Discourse-based lexical anticipation

The nature and contextual basis of
predictions in language comprehension

Cover design by Martijn Wokke & Marte Otten
Printed by Print Partners Ipskamp B.V., Enschede

The research presented in this thesis was supported by an NWO Innovation Impuls Vidi grant to Jos van Berkum.

UNIVERSITY OF AMSTERDAM

NWO
Netherlands Organisation for Scientific Research

Discourse-based lexical anticipation

The nature and contextual basis of
predictions in language comprehension

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college
voor promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op dinsdag 1 april 2008, te 12:00 uur

door

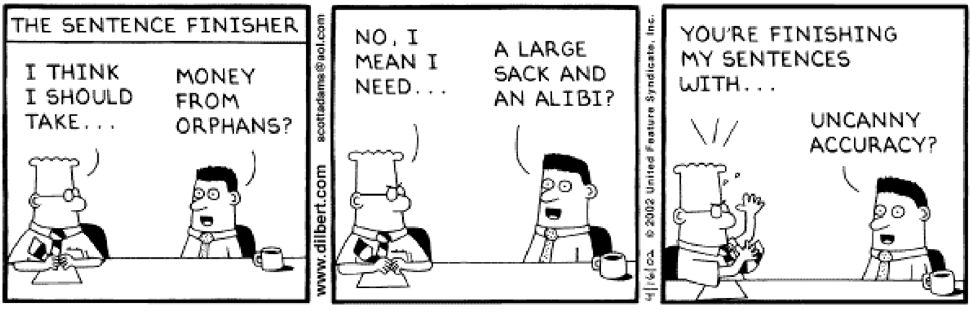
Marte Otten

geboren te Breda

Promotiecommissie

Promotor:	Prof. dr. J.G.W. Raaijmakers
Co-promotor:	Dr. J.J.A. van Berkum
Overige Leden:	Prof. dr. G. Kempen Prof. dr. A.E. Baker Prof. dr. A.M.B. de Groot Prof. dr. J.M.W. Murre Dr. D.J. Chwilla Dr. J.C.J. Hoeks

Faculteit der Maatschappij- en Gedragswetenschappen


THE SENTENCE FINISHER
I THINK I SHOULD TAKE...
MONEY FROM ORPHANS?
scottadams@aol.com
www.dilbert.com
NO, I MEAN I NEED...
A LARGE SACK AND AN ALIBI?
© 2002 United Feature Syndicate, Inc.
YOU'RE FINISHING MY SENTENCES WITH...
UNCANNY ACCURACY?

Contents

Chapter 1	**Introduction**	1
Chapter 2	**Discourse-based word anticipation during language processing: Prediction or priming?**	11
Chapter 3	**Great expectations: Specific lexical anticipation influences the processing of spoken language**	43
Chapter 4	**The role of working memory capacity in lexical prediction**	59
Chapter 5	**Does appearance matter? A self-paced reading study exploring whether linguistic predictions extend to the visual form**	83
Chapter 6	**What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse dependent N400 effect**	99
Chapter 7	**General discussion and summary**	121
	References	129
	Appendices	145
	Nederlandse samenvatting	151
	Dankwoord	159

Chapter 1

Introduction

> "Each word is doubly awakened; once from without by the lips of the talker, but already before that from within by the premonitory processes irradiating from the previous words, and by the dim arousal of all processes that are connected with the 'topic' of the talk."
>
> *[William James, The Principles of Psychology, 1890]*

Prediction in language comprehension

Language comprehension is highly incremental: when we hear or read a sentence or a combination of sentences, we immediately and continuously extract meaning. Concurrent use of the extracted meaning to anticipate how the sentence or story will continue could benefit the comprehender. If the new, continuously incoming information matches the expectancy, comprehension can be limited to rudimentary processing. If the new information does not match the prediction, this can provide the language comprehension system with an early 'flag' so that attention can immediately be directed to processing this unexpected part of the sentence.

But can we indeed make such predictions? Our own every-day experience shows that we can sometimes finish other peoples sentences, if our conversational partner falters mid-sentence for some reason. But, do these sentence-completions also occur when time is limited, concurrent with our normal comprehension process? After reading a sentence like "*The UPS man delivered the...*" the word *package* is processed more easily than the less expected word *box*: it requires shorter naming latencies (Hess, Foss, & Carroll, 1995; Traxler & Foss, 2000) and is read faster (Cook & Myers, 2004; Morris, 1994). These effects could result from the fact that the reader or listener has pre-activated the word *package*, based on the constraining sentence that they have just read. However, these effects could also occur simply because the meaning of *package*, once unlocked by having seen the word, is more easily integrated in the context than *box*, for example because *package* fits the "UPS delivery"-scenario better.

Using innovative experimental designs, several researchers have shown in recent years that people sometimes indeed pre-activate predictable words that are likely to follow. In an experiment by Kamide, Altmann and Haywood (2003) participants were looking at a scene such as the one depicted in Figure 1.1. When they heard the sentence "*The woman will spread …*" their gaze shifted to the only spreadable object in the display, namely the butter. Importantly, this happened before they had actually heard the word butter. In

Figure 1.1 *Example of a display used in (Kamide, Altmann, & Haywood, 2003).*

a less constraining context ("*The woman will slide …* ") no such preference for a specific object was detected. This indicates that the listeners were thinking about butter before the word "butter" had actually been pronounced, i.e. that they predicted that the word butter would follow. Of course, the visual display already limited the possibilities for the words that could follow. So, even though the results are certainly suggestive, this experiment does not provide definite evidence that prediction occurs in normal language comprehension.

More direct evidence for online prediction in language comprehension was provided by experiments by Van Berkum and colleagues (2005). Most of you who read this introduction are native speakers of Dutch. It could be that after you have read

De inbreker wist precies waar hij de geheime familiekluis moest zoeken. Deze bevond zich namelijk achter een…

you are thinking about the word *schilderij* (see Figure 1.2 for an approximate English translation of this story). At least, an offline cloze test indicates that the majority of readers finishes this story with this word. Van Berkum et al. looked for prediction-related ERP effects not on the noun, but on an adjective *preceding the noun*. In Dutch, adjectives in indefinite singular noun phrases have a suffix that depends on the arbitrary, lexically memorized gender (Van Berkum, 1996, Ch. 2) of the noun they precede. Adjectives that modify a common-gender noun carry an –e suffix (e.g., "*oude boekenkast*"), whereas adjectives modifying a neuter-gender noun are not inflected (e.g., "*oud schilderij*"). Van Berkum et al. (2005) reasoned that to the extent that listeners strongly anticipated a specific noun (*schilderij*), an adjective with a mismatching gender suffix ("*oude*") would come as an unpleasant surprise, and might as such elicit a differential ERP effect relative to a prediction-consistent control ("*oud*"). Adjectives with a prediction-inconsistent inflection indeed elicited a differential event-related potential (ERP, see Box 1): a small but significant positivity emerging in the 50-250 ms after acoustic onset of the

Box 1: On-line measures of language comprehension

Studying the way listeners finish other peoples faltering sentences during conversation, or how readers finish an incomplete sentence in a pencil-and-paper cloze test seems helpful to examine whether people actually pre-activate a word while they are reading or listening. However, in both cases people are prompted, either by instruction or by social customs, to think about a possible continuation, and they have relatively much time to do so. To evaluate the *ongoing* process of language comprehension and the predictive processes involved, these processes need to be monitored *on-line*. For the studies reported in this thesis we have used two different methods.

Self-paced reading

In this paradigm, participants read sentences or stories word for word (or phrase by phrase, or sentence by sentence). The participant can signal that he/she has finished reading a word by pressing a button, so that the word on the screen is replaced by the next word. This provides a reading time for each individual word. These reading times are thought to reflect the underlying cognitive processes associated with the processing of that word.

Event-related potentials

The electrical activity of the brain can be recorded with electrodes that are placed on the scalp. This record of fluctuating voltage across time is the electroencephalogram (EEG). The event-related potential or ERP is that part of the signal that is related to an external event. At the scalp, such an ERP is much smaller (5-10 μV) than the background EEG (50-100 μV). By averaging several ERPs that are elicited by similar stimuli, most of the background EEG and environmental noise can be averaged out. The rationale is that the event-related activity is time-locked to the appearance of the stimulus, but the background EEG and environmental noise are not. In the averaging process, unrelated noise present from the individual trials will cancel each other. The resulting average ERP waveform reflects the sensory and cognitive brain-processes related to the stimulus, up to the millisecond.

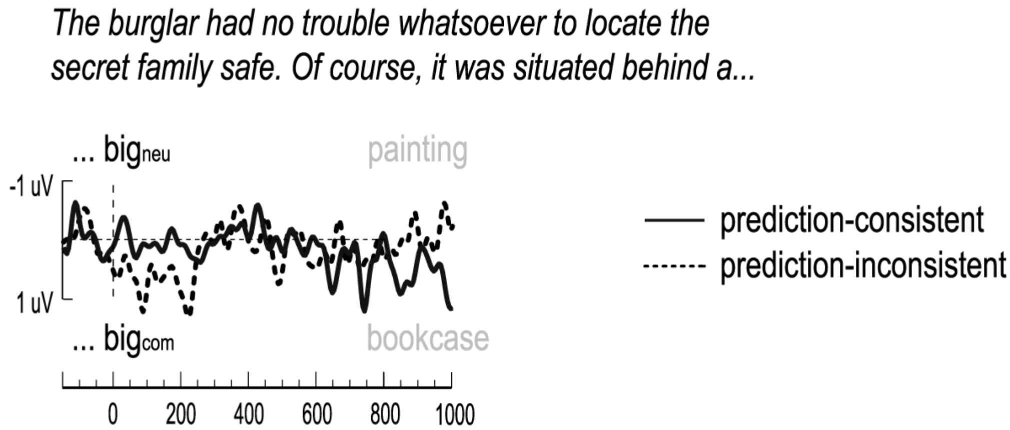


Figure 1.2 *ERPs to spoken adjectives with an inflection that was in line (solid line) or inconsistent (dotted line) with the gender of the predictable noun, timelocked to the onset of that inflection (based on data from Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005).*

inconsistent inflection (see Figure 1.2). This ERP effect occurred before any noun had been presented, at a point in time where both gender inflections were equally correct. Van Berkum et al. therefore took this effect as evidence that their listeners had indeed pre-activated a specific word and its lexical features, like gender, based on discourse information. In a follow-up self-paced-reading study (Van Berkum et al., 2005, Experiment 3), prediction-inconsistent adjectives also slowed down the reading process.

A series of studies by Wicha and colleagues (Wicha, Bates, Moreno, & Kutas, 2003; Wicha, Moreno, & Kutas, 2003, 2004) with Spanish single sentence stimuli, showed that specific lexical prediction is not limited to longer stories or spoken materials. Prediction-inconsistent gender-marked determiners elicited an enhanced negativity between 300-500 ms (spoken sentences: Wicha, Bates et al., 2003; written sentences: Wicha, Moreno et al., 2003). In these two experiments the expected and unexpected articles were followed by line drawings of the expected or an unexpected concept. When all the stimuli were presented in writing, unexpected articles evoked an enhanced positivity between 500 and 700 ms (Wicha et al., 2004). In an experiment that used the fact that in English indefinite determiners differ depending on the initial phoneme of the word that follows (a/ an), DeLong, Urbach and Kutas (2005) found that determiners that were not in line with the expected word elicited a negative shift between 300 and 500 ms compared to the prediction-consistent determiners.

Taken together, these results suggest that we are able to make very specific predictions when we are processing incoming language. The experiments collected in this dissertation explore the nature and extent of these specific lexical predictions.

The contextual basis of prediction

Lexical and scenario-based association

The results reviewed above show that people pre-activate lexical information of words that they deem likely to follow. One relevant process underlying these predictions could be automatic activation of lexical-semantic information by words in the discourse (as already suggested by William James, as you van see in the quote at the beginning this chapter). According to the spreading activation account (Collins & Loftus, 1975; Dell, 1986; Meyer & Schvaneveldt, 1976; Swinney, 1979) activation feeding from one representation in the lexicon to other representations that share a connection causes priming (i.e. facilitated processing of the related word). *UPS* can thus prime *package* because the two words co-occur frequently, and this is captured in the strength of the connection between them. The greater the strength of the connection, the more activation is fed from one representation to the other when one becomes activated.

There is extensive evidence that lexical association facilitates processing. Words are processed faster and more accurately when they follow a related prime (see Neely, 1991 for a review) and the N400 is attenuated for these related words (cf. Bentin, McCarthy, & Wood, 1985). The priming effects of a related word are also visible when the word pairs are presented as part of a coherent sentence (Camblin, Gordon, & Swaab, 2007; Carroll & Slowiaczek, 1986; Hoeks, Stowe, & Doedens, 2004; Van Petten, 1993; Van Petten, Weckerly, McIsaac, & Kutas, 1997).

Automatic pre-activation is not necessarily based purely on lexical priming. Models of text comprehension and memory suggest that the words in a text can also provide semantic constraints via the activation of related information stored in long term memory (Kintsch, 1988; McKoon & Ratcliff, 1992; Sanford, 1990). When you read about a UPS delivery, this could activate not only those words in your lexicon that are related to the word UPS, but also the world-knowledge about deliveries that is stored in your long-term memory (i.e. the man from UPS comes over to your house in his brown van, rings the bell and asks you to sign for your package). According to the 'resonance model' (Cook, Halleran, & O'Brien, 1998; Myers & O' Brien, 1998;

Myers, O'Brien, Albrecht, & Mason, 1994), individual concepts from the linguistic input send out a signal to long term memory. Concepts in memory then resonate as a function of their relatedness to the input, based on the overlap between the semantic and contextual features of the concepts involved. Eventually, those concepts that have the highest level of activation enter working memory. The resonance process is assumed to be fast-acting and autonomous (or "dumb" (Myers & O' Brien, 1998))

Message-level processing vs. automatic activation

Comprehending the actual meaning of a sentence is much more than extracting the scenario-relevant gist, or a simple addition of individual words. We are able to understand that "*The UPS man delivered the ...*" is fundamentally different from "*The man delivered UPS the ...*" even though the same words are present. It is indisputable that we are able to extract the complete and correct meaning of both these utterances. But is this complete message-level representation, recomputed with each new piece of information, also what drives the prediction of upcoming words? Or are these anticipatory processes completely based on the abovementioned automatic activation? The evidence for the role of message-based processing in early language comprehension is mixed.

Several experiments have shown that the message of an utterance can overrule the facilitating effects of the primes that are present in that sentence or story (Morris, 1994; Traxler, Foss, Seely, Kaup, & Morris, 2000; Van Petten, Coulson, Rubin, Plante, & Parks, 1999). For example, Morris (1994) showed his participants sentences that contained several primes. When the message of the entire sentence was in line with the individual primes, such as in "*The gardener talked as the barber trimmed the moustache.*", the related word *moustache* was read faster. In sentences where the actual message of the sentence did not support the appearance of the primed word, i.e. "*The gardener talked to the barber and trimmed the moustache*", this facilitatory effect disappeared.

