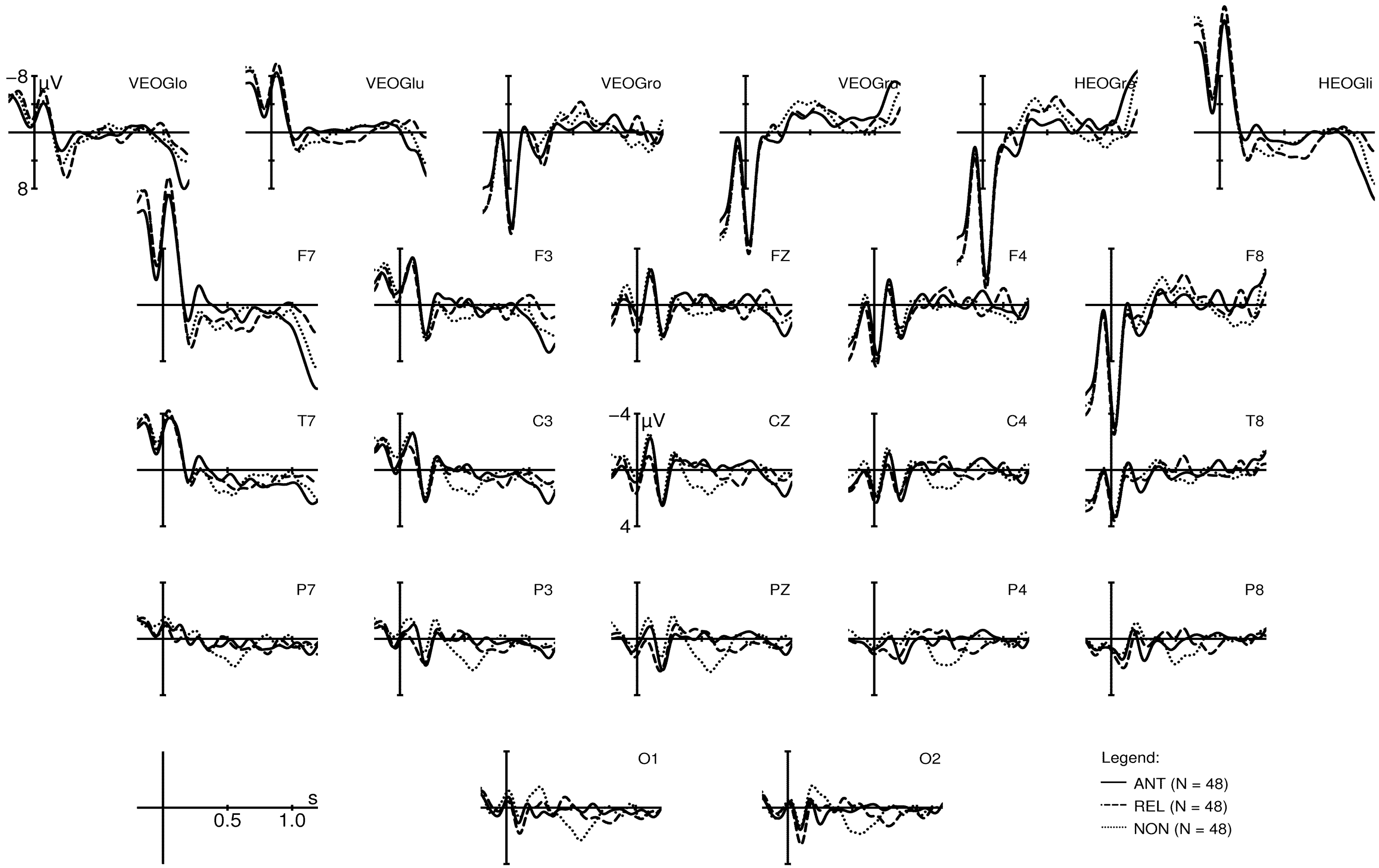
Supplementary figure for Experiment 1. Grand-average ERPs time-locked to the onset of the critical word.



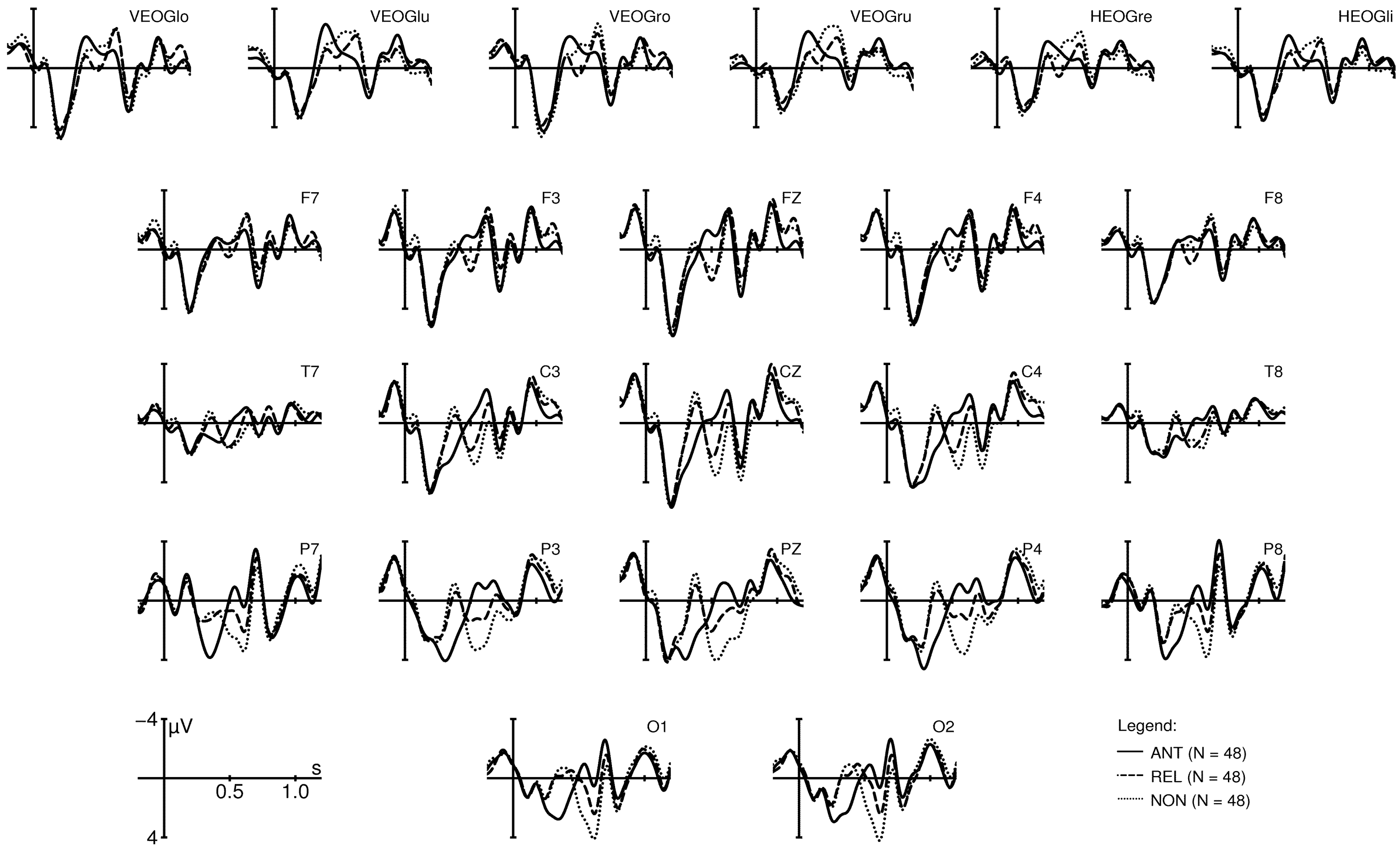
Supplementary figure for Experiment 2. Grand-average ERPs time-locked to the onset of the first fixation on the target word. Note that the amplitude scale for the EOG electrodes has been adapted to account for the higher amplitude ranges owing to the natural reading setting.



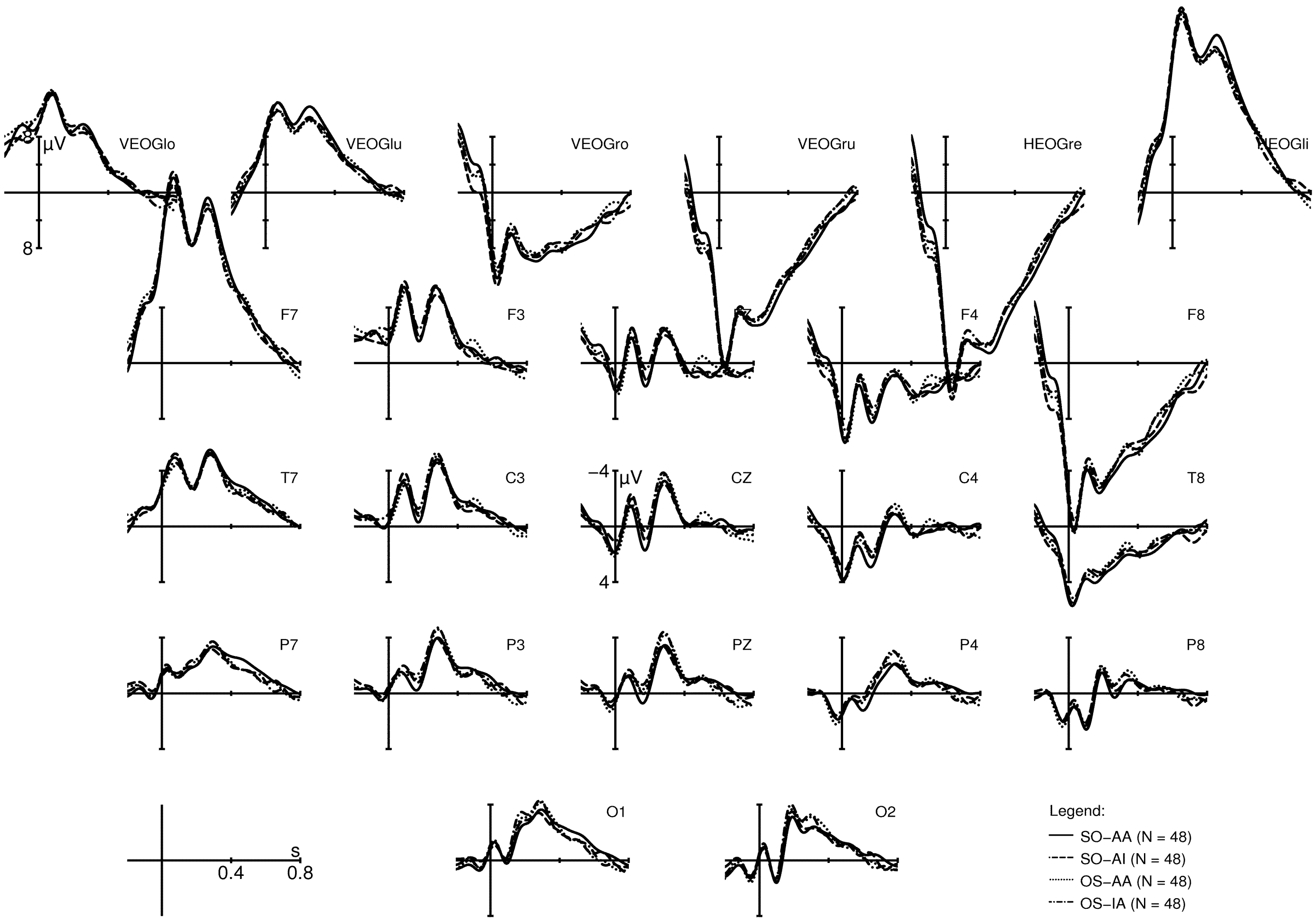
Supplementary figure for Experiment 2. Grand-average ERPs time-locked to the onset of the last fixation before the target word. Note that the amplitude scale for the EOG electrodes has been adapted to account for the higher amplitude ranges owing to the natural reading setting.