However, other experiments have shown that scenario-mediated or lexical association can play a role in discourse comprehension even when the activated information is irrelevant to, or at odds with, the actual message of that discourse (Duffy, Henderson, & Morris, 1989; Garrod & Terras, 2000; O'Seaghdha, 1997). Garrod and Terras (2000), for example, showed that the word "*pen*" is initially just as effectively integrated when presented in a sentence following "*The teacher wrote a letter*" as it is after the sentence "*The teacher wrote the exercise on the blackboard*". Only in regression path analysis and second pass reading times a significant difference was observed between the appropriate and inappropriate contextual message. This indicates that participants did not at first notice the message-level incongruence of "pen", which in turn suggests that in some cases (scenario-mediated) automatic activation can overrule the actual message of the discourse.

It thus seems that both message-level processing and lexical/ scenario-based association play a role when we try to understand and integrate incoming linguistic information. This suggests that both levels of analysis can play a part in predictive processing. In the following chapters, we explore whether prediction is purely based on automatic activation by lexical associations or related scenarios, or whether the actual message of the discourse also plays a role.

Outline of this dissertation

The central question addressed in this dissertation concerns the contextual basis of specific lexical predictions: are these predictions a simple by-product of relatively "dumb" automatic activation processes, or are they rooted in a more comprehensive representation of the predictive discourse?

In chapter 2, two ERP experiments are reported that approach this question from two different angles. The first experiment tests whether unexpected and incoherent words are interpreted as more unexpected and incoherent when a reader has made a (message-based) prediction. In the second experiment inflected adjectives that were not in line with the gender of the predicted word were used to test for the presence of predictions (as in Van Berkum et al., 2005). These prediction-probes were presented in stories that were either message-level constraining or that contained the same potential prime words, but were not message-level predictive. A replication of this second experiment with spoken materials is reported in chapter 3. The ERP experiment reported in Chapter 4 focuses on possible differences between people with larger and smaller working memory capacity in their ability to a) predict upcoming words and b) make message-based predictions.

Previous experiments have shown that specific lexical predictions involve semantic and lexical features of the upcoming word (Van Berkum et al., 2005; Wicha et al., 2004) and perhaps also acoustic properties (DeLong et al., 2005). Chapter 5 explores, in a self-paced reading paradigm, whether specific linguistic predictions can also include the exact visual manifestation of the predicted word.

The ERP study reported in chapter 6 investigates whether the classical difference in N400 for expected and unexpected words is completely based on (scenario-based) priming, or whether message-based processes are relevant as well.

Finally, in chapter 7 all findings are briefly summarized and discussed.

Chapter 2

Discourse-based word anticipation during language processing: Prediction or priming?

Language is an intrinsically open-ended system. This fact has led to the widely shared assumption that readers and listeners do not predict upcoming words, at least not in a way that goes beyond simple priming between words. Recent evidence, however, suggests that readers and listeners do anticipate upcoming words 'on the fly', as a text unfolds. In two event-related potentials (ERP) experiments, we examined whether these predictions are based on the exact message conveyed by the prior discourse, or on simpler word-based priming mechanisms. Participants read texts that strongly supported the prediction of a specific word, mixed with non-predictive control texts that contained the same prime words. In Experiment 1A, anomalous words that replaced a highly predictable (as opposed to a non-predictable but coherent) word elicited a long-lasting positive shift, suggesting that the prior discourse had indeed led people to predict specific words. In Experiment 1B, adjectives whose suffix mismatched the predictable noun's syntactic gender elicited a short-lived late negativity in predictive stories but not in prime control stories. Taken together, these findings reveal that the conceptual basis for predicting specific upcoming words during reading is the exact message conveyed by the discourse, and not the mere presence of prime words.

Introduction

The brave knight saw that the dragon threatened the benevolent sorcerer. Quickly he reached for his …

When readers are asked to complete the above mini-story in a paper-and-pencil test, most of them will write down sword before adding anything else. In psycholinguistics, such completion or 'cloze' tests are routinely used to validate the experimental stimuli for some bigger experiment. However, the convergence on a word like sword by itself raises a very interesting issue. If a large number of respondents in a cloze test converge in believing that sword is a very good way to continue this particular incomplete story, this means that the story at hand is relatively predictable at that point. This opens up the interesting possibility that when people are reading through a text, such as a novel or newspaper article, they might actually predict specific upcoming words as the text unfolds. In two event-related brain potential (ERP) experiments, we examine whether such rapid 'on-the-fly' word anticipation is a natural part of text comprehension, and, if it is, how exactly readers generate these predictions.

Of course, *sword* is not the only possible continuation of our example story – the knight could also be reaching for his lance, longbow or dagger in his attempt to rescue the sorcerer. In fact, it will usually not be the case that an incomplete piece of discourse allows for *just one* acceptable word, even in highly constraining contexts. Because of this open-ended character of language, many linguists and psycholinguists tend to be highly sceptical about the existence of discourse-based word prediction. Commenting on a notable exception (Elman, 1990) Jackendoff (2002) recently expressed the general attitude to prediction-oriented research: "One might well predict that what comes after "The big star's beside a little…" is likely to be a noun (though it might be *blue* or *very old*), but that still leaves open some tens of thousands of choices". In other words, the prediction of upcoming words can not really amount to anything useful. One consequence of this prevailing point of view is that so far very few researchers have empirically examined the issue. As a result, textbooks on language comprehension are silent on anticipation and prediction. In all, the idea that readers or listeners might sensibly anticipate or predict specific upcoming words as language unfolds is generally considered to be a non-starter.

The evidence from psycholinguistic experiments, however, suggests otherwise. First, people anticipate upcoming syntactic structure (Kamide et al., 2003; Van Berkum, Brown, & Hagoort, 1999b), and can make predictions about the grammatical role of an upcoming word (Altmann, van Nice, Garnham, & Henstra, 1998; Lau, Stroud, Plesch, & Phillips, 2006). Second, a considerable body of literature shows that readers make predictions about upcoming meaning. A sufficiently constraining sentence or text can lead people to make predictive inferences as language unfolds, so that if they for instance read about a porcelain vase falling from a 20-story building, they infer that it will probably break (Calvo, Castillo, & Estevez, 1999; Campion, 2004 ; Fincher Kiefer, 1993, 1995, 1996; Graesser, Singer, & Trabasso, 1994; Keefe & McDaniel, 1993; Linderholm, 2002; McKoon & Ratcliff, 1992; Murray & Burke, 2003; Murray, Klin, & Myers, 1993; Schmalhofer, McDaniel, & Keefe, 2002). In similar vein, event-related brain potentials research has shown that high-constraint sentences such as “The vegetarian never ate… ” can lead people to expect words from a very specific semantic field such as type of meat (Federmeier & Kutas, 1999a, 1999b; Federmeier, McLennan, De Ochoa, & Kutas, 2002; Kutas & Federmeier, 2000; Nieuwland & Van Berkum, 2006b). ERP research also suggests that readers sometimes expect specific additional information about particular discourse entities (e.g., expect a sentence like “David praised Linda because…” to continue with something about Linda (Van Berkum, Koornneef, Otten, & Nieuwland, 2007)). Although none of these findings provides direct evidence for the prediction of specific words, they do reveal that the comprehension system continuously extrapolates its unfolding syntactic and conceptual analysis, in ways that could also lead to the prediction of a specific upcoming word.

Recently, several ERP studies have shown that when people are listening to or reading through short texts such as the above example about the knight, they indeed also make rapid predictions about *very specific* upcoming words, as the text unfolds. In a study by Van Berkum et al. (2005, experiment 1) participants heard short stories like “The burglar had no trouble locating the secret family safe. Of course it was situated behind a…”, which had been designed to support the prediction of a specific noun (e.g., “painting”). Van Berkum et al. looked for prediction-related ERP effects not on the noun, but on an adjective *preceding the noun*. In Dutch, adjectives in indefinite noun phrases have a suffix that depends on the arbitrary, lexically memorized gender (Van Berkum, 1996, Ch. 2) of the noun they precede. Adjectives that modify a common-gender noun carry an –e suffix (e.g., “oude boekenkast”, “old[e]$_{com}$ bookcase$_{com}$”), whereas adjectives modifying a neuter-gender noun are not inflected (e.g., "oud schilderij", "old[Ø]$_{neu}$ painting$_{neu}$"). Van Berkum et al. (2005) reasoned that to the extent that listeners strongly anticipated a specific noun (e.g., "painting$_{neu}$"), an adjective with a mismatching gender suffix (e.g., "old[e]$_{com}$") would come as an unpleasant surprise, and might as such elicit a differential ERP effect relative to a prediction-consistent control (e.g., "old[Ø]$_{neu}$").

Adjectives with a prediction-inconsistent inflection indeed elicited a differential brain potential response: a small but significant positivity emerging in the 50-250 ms after acoustic onset of the inconsistent inflection. This ERP effect occurred before any noun had been presented, at a point in time where both gender inflections were equally correct. Van Berkum et al. therefore took this effect as evidence that their listeners had indeed pre-activated a specific word and its lexical features, like gender, based on discourse information. In a follow-up self-paced-reading study (Van Berkum et al., 2005, Experiment 3), prediction-inconsistent adjectives also slowed down the reading process. Furthermore, ERP-research by Wicha and colleagues (Wicha, Bates et al., 2003; Wicha, Moreno et al., 2003; Wicha et al., 2004) and DeLong and Kutas (2005) has shown that these predictive processes also play a part in the processing of single sentences.

In all, the recent evidence converges to suggest that, when comprehending sufficiently constraining yet natural fragments of discourse, listeners and readers do anticipate upcoming words 'on the fly', as the text unfolds. This means that the prevailing scepticism with respect to the prediction of specific upcoming words may well be unwarranted. However, an important issue about the nature of such prediction remains to be resolved. In prior research on discourse-based word prediction (Van Berkum et al., 2005), the wider discourse was assumed to support the anticipation of specific upcoming words via a *message-level representation*, i.e., the exact communicative message of the story up to the critical word. However, a discourse context that is highly predictive towards a specific target word will often also contain primes that are (mildly to strongly) related to that target. When looking at our earlier example ("The brave knight saw that the dragon threatened the benevolent sorcerer."), it could be that the word *sword* is activated as a result of the occurrence of the strongly related word "knight", or the co-occurrence of "knight", "brave" and "dragon", whereas the alternative, *lance*, is less, or not at all, supported by the prime words in the context. This opens up the possibility that the discourse-based prediction effects observed by Van Berkum et al. (2005), as well as the sentential effects observed by Wicha et al. (2004) and DeLong et al. (2005), reflect some form of automatic activation or priming of the expected word based on the presence of related words in the context, rather than true prediction based on the message conveyed so far. We examined this possibility in two experiments.

Experiment 1A[1]

The experiments discussed before (DeLong et al., 2005; Van Berkum et al., 2005; Wicha et al., 2004) reveal that the processing of adjectives and articles that precede a noun can be affected by text-based expectations about an upcoming noun. As such, they clearly indicate that readers and listeners can exploit the (message- and/or word-level) context to predict upcoming words. However, the experimental paradigms used so far have an important limitation: they strongly depend on language-specific features in order to test for the prediction of words. The design used by Van Berkum et al. (2005) as well as Wicha et al. (2004) requires a language with a grammatical gender system (and, in addition, one in which the gender of the words used in the experiment is arbitrary, and can not be derived from word meaning). De Long et al. relied on the *a/an* alternation, an idiosyncratic feature of the English language. Although other languages with similar features exist, such agreement-based designs do impose an important practical constraint on research into the mechanisms of word prediction.

One might think that word prediction should above all be measurable at the predicted noun itself, relative to some less expected control noun. And indeed there is extensive evidence that the N400, a negative deflection in the ERP peaking approximately 400 ms after the presentation of the stimulus, is larger for unexpected (or 'low-cloze') words compared to expected (or 'high-cloze') words. This effect has been observed in single sentences (Kutas & Hillyard, 1984; Van Petten et al., 1999) as well as in short texts (Otten & Van Berkum, 2007; Van Berkum et al., 2005). The difference in N400 amplitude could reflect a direct processing benefit of discourse-based word prediction. However, unexpected words do not only differ from expected words in terms of their expectedness, but also in how well they fit the wider context regardless of prediction. In our knights-and-dragons example it is possible that *sword* is simply easier to integrate within the given context than *lance* because of what we know about the world, even when a reader or listener

[1] The two experiments we present in this paper were run concurrently in one experimental session. The obvious differences in the critical manipulation and logic, however, necessitate a separate presentation.

forms absolutely no prediction about whether *sword* might follow. Van Berkum, Zwitserlood, Hagoort and Brown (2003, Figure 3) have shown that ease of integration indeed plays a role in language processing, and influences N400 amplitude. In low constraint stories where the cloze value was below 5% (mean cloze probability of 1%) contextually incoherent words evoked a larger N400 compared to contextually coherent words. In this case, the incoherent words did not violate a (specific) prediction or expectation, because the discourse was not predictive enough to support such a prediction. The effect thus reflects the difference in ease of integration for the coherent and incoherent noun.

It is more difficult to test for the effects of prediction without potentially introducing some form of facilitation through contextual integration. To isolate the effects of prediction, the conditions that are compared need to be identical with respect to the ease of integration of the studied target word(s). As a consequence, simply comparing behavioural or electrophysiological measures to expected and unexpected words which are presented in the same (predictive) context can not be taken as unequivocal evidence for lexical anticipation, because expected words are also likely to be easier to integrate in the context. The same argumentation applies when studying the influence of lexical prediction through comparing a predictive and a non-predictive context. Presenting a word that is more expected in one context than in the other will automatically bring along a larger ease of integration in the context where the word is expected.

In the present experiment we circumvented this potential confound of ease of integration by using anomalies. Anomalous words can probe for lexical anticipation without a 'post-lexical' confound, since anomalous words are by definition impossible to integrate. Following a similar, single sentence experiment by Hoeks, Stowe and Doedens (2004) we therefore replaced the expected words with anomalies in predictive contexts, and in non-predictive control stories that were matched on potential prime words.

As can be seen in the example in Table 2.1, the same anomalous word (e.g., "stove") appeared in two discourses which have different levels of contextual constraint. In the Predictive Context the text up to the anomalous word (i.e., up to and including the article "a") strongly suggests a particular completion (*disco*), one that critically hinges on the first sentence. In the so-called prime control context, this context sentence had been modified such that the resulting text no longer strongly suggested a particular word. To make sure that any differential effects of the predictive context would hinge on a message-level representation of the context, and not on a low-level intra-

Table 2.1. An example of the items used in experiment 1A, in the original Dutch version and an approximate English translation. The contextually inappropriate noun is inserted at the location of the expected noun (which in this case is "disco").

Predictive context	Prime control context
Sylvie en Joanna hebben zin om te dansen en te sjansen vanavond. Daarom gaan ze naar een **oven** [disco] waar ze ook erg lekkere cocktails hebben.	Na al het dansen hebben Sylvie en Joanna geen zin meer om te sjansen vanavond. Daarom gaan ze naar een **oven** [disco] waar ze ook erg een rustige chill-out zone hebben.
*Sylvie and Joanna really feel like dancing and flirting tonight. Therefore they go to a **stove** [disco] where they also make very nice cocktails.*	*After all the dancing Joanna and Sylvie really don't feel like flirting tonight. Therefore they go to a **stove** [disco] where they also have a nice and quiet chill-out zone*

lexical priming mechanism, the same potential prime words (e.g., "dancing", "flirting" and "tonight") occurred in both the predictive and the prime control context.

By comparing the ERP evoked by anomalies in a predictive context with equally anomalous words that appear in a non-predictive prime control context we can test whether readers use the message of the discourse to generate expectations about the remainder of the story. Readers that are presented with a story about two girls that go to a stove will generally have hard time incorporating the word "stove" into the preceding discourse. However, if they have already formed a rather stable anticipation about how the story will continue then "stove" is not only generally incoherent, it also mismatches the more or less specific prediction. If this is indeed the case, and if the predictions are based on the message of the discourse, then anomalies that appear in a predictive context should evoke a different ERP than anomalies in a prime control context, because the anomalies that also mismatch a prediction require differential processing. If predictions are not generated online, or if they rely on automatic activation rather than the message of the preceding discourse, then we should observe no difference between the ERP for anomalies in a predictive or a non-predictive context.

Methods Experiment 1A

Participants

24 right-handed native speakers of Dutch (19 female participants, mean age 22, range 18-36 years) participated in the experiment, as part of a course requirement. None had any neurological impairment, had experienced any neurological trauma, or used neuroleptics. None of the participants had participated in the pre-tests conducted during the material construction phase.

Materials

The critical stimuli in this experiment were 80 mini-stories of two sentences each, which were mixed with the 160 mini-stories reported in this paper as experiment 1B. As exemplified in Table 2.1, every story had a highly predictive variant as well as a prime control variant. Both involved the same anomalous critical noun (e.g., "stove") embedded in the same local 'carrier sentence' context (e.g., "Therefore they go to a …"), and this noun was followed by at least two identical words (e.g., "where they also…"). In the critical predictive condition, these carrier sentences were preceded by a context sentence which supported the prediction of a specific word in the second sentence. In the prime control condition, the message was changed such that it was much less predictive at the noun position, while keeping the same potential prime words. These control stories were created using different strategies, which are illustrated by the examples in Appendix 1.

<u>Pretest: Cloze test.</u> Prior to the EEG experiment we checked our predictiveness manipulation in a pencil-and-paper story completion or 'cloze' test. 66 participants (native speakers of Dutch) were shown the mini-stories up to the indefinite article (and thus not including the critical target noun), and they were asked to complete the second sentence. The stories were distributed across lists such that each subject completed each story in only one of its two versions. The 80 critical items were intermixed with highly predictive as well as neutral filler items (among which the items used in experiment 1B). For each item we calculated the cloze value of the predictable word (the proportion of participants who used this word), for both the predictive and the prime control condition. Whereas predictive stories had a relatively high average cloze value (mean = 0.65, sd = 0.18, ranging from 0.40 to 0.95), prime control stories had a much smaller average cloze value (mean = 0.35, sd = 0.15, ranging from 0.00 to 0.50). The difference in cloze value between predictive and control stories was always at least 0.20.

Prestest: Plausibility rating. We conducted an independent rating task to support our assumption that critical nouns were equally anomalous in the predictive and the prime control context. Respondents were 32 native speakers of Dutch who had not participated in the preceding cloze tests. Participants were shown the mini-stories up to and including the anomalous target word. The items were distributed across lists such that each subject rated each story in only one of its variants. The 80 anomalous stories were mixed with 80 similar filler stories ending with nouns that were reasonably to completely acceptable within that context. Participants were asked to rate the acceptability of the last word of each story (the noun) within the preceding context on a scale ranging from 1 (highly anomalous) to 7 (completely acceptable). As intended, the anomalous noun was perceived to be equally anomalous in predictive and prime control stories (average rating of predictive stories: 1.60 (sd = 1.28), average rating of prime control stories: 1.53 (sd = 1.19)).

The 80 items of experiment 1A (40 for each of the conditions shown in Table 2.1) were randomly mixed with the 160 items of experiment 1B. A second list was created by rotating the conditions. Each participant was shown one of these two lists, so that one participant saw all 80 critical stories, but never in more than one condition.

Procedure, EEG recording and analysis

Each participant saw 240 stories, 80 of which were critical for the current experiment. The 160 filler-items did not structurally differ from the experimental items up to the critical region, and were therefore not indicative of whether an anomaly or a congruent continuation would follow (for examples of the stimuli, see appendices 1 and 2). The participants were asked to read for comprehension and were not required to perform any other task. The electroencephalogram (EEG) was recorded from 30 electrode sites, mounted in an elastic cap, each referenced to the left mastoid. Blinks and vertical eye-movements were registered by placing an electrode under the left eye, also referenced to the left mastoid. Electrode impedance was kept below 5 kOhms during the experiment. The EEG was amplified with BrainAmps amplifiers (BrainProducts, München), band-pass filtered at 0.03 Hz-100 Hz and sampled with a frequency of 500 Hz.

During the comprehension task the participants sat in a comfortable chair in a moderately lit room. The stimuli were presented in black 36 point courier new font on a white background on a fast TFT display (Iiyama TXA 3834 MT) positioned approximately 80 cm away form the subject. Before each trial, a fixation cross was shown in the centre of the screen for 2.5 s. Participants were instructed to avoid blinks and eye-movement when the words were presented on screen, and were encouraged to blink when the fixation cross was shown. To signal the start of each trial to the subject, a beep was presented 1 s before the onset of the first word. The stories were then presented word by word. To make this presentation more natural, we designed a *Variable Serial Visual Presentation* (VSVP) procedure in which the presentation duration of each non-critical word varied with its length and position in the sentence (see also Nieuwland & Van Berkum, 2006a; Otten & Van Berkum, 2007; Van Berkum et al., 2007). Non-critical word duration consisted of a standard offset of 187 ms plus and additional 27 ms per letter (with an upper bound of 10 letters for each word). In the present experiment, durations varied from 214 ms for a one-letter word to 450 ms for words consisting of ten or more letters. Between words, the screen went blank for a standard duration of 106 ms. Furthermore, the presentation duration of clause-final words preceding a comma was prolonged with an additional 200 ms. In addition, presentation time for sentence-final words was extended with an extra 293 ms, followed by a 1 s pause until the next sentence began. These various parameters were based on natural reading times (Haberlandt & Graesser, 1985; Legge, Ahn, Klitz, & Luebker, 1997), a subjective assessment of the naturalness of the resulting presentation, and technical constraints imposed by the video refresh rate. For the materials at hand, the average presentation time for all words (including critical words) was 326 ms.

Note that to the extent that critical words, or words close to the critical word, differ in average length, the above procedure will induce unintended shifts in the ERP waveforms (particularly the early components associated with visual word onset and offset). To avoid spurious ERP effects due to these shifts, words whose early components fall in the critical EEG epoch (or baseline) should therefore be equated across condition on their presentation time. In the present study, the critical anomalous words and the two words that followed were presented with a fixed duration of 376 ms, based on the average critical word length across all stories. The word just before the anomalous word was always the same 3-letter indefinite article presented for 268 ms, again with the standard 106 ms interword interval. Participants did not consciously notice the alternation between completely variable and semi-fixed word duration presentation within a single story.

The EEG signals were re-referenced off-line to the average of right and left mastoids. Blinks and eye movements were removed from the data using a procedure based on Independent Component Analysis (ICA) as described by

Jung et al. (Jung, Makeig, Humphries et al., 2000; Jung, Makeig, Westerfield et al., 2000). After that the data were segmented in epochs lasting from 500 ms before critical word onset until 1200 ms after critical word onset. After baseline-correcting the signals by subtracting mean amplitude in the 150 ms preceding critical word onset, we eliminated segments in which the signal exceeded ±100 µV, or which featured a linear drift of more than ±40 µV beginning before the onset of the critical word. As a result 15% of all trials were deleted in both conditions. For each participant the remaining trials (between 61% to 100% of the original amount over participants) were averaged per condition.

Repeated-measure ANOVAs were used to compare the ERPs to

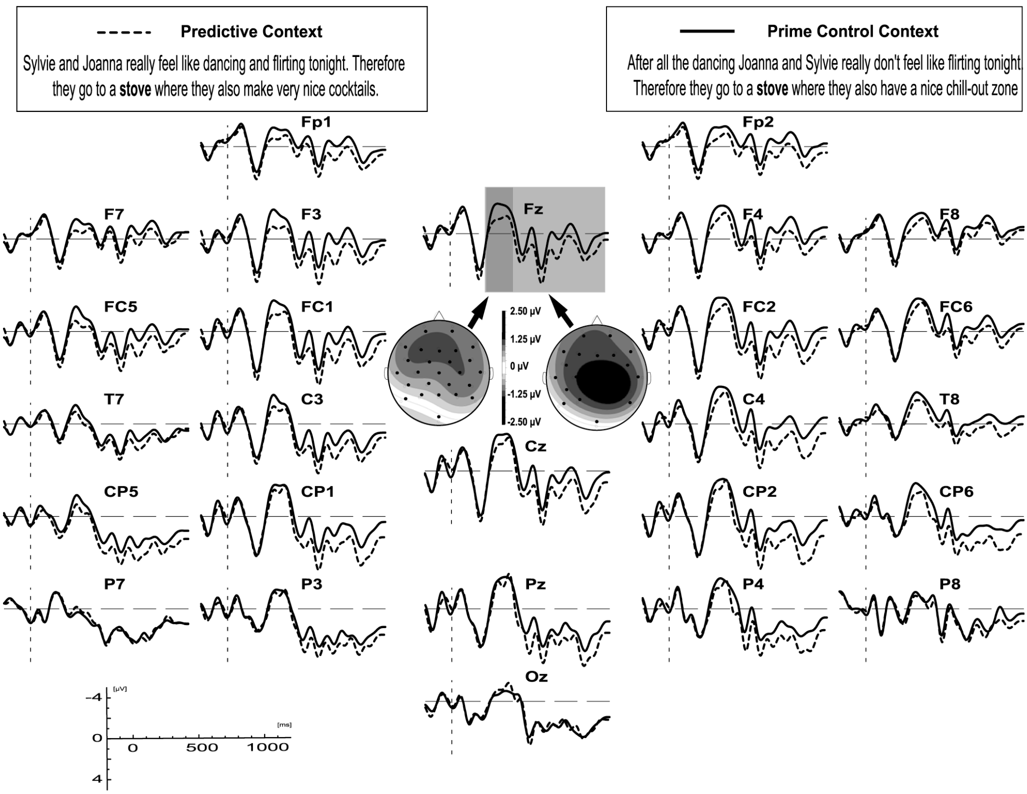


Figure 2.1 *Grand average ERPs elicited by anomalous nouns (right panel) in the predictive (dotted line) and the prime control context (solid line). The two windows of analysis (300-500 ms and 500-1200 ms) are highlighted. The scalp distributions corresponding to the effect of discourse type (predictive discourse – prime control discourse) are depicted for both time intervals. Note that in this and all following figures positive voltage is plotted downwards.*

anomalous words in the predictive context and the control context condition. To assess not only the main effect of context type, but also the possible interaction with electrode position we employed an ANOVA crossing Context (predictive vs. prime control) with a Hemisphere (left/ right) by Anteriority (anterior/ posterior) factor. This analysis involved four quadrants: (1) left-anterior, comprising FP1, F3, F7, F9, FC1, FC5 and FT9; (2) right-anterior, comprising FP2, F4, F8, F10, FC2, FC6 and FT10; (3) left-posterior, comprising C3, T7, CP1, CP5, P3 and P7; (4) right-posterior, comprising C4, T8, CP2, CP6, P4 and P8. Univariate F tests with more than one degree of freedom in the numerator were adjusted by means of the Greenhouse-Geisser or Huynh–Feldt correction where appropriate. Uncorrected degrees of freedom and corrected *P*-values are reported.

Results Experiment 1A

Figure 2.1 displays the ERP evoked by words presented in a prime control context (solid line) and in a predictive context (dotted line) for all electrodes. Anomalous words embedded in highly predictive stories elicit a positive (i.e. downward) shift, starting at approximately 300 ms and lasting until the end of the recording epoch (1200 ms) relative to the same words in prime control stories.

As Figure 2.2 shows, the effect observed for the anomalies is not the result of a difference in overall processing difficulty between the prime control

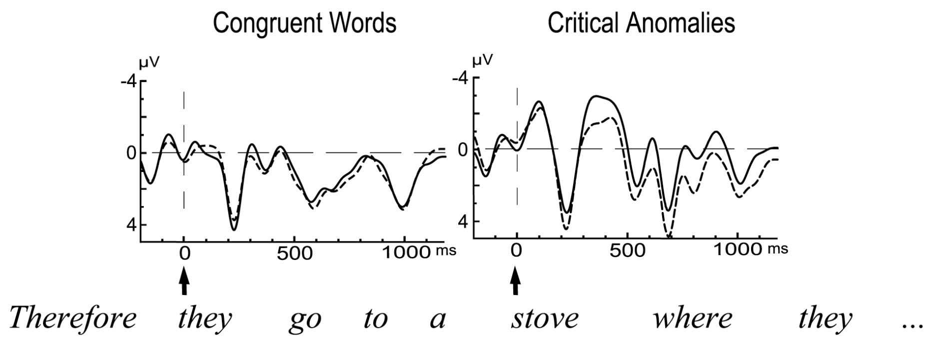


***Figure 2.2** Grand average ERPs elicited by anomalous nouns (right panel) and by congruent words preceding the anomaly (left panel) in the predictive (dotted line) and the prime control context (solid line) for electrode Fz. The differential effect observed at the anomalies is not present in the words preceding the anomaly, indicating that this ERP effect reflects the absence/ presence of a prediction and not an overall difference in processing difficulty between the predictive and prime control context.*

stories and the predictive stories. It is clear that words that *precede* the anomaly in the second sentence (excluding the first word of the second sentence and the indefinite determiner) do not differ in their brain potentials depending on whether they are embedded in a predictive or a control discourse, whereas the critical anomalies do.