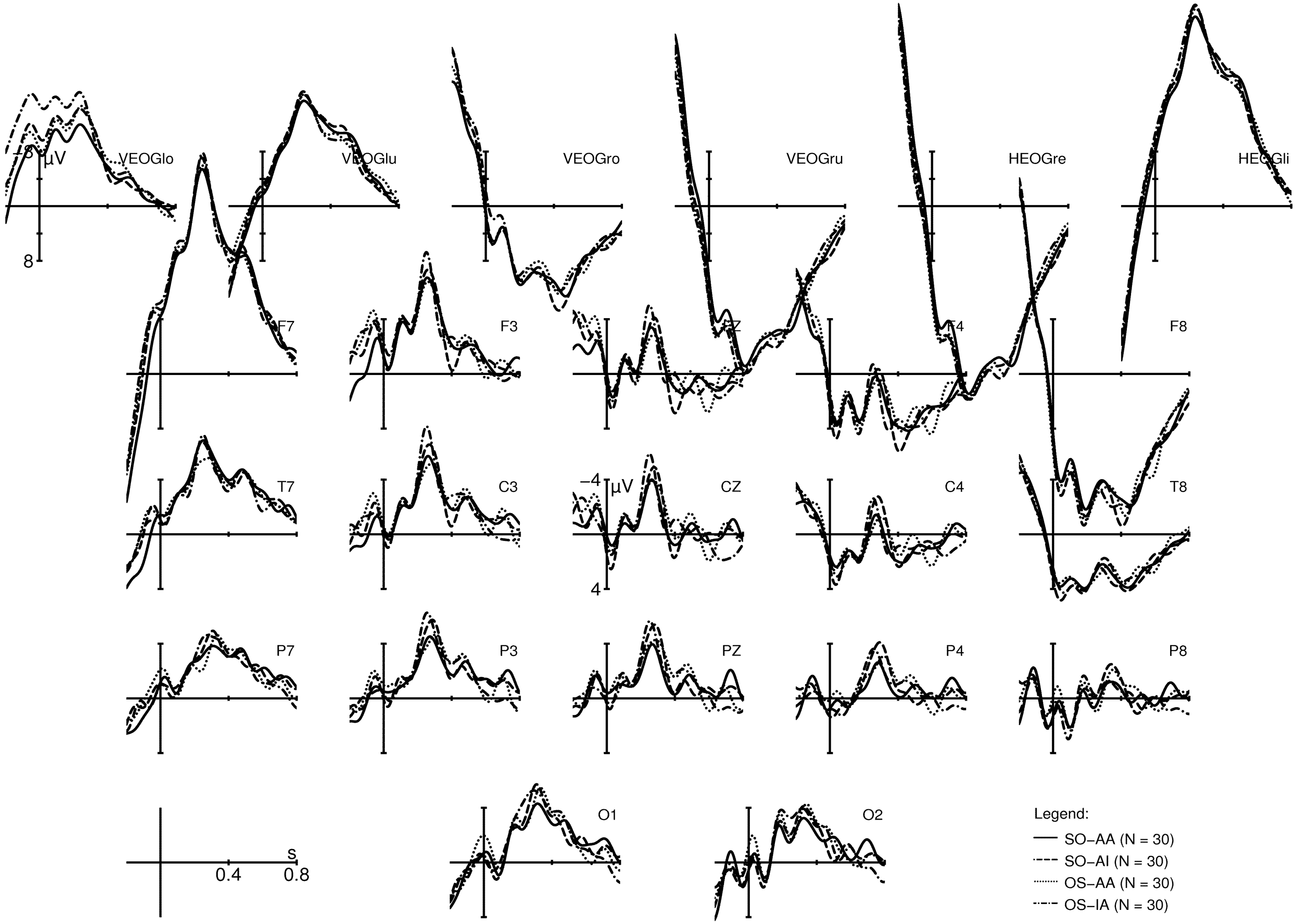
Supplementary figure for Experiment 2. Grand-average ERPs time-locked to the onset of the single fixation on the target word. Note that the amplitude scale for the EOG electrodes has been adapted to account for the higher amplitude ranges owing to the natural reading setting.



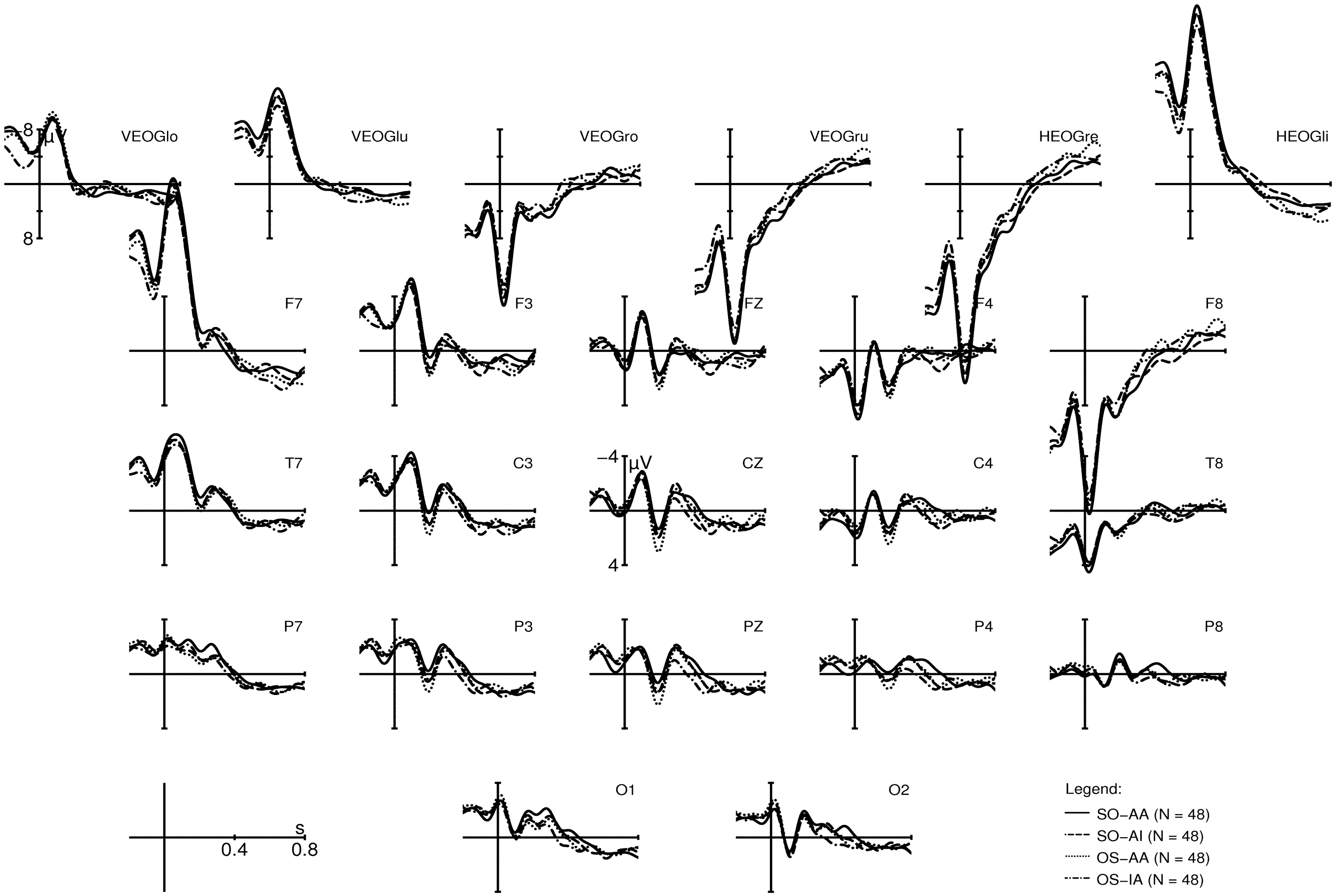
Supplementary figure for Experiment 3. Grand-average ERPs time-locked to the onset of the target word.



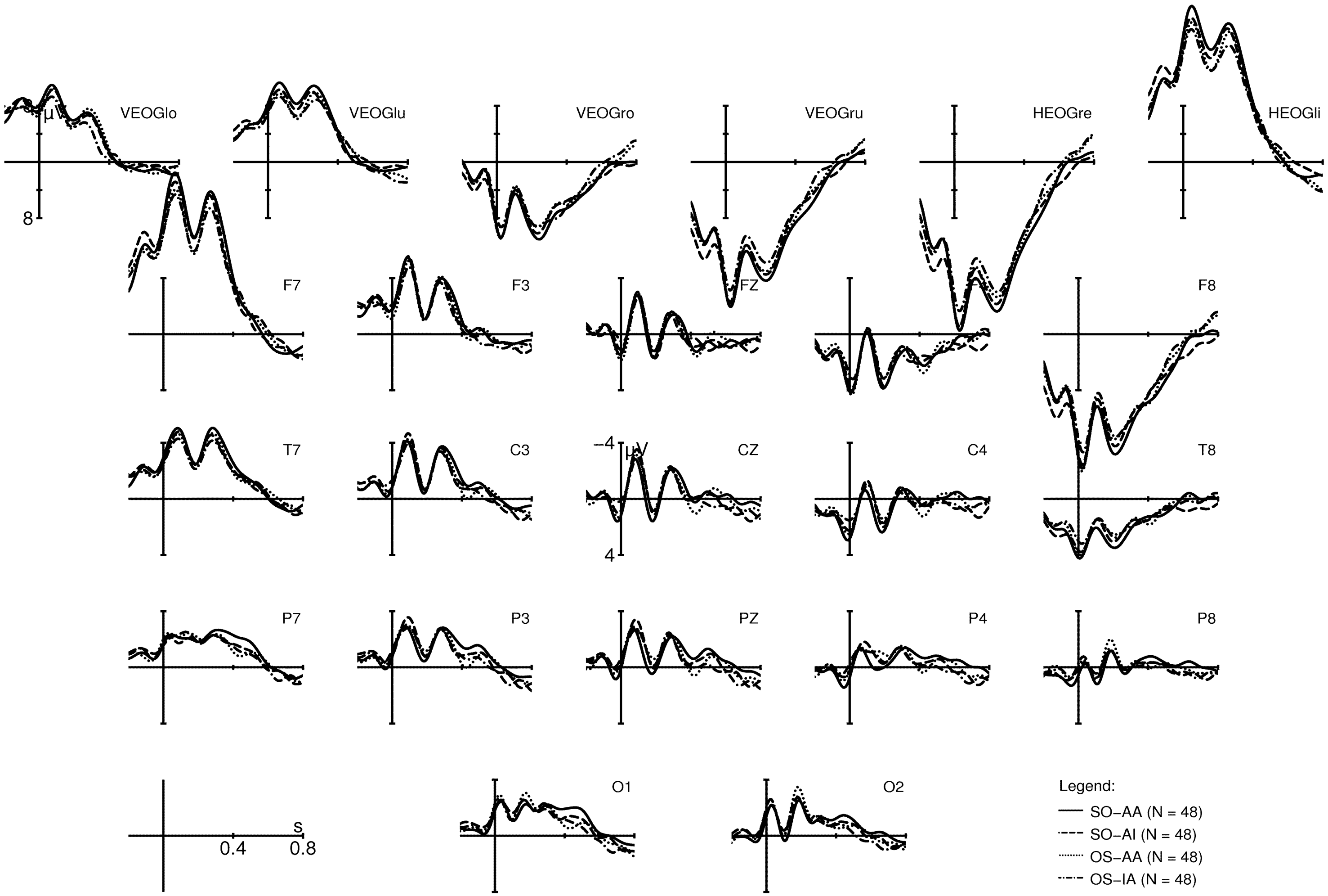
Supplementary figure for Experiment 4. Grand-average ERPs time-locked to the onset of the first fixation on the first NP. Note that the amplitude scale for the EOG electrodes has been adapted to account for the higher amplitude ranges owing to the natural reading setting.