Since the scalp distribution of the prediction effect shows a transition from a predominantly frontal distribution to a more central distribution at about 500 ms we analysed the effect using mean amplitudes from two consecutive time windows: 300 to 500 ms and 500 to 1200 ms. Between 300 and 500 ms anomalies presented in a predictive context elicited a significantly more positive ERP than anomalies in the prime context ($F(1, 23) = 4.30$; $p = .05$). The seemingly frontal scalp distribution of this effect, as visible in figure 2.1, is not backed up by a significant interaction between Context and the factor Anterior-Posterior ($F(1, 23) = 2.10$, $p = .16$). Between 500 to 1200 ms anomalies presented in a predictive context also showed a larger positivity than anomalies in the prime context ($F(1, 23) = 8.01$; $p = .01$). This late effect of context type did not interact with either the factor Posterior-Anterior ($F(1, 23) = .80$, $p = .38$) or Hemisphere ($F(1, 23) = 1.34$, $p = .26$).

Discussion Experiment 1A

Anomalies presented in a predictive context evoked a different ERP compared to the same, anomalous word presented in a prime control context. When the anomaly replaced a highly predictable word, it resulted in a long-lasting positive shift which was not present in the prime control context. As the presented noun was judged equally anomalous in both cases, this differential processing effect is unlikely to be due to a difference in 'post-lexical' ease of integration. Furthermore, it cannot be the result of an overall difference in processing difficulty between predictive and prime control stories, because in that case, 'neutral' words which preceded the anomalous noun should have elicited the same differential effect across condition. The ERP effect observed at anomalous nouns is therefore most consistent with the claim that comprehenders can use the unfolding discourse to rapidly form predictions about upcoming words. Furthermore, because prime control stories contained the same potential prime words as predictive stories, this anticipation effect can not easily be explained as a consequence of some form of low-level convergent lexical priming. In all, our findings thus suggest that readers make lexical predictions as the story unfolds, and that they base their predictions on the message of the discourse.

The above reasoning hinges on our assumption that highly anomalous nouns that were rated as (on average) equally anomalous across conditions in a pretest will be equally difficult to integrate after they have been read. However, as pointed out by a reviewer, it is possible that when faced with an anomalous word in a predictive and thus more informative context, readers might try harder to make sense of the word at hand. Indeed, if "My girlfriends really felt like dancing last night so they went to a stove" would be what a teenage daughter is telling her dad, the latter might well entertain the possibility that "stove" is current teenage slang for a disco. We can not fully exclude this possibility. What we can say, though, is that to native speakers of Dutch, the large majority of our anomalies did not lend itself to such communicative reinterpretation (see Table 2.1 and Appendix 1 for some examples). Furthermore, note that what might inspire this particular dad to try to restore coherence is exactly the fact that he had anticipated a very different word. In all, we think that an anticipatory account of our current findings is the more likely one. And, as will be seen below, a post-lexical reinterpretation account also cannot explain the findings of experiment 1B.

The design of experiment 1A closely resembles the design of an experiment by Hoeks et al. (2004). The nature of the observed ERP effect, however, does not. Whereas anomalous words in highly predictive contexts have elicited a negativity in the Hoeks et al. (2004) study, in the present experiment they elicited a widespread positivity from about 300 ms onward. The reason for this divergence might lie in the fact that in the Hoeks et al. (2004) study contextual constraint was manipulated by a syntactic change, and not at a semantic level. This has also been suggested by Federmeier and colleagues, who in a recent experiment (Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007) observed that unexpected (albeit not anomalous) nouns presented in a predictive context did not modify the N400 compared to words presented in less constraining context, but instead elicited a larger positivity from 500 to 900 ms after stimulus onset. Since the design used by Federmeier et al. (2007) very much resembles the design of the present experiment, the positivities observed in the two experiments may well reflect the same underlying process. Furthermore, the fact that our effect emerged somewhat earlier in the EEG signal than the effect observed by Federmeier et al. could plausibly reflect the increased processing load imposed by an unexpected *anomalous* word, as compared to that imposed by an unexpected sensible word.

Both positivities might be related to the positive deflection that sometimes follows – or even replaces – the N400 effect elicited by semantic violations (see Kuperberg, 2007 for review; Kuperberg et al., 2003; Munte, Heinze, Matzke, Wieringa, & Johannes, 1998; Van Herten, Kolk, & Chwilla, 2005)[2]. This positivity has been argued to reflect a reanalysis of the context in an attempt to re-integrate inconsistencies in the context (Munte et al., 1998), or attempts to resolve a conflict between various combinatorial processing streams (Kuperberg, 2007). If so, then our findings would suggests that anomalies that violate predictive contextual constraints require more adjustments than anomalous words that do not. Moreover, the sustained nature of our effect would then indicate that this revision is –as one might expect– a lengthy and difficult process.

Experiment 1B

Experiment 1A examined discourse-dependent lexical prediction by means of an experimental paradigm that does not depend on specific syntactic (Van Berkum et al., 2005; Wicha et al., 2004) or phonological (DeLong et al., 2005) properties of a language. This paradigm allowed us to control for the potential effects of word-based priming, but did not allow us to examine whether such low-level mechanisms indeed make their own contribution to discourse-dependent prediction. Furthermore, although the results of Experiment 1A show that readers generate anticipations online, the results are not informative about the specificity of these predictions, neither in terms of the level of representation at which they are made, nor in terms of the specific entity being anticipated. In this experiment we therefore return to the more selective gender-dependent paradigm (Van Berkum et al., 2005; Wicha et al., 2004), which probes for predictions using the specific lexical gender of the predictable word.

Half of the items in Experiment 1B were so-called predictive stories, designed to support the prediction of a specific Dutch noun (e.g., "sword" in the example story in Table 2.2). To probe whether readers actually predicted this noun before it came along, we first presented a gender-inflected adjective, with a gender that was consistent or inconsistent with the discourse-predictable noun. As in the Van Berkum et al. (2005) study, we expected to see a differential ERP effect for adjectives with a gender-inflection that was inconsistent with the gender of the expected noun (e.g., "oude (old[e]$_{com}$)" in Table 2.2) compared to consistent adjectives. In Dutch, the gender of a noun is

[2] Note, however, that the scalp distribution of the positivity that follows the N400 is usually more posterior than the frontal/ central distribution of the effect observed here.

Table 2.2. An example of the items used in experiment 1B in all four conditions, in the original Dutch version and an approximate English translation.

Predictive context	**Prime control context**
De koene ridder zag dat de draak de goede tovenaar bedreigde. Hij pakte snel een **groot** maar best wel oud zwaard/ **grote** maar best wel oude lans en doodde het vuurspuwende beest. *The brave knight saw that the dragon threatened the benevolent sorcerer. Quickly he reached for a **big[Ø]**$_{neu}$ but rather old sword$_{neu}$ / **big[e]**$_{com}$ but rather old lance$_{com}$ and killed the fire-breathing beast.*	De goede tovenaar zag dat de draak de koene ridder bedreigde. Hij pakte snel een **groot** maar best wel oud zwaard/ **grote** maar best wel oude lans en doodde het vuurspuwende beest. *The brave knight saw that the dragon threatened the benevolent sorcerer. Quickly he reached for a **big[Ø]**$_{neu}$ but rather old sword$_{neu}$ / **big[e]**$_{com}$ but rather old lance$_{com}$ and killed the fire-breathing beast.*

an arbitrary lexical-syntactic feature that cannot be derived from the semantic features of the accompanying concept, and that must instead be stored as part of the lexical memory for that specific word (see Van Berkum, 1996, Ch. 2). Therefore, a gender-dependent anticipation effect on adjectives would demonstrate that people were indeed anticipating the specific noun at hand, and not just anticipating the general semantic field. Whether in predictive written stories this inflection-elicited 'lexical prediction effect' would be identical to the very early positivity that such inflections had elicited in predictive *spoken* stories (Van Berkum et al., 2005) remains to be seen. In fact, establishing this was one of the objectives of this study.

For the other half of the critical items, so-called prime control stories, we changed the predictive stories such that the message-level representation was completely different and much less predictive, while preserving any potential prime words. As illustrated in the prime control example, neither the previously expected noun ("sword") nor the previously unexpected noun ("lance") is particularly highly expected (nor, in fact, was any other word). We reasoned that if the differential ERP effect elicited by a prediction-inconsistent adjective inflection in the predictive condition is solely based on word-word priming, then these inflections should elicit the same effect in the prime control condition. On the other hand, if the lexical prediction effect in predictive stories critically hinges on the entire message conveyed by the discourse up to that point, no such effect should be observed in prime control stories. Establishing this was the main objective of Experiment 1B.

Method Experiment 1B

Participants

Since experiment 1B was run concurrently with experiment 1A, see experiment 1A for subject details.

Materials

The stimuli in this experiment were 160 mini-stories, consisting of a context sentence followed by the target sentence. For each item a predictive as well as a non-predictive context sentence was created, both containing the same prime words. In the predictive condition the stories were designed to suggest a specific 'discourse-predictable' noun right after the indefinite article in the target sentence (the second sentence), whereas in the prime control condition, no specific noun was expected at that point. To make sure that their gender could not be retrieved in any other way than via retrieval of the noun itself, all critical nouns were monomorphemic (for some morphologically complex Dutch nouns, gender can in fact be derived from specific parts of the word, see Van Berkum, 1996). A gender-inflected critical adjective always followed the indefinite article. In Dutch indefinite noun phrases, adjectives that modify a common-gender noun take an –e inflection, whereas adjectives that modify a neuter-gender noun take no inflection. The adjective could therefore be either consistent (carrying an inflection that agreed with the gender of the predicted noun) or inconsistent (carrying an inflection that did not agree with the gender of the predicted noun). Note that although the status of the adjective could be inconsistent relative to the predicted noun, at this point in the story both variants of the adjective were fully correct. Prediction-consistent adjectives were always followed by discourse-predictable nouns. However, to avoid grammatical violations later in the sentence, prediction-inconsistent adjectives were always followed by a coherent but much less expected alternative noun, with a gender that matched the inflection. Across the 160 items, 98 expected nouns had common gender, and 62 had neuter gender. At least 3 words separated the first critical adjective from the (un)expected noun (a second adjective and at least 2 words separating first and second adjective). See Table 2.2 for an example story.

<u>Pre-test: Cloze test.</u> The difference in predictability between predictive and prime control stories was determined in a pencil-and-paper cloze test, prior to

the EEG-experiment. For the predictive version of an item, at least 50 % of the participants used the discourse predicted noun, resulting in an average cloze value of 0.74 (sd = 0.14, ranging from .53 to 1.00) across all predictive stories. For the non-predictive prime control version of each item the response percentage for the discourse predicted noun, or any other alternative, was below 30 %, which on average resulted in a cloze value of 0.18 (sd = 0.15, ranging from 0.00 to 0.30). The cloze value for the unexpected target word was 0.03 in both the predictive (sd = .06) and the prime control stories (sd = .07).

The 160 items of experiment 1B (40 for each of the four conditions shown in Table 2.2) were randomly mixed with 80 filler items used for experiment 1A. By rotating the conditions in this list, three more lists of stimuli were created. Each of the four lists contained all 160 experimental stimuli, 80 stories in the predictive context version and 80 with a prime control context. 40 of the 80 predictive items and 40 of the 80 prime control items contained the expected noun (and therefore the expectedly inflected adjectives) while the remaining 40 ended with an unexpected noun (and the unexpectedly inflected adjectives). Each participant was shown one of these four lists of stimuli, so that one participant saw all the stimuli, but never in more than one condition.

Procedure, EEG recording and analysis.

Each participant saw 240 stories, 160 critical stories and 80 currently non-critical stories. Stimulus presentation, EEG recording and EEG processing were all identical to experiment 1A, with the exception that all words from the critical adjective up to and including the noun were shown with a standard presentation rate of 346 ms, based on the average length of all critical words (6 characters). As a result of the rejection of trials that contained (drift) artefacts on average 13% of the trials were lost in each of the four context conditions. For individual participants loss of trials varied from 2% to 38%.

Repeated-measure ANOVAs were used to compare the ERPs to (un)expectedly inflected adjectives in a predictive context and a control context. To assess not only the effects of expectedness and context type, but also the possible interaction with electrode position we employed an ANOVA crossing Context (predictive vs. prime control) with a Hemisphere (left/ right) by Anteriority (anterior/ posterior) factor. This analysis involved four quadrants: (1) left-anterior, comprising FP1, F3, F7, F9, FC1, FC5 and FT9; (2) right-anterior, comprising FP2, F4, F8, F10, FC2, FC6 and FT10; (3) left-posterior, comprising C3, T7, CP1, CP5, P3 and P7; (4) right-posterior, comprising C4, T8, CP2, CP6, P4 and P8. Univariate F tests with more than one degree of freedom in the numerator were adjusted by means of the

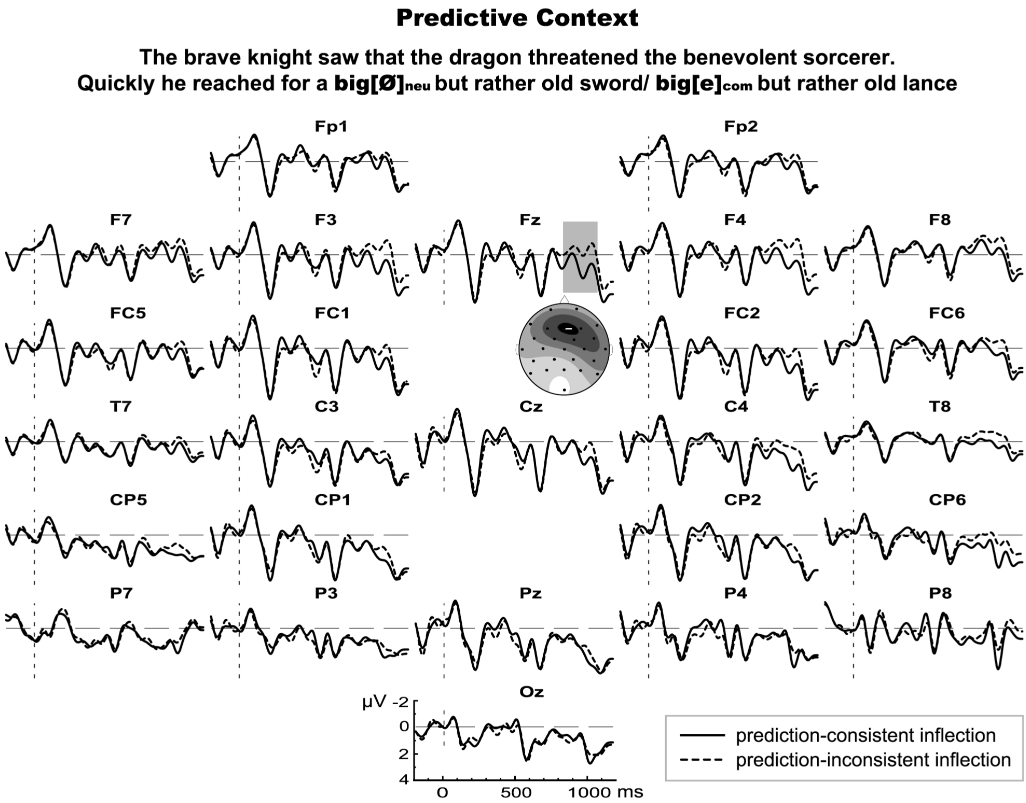


Figure 2.3 *Grand average ERPs elicited by the prediction-consistent (solid line) and prediction-inconsistent (dotted line) adjectives in the predictive condition. The time window where prediction inconsistent adjectives show a significant difference in the predictive context (900 - 1200 ms) is highlighted. The corresponding scalp distribution shows the effect of a prediction mismatch (inconsistent adjectives - consistent adjectives).*

Greenhouse-Geisser or Huynh–Feldt correction where appropriate. Uncorrected degrees of freedom and corrected *P*-values are reported.

Results Experiment 1B

Figure 2.3 and 2.4 display the grand average ERPs on all electrodes time-locked to the onset of the critical adjective for the consistently inflected adjectives (solid line) and the inconsistently inflected adjectives (dotted line) presented within the predictive context (Figure 2.3) and the prime control context (Figure 2.4). When read in a predictive context, prediction-consistent

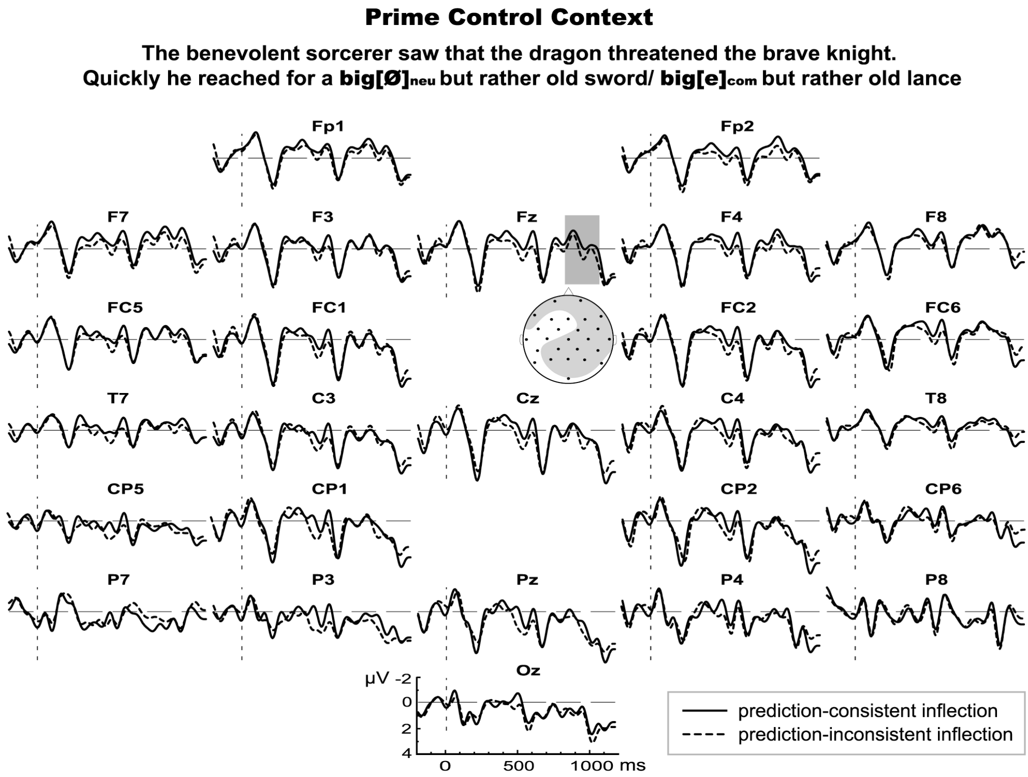


Figure 2.4 *Grand average ERPs elicited by the prediction-consistent (solid line) and prediction-inconsistent (dotted line) adjectives in the prime control condition. The time window where prediction inconsistent adjectives show a significant difference in the predictive context (900 - 1200 ms) is highlighted. The corresponding scalp distribution shows the effect of a prediction mismatch (inconsistent adjectives - consistent adjectives).*

adjectives clearly evoke a different ERP from prediction-inconsistent adjectives, emerging between 800 and 900 ms after the onset of the adjective. In the time-window between 900 and 1100 ms inconsistent adjectives evoke a more negative-going wave than the consistent adjectives (Figure 2.3). This difference is not present in the prime control condition (Figure 2.4).

The statistics corroborate this observation with a significant interaction of adjective consistency with context type between 900 and 1100 ms ($F(1,23) = 13.02$, $p = .001$). Post-hoc tests for this time-window show that there is indeed a difference between the consistent and inconsistent adjective in the predictive context ($F(1,23) = 5.23$, $p = .03$) but not in the prime control context ($F(1,23) =$

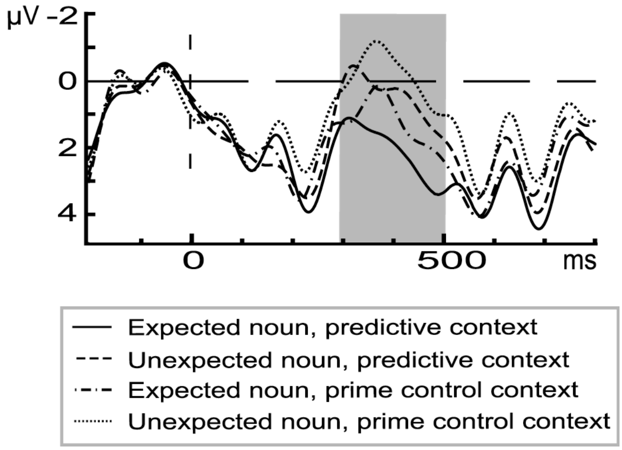


Figure 2.5 *Grand average ERPs elicited by the expected and unexpected nouns in the predictive and prime control context for electrode Pz. Unexpected nouns evoke a larger N400 between 300 and 500 ms in both the predictive and the prime control context.*

2.96, p = .14). Although the scalp distributions in 2. 3 suggest that the message-based effect of expectancy is frontally dominant, there is no interaction with hemisphere in the quadrant ANOVA (F(1,23) = 0.57, p = .81).

As can be seen in figure 2.4, inconsistent adjectives in prime control stories also seem to differ from consistent adjectives in an earlier time window, between 400 and 600 ms. A tiny effect can also be observed around this time in the predictive stories (Figure 2.3). The statistics reveal an overall main effect of expectancy in this time window that is just significant (F(1,23) = 4.47, p = .05), and that does not reliably interact with context (F(1,23) = 1.46, p = .24). This early positivity could be taken to reflect some processing consequence of prime-based anticipations being made in both types of stories. However, in view of the marginal significance of the effect, its small magnitude relative to other, non-significant fluctuations in the ERP signals, and the poor match between the statistics (suggesting, if anything, a main effect) and the ERP waveforms (which do not really suggest a comparable ERP effect across story types), we refrain from associating strong claims to these early deflections.

For completeness we also report the data from the nouns that follow the adjectives. Figure 2.5 shows the ERPs evoked by expected and unexpected nouns in a predictive and a prime control context. Unexpected nouns embedded in a predictive context still evoke a large N400 between 300 and 500 ms, even though they are preceded by a prediction-inconsistent adjective.

Furthermore, and surprisingly, this difference between expected and unexpected nouns is still present in the prime control context, where message-based expectancies for the two types of nouns are nearly identical. These observations are reflected by a main effect of predictability between 300 and 500 ms ($F(1,23) = 16.46$, $p = .00$), but no interaction between the predictability of the nouns and the context in which they occur ($F(1,23) = 0.63$, $p = .44$).

Discussion Experiment 1B

Adjectives with an inflection that was formally correct but did not match the gender of a discourse-predictable noun elicited a differential ERP wave around 900-1100 ms after adjective onset compared to consistent adjectives. Because the critical adjective and the later noun were always separated by at least three words (i.e., at least 1800 ms), this effect can not be attributed to the (un)expectedness of that noun. Furthermore, the only difference between prediction-inconsistent and consistent adjectives was whether or not they agreed with the grammatical gender of the discourse-predictable noun. This ERP effect therefore provides clear evidence for the fact that readers anticipate *specific* upcoming words, pre-activating the specific semantic as well as syntactic properties of the words. This effect is not present in a prime control discourse. This strongly indicates that specific lexical predictions draw upon a *message-level representation* of the unfolding discourse, rather than upon some form of word-based priming. That is, prediction is not the result of relatively low-level word-based priming mechanisms (such as automatic spreading activation in a lexical-semantic network, or ASA-mediated scenario priming), but involves a more sophisticated message-level mechanism that can take into account the actual nuances of the preceding discourse.

The effect of expectancy in the predictive sentences emerges relatively late, at about 900 ms after onset of the critical adjective. In the Van Berkum et al. (2005) spoken-language study, however, the processing consequences of a disconfirmed lexical prediction showed up in ERPs much earlier, within about 50-250 ms. Note that in the latter experiment, the ERPs were time-locked to the onset of the gender-(in)consistent inflectional suffix, i.e., right at the start of the critical cue, and towards the end of the adjective. In our current written-language study, ERPs were instead time-locked to the visual onset of the entire adjective, which means that at time 0, readers still had to recognize the whole word and strip the critical cue away from this word. Furthermore, with spoken language materials, listeners actually receive acoustic cues to the nature of the upcoming inflection well before the formal acoustic onset of the inflection itself (see Van Berkum et al., 2005, for discussion). These two

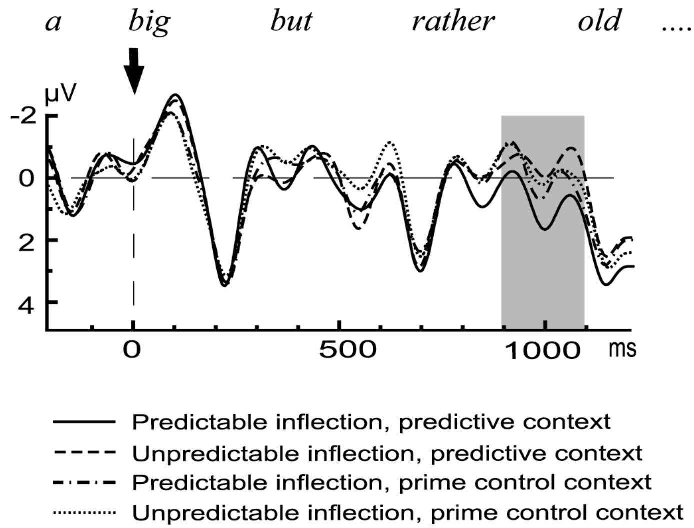


Figure 2.6 *Grand average ERPs elicited by the prediction-consistent and prediction-inconsistent adjectives in the predictive and prime control context for electrode Fz. Prediction-inconsistent adjectives elicit a late positive deflection between 900-1100 ms, in the predictive context only.*

differences in how the ERP time-locking point relates to the time at which the critical cue will become available to the comprehender can to some extent explain why the spoken-language effect observed by Van Berkum et al. was so much earlier.

The ERP waves for the predictable and unpredictable adjectives, as plotted in Figure 2.3 and 2.4, suggest that unpredictable adjective-inflections evoke a more negative inflection compared to the predictable inflections. However, when the ERPs of all four conditions are plotted together (see Figure 2.6) it is immediately clear that it is the predictable adjective that evokes a more positive deflection between 900 and 1200 ms, compared to the other conditions. This would suggest that the underlying process that elicits this ERP effect is not so much the mismatch between the expected inflection and the actually perceived inflection, but instead that the effect is based on the match between the prediction and the incoming information.

Although the same process, namely matching incoming information onto specific lexical predictions, seems to underlie the effects observed in experiment 1A and 1B, the actual ERP effects are clearly different for the two experiments, both with regard to the timing and to the nature of the components involved. At first sight, this might seem surprising. However, even though the effect is based on a comparison between predicted and observed lexical-semantic information, the nature of the actual (mis)match that seems to underlie the effect is very different. The long-lasting positive shift observed in Experiment 1A is a clear consequence of the repair processes due to the *semantic violation* of the prediction, whereas the shorter-lasting late negativity seems to result from the *syntactic match* between the expected and perceived gender of the word. The differential nature of the underlying cognitive processes could very well explain the difference in the observed ERP effects.

Previous results by DeLong and colleagues (2005) suggest that seeing information that mismatches our prediction does not immediately result in the adjustment of that prediction: after seeing an article that was inconsistent with the prediction, e.g. seeing "an" when "kite" was the expected noun, the unexpected noun "aeroplane" still evoked a larger N400 than the expected noun. Figure 2.4 shows that in the present experiment unexpected nouns evoked a larger N400 compared to expected nouns when the discourse had a constraining message. However, in the prime control condition this difference in N400 was also present. This could be taken to suggest that the N400 does not reflect message-level expectancy or integration, but that, in contrast to the prediction effect, it reflects integrative or predictive processes related to word-based priming. Results from a recent experiment that used the same stories without the critical preceding adjectives (Otten & Van Berkum, 2007) suggest that the discourse-based N400 effect can not be solely attributed to processes reflecting automatic activation. The adjectives that precede the noun thus seem to critically modify the expectations in the prime control condition. A follow-up cloze test (40 participants) that included the inflected adjectives confirms this idea. In the prime control context the cloze probability for predicted nouns was 18% before reading the inflected adjective, but this probability rose to 47% after reading the adjectives (the probability for the unpredictable noun remained relatively low at 15%). This suggests that the difference in N400 amplitude between predictable and unpredictable words in predictive and prime control contexts does not result from differences in word-based priming, but differences in discourse-based expectancies.

General discussion

In two ERP experiments, we examined whether predictions of upcoming words in the text are based on a precise message-level representation of the discourse up to that point, or whether a simpler word-based priming mechanism is at work. Van Berkum and colleagues (2005) argued that their evidence for discourse-based lexical anticipation was difficult to explain without assuming the involvement of message-level representations. In contrast, DeLong and colleagues (2005) have suggested that the prediction of specific upcoming words is based on single words or specific combinations of words, via the activation of lexical-semantic and encyclopedic knowledge in semantic memory. The latter view suggests that it is not the exact message of the current discourse that enables prediction, but only the individual words present in that discourse, and the conceptual scenarios suggested by them.

Our findings strongly indicate that the latter can not be the whole story, and that discourse-based lexical prediction requires more than a simple word-based priming mechanism. In Experiment 1A, the presentation of an anomalous word in highly predictable stories elicited processing consequences in the ERPs that were uniquely dependent on message-level constraint. In Experiment 1B, grammatically and semantically correct adjectives whose gender inflection mismatched the gender of the discourse-predictable noun elicited a differential ERP effect only in stories in which predictability hinged on strong message-level constraint; in prime control stories in which the same words conveyed a much less constraining message, the ERP effect disappeared. In line with our earlier account (Van Berkum et al., 2005), both experiments thus suggest that people make use of the *exact message* of the discourse to anticipate upcoming words.