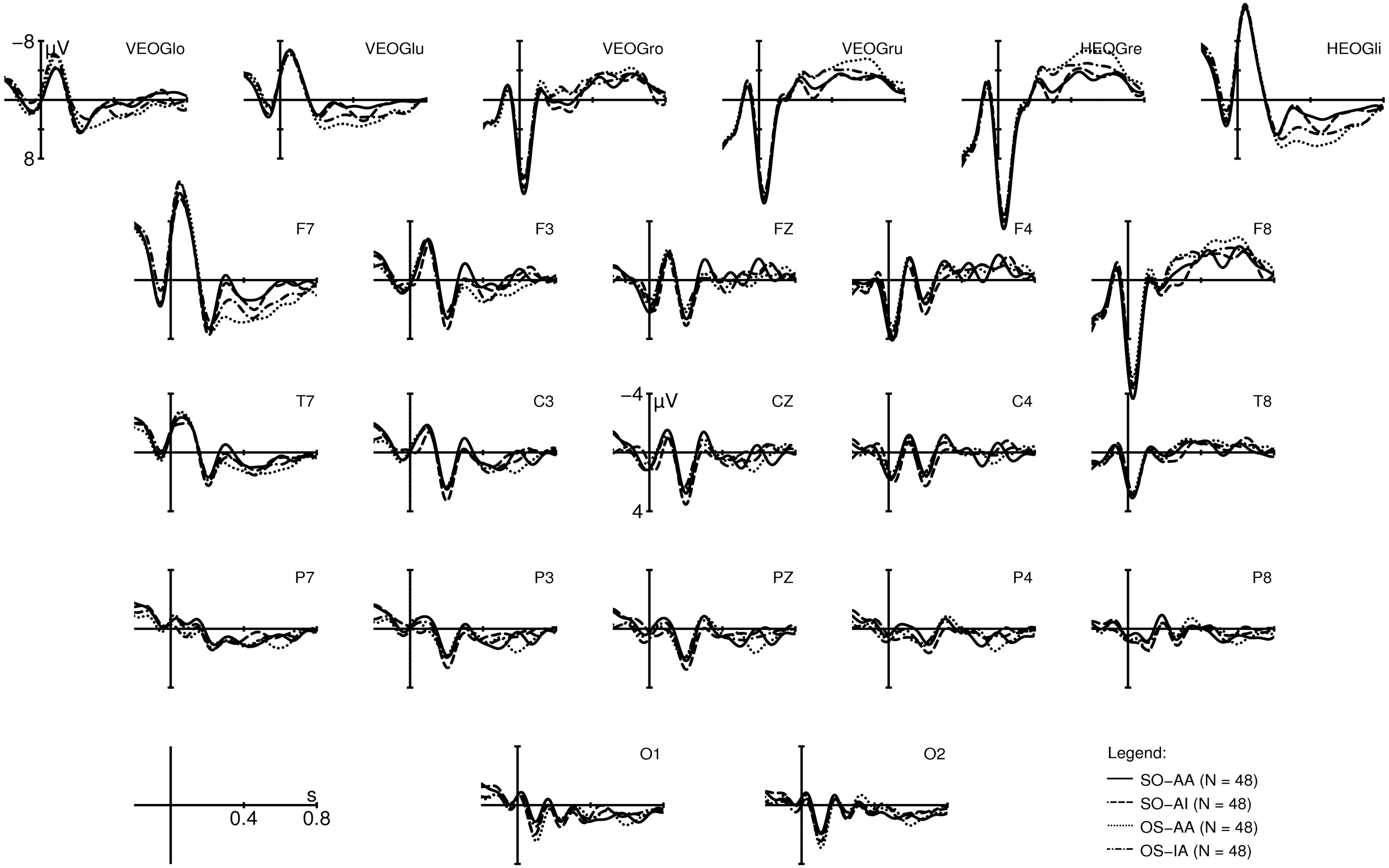
Supplementary figure for Experiment 4. Grand-average ERPs time-locked to the onset of the last fixation before the first NP. Note that the amplitude scale for the EOG electrodes has been adapted to account for the higher amplitude ranges owing to the natural reading setting.



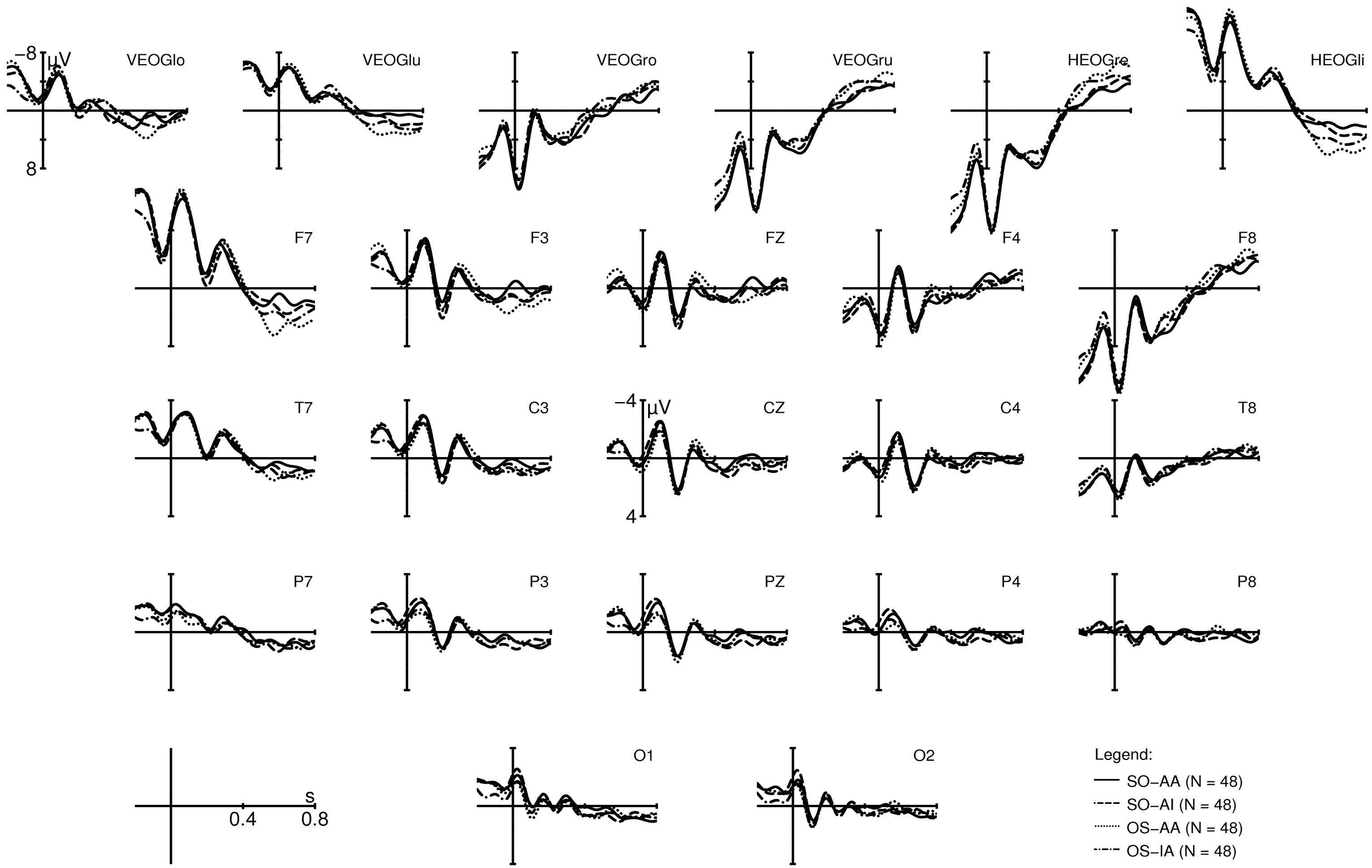
Supplementary figure for Experiment 4. Grand-average ERPs time-locked to the onset of the first fixation on the second NP. Note that the amplitude scale for the EOG electrodes has been adapted to account for the higher amplitude ranges owing to the natural reading setting.