The anomalous words of experiment 1A (e.g. “stove” in Table 2.1) were judged to be equally anomalous independent of whether the context did or did not support a specific prediction at that point. As discussed before, the differential ERP effect elicited by anomalies in a predictive discourse can therefore not easily be explained by differences in ‘post-lexical’ integration. Of course, anomalous words do perturb the comprehension system, and one might thus argue that our findings do not generalize to normal text comprehension. We think such an argument would be too simple, for although the differential ERP effect in Experiment 1A might well reflect recovery processes that would not occur with coherent words, the very fact that such recovery occurs – and does so *differentially* – is testimony to the fact that the system is in its normal mode of operation and is interpreting language as it comes in. Furthermore, Experiment 1B, the logic of which does *not* depend on a plausibility pretest or assumptions about the processing of an anomaly, also provides evidence for message-based anticipation. The differential reaction to anomalies in experiment 1A shows that people have a more clear idea how the story will develop when the discourse is constraining, but the design gives no clear indication about what these anticipations are. In contrast, experiment 1B shows that the predictions are highly specific, since both the predictable word and its lexical gender are pre-activated. Taken together both experiments clearly point in the same direction: people can rapidly use a message-level representation of the prior discourse to anticipate very specific upcoming words, as the text unfolds.

Note that we are not claiming that a message-based mechanism is the *only* possible mechanism that supports discourse-based lexical prediction. The design of Experiment 1A did not allow us to study the independent effect, if any, of simpler word-based priming. Also, whereas the design of Experiment 1B did allow for this, we had derived the prime control stories from their predictive counterparts, and had not deliberately used strong associative or semantic prime words in the latter. The absence of a differential effect in prime control stories in Experiment 1B therefore does not provide compelling evidence against additional word-based priming in text comprehension. What it does unequivocally show, though, is that people can predict upcoming words in a way that can not be reduced to simple priming.

At this point, researchers who usually test for (pre)activation with probe word methodology or similar 'content-sensitive' measures may well feel that we can not really say anything about the exact word that is being predicted. The adjective that mismatches the gender of the predicted word *sword* also mismatches thousands of other Dutch nouns. Nevertheless, each of our critical stories was designed to support the prediction of a specific word, and with a cloze test we estimated the probability that any one reader could in principle entertain that word at the relevant point in the unfolding story. Our critical manipulations, such as whether the preceding adjective did or did not agree with the gender of a specific word in Experiment 1B, were defined *relative to this word*. The fact that these manipulations matter (and that obvious confounds were ruled out) therefore supports the claim that readers were not just anticipating any word, but were predominantly anticipating that word, on the majority of critical trials, for the majority of subjects. Of course, with items whose predictability is less than 100%, a reader will sometimes anticipate another specific word, or even no specific word at all. Nevertheless, on a sufficient number of occasions, our subjects must have had our pre-chosen predictable words in mind.

Previous research on anticipation in discourse comprehension typically did not focus on the prediction of specific words, but on more general anticipations of likely events or consequences, e.g. predictive inferences. Research on predictive inferencing has shown that people can pre-activate a likely consequence or event on-line if the context is considerably constraining. The present and previous experiments show that people are also able to use their model of the discourse to pre-activate specific words. Before we explore the potential link between predictive inferences and discourse-based lexical predictions, we first examine how the latter might actually come about.

In line with Van Berkum et al. (2005), we suggest that what underlies specific prediction is the result of convergent predictions being made at several levels of unfolding structure. It is well known that language comprehenders compute the syntactic and conceptual analysis of the incoming language *incrementally* and in parallel (see Jackendoff, 2002; Jackendoff, 2007 for an overall framework and; Vosse & Kempen, 2000 for an explicit computational model of the syntactic side of things). As a consequence, at any point in an unfolding sentence, readers and listeners have at their disposal a partial syntactic and conceptual analysis of the preceding sentence fragment (and, in the conceptual analysis, the relevant wider discourse context, see Van Berkum, Brown, & Hagoort, 1999a; Van Berkum, Brown, Hagoort, & Zwitserlood, 2003). Each of these partial representations can by itself suggest what might come next. For example, in "Quickly he reached for a ...", various aspects of the syntax all suggest that a singular noun is about to follow. Furthermore, in the context of "The brave knight saw that the dragon threatened the benevolent sorcerer.", there is a fair chance that what will be reached for is the typical weapon that knights tend to slay dragons with, a fact that we know people to exploit in reading (Calvo & Castillo, 1996; Cook, Limber, & O'Brien, 2001; Fincher Kiefer, 1993; McDaniel, Schmalhofer, & Keefe, 2001; Murray et al., 1993).

Although these predictions arise at different levels of representation, it is not difficult to see how they might come together and converge onto a specific word. As laid out by Jackendoff (2002; 2007; see also Kempen & Huijbers, 1983; Levelt, 1989), an individual lexical item like *sword* consists of bits of orthographic, phonological, syntactic, and conceptual information, bundled together into a single multi-leveled structure. If people actually read or hear *sword*, i.e., strongly activate its orthography or phonology, the associated fragments of syntactic and conceptual structure are activated and merged ('unified') with the syntactic and conceptual analyses constructed for the language input so far. However, within the same framework, the preceding syntactic and conceptual context can, if sufficiently constraining, also each *pre-activate* the relevant bits of structure, resulting in the prediction of the related lexical item. In this case, it is the convergent pull of syntactic and message-based conceptual constraints that activates a particular word, and not the orthographic or phonological input. This account suggests that predictions are made continuously, and in a graded fashion, a view that is consistent with findings from DeLong et al. (2005).

With this general framework in place, how do predictive inferences fit in? We know that readers and listeners construct a situation model of what is being written or spoken about (e.g., a delicate vase dropping from a twenty-story building), and predictive inferences can be viewed as reasonable inferences about what might happen next in the world captured in this situation model (e.g., the vase will break when hitting the ground; see McDaniel, Schmalhofer and Keefe (2001; Schmalhofer et al., 2002) for exactly this view). In ongoing communication, such inferences will often lead people to anticipate what might be talked about next. However, although they will often feed into predictions about upcoming communication, predictive inferences are by no means *equivalent* to the latter. If "So with a single well-aimed throw, he propelled the delicate vase through the open window" is the last line of a short novel, for instance, readers will definitely not predict another sentence continuing the story, nor will they thus predict that the writer will use the words "break" or "broken" therein. It is only in the context of further communication (e.g., an unfolding sentence like "Of course, when it hit the ground the vase…") that predictive inferences can lead people to anticipate specific words, presumably via forces impinging upon the conceptual structure being constructed for the unfolding communication. Note, furthermore, that upcoming words can also be predicted in the absence of predictive inferences: after reading *"The moon revolves around the…"* most people will anticipate the word "earth" without making any predictive inference whatsoever. In all, although predictive inferences and discourse-based lexical predictions will in practice often go together, they are not equivalent, nor bound to each other in principle.

After exploring the various constraints involved, one question remains: how do lexical predictions actually come about? One possibility is that anticipation is an intrinsic consequence of the nature of syntactic and conceptual representation. In so-called lexicalized grammars, for example, the lexical representation of a word like "a" can, beyond specifying that it is a determiner, also contain a small NP structure ('treelet') in which the obligatory slot for a head noun is yet to be filled (Jackendoff, 2007; Vosse &

Kempen, 2000). Analogously, the semantic representation of this determiner can contain an as yet to be filled obligatory slot for a specific entity. The activation of such small bits of partially instantiated structure can be said to generate a prediction without requiring any additional machinery. Another intriguing possibility is that, it may not be our language *comprehension* system, but our language *production* system that underlies the process of prediction (see also Garrett, 2000; Pickering & Garrod, 2007; Van Berkum et al., 2005). Further research will have to clarify which of these accounts best characterizes the processes underlying the discourse-based anticipation of specific upcoming words.

Conclusion

In the late 1950s, Chomsky and his colleagues demonstrated that human language is a generative system, a system that allows us to communicate an infinite number of things by finite means. In psycholinguistics, the discovery that language is an intrinsically open-ended system has led to the widely shared assumption that readers and listeners can not and therefore do not predict upcoming words, at least not in a way that goes beyond simple priming between words. Our ERP findings show that this apparently reasonable assumption is wrong. We have provided clear evidence for the fact that readers do not always passively wait for upcoming input, but can make intelligent guesses about the words they might soon encounter, based on the message conveyed by the discourse so far.

These observations converge with and extend other recent evidence on discourse-and sentence-based word prediction (DeLong et al., 2005; Van Berkum et al., 2005; Wicha et al., 2004). They also cohere with evidence for other forms of anticipation during language comprehension. Predictive inference research has demonstrated that people can anticipate specific developments in the scenario described (e.g., that a porcelain vase dropped on the floor will probably break). Other studies have shown that readers and listeners make predictions as to whom or what will be referred to next (Altmann & Kamide, 1999; Nieuwland & Van Berkum, 2006b). Of course, with tens of thousands of nouns in the language, knowing that a determiner will usually be followed by a noun doesn't tell you all that much. However, if you know that it is going to refer to something edible, or to what is left of a vase that just dropped from a 20-story building, things may well lighten up. We suggest that the human language comprehension system has the talent to combine such diverse constraints, as well as the boldness to use them to look ahead.

Chapter 3

Great expectations: Specific lexical anticipation influences the processing of spoken language

Recently several studies have shown that people use contextual information to make predictions about the rest of the sentence or story as the text unfolds. Using event related potentials (ERPs) we tested whether these on-line predictions are based on a message-level representation of the discourse or on simple automatic activation by individual words. Subjects heard short stories that were highly constraining for one specific noun, or stories that were not specifically predictive but contained the same prime words as the predictive stories. To test whether listeners make specific predictions critical nouns were preceded by an adjective that was inflected according to, or in contrast with, the gender of the expected noun. When the message of the preceding discourse was predictive, adjectives with an unexpected gender inflection evoked a negative deflection over right-frontal electrodes between 300 and 600 ms. This effect was not present in the prime control context, indicating that the prediction mismatch does not hinge on word-based priming but is based on the actual message of the discourse. This shows that people rapidly make very specific predictions about the remainder of the story, as the story unfolds. These predictions are not simply based on word-based automatic activation, but take into account the actual message of the discourse.

Introduction

"In this branch house of ours, Handel, we must have a--"
I saw that his delicacy was avoiding the right word, so I said, "A clerk."
"A clerk. And I hope it is not at all unlikely that he may expand into a partner. "
[Great Expectations, Charles Dickens]

In this short exchange, we can see that Pip, the main character of the novel (who is addressed here as Handel by his good friend Herbert), generates not only great expectations but small ones as well. The expectation at hand does not refer to his hopes and plans for the future. Pip merely anticipates how the sentence that his friend is hesitantly uttering will end. This form of prediction – the temptation to finish a slow speaker's sentence– is one we are probably all familiar with in everyday life. Recent event related potential (ERP) studies have shown that predictive processes in language comprehension are not limited to instantiations where the speaker falters. Predictions about the continuation of a sentence or story are actually made regularly and on the fly (DeLong et al., 2005; Van Berkum et al., 2005; Wicha et al., 2004). One important issue is whether these predictions are initiated by a relatively simple automatic activation process, based on (a combination of) individual words in the discourse, or whether they are based on a more thorough understanding of the message of the discourse. In this spoken language ERP experiment we explored which of these constraints actually trigger specific lexical predictions.

Although strong and influential arguments have been made against anticipation in language processing based on the inherent open-ended character of language (cf. Chomsky, 1957), a multitude of psycholinguistic experiments suggests that people do use context to form expectations about the language utterance that is still to follow. These predictive processes pertain to the grammatical role of words (Altmann et al., 1998; Lau et al., 2006) but also to inferences about the general syntactic (Kamide et al., 2003; Van Berkum, Brown et al., 1999b) and semantic (Calvo et al., 1999; Campion, 2004 ; Fincher Kiefer, 1996; Graesser et al., 1994; Keefe & McDaniel, 1993; Linderholm, 2002; McKoon & Ratcliff, 1992; Murray & Burke, 2003; Schmalhofer et al., 2002) content of the utterance (see also (Pickering & Garrod, 2007) for a short review). Recent ERP studies have shown that people furthermore use their rapid syntactic and semantic analysis of the discourse to anticipate *specific* words, in spoken (Van Berkum et al., 2005; Wicha, Bates et al., 2003) as well as written language (DeLong et al., 2005; Wicha, Moreno et al., 2003; Wicha et al., 2004).

DeLong and colleagues (2005) have shown that these specific lexical predictions are stronger as the context is more constraining. This contextual constraint, however, can have its predictive effect at two different levels. Predictions could arise from a relatively simple priming process, by which individual words activate lexical-semantic and world knowledge in semantic memory (Beeman, Bowden, & Gernbacher, 2000; DeLong et al., 2005). On the other hand, it is well known that our understanding of spoken or written language does not rely on a compilation of disjoint words: we form a comprehensive structured model of the discourse (Clark, 1996; Kintsch, 1998; Zwaan & Radvansky, 1998), combining contextual information through rapid syntactic and semantic analysis (MacDonald, Pearlmutter, & Seidenberg, 1994; Tanenhaus & Trueswell, 1995), which includes not only the local but also the wider context (Hagoort, Hald, Bastiaansen, & Petersson, 2004; Van Berkum, Hagoort, & Brown, 1999; Van Berkum, Zwitserlood et al., 2003). Van Berkum and colleagues (2005) have therefore suggested that it is more likely that specific lexical predictions are based on an extensive, message-level representation of the discourse than on automatic activation by (a set of) individual words.

To test whether specific lexical predictions are based on the actual message of the discourse or related to some simpler form of word-based priming, we designed predictive stories as well as so-called prime control stories. The predictive stories had a message-level content that supported the prediction of a specific Dutch noun. In the predictive story in Table 3.1, for example, the word "cross" would indeed be the most sensible and 'expected' continuation at that point, (confirmed by the fact that in a completion test, the large majority of Dutch readers would use "cross" to continue the story at this point). However, note that words like "religious" and "grandparents" are themselves also (mildly) related to "cross", via simple lexical associations (religious – cross) and possibly also scenario-mediated associations (religious grandparents – cross).

To uncover the potential contribution of such simpler priming mechanisms to discourse-based lexical prediction, the prime control stories contained the same potential prime words as the predictive stories but had a completely different and much less predictive message-level representation. As illustrated in the prime control example in Table 3.1, neither the previously expected noun ("cross") nor the previously less expected noun ("crucifix") is particularly expected (nor, in fact, is any other word) but the possible prime words (i.e. "grandparents" and "religious") are still present in the preceding context sentence. Thus the message-level constraint of the context is low for

Table 3.1. Example story in the original Dutch version and an approximate English translation, across all four conditions. Critical adjectives are printed in bold face, and critical nouns are printed in italics (in the Dutch example) or regular letters (English example). The grey word between brackets indicates the predictable word at that point in the story.