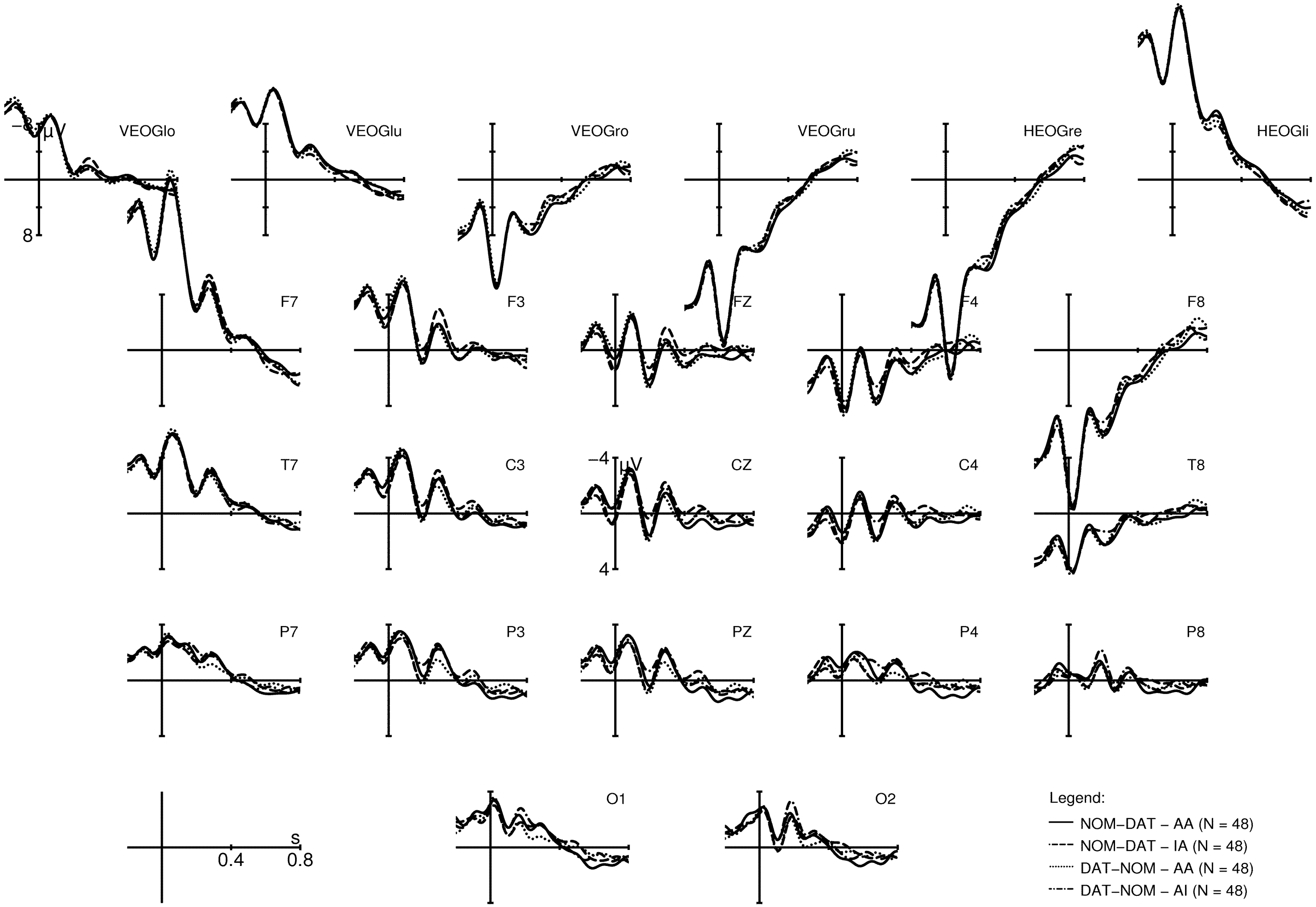
Supplementary figure for Experiment 4. Grand-average ERPs time-locked to the onset of the last fixation before the second NP. Note that the amplitude scale for the EOG electrodes has been adapted to account for the higher amplitude ranges owing to the natural reading setting.



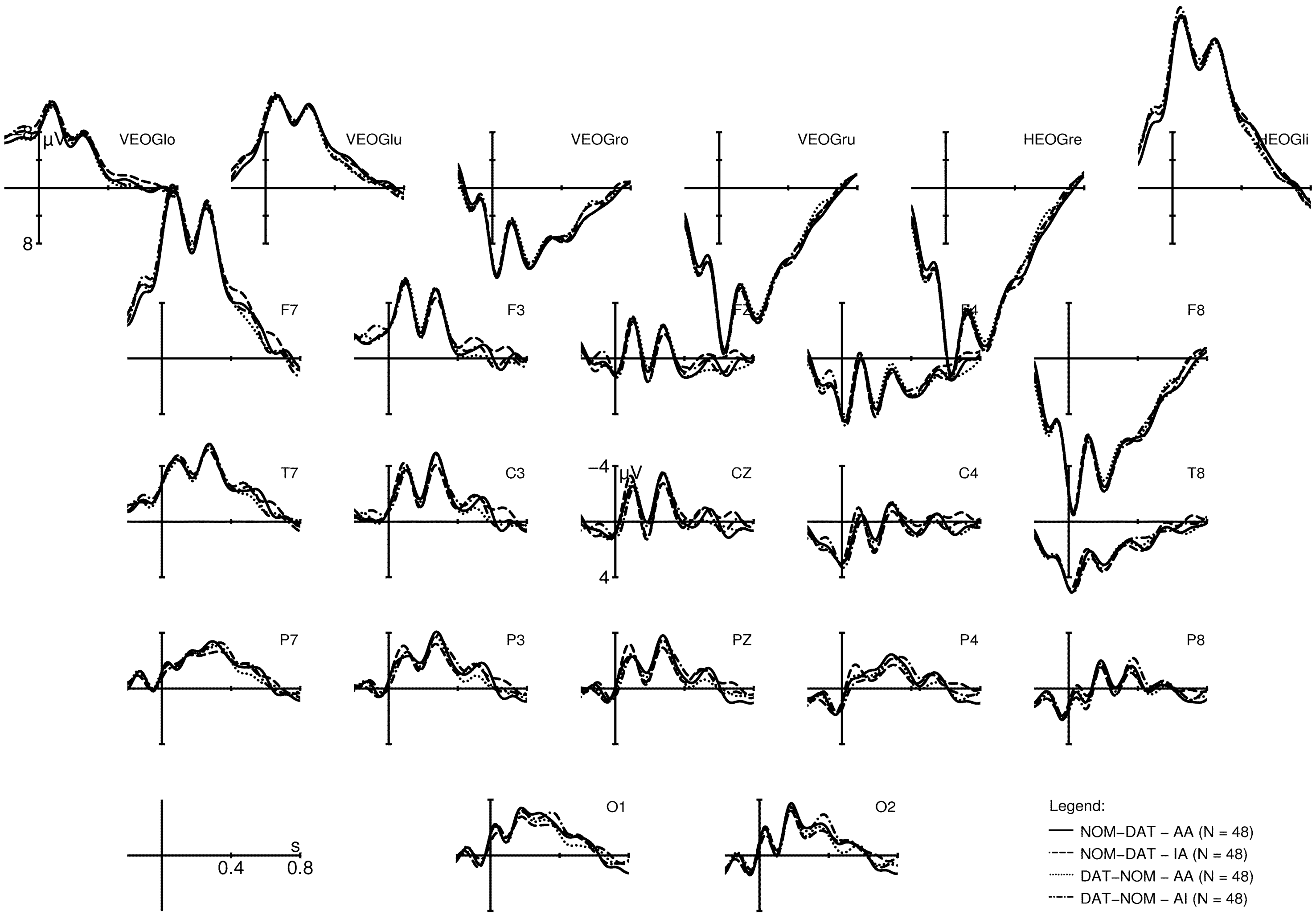
Supplementary figure for Experiment 4. Grand-average ERPs time-locked to the onset of the first fixation on the verb. Note that the amplitude scale for the EOG electrodes has been adapted to account for the higher amplitude ranges owing to the natural reading setting.



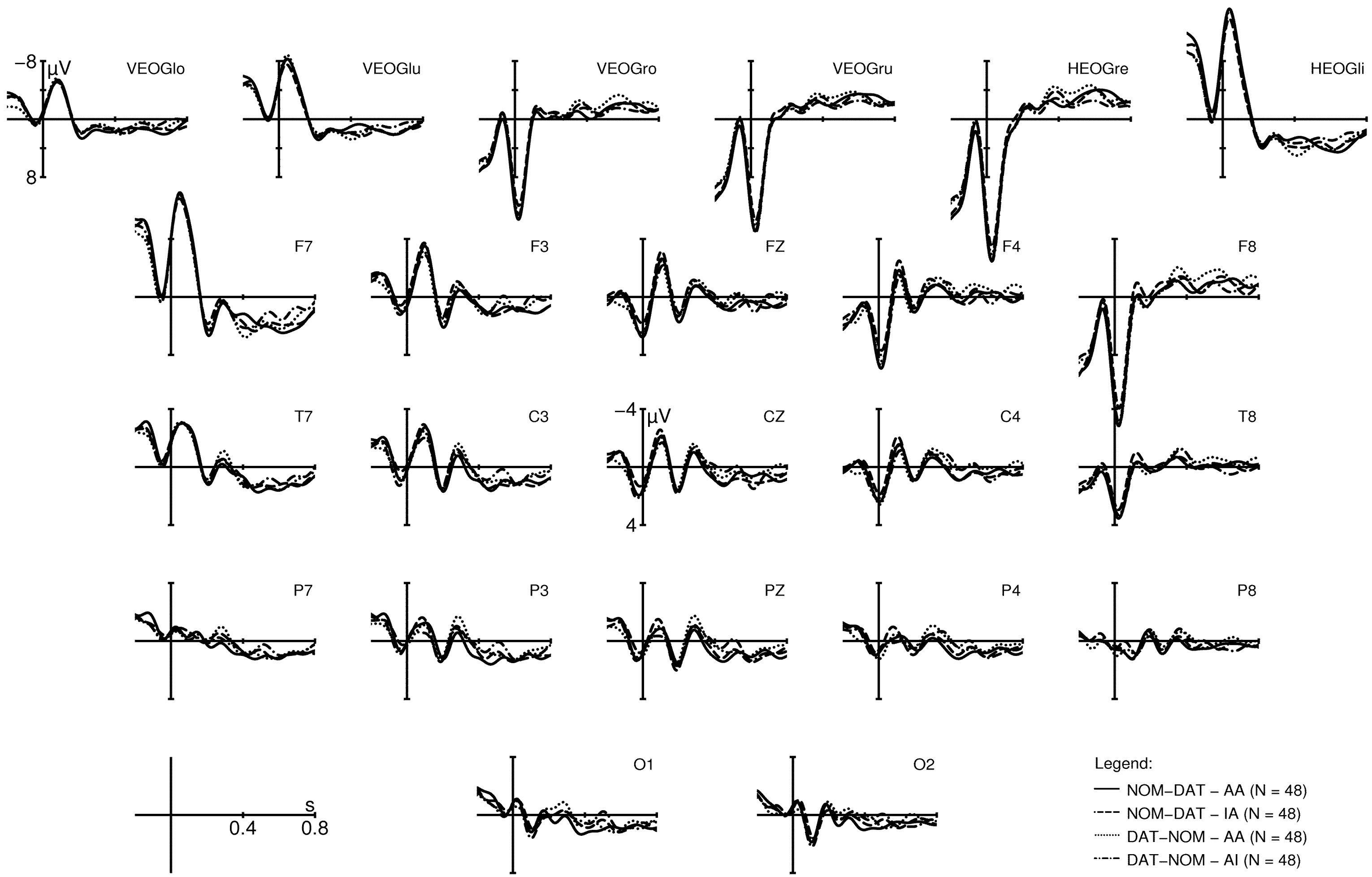
Supplementary figure for Experiment 4. Grand-average ERPs time-locked to the onset of the last fixation before the verb. Note that the amplitude scale for the EOG electrodes has been adapted to account for the higher amplitude ranges owing to the natural reading setting.



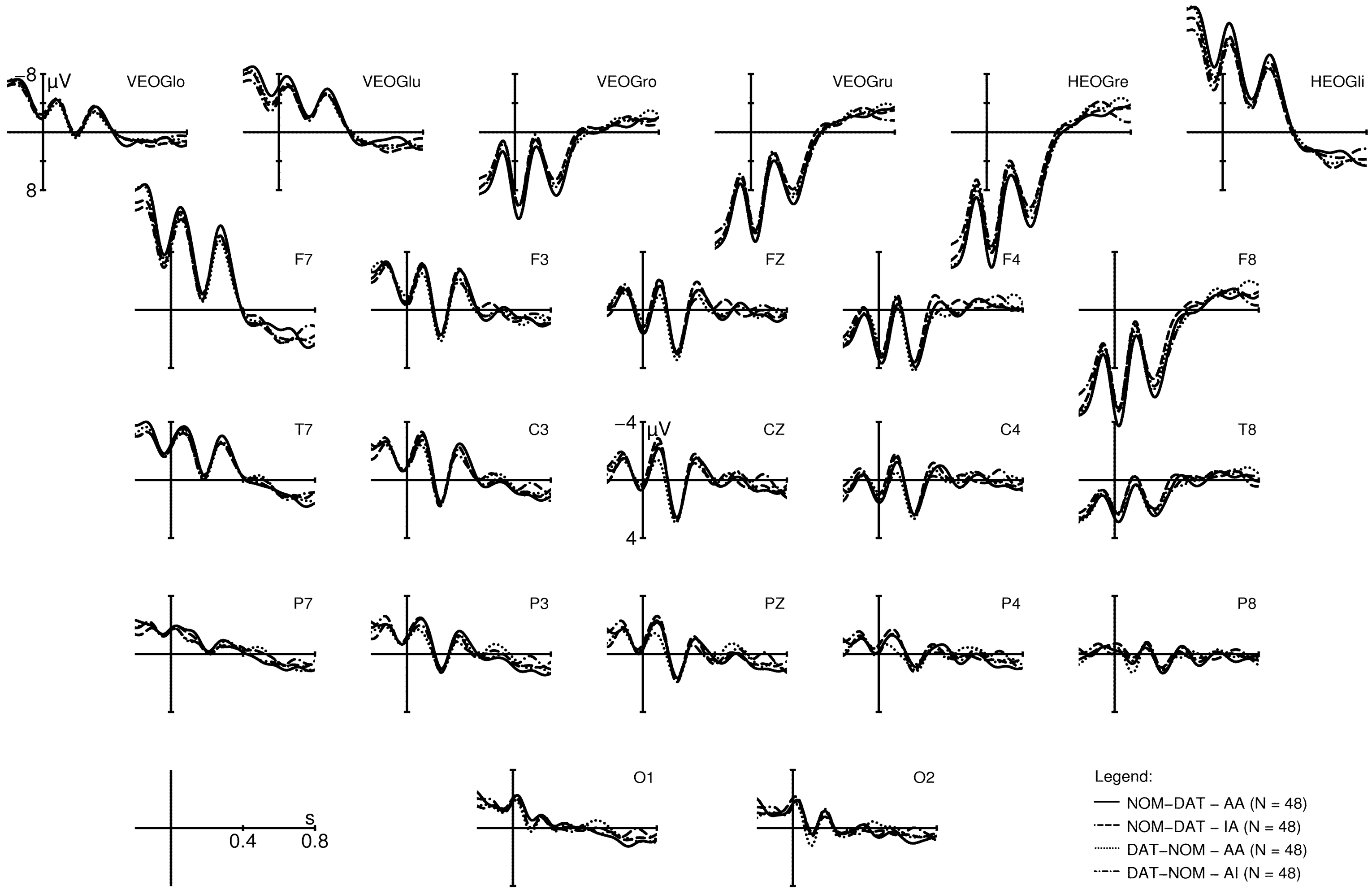
Supplementary figure for Experiment 5. Grand-average ERPs time-locked to the onset of the first fixation on the first NP. Note that the amplitude scale for the EOG electrodes has been adapted to account for the higher amplitude ranges owing to the natural reading setting.



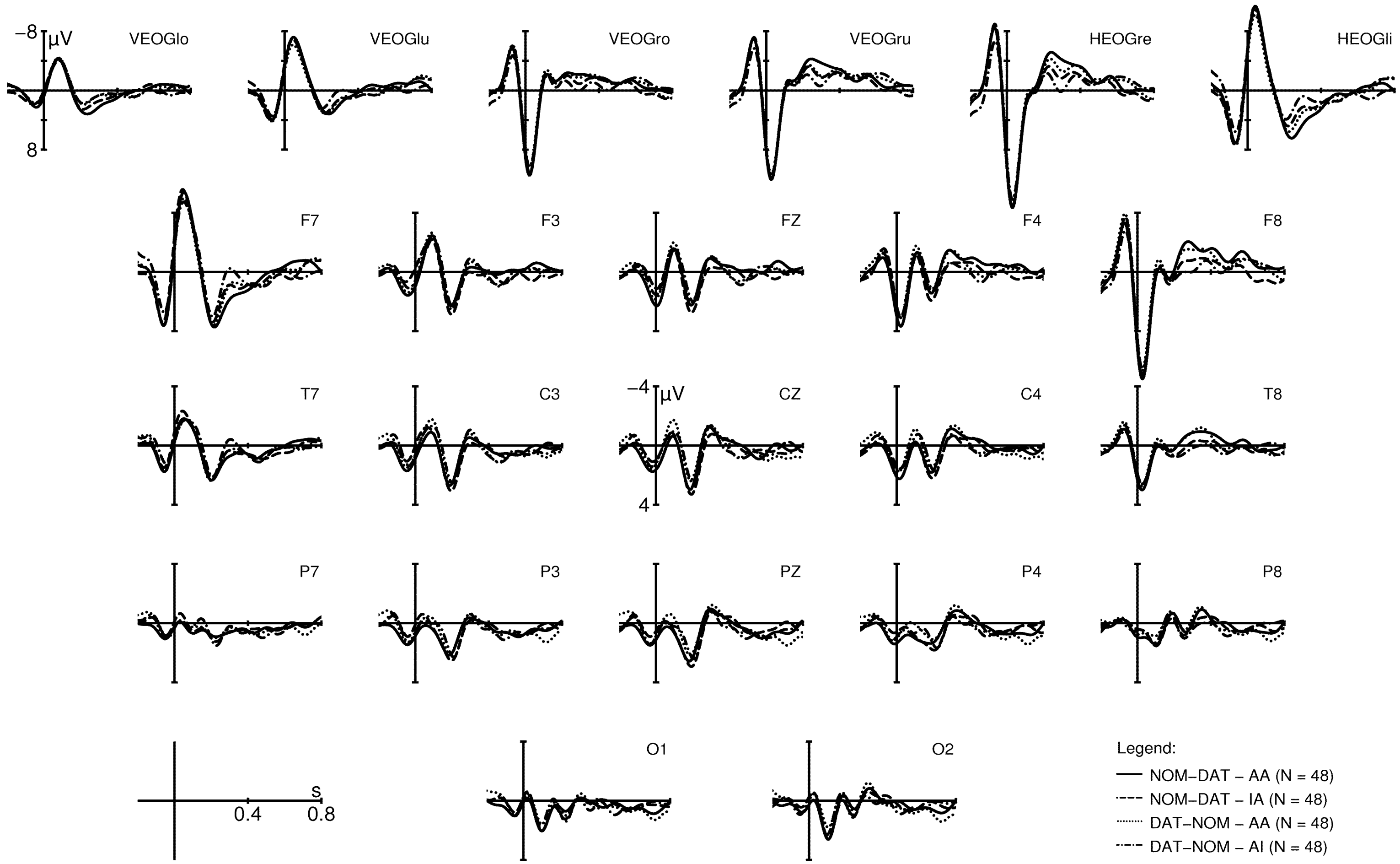
Supplementary figure for Experiment 5. Grand-average ERPs time-locked to the onset of the last fixation before the first NP. Note that the amplitude scale for the EOG electrodes has been adapted to account for the higher amplitude ranges owing to the natural reading setting.



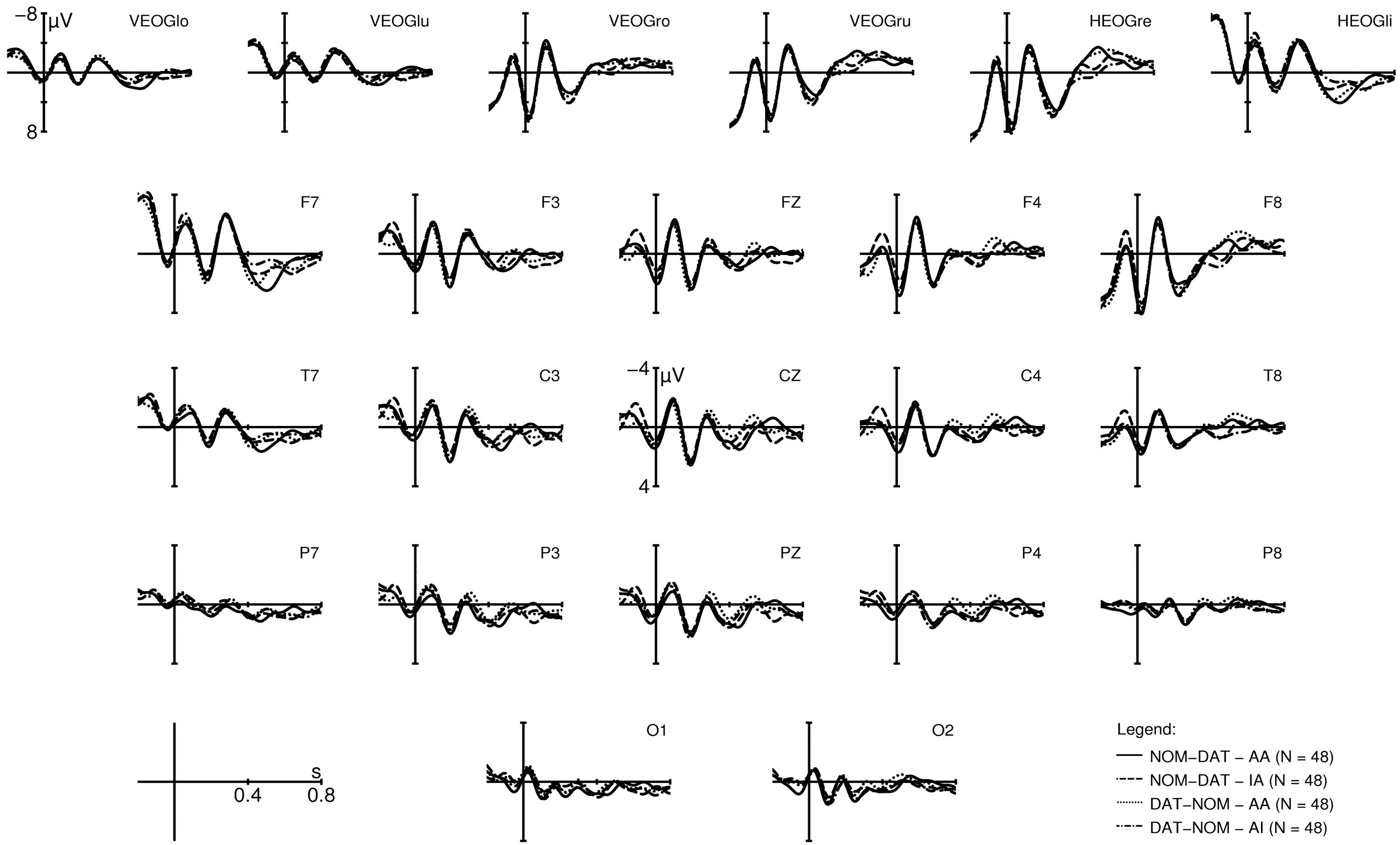
Supplementary figure for Experiment 5. Grand-average ERPs time-locked to the onset of the first fixation on the second NP. Note that the amplitude scale for the EOG electrodes has been adapted to account for the higher amplitude ranges owing to the natural reading setting.



Supplementary figure for Experiment 5. Grand-average ERPs time-locked to the onset of the last fixation before the second NP. Note that the amplitude scale for the EOG electrodes has been adapted to account for the higher amplitude ranges owing to the natural reading setting.