Predictive context	**Prime control context**
Mijn opa en oma zijn erg religieus. Boven hun bed hangt een [kruis] (1) **groot** en nogal dramatisch *kruis* (2) **grote** en nogal dramatische *crucifix* aan de muur, en verscheidene schilderijen van heiligen.	Mijn opa en oma zijn niet erg religieus. Boven hun bed hangt een […] (1) **groot** en nogal dramatisch *kruis* (2) **grote** en nogal dramatische *crucifix* aan de muur, maar dat is een erfstuk.
My grandfather and grandmother are very religious. Above the head of their bed hangs a [cross$_{neu}$] *(1) **big$_{neu}$** and rather dramatic* cross *(2) **big$_{com}$** and rather dramatic* crucifix *on the wall, together with several paintings of saints.*	*My grandfather and grandmother are not very religious. Above the head of their bed hangs a […]* *(1) **big$_{neu}$** and rather dramatic* cross *(2) **big$_{com}$** and rather dramatic* crucifix *on the wall, but that is a family heirloom.*

the prime control context, but the prime-based constraint is identical for the predictive and the prime control context.

To probe whether readers actually predicted the expected noun *before* it came along, we first presented a gender-inflected adjective, with a gender that was consistent or inconsistent with the discourse-predictable noun. In Dutch, adjectives in indefinite noun phrases have a suffix that depends on the arbitrary, lexically memorized gender (Van Berkum, 1996) of the noun they precede. Adjectives that modify a common-gender noun carry an –e suffix (e.g., "grote crucifix", "big$_{com}$ crucifix$_{com}$"), whereas adjectives modifying a neuter-gender noun are not inflected (e.g., "groot kruis", "big$_{neu}$ cross$_{neu}$").

If listeners strongly anticipate a specific noun, an adjective with a mismatching gender suffix will come as an 'unpleasant' surprise compared to the prediction-consistent adjective. As in previous studies that have employed probes (gender-inflected adjectives in Dutch (Van Berkum et al., 2005), gender-marked articles in Spanish (Wicha, Bates et al., 2003; Wicha, Moreno et al., 2003; Wicha et al., 2004) and the a/an distinction in English (DeLong et al., 2005)) to test for prediction, we expected that adjectives with an inconsistent adjective inflection would elicit a different ERP effect compared to consistent adjectives. The exact electrophysiological consequences of a prediction mismatch, however, have not been clearly established. Phonological or gender-related information that mismatches a prediction can elicit an increase in N400 amplitude (DeLong et al., 2005; Wicha, Moreno et al., 2003). However, prediction mismatches can also elicit negative ERP effects with a timing and scalp distribution that clearly differs from a standard N400 (Otten & Van Berkum, 2007; Wicha, Bates et al., 2003), as well as positive ERP effects (Van Berkum et al., 2005; Wicha et al., 2004). Because the sources of this variability are as yet not understood, the exact nature of the ERP effect to a prediction mismatch was difficult to predict. However, since the majority of the experiments have yielded negative ERP effects as a response to information that (implicitly) contradicts a prediction, it seemed most likely that unexpectedly inflected adjectives would also elicit a more negative ERP. If the differential ERP effect elicited by a prediction-inconsistent adjective inflection in the predictive condition is solely based on word-word priming, we should observe the same effect in the prime control condition. On the other hand, if the lexical prediction effect in predictive stories critically hinged on the entire message conveyed by the discourse up to that point, no such effect should be observed in prime control stories.

Methods

Participants

32 right-handed native speakers of Dutch participated in the experiment as part of a course requirement. Three participants were excluded from analysis because more than 50% of the critical trials had to be deleted due to artefacts (see below). Of the remaining 29 participants 17 were male. Mean age over participants was 23 years, ranging from 18 to 33.

Materials

The critical stimuli were 160 naturally spoken two-sentence mini-stories, consisting of a context sentence followed by the target sentence. For each item we created a predictive context sentence, that was constraining at a message level, as well as a prime control context sentence, that contained the same prime words but was *not* predictive at the message level. We employed several different strategies in creating the prime control sentences, which are based on the original, predictive, sentence. A selection of stimuli in Appendix 2 illustrates these different strategies: negation (see also the example stimulus in Table 3.1), adding words, deleting/ replacing (non-prime) words, or changing the order of the words.

In a pencil-and-paper “cloze test”, 66 participants were shown the stories up to and including the indefinite determiner, and were asked to finish these stories. At least 50 % of the participants used the same noun when the context was predictive, resulting in an average cloze value of 0.74 for the predicted noun (sd = 0.14, ranging from .53 to 1.00) across all predictive stories. For the non-predictive prime control version the response percentage for predicted noun, or any other alternative, was below 30 %, which on average resulted in a cloze value of 0.18 (sd = 0.15, ranging from 0.00 to 0.30). The cloze value for the unexpected target word was 0.03 in both the predictive (sd = .06) and the prime control stories (sd = .07).

In both the predictive and the prime control condition the target sentence could contain the predicted word or an unexpected but still completely coherent alternative. The critical stimuli in this experiment, however were not the (un)expected nouns, but the gender-inflected adjectives that preceded each critical noun. In Dutch indefinite noun phrases, adjectives that modify a common-gender noun take an -e inflection, whereas adjectives that modify a neuter-gender noun take no overt inflection. Adjectives could be consistent or inconsistent relative to the gender of the predicted noun, but at the time that listeners heard these adjectives both variants of the adjective did not pose an overt violation. Furthermore, to avoid grammatical violations later in the sentence, prediction-inconsistent adjectives were always followed by a coherent but much less expected alternative noun, with a gender that matched the inflection. Across the 160 items, 98 expected nouns had common gender, and 62 had neuter gender. At least 3 words separated the first critical adjective from the (un)expected noun (a second adjective and at least 2 words separating first and second adjective). Expected and unexpected nouns were not exactly matched on length or frequency. The mean length of the expected and unexpected noun was respectively 6.1 (sd = 2.3) characters and 7.4 characters (sd = 2.3), and the mean frequency for expected and unexpected nouns was respectively 32.2 (sd = 53.7) and 26.7 (sd = 96.0) per 1 million, as stated in the Celex database. A list with all critical items (in Dutch) can be obtained from the first author.

Each story was recorded in four different versions (predictive context – expected inflection/ noun, predictive context – unexpected inflection/ noun, prime control context – expected noun, prime control context – unexpected noun), by the same female speaker, at normal rate and intonation. The average duration of the critical words was 513 ms for the adjective (range 243 – 924 ms) and 524 ms for the noun (range 170 – 990 ms). The onset of the noun was separated from the onset of the first adjective by 1580 ms on average (range 791-2725 ms). The end of the sentence on average came 2751 ms after the onset of the critical noun (range 1483 – 6070 ms).

Four different trial lists were used. The first list was created by pseudorandomly mixing the 160 critical items (40 for each of the four conditions shown in Table 3.1) with 90 filler items, so that each participant heard all the stimuli in only one condition. Three more lists were created by rotating the conditions in the original first list.

Procedure and EEG recording

The total of 250 items were divided in 10 blocks, separated by a pause. Each trial was separated from the next by a 5 sec silence and was preceded by a short warning tone. Total time-on-task was approximately eighty minutes. Participants were seated in front of two loudspeakers, and were informed that they would be listening to short stories. They were instructed to listen for comprehension and minimize movement. No additional task demands were imposed.

The electroencephalogram (EEG) was recorded from 30 electrode sites (FP1, FP2, F9, F7, F3, Fz, F4, F8, F10, FT9, FC5, FC2, FC6, FC1, FT10, T7, C3, Cz, C4, T8, CP5, CP1, Cp2, Cp6, P7, P3, Pz, P4, P8 and Oz), mounted in an elastic cap, each referenced to the left mastoid. Blinks and vertical eye-movements were registered by placing an electrode under the left eye, also referenced to the left mastoid. The EEG was amplified with BrainAmps amplifiers (BrainProducts, München), band-pass filtered at 0.03 Hz-100 Hz and sampled with a frequency of 500 Hz. The EEG signals were re-referenced off-line to the average of right and left mastoids. Blinks and eye movements were removed from the data using a procedure based on Independent Component Analysis (ICA) as described by Jung et al. (Jung, Makeig, Humphries et al., 2000; Jung, Makeig, Westerfield et al., 2000).

We timelocked the ERPs to the onset of the critical adjective and noun. After baseline correcting (by subtraction) the waveforms of the individual trials relative to the relevant 200-ms pre-stimulus baseline intervals, we computed average waveforms for each subject and condition relative to the estimated acoustic onset of the first adjective and the noun that followed. Because the earliest (un)expected nouns, signifying an overt (mis)match with the predicted noun, began at about 800 ms after the onset of the critical adjective, we analysed the ERPs evoked by the adjectives in a time-interval from 0 to 800 ms. To avoid spurious effects due to the sentence offset, the window of analysis for the nouns ranged from 0 to 1500 ms after noun onset.

Segments in which the signal exceeded ±75 μV, or which featured a linear drift of more than ± 50 μV, beginning before the onset of the critical word, were eliminated. For three subjects the data loss exceeded 50% (respectively 67%, 79% and 91%, averaged over all conditions and critical words), and therefore these subjects were excluded from further analysis. For the remaining 29 subjects 23% of the trials was deleted (ranging between subjects from 2% to 48%). The proportion of deleted trials did not differ across conditions.

Analyses

To assess not only the effects of consistency and context type, but also the possible interaction with electrode position the ERPs elicited by adjectives and nouns were evaluated in an ANOVA crossing Consistency (prediction consistent/ prediction-inconsistent), Context (predictive/ prime control), Hemisphere (left/ right) and Anteriority (anterior/ posterior). This analysis thus involved four quadrants: (1) left-anterior, comprising FP1, F3, F7, F9, FC1, FC5 and FT9; (2) right-anterior, comprising FP2, F4, F8, F10, FC2, FC6 and FT10; (3) left-posterior, comprising C3, T7, CP1, CP5, P3 and P7; (4) right-posterior, comprising C4, T8, CP2, CP6, P4 and P8. Effects on the midline electrodes (Fz, Cz, Pz and Oz) were assessed in a separate ANOVA crossing the factors Context, Consistency and Electrode position. F tests with more than one degree of freedom in the numerator were adjusted by means of the Greenhouse-Geisser or Huynh–Feldt correction where appropriate. Uncorrected degrees of freedom and corrected P-values are reported.

Results

Figure 3.1 shows the ERPs by prediction-consistent and prediction-inconsistent adjectives in a predictive context, timelocked to the onset of the inflected adjective. Adjectives carrying an inflection inconsistent with the gender of the expected noun evoke a negativity on the right frontal electrodes compared to consistent adjectives, starting at about 300 ms and lasting until 600 ms after the onset of the adjective. Crucially, when the inconsistent

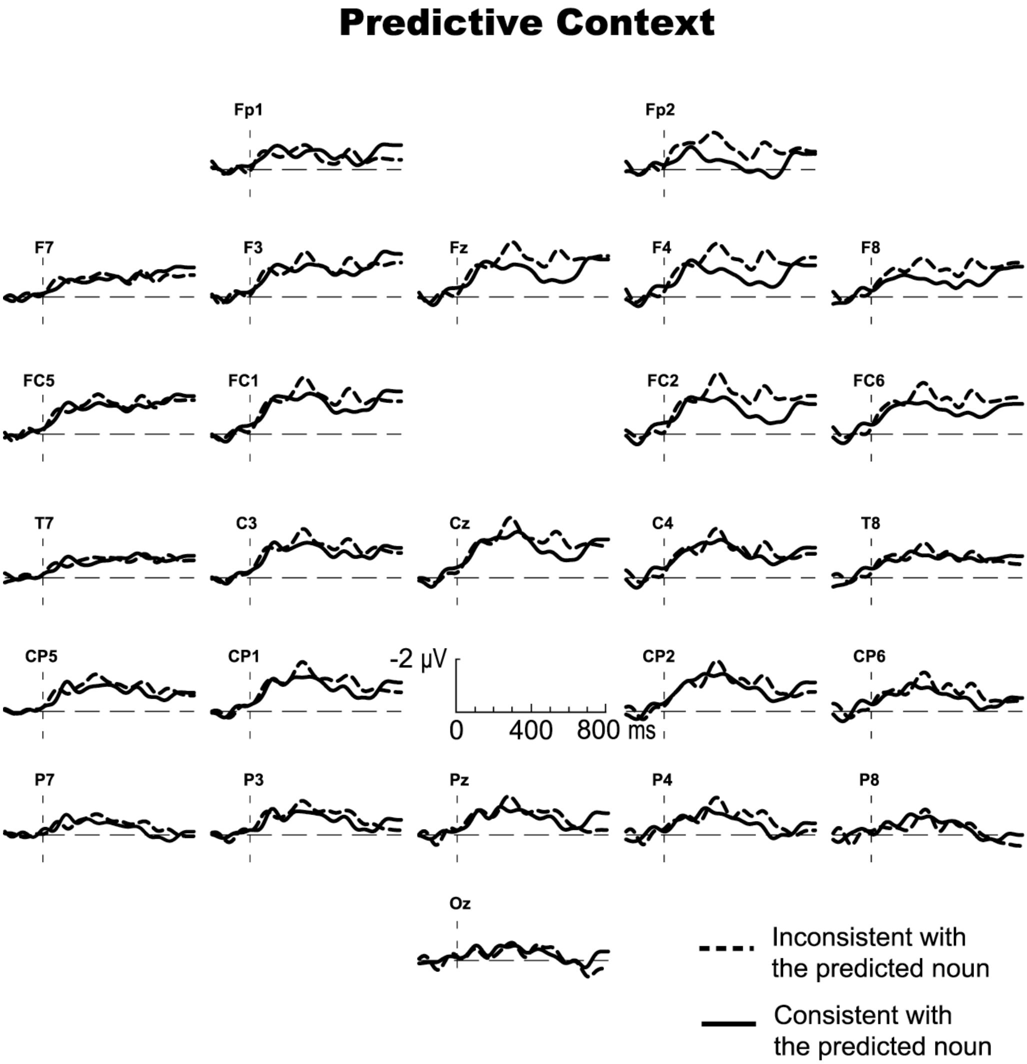


Figure 3.1 *Grand average ERPs elicited by the critical adjectives in a predictive context. Black lines represent the response to adjectives bearing an inflection that is consistent with the gender of the predicted noun; red lines represent responses to gender inconsistent adjectives. The ERPs are timelocked to the onset of the adjective, and are filtered (8 Hz high cut-off, 48 dB/oct) for presentation purposes only.*