Supplementary figure for Experiment 5. Grand-average ERPs time-locked to the onset of the first fixation on the verb. Note that the amplitude scale for the EOG electrodes has been adapted to account for the higher amplitude ranges owing to the natural reading setting.



Supplementary figure for Experiment 5. Grand-average ERPs time-locked to the onset of the last fixation before the verb. Note that the amplitude scale for the EOG electrodes has been adapted to account for the higher amplitude ranges owing to the natural reading setting.

Abstract (English)

The question of how language comprehension can best be measured in real time has led to a steady development of psycholinguistic and neurolinguistic methods over the past decades. One of the current debates is concerned with the temporal paradox that arises when data from the two most widely used methods, namely the monitoring of readers' eye movements during reading and the extraction of event-related potentials (ERPs) from the human electroencephalogram (EEG), are compared with one another. For example, while the duration of the first fixation is increased to about 300 ms for infrequent words, the corresponding ERP component, the N400, has its maximum peak at about 400 ms post word-presentation onset. Thus, eye movement measures show effects of lexical frequency about 100 – 150 ms earlier than the ERP measure. This temporal divergence is hard to reconcile with the idea that neuronal effects of comprehension difficulty should not only have a correlate in the behavioral output (e.g., the eye movement pattern), but should actually trigger, and thus precede, it.

The debate about this temporal paradox has been based largely on data that were obtained from separate EEG and eye movement experiments. This is disadvantageous for two reasons. On the one hand, differences in stimuli, participants or other general environmental factors could lead to unwanted (and sometimes unrecognized) biases in the comparison of eye movement patterns and ERPs. Natural reading, on the other hand, is remarkably different from both the auditory modality and the rapid serial visual presentation (RSVP), which have been used in prior EEG studies. This difference lies in the parafoveal preview of upcoming words. Because of the unique availability of a parafoveal preview in natural reading (i) linguistic information is processed in parallel to a certain extent, (ii) parafoveally obtained information is always degraded compared to foveal information, and (iii) the allocation of

attention can be distributed across several words (in contrast to focused attention in the other two modalities).

The present thesis set out to study the temporal paradox described above in more detail by combining the ERP and the eye-tracking methods in one experiment. Eye movements and ERPs were collected concurrently while participants read whole sentences. The main aim was to investigate the reliability of two endogenous ERP components indexing language-related processes, the N400 and, in particular, the P300. Although relatively few psycholinguistic studies have examined the role of the P300 in language comprehension, it nevertheless appears suited to investigate whether and how natural reading influences neuronal indices of information processing. It is well-known from the literature that the P300 is sensitive to the contrast between parallel vs. serial processing, the allocation of attention (focused vs. distributed), and to the perceptual quality of the experimental stimuli. The P300 should thus interact with the natural reading modality, i.e. the properties of natural reading as described above should have a measurable impact on the P300. As for the N400, latency was of main interest. If parafoveal information in reading is relevant for the recognition of the upcoming word, the latency of the N400 might change considerably compared to previous findings (especially those from RSVP experiments). Since the N400 is less susceptible to the perceptual quality of stimuli, the appearance of a parafoveally induced N400 was also conceivable.

The five experiments presented in this thesis examined linguistic structures that lend themselves to the investigation of both the P300 and the N400, in terms of making predictions about their occurrence. Experiment 1 through 3 used the antonym paradigm (*X is the opposite of Y*) in which not only the contrast between the antonym and any non-antonym (e.g., *schwarz ist das Gegenteil von* *<u>weiß</u>*; 'black is the opposite of white') was of interest, but also the lexical relatedness of the non-antonyms to the expected target antonym. A non-antonym was either related (e.g., *schwarz ist das Gegenteil von* *<u>blau</u>*; 'black is the opposite of blue') or unrelated (e.g., *schwarz ist das Gegenteil von nett*; ‘black is the opposite of nice’) to the expected antonym. As predicted, P300 and presentation modality interacted with one another. Experiment 1 with auditory presentation and Experiment 3 with RSVP revealed almost identical results, while natural reading in Experiment 2 showed a significantly different pattern. With auditory presentation and RSVP, the expected antonyms exhibited a P300, while both related and unrelated non-antonyms elicited an N400 in the same time window. A P300 followed the N400 for both non-antonym types. In natural reading, there was neither a parafoveal (for the last fixation before the critical word) nor a foveal (for the first fixation on the critical word) P300 to antonyms, whereas the biphasic N400-P300 pattern for non-antonyms was observed foveally on the critical word. There was moreover a parafoveal priming effect for antonyms and related non-antonyms as revealed by the reduced N400 amplitudes in these two conditions. None of these N400 components exhibited a latency outside the previously reported range of N400 latencies.

These effects were interpreted as follows. The parafoveal N400 priming effect was taken to reflect automatic spreading activation processes between neighboring entries in the mental lexicon, so that not only the word form of the expected antonym was pre-activated by the context, but also those of semantic related non-antonyms. The foveal N400 effect, by contrast, reflects the well-known cloze-probability effect, in which the N400 amplitude of contextually unexpected words is increased. The P300 was interpreted as a marker for the categorization/ evaluation of the target word as either the expected antonym or a non-antonym. The P300 is inhibited because pre-lexical information (i.e., orthographic forms) is acquired in the parafovea; semantic information is only available foveally, that is via the first fixation on the critical word. So, at each fixation, the cognitive system attempts to make a decision on the perceived input (i.e., to categorize it), but pieces of information from the respective other domain are missing. In the parafovea, semantic information is not yet available and in the fovea, the categorization based on parafoveal orthographic information is already terminated. Although the input has been evaluated and categorized, both decisions are characterized by a high level of uncertainty.

This finding suggests that the mechanisms underlying the P300 depend on a partial overlap in the processing stages of word recognition. A serial processing routine (which perhaps involves the results from one stage being added to the next stage) does not appear to induce a P300 for highly expected words – at least, when these words are predicted on two different linguistic domains. Both the pre-lexical orthographic form and the lexical meaning of the expected antonym are predicted, so that the visual input has to meet both predictions simultaneously if the categorization of antonyms is to be complete and certain. The temporal dissociation of pre-lexical word form information and word semantics prohibits such an overlap. Since non-antonyms – regardless of their lexical relatedness to the context or prime word – are unexpected, the categorization process can begin only after their lexical processing has been completed. In other words, the categorization of any non-antonym presupposes that the meaning of these words and their appropriateness in the sentence context are known, thereby delaying the P300 to the post-N400 time range for these words. The emergence of a parafoveal N400 component and the disappearance of the P300 are the two most important findings of Experiment 2.

Two additional experiments using different stimuli were carried out to investigate the reliability of the parafoveally induced N400. Experiment 4 examined the comprehension of case-ambiguous preverbal arguments (e.g. *Dass <u>Elfriede Statisten vergisst</u>, hat viele meist verärgert* 'It annoyed many people that Elfriede forgets movie extras'). Although the clause-final verb eventually disambiguated the case marking of both nominal phrases (NPs), there was an additional cue about the order of subject and object, which occurred right at the first NP position. The NP type (specific proper name vs. non-specific bare plural) was unambiguously associated with one grammatical function, i.e. the proper name always served as the subject of the clause, regardless of its position in first or second argument position (*dass Elfriede Statisten vergisst* vs. *dass Statisten Elfriede vergisst* 'that Elfriede forgets movie extras'). Beside the word order manipulation, the animacy of the bare plural was varied such that it was either animate or inanimate (e.g. *Dass Elfriede Statisten vergisst* vs. *Dass Elfriede Tabletten vergisst* 'that Elfriede forgets pills'). At the position of the second argument, both factors elicited robust N400 effects. Two animate arguments induced an interference effect, visible in an enhanced N400 amplitude. Subject-initial clauses also elicited an N400 effect at this position because the initial intransitive reading of the sentence beginning with *dass Elfriede* ('that Elfriede') had to be revised. At the verb, there was a late positivity reflecting the animacy-induced interference effect and a late positivity for object-initial clauses. The second positivity was interpreted as belonging to the P300 family because it reflected the well-formedness evaluation of object-initial clauses. Importantly, none of the above effects were due to parafoveal processing; they were consistently found for the first fixation on a word.

Experiment 5 was carried out to test whether the case ambiguity in Experiment 4 increased foveal processing load, thereby reducing the parafoveal preview of the upcoming word and the likelihood of detecting parafoveal ERP effects. The order between a nominative and a dative argument, embedded in a passive clause, was varied (*Gestern wurde das Pferd dem Reiter verkauft, erklärte Anna* vs. *Gestern wurde dem Reiter das Pferd verkauft, …* 'Anna explained that the horse was sold to the equestrian yesterday') and the animacy of the nominative argument was manipulated as well (*Gestern wurde das Pferd dem Reiter verkauft* 'yesterday the horse was sold to the equestrian' vs. *Gestern wurde der Sattel dem Reiter verkauft* 'yesterday the saddle was sold to the equestrian'). The results showed no ERP effects due to the word order manipulation. The N400 amplitude was, however, increased for animate nominative arguments. This effect was attributed to the conflict between different types of prominence information. Both nominative case and animacy are in conflict with the assigned undergoer role (the most patient-like role) because animate nominative arguments are preferably linked to the actor role (the most agent-like role) in German. While this conflict was found for the first fixation on the second argument position, it was already observed parafoveally, i.e. for the last fixation preceding the first argument position (in nominative-before-dative structures). In sum, these results suggest that parafoveal N400 effects do occur in natural reading, provided that the fixation providing the crucial information is located on parts of the sentence that do not cause processing difficulty themselves. This precondition was met in Experiments 2 and 5, but not in Experiment 4. It also appears to be the case that, depending on the linguistic context, the parafoveal N400 can be linked to both pre-lexical and lexical-semantic processing.

With respect to the correlation between eye movements and ERPs the present thesis revealed the following pattern. Foveally induced ERPs correlated with increased fixation durations in several eye movement measures. Foveal N400 effects due to lexical-semantic manipulations tended to correlate with an increase of the corresponding first fixation on the target word, whereas N400 effects due to word order variations rather led to refixations of the word, without affecting the duration of the first fixation. In contrast to the foveal effects, the parafoveal N400 effects did not translate into longer fixation times. This pattern suggests that eye movements in reading do not simply reflect processing difficulty. If this were the case, then a parafoveal N400 should have led to increased fixation durations. Rather, it appears that effects in the eye movement pattern result from a dynamic interplay of new information uptake and the inhibition of further information uptake. Given that the information acquired during the current fixation is not helpful in resolving the processing difficulty and given the evidence that the upcoming sentence context could provide useful information, an increase in the duration of the current fixation is rather unlikely. This would explain why the durations of the fixations associated with the parafoveal N400 effects did not reveal equivalent effects. If, on the other hand, processing difficulties arise which could be resolved during the current fixation, but would need more time, longer fixation durations or a higher probability of refixations on the word will result. Interestingly, comprehensive measures of eye movements appeared to reflect general cognitive processing in addition to purely linguistic processing. Experiment 2, for example, revealed that P300 latency and amplitude variation seemed to correlate with similar effects in comprehensive eye movement measures (e.g., total reading time on a word).

Finally, the present dissertation has raised several open questions which remain to be answered in future research. The present co-registration experiments could not conclusively explain which linguistic manipulation causes longer (first) fixation durations on the critical word and which one leads to more refixations (e.g., longer first pass time). Moreover, the results from Experiment 5 confirm previous findings from separate ERP and eye movement studies, which revealed that evidence from both measures does not necessarily converge – even for highly similar stimuli. Although there was no ERP effect for the initial dative argument in dative-before-nominative structures, the eye movement pattern evidenced a disadvantage for this word order (relative to the nominative-before-dative structure). Brain responses and (behavioral) oculomotor responses thus appear to be sensitive to somewhat different aspects involved in processing certain structures (e.g., grammatical rules vs. structural frequency). Nevertheless, this experiment demonstrated that both the N400 and the P300 are reliable markers of language processing, both of which, however, interacted to some extent with the presentation method used. Some of the mechanisms assumed to underlie the N400 could be better dissociated from one another in natural reading than in other modalities. The inhibition of the P300 to highly expected words showed what kind of domain-general cognitive processes are important in language comprehension – in addition to purely lexical word recognition.

Abstract (German)

Die Frage, wie Sprachverstehen in Echtzeit gemessen werden kann, hat in den letzten Jahrzehnten zu einer stetigen Weiterentwicklung psycho- und neurolinguistischer Methoden geführt. Eine der gegenwärtigen Methodendebatten ist bestimmt durch das zeitliche Paradox, das der Vergleich der beiden wohl am häufigsten genutzten Methoden, der Registrierung von Blickbewegungen beim Lesen und der Ermittlung Ereignis-korrelierter Potenziale (EKP) im Elektroenzephalogramm (EEG), mit sich bringt. Dieses Paradox ergibt sich aus dem Umstand, dass, vor allem bei lexikalisch-semantischer Verarbeitung, die Blickbewegungen Effekte zeigen, die im Durchschnitt 100-150 ms vor dem entsprechenden EKP-Effekt liegen. Ein Wort mit einer niedrigen lexikalischen Frequenz führt beispielsweise zu einer Erhöhung der Dauer der ersten Fixation auf diesem Wort, sodass diese etwa 300 ms andauert. Ein häufig auftretendes Wort weist demgegenüber eine Fixationszeit von durchschnittlich 250 ms auf. Die N400, eine EKP-Komponente, die sensitiv für lexikalische Frequenz ist, hat ihren stärksten Ausschlag jedoch erst um 400 ms nach Präsentationsbeginn des Wortes. Diese zeitliche Inkongruenz ist schwer in Einklang zu bringen mit der Hypothese, dass neuronale Verarbeitungsprozesse sich nicht nur in behavioralem Output wiederfinden lassen sollten, sondern diesen auch bedingen.

Die bisherige Debatte über dieses zeitliche Paradox fußte nahezu gänzlich auf Daten, die aus separaten Blickbewegungs- und EKP-Experimenten gewonnen wurden, was eine fundierte Analyse im Grunde nicht zulässt. Zum einen besteht die Möglichkeit, dass Unterschiede in Stimulusmaterial, Probandenauswahl oder anderen allgemeinen Umgebungsfaktoren zu unerwünschten (und unentdeckten) Verzerrungen im Vergleich führen. Andererseits weist natürliches Lesen zum Teil fundamental andere Eigenschaften auf als die auditive Modalität oder die rapid serial visual presentation (RSVP), d.h. die visuelle Einzelwortpräsentation von Sätzen, die in EEG-Experimenten bisher zum Einsatz kamen.

Hier spielt der sogenannte „parafoveal preview“, die Vorschau auf parafoveale (d.h. demnächst noch zu fixierende) Wörter, eine besondere Rolle. Durch diese Vorschau ist die Verarbeitung linguistischer Informationen im Leseprozess partiell parallel, parafoveale Informationen sind, im Gegensatz zu fovealen Informationen, stets perzeptuell verschlechtert und die Verteilung von Aufmerksamkeit kann sich über mehrere Wörter erstrecken. Diese Merkmale unterscheiden natürliches Lesen sowohl von auditiver als auch von RSVP-Präsentation.

Die vorliegende Dissertation versucht das beschriebene zeitliche Paradox zu untersuchen, indem die EKP-Methode und die Blickbewegungsregistrierung in einem Experiment mit einander gekoppelt wurden. Die Augenbewegungen und das EEG der Versuchspersonen wurden parallel aufgezeichnet, während diese Sätze lasen, die wie in einem Blickbewegungsexperiment als Ganzes präsentiert wurden. Die zugrunde liegende Hauptfrage war, ob bislang gültige EKP-Komponenten der Sprachverarbeitung auch unter natürlichen Lesebedingungen Bestand haben würden. Als zu untersuchende Komponenten wurden die P300 und die N400 ausgewählt. Obwohl die P300 bislang vergleichsweise wenig Beachtung in der Psycholinguistik gefunden hat, erscheint sie doch besonders dafür geeignet, den Einfluss der natürlichen Leseumgebung auf neuronale Verarbeitungskorrelate zu untersuchen. Aus der psychophysiologischen Forschung ist bekannt, dass die P300-Komponente auf parallele Verarbeitungsprozesse (vs. serielle Prozesse), auf die Art und Weise, wie Aufmerksamkeit verteilt wird (selektiv fokussiert vs. breit verteilt) und auf die perzeptuelle Qualität von Stimuli reagiert. Die P300 sollte also mit der Präsentationsmodalität interagieren, d.h. die oben beschriebenen Eigenschaften des natürlichen Lesens sollten einen messbaren Einfluss auf die P300 besitzen. Hinsichtlich der N400 war besonders ihre Latenz von Interesse. Falls parafoveale Informationen relevant für die Verarbeitung eines Wortes sind, so könnte sich die Latenz der N400 verändern. Da die N400-Komponente weniger anfällig für die perzeptuelle Qualität der Stimuli ist, ist die Existenz von parafovealen N400-Komponenten während des Lesens ebenfalls vorstellbar.

Die fünf vorgestellten Experimente untersuchten Strukturen, für die auf Basis bestehender Daten spezifische Vorhersagen über das Auftreten einer P300 oder N400 gemacht werden konnten. Die Experimente 1 bis 3 nutzten ein Antonymparadigma (*X ist das Gegenteil von Y*), bei dem neben dem Kontrast zwischen dem vorhergesagten Antonymwort (z.B. *schwarz ist das Gegenteil von weiß*) und einem stattdessen präsentierten Nicht-Antonym auch die lexikalische Relatiertheit innerhalb der Gruppe der unerwarteten Nicht-Antonyme von Interesse war. Ein Nicht-Antonym war entweder relatiert (z.B. *schwarz ist das Gegenteil von blau*) oder unrelatiert (z.B. *schwarz ist das Gegenteil von nett*) zum erwarteten Antonym. Wie vorhergesagt interagierte die P300 mit der gewählten Präsentationsmethode. Das auditive Experiment 1 und das mit der RSVP-Methode durchgeführte Experiment 3 zeigten nahezu identische Ergebnisse, während die natürliche Leseumgebung in Experiment 2 zu einem deutlich veränderten EKP-Muster führten. Bei auditiver und bei RSVP-Präsentation zeigten die durch den Satzkontext hochgradig erwarteten Antonyme eine P300, während relatierte und unrelatierte Nicht-Antonyme eine N400 im gleichen Zeitfenster aufwiesen. Dieser N400-Effekt war gefolgt von einer P300 für die Nicht-Antonyme. Beim Lesen zeigte sich weder parafoveal (für die letzte Fixation vor dem kritischen Wort) noch foveal (für die erste Fixation auf dem kritischen Wort) eine P300 für Antonyme, wohingegen sich das N400-P300 Muster für Nicht-Antonyme foveal auf dem kritischen Wort zeigte. Parafoveal zeigte sich außerdem ein Priming-Effekt für Antonyme und relatierte Nicht-Antonyme, der in beiden Fällen zu reduzierten Amplituden der N400 führte. Keine dieser N400-Komponenten wies eine Latenz auf, die deutlich verschieden von der bislang berichteten durchschnittlichen Latenz der N400 war.

Es wurde argumentiert, dass die parafovealen N400-Priming-Effekte automatische „spreading activation"-Prozesse zwischen benachbarten Einträgen im mentalen Lexikon widerspiegeln, die dazu führen, dass die durch den Kontext hervorgerufene Voraktivierung der Wortform des erwarteten Antonyms auf semantisch relatierte Nicht-Antonymformen überspringt. Die foveale N400 zeigt hingegen den aus der Literatur bekannten „cloze probability“-Effekt, wonach ein kontextuell unerwartetes Wort die Amplitude der N400 erhöht. Die P300 wurde als Indikator für die Kategorisierung/ Evaluation des Zielwortes als Antonym oder Nicht-Antonym interpretiert. Die P300 wird blockiert, da prä-lexikalische Informationen (orthographische Wortformen) parafoveal verfügbar sind, semantische Informationen aber erst foveal über die erste Fixation auf einem Wort aufgenommen werden. Das bedeutet, dass zwar zu jedem Fixationspunkt ein Kategorisierungsversuch unternommen wird, dass aber jeweils Informationen aus der prä-lexikalischen oder der semantischen Ebene nicht verfügbar sind. Parafoveal stehen semantische Informationen noch nicht zur Verfügung und foveal ist die Kategorisierung aufgrund von parafovealen orthographischen Informationen bereits abgeschlossen. Somit basiert die Kategorisierung des Antonyms auf unvollständigen Informationen, was die Entscheidung mit einem ungewissen Status versieht.

Dieses Ergebnis deutet darauf hin, dass die der P300 zugrunde liegenden Mechanismen auf eine partielle Überlappung in der Verarbeitung der verschiedenen Ebenen der Worterkennung angewiesen sind und dass eine serielle Verarbeitungsweise (mit eventueller Addierung der Ergebnisse von vorherigen Ebenen) nicht zu einer P300 für hochgradig vorhergesagte Wörter führt – zumindest wenn ein Wort auf mehr als nur einer Ebene vorhergesagt wird. Sowohl die Wortform als auch die Bedeutung der Antonyme wird vorhergesagt, sodass der visuelle Input die Vorhersagen in der Form- und der Bedeutungsebene gleichzeitig erfüllen muss, damit die Kategorisierung der Antonyme vollständig und gewiss ist. Die zeitliche Trennung von Informationen zur Wortform und Wortbedeutung während des Lesens verhindert dies allerdings. Da Nicht-Antonyme, unabhängig von ihrer semantischen Relatiertheit zum Prime, nicht vorhergesagt sind, werden die Kategorisierungsprozesse erst angestoßen, wenn ihre lexikalische Verarbeitung abgeschlossen ist bzw. die Kategorisierung von sämtlichen Nicht-Antonymen kann überhaupt erst beginnen, wenn ihre Bedeutung und ihre Angemessenheit im Satzkontext bekannt sind. Daher folgt die P300 der N400 in diesen Fällen. Das Auftreten einer parafovealen N400 und das Ausbleiben der P300 für Antonyme sind die wichtigsten Ergebnisse dieses Experiments.

Zwei zusätzliche Experimente mit unterschiedlichen Stimuli wurden durchgeführt, um die Reliabilität der parafoveal erzeugten N400 zu überprüfen. Experiment 4 untersuchte Strukturen, in denen die präverbalen Argumente kasusambig waren (z.B. *Dass Elfriede Statisten vergisst, hat viele meist verärgert*). Erst das Verb machte die Zuweisung von Nominativ und Akkusativ zu den jeweiligen Verbargumenten eindeutig. Unabhängig von der Auflösung der Kasusambiguität sorgte die zusätzlich eingeführte Korrelation zwischen Typ der Nominalphrase (spezifischer Eigenname vs. unspezifische generische Nominalphrase) und grammatischer Funktion dafür, dass die beiden kasusambigen Nominalphrasen hinsichtlich ihrer Subjekt- oder Objektfunktion eindeutig interpretierbar waren. So war beispielsweise der Eigenname immer das Subjekt des Satzes, er stand aber entweder an erster oder an zweiter Position (*dass Elfriede Statisten vergisst* vs. *dass Statisten Elfriede vergisst*). Neben der Wortabfolge wurde die Belebtheit des Objekts variiert (*dass Elfriede Statisten vergisst* vs. *dass Elfriede Tabletten vergisst*). An der Position des zweiten Arguments zeigten sich N400-Effekte für beide Manipulationen. Zwei belebte Argumente erzeugten einen Interferenzeffekt (erkennbar an einer erhöhten N400-Amplitude) und ein satzinitiales Subjekt führte zu einer Erweiterungs-N400 auf der folgenden Objektposition, da das anfängliche Fragment *dass Elfriede* zunächst als intransitiv interpretiert wurde. Auf dem Verb zeigte sich eine späte Positivierung für den belebtheitsbedingten Interferenzeffekt sowie eine späte Positivierung für objektinitiale Sätze. Letztere wurde als P300-Komponente klassifiziert, da sie die Evaluation der Wohlgeformtheit von objektinitialen Sätzen reflektiert. Interessanterweise waren alle diese Effekte auf foveale Verarbeitungsprozesse beschränkt. Es wurden keine parafoveale N400 oder eine Positivierung gefunden.

Ein weiteres Experiment wurde durchgeführt, um die Hypothese zu überprüfen, dass die Kasusambiguität zu erhöhtem fovealen Aufwand geführt haben könnte, welcher wiederum die parafoveale Vorschau auf das nächste Wort verhindert hat. In Experiment 5 wurden die Abfolge eines Nominativ- und eines Dativarguments in Passivstrukturen (*Gestern wurde das Pferd dem Reiter verkauft, erklärte Anna* vs. *Gestern wurde dem Reiter das Pferd verkauft, ...*) sowie die Belebtheit des Nominativarguments (*Gestern wurde das Pferd dem Reiter verkauft* vs. *Gestern wurde der Sattel dem Reiter verkauft*) variiert. Es zeigten sich keinerlei EKP-Effekte für die Wortstellungsvariation. Die Amplitude der N400 war jedoch erhöht, wenn ein belebtes Nominativargument auftrat. Dieser Effekt wurde auf widersprüchliche Prominenzinformationen zurückgeführt. Sowohl die Kasusinformation als auch die Belebtheit stehen mit der Zuweisung der Undergoerrolle (bzw. Patiens) für dieses Argument in Konflikt, da belebte Nominativargumente im Deutschen bevorzugt Actor-Argumente (bzw. Agens) sind. Während sich dieser Konflikt auf der zweiten Argumentposition foveal zeigte, war er für die erste Argumentposition (also in Nominativ-vor-Dativ-Strukturen) bereits parafoveal, d.h. für die letzte Fixation vor dem Nominativargument, erkennbar. Zusammengefasst sprechen diese Ergebnisse dafür, dass parafoveale N400-Effekte im natürlichen Lesen auftreten, wenn die Fixation, die die relevanten Informationen liefert, sich nicht auf Satzteilen befindet, welche selbst Verarbeitungsschwierigkeiten verursachen. Diese Voraussetzung war in Experiment 2 und 5 erfüllt, jedoch nicht in Experiment 4. Ob die parafoveale N400 mit prälexikalischen oder semantischen Verarbeitungsschritten korreliert, scheint zudem abhängig vom gewählten Kontext zu sein.

Hinsichtlich der Korrelation zwischen EKP- und Blickbewegungseffekten ergaben die kombinierten Messungen folgendes Gesamtergebnis. Foveale EKP-Effekte korrelierten meist mit erhöhten Fixationszeiten in diversen Fixationsmaßen. Es bestand die Tendenz, dass N400-Effekte, die auf lexikalisch-semantische Manipulationen zurückzuführen waren, mit erhöhten Zeiten für die erste Fixation auf dem entsprechenden Wort einhergingen.

Wortstellungsmanipulationen führten zu vermehrten Fixationen auf dem kritischen Wort, während die Dauer der ersten Fixation nicht betroffen war. Die parafovealen N400-Effekte waren nicht von Änderungen in den entsprechenden Fixationszeiten begleitet. Dieses Muster deutet darauf hin, dass Blickbewegungen beim Lesen nicht nur kognitive Verarbeitungsschwierigkeiten widerspiegeln. Wäre dies so, hätte auch die parafoveale N400 zu erhöhten Fixationszeiten führen müssen. Es scheint stattdessen eher so zu sein, dass Effekte in den Augenbewegungen aus einem dynamischen Zusammenspiel von neuer Informationsaufnahme und Informationsstopp resultieren. Wenn die aus der aktuellen Fixation gewonnenen Informationen für die Überwindung einer Verarbeitungsschwierigkeit nicht ausreichen und wenn damit gerechnet werden kann, dass der weitere Satzkontext hilfreiche Informationen bereithält, so ist die Verlängerung der Fixation eher unwahrscheinlich. Diese Annahme erklärt die ausbleibenden parafovealen Blickbewegungseffekte. Wenn andererseits Verarbeitungsschwierigkeiten auftreten, die lokal lösbar sind, aber mehr Zeit erfordern, so zeigen sich erhöhte Fixationszeiten (oder Refixationen auf dem kritischen Wort). Interessanterweise schienen umfassendere Blickbewegungsmaße nicht nur mit linguistischen Verarbeitungsschritten zu variieren, sondern auch zusätzlich mit allgemein kognitiven Prozessen. Experiment 2 zeigte beispielsweise das Veränderungen in der Latenz und der Amplitude der P300 von Änderungen im Muster umfassender Blickbewegungsmaße (z.B. die Gesamtzeit, für die ein Wort fixiert wird) begleitet war.

Abschließend bleibt festzuhalten, dass die vorliegende Dissertation auch eine Reihe an offenen Fragen aufwirft. Die vorgestellten Experimente konnte die Frage, welche Manipulation zur Verlängerung der ersten Fixation auf dem kritischen Wort und welche zu weiteren Refixationen führen, nicht abschließend klären. Außerdem bestätigen die Ergebnisse von Experiment 5 frühere Befunde aus der Literatur, wonach neuronale Korrelate und Blickbewegungskorrelate des Sprachverstehens nicht immer konvergieren. Während es für

Dativ-vor-Nominativ-Strukturen in Experiment 5 keinen sichtbaren EKP-Effekt für den initialen Dativ gab, zeigten die Blickbewegungen an dieser Position einen Nachteil für diese Abfolge gegenüber der Nominativ-vor-Dativ-Struktur. Neuronale und die (behavioralen) okulomotorischen Reaktionen auf Probleme in der Sprachverarbeitung scheinen somit auf teilweise unterschiedliche Aspekte im Verlauf der Verarbeitung bestimmter Strukturen anzusprechen (z.B. grammatische Regeln vs. strukturelle Frequenz). Trotz dieser offenen Fragen, die in weiteren Experimenten zu beantworten sind, konnten die vorliegenden Experimente belegen, dass sowohl die N400 als auch die P300 reliable Indikatoren von Sprachverarbeitungsprozessen sind. Beide Komponenten interagieren jedoch mit der gewählten Präsentationsmethode. Einige der N400 zugrunde liegenden Prozesse können im natürlichen Lesen besser voneinander unterschieden werden als in auditiven oder RSVP-Experimenten und das Ausbleiben der P300 für vorhergesagte Wörter zeigt an, welche allgemein kognitiven Verarbeitungsschritte im Sprachverstehen zusätzlich zur rein sprachlichen Worterkennung eine Rolle spielen.

Curriculum vitae

Name:	Franziska Kretzschmar
Geburtsdatum:	25. Dezember 1980
Geburtsort:	Potsdam
Nationalität:	Deutsch

Ausbildung

09/2010	Promotion zum Dr. phil. an der Philipps-Universität Marburg
02/2010 – 05/2010	Stipendium zur Fertigstellung der Promotion am Lehrstuhl für Neurolinguistik von Prof. Dr. Ina Bornkessel-Schlesewsky an der Philipps-Universität Marburg
02/2007 – 01/2010	Promotionsstipendium der Deutschen Forschungsgemeinschaft (DFG) im Rahmen des Graduiertenkollegs „Neuronal Representation and Action Control“ (GK 885/1)
10/2003 – 12/2006	M.A. (Magister Artium); Hauptstudium in den Fächern Deutsche Sprache und Literatur sowie Politikwissenschaft an der Philipps-Universität Marburg
10/2002 – 07/2003	Erasmusstipendium an der University of Aberdeen; Fächer: German Studies und Politics and International Relations
10/2000 – 09/2002	M.A. (Magister Artium); Grundstudium in den Fächern Germanistik und Politikwissenschaft an der Ernst-Moritz-Arndt-Universität Greifswald

Wissenschaftlicher Werdegang

seit 06/2010	Wissenschaftliche Mitarbeiterin am Department of English and Linguistics an der Johannes Gutenberg-Universität Mainz
2009 – 2011	Lehrbeauftragte im Bereich Allgemeine Sprachwissenschaft und im Studium Generale (Stiftungsprofessur „Sprache und Gehirn“ 2010) an der Johannes Gutenberg-Universität Mainz
02/2007 – 09/2010	Doktorandin an der Philipps-Universität Marburg

Publikationen

Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., Roehm, D., & Schlesewsky, M. (2011). Think globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and Language, 3,* 133-152.

Kretzschmar, F., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). Parafoveal vs. foveal N400s dissociate spreading activation from contextual fit. *Neuroreport 20*, 1613-1618.

Kretzschmar, F., Bornkessel-Schlesewsky, I., Staub, A., Roehm, D., & Schlesewsky, M. (in press). Prominence facilitates ambiguity resolution: On the interaction between specificity and word order in syntactic reanalysis. In M. J. A. Lamers & d. P. Swart (Eds.), *Case, Word Order, and Prominence: Psycholinguistic and theoretical approaches to argument structure.* Berlin: Springer.

Selbständigkeitserklärung

Hiermit erkläre ich, dass ich meine Dissertation mit dem Titel
„The electrophysiological reality of parafoveal processing: On the validity of language-related ERPs in natural reading“
selbständig ohne unerlaubte Hilfe angefertigt, keine anderen als die angegebenen Hilfsmittel verwendet und alle Stellen, die anderen Quellen dem Sinn nach entnommen sind, durch Angabe der Herkunft kenntlich gemacht habe. Alle wörtlich entnommenen Stellen habe ich als Zitate gekennzeichnet.

Die Dissertation hat in ihrer jetzigen oder einer ähnlichen Form weder ganz noch in Teilen einer in- oder ausländischen Hochschule zu Prüfungszwecken vorgelegen.

________________________ ______________________________

Ort, Datum Franziska Kretzschmar