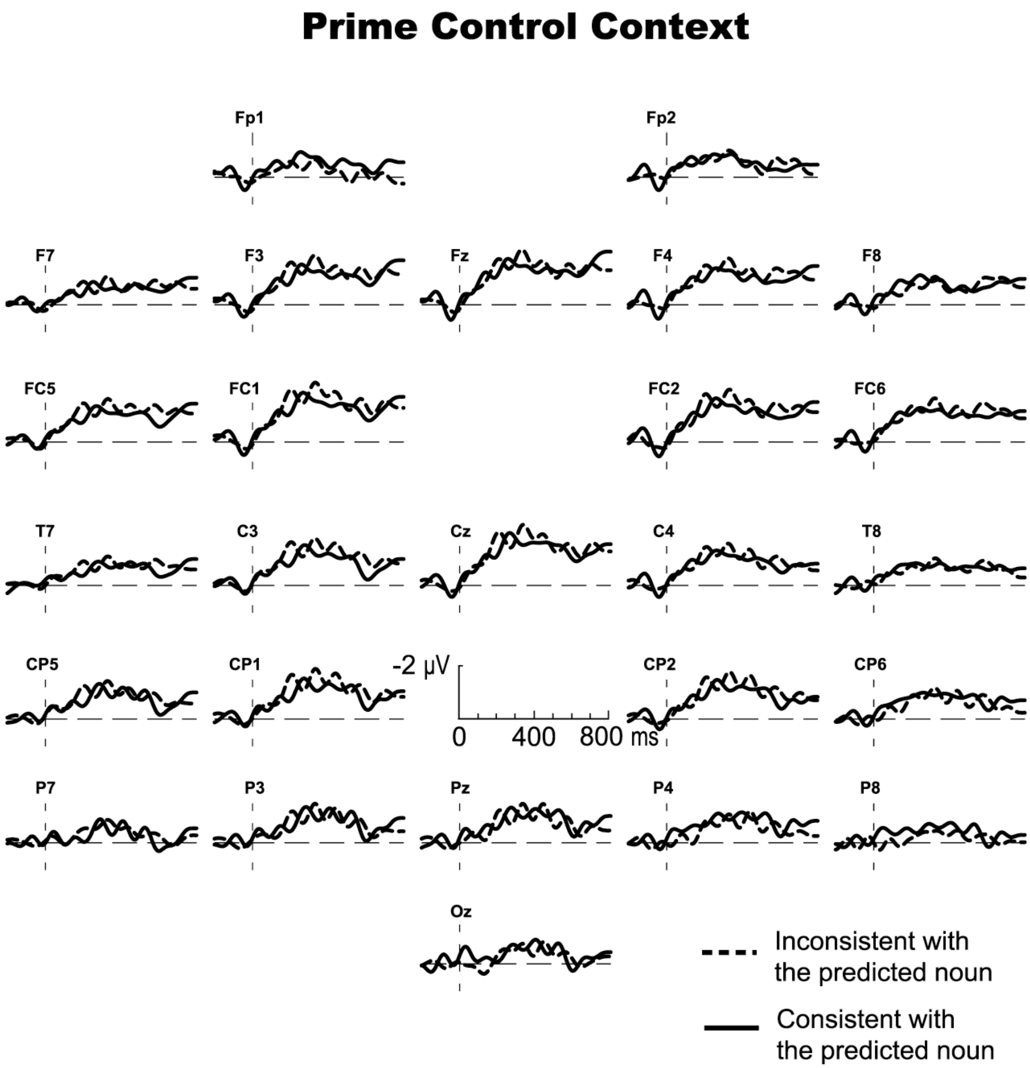


Figure 3.2 *Grand average ERPs elicited by the critical adjectives in a prime control context. Black lines represent the response to adjectives bearing an inflection that is consistent with the gender of the predicted noun; red lines represent responses to gender inconsistent adjectives. The ERPs are timelocked to the onset of the adjective, and are filtered (8 Hz high cut-off, 48 dB/oct) for presentation purposes only.*

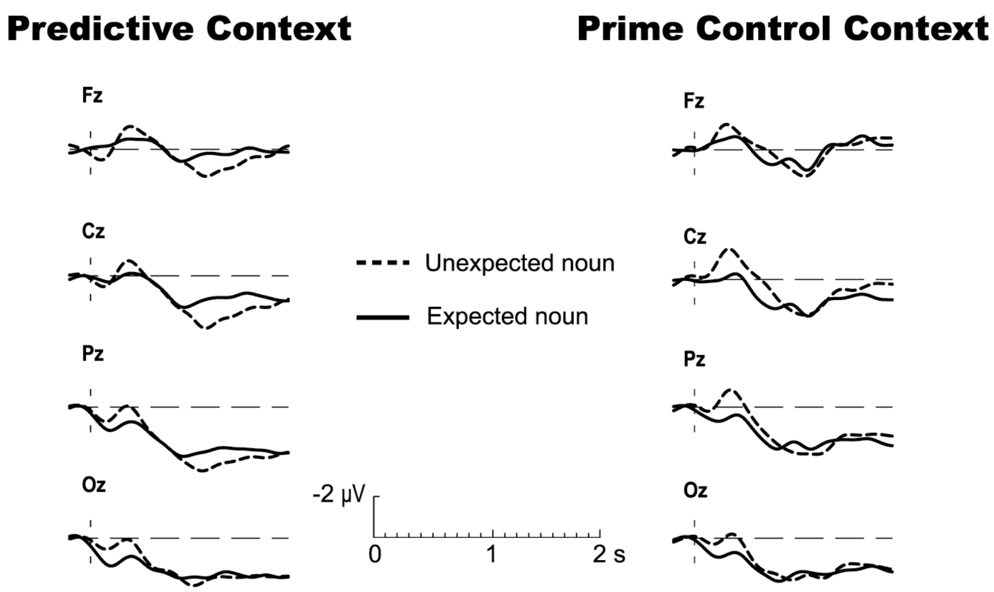


Figure 3.3 *Grand average ERPs elicited by the critical nouns in a predictive context and prime control context over the midline electrodes. Black lines represent the ERP to the predicted nouns; red lines represent the ERP to unexpected but still congruent nouns. The ERPs are timelocked to the onset of the noun, and are filtered (4 Hz high cut-off, 48 dB/oct) for presentation purposes only.*

For completeness, Figure 3.3 shows the ERPs elicited by the nouns that follow the critical adjectives in both types of context. Unexpected nouns that follow a predictive context evoke a larger N400 between 200 and 600 ms, as well as a positivity that emerges at around 900 ms, and remains until 1600 ms after word onset. When the same nouns follow a prime control context the N400 effect is still present, but this effect is not followed by a later positive deflection.

The amplitude of the N400 differs significantly for expected and unexpected nouns between 200 and 600 ms ($F(1,28) = 7.3$; $p = .01$). The effect does not reliably differ between predictive and prime control context ($F(1,28) = 1.6$; $p = .22$). The later widespread positive component elicited by unexpected nouns is reflected in a significant interaction between expectedness and context-type ($F(1,28) = 4.7$; $p = .04$) between 1000 and 1500 ms. Post-hoc tests show that this positive shift is only present in the predictive stories ($F(1,28) = 10.3$; $p = .003$), and not in the prime control stories ($F(1,28) = 0.6$; $p = .82$).

Discussion

After listening to a constraining discourse whose message suggests a plausible upcoming noun, an adjective with an inflection that is not in line with the gender of the predictable noun elicits a differential ERP effect compared to the adjectives that are consistent with the gender of this noun. Importantly, at this point in the story both gender-inflections are semantically and syntactically correct, since no noun has been shown yet. This ERP effect can therefore only be attributed to a mismatch of the observed gender with the gender of the predictable noun, indicating that listeners have already activated (the gender of) the word they think will follow. This finding thus confirms earlier claims that people use the cues provided by the sentential context or wider discourse to anticipate upcoming words (DeLong et al., 2005; Kamide et al., 2003; Van Berkum et al., 2005; Wicha, Bates et al., 2003; Wicha, Moreno et al., 2003; Wicha et al., 2004).

Crucially, the effect of prediction mismatch was absent in the prime control context, which contained the same prime words but did not support a lexical prediction at the message level. This shows that specific lexical predictions of the type observed here are not based on a simple word-based automatic priming process, but critically rely on the precise message-level content. In other words, it is the exact message that counts here, and not the compilation of individual words. Note that we did not specifically include strong primes into the discourse. We therefore can not exclude that in the presence of such primes, prime-induced predictions can also arise independently of the message of the surrounding discourse. The present results, however, clearly show that stories of the type used here induce predictions that are based on the actual message of the preceding context.

The observed electrophysiological consequence of a prediction mismatch resembles other effects of prediction mismatch that have been reported in previous studies in polarity and timing and, to a lesser extent, scalp distribution (DeLong et al., 2005; Wicha, Bates et al., 2003; Wicha, Moreno et al., 2003). Although our effect resembles the standard N400 effect in timing and polarity, the scalp distribution of the prediction mismatch effect does not resemble the standard distribution of the centro-parietal N400 effect. Therefore, we are reluctant to interpret the present effect as a canonical N400 effect. At the same time, though, the timing of the ERP effect and the critical involvement of high-level meaning are consistent with the idea that at least some of the neural generators that underlie the canonical N400 effect might also be activated when people hear something that (indirectly) mismatches their prediction.

In addition to these central results, Figure 3.3 shows that unexpected nouns evoked a larger N400 followed by a relatively long-lasting positive shift in the predictive context, compared to expected nouns. In the prime control condition the N400 effect was also present, whereas the late positive shift disappeared.

Within the domain of language processing, late positive components are often related to syntax-based reanalysis (Friederici, 1997; Osterhout, 1994). However, since the unexpected nouns in the present experiment are not incongruous at any level, they are not very likely to induce re-analysis of earlier syntactic assignments. An alternative possibility might be that the observed positivity reflects the processing of improbable events (Coulson, King, & Kutas, 1998). What this leaves to be explained, however, is why a similar late positivity was not observed in other studies with semantically unexpected words (e.g. Hoeks et al., 2004).

In contrast to the late positivity, the N400 is present in both predictive and prime control context. This pattern of results seems to suggest that the N400 does not reflect message-level expectancy or integration, but rather integrative or predictive processes related to word-based priming. However, results from a recent experiment where participants were presented the same stories without the critical preceding adjectives (Otten & Van Berkum, 2007) suggest that the discourse-based N400 effect cannot be solely attributed to processes reflecting automatic activation.

A possible explanation for the currently equivalent N400 effects in predictive and prime control stories might lie in the design of the stimuli. In the present experiment the unexpected nouns differ not only from the expected nouns in their level of expectancy and contextual fit, but also in their length and frequency. Hence, the larger N400 for unexpected nouns in both predictive and prime control context could to some extent be attributed to other factors than message-level expectancy. Furthermore, the cloze values used in this experiment reflect the expectancies that readers or listeners have right at the indefinite article. The adjectives, however, contain additional cues to the nature of the noun that might follow, which will critically alter expectations. As a result, the interpretation of the ERP effects evoked by the nouns is necessarily tentative.

Conclusion

We have shown that listeners use the information from the context to make predictions about what is to come next, confirming previous research on specific lexical prediction. Furthermore, in a natural discourse these predictions are not based on simple automatic activation processes, but on the exact message of the discourse. People are thus not only capable of rapidly extracting the full meaning of a discourse, but they can also use this knowledge to anticipate what might come next in the story, down to the level of specific upcoming words.

Chapter 4

The role of working memory capacity in lexical prediction

When reading a constraining story, people can anticipate how the story will continue, up to a very specific level. In this experiment we used event-related potentials to test whether readers with low working memory capacity (WMC) would differ from high WMC readers in their capability to make these on-line linguistic predictions (because of their lack of cognitive resources) or in the contextual basis of these predictions (because of their inability to suppress automatic, prime-activated predictions). High and low WMC participants were shown stories that were highly constraining for one specific noun, or stories that were not specifically predictive but contained the same prime words as the predictive stories. To test whether listeners made specific predictions, critical nouns were preceded by a determiner with a gender that was in line with, or contrasted with, the gender of the expected noun. Both high and low WMC readers showed an early negative deflection (300 - 600 ms) for unexpected compared to expected determiners, which was not present in the prime control condition. This shows that both groups can use the message of a predictive discourse to anticipate with which word(s) a story will continue. This early deflection was followed by a later negativity (900 - 1500 ms), but only for the low WMC participants. This suggests that WMC influences how readers process prediction-inconsistent information.

Introduction

Many studies have shown that people use contextual information to anticipate up to a very specific level which words will come next when a context or discourse is sufficiently constraining (DeLong et al., 2005; Van Berkum et al., 2005; Wicha, Bates et al., 2003; Wicha, Moreno et al., 2003; Wicha et al., 2004). The results of two recent experiments (Otten, Nieuwland, & Van Berkum, 2007; Otten & Van Berkum, in press) indicate that specific lexical predictions do not stem from automatic activation through individual words in the context, but that they instead are based on the message of the preceding discourse. In the present experiment we explored the relationship between specific lexical prediction and working memory capacity (WMC), exploring whether individuals with a less extensive WMC are also capable of online lexical prediction and, if they are, whether these predictions are based on the actual message of the discourse or on automatic activation.

In most models of text comprehension working memory plays an important role (cf. Caplan & Waters, 1999; Kintsch, 1998; Kintsch, Patel, & Ericsson, 1999). Working memory is usually defined as a limited capacity mental workspace where information is simultaneously maintained and processed (cf. Baddeley, 2003). In 1980 Daneman and Carpenter introduced the reading span task to quantify individual differences in the capacity of this memory system. To determine reading span, participants read aloud an increasing number of sentences, and are asked to recall the last word of each sentence. The maximum number of sentences for which recollection is perfect is taken as a measure of WMC. There is ample experimental evidence that WMC is related to successful language comprehension. For example, several experiments have shown a relation between WMC and syntactic processing (Daneman & Carpenter, 1980; King & Just, 1991; MacDonald, Just, & Carpenter, 1992) as well as semantic processing (Budd, Whitney, & Turley, 1995; Daneman & Carpenter, 1983; Singer, Andrusiak, Reisdorf, & Black, 1992). Here we will outline in what way differences in working memory could directly influence if (and if so, how) readers make specific lexical predictions.

There are two factors related to WMC that could have a direct relation with making on-line lexical predictions. First, differences in WMC have often been ascribed to differences in the availability and allocation of the resources that are necessary to store and manipulate information in working memory (Daneman & Carpenter, 1980, 1983; Just & Carpenter, 1992; Just, Carpenter, & Keller, 1996). Within this framework, low WMC individuals are hypothesized to have less resources, which will put them at a disadvantage in highly WM-demanding tasks like language comprehension. More recently, however, it has been suggested that differences in WMC actually reflect differences in the ability to control attention in order to maintain or suppress information (Engle, 2002). According to this theory, problems in language processing are due to competing demands and the relative inability of low WMC individuals to suppress information. Both a lack of resources and the inability to suppress related but irrelevant information could cause low WMC readers to make less, or less relevant, predictions

An indication that the lack of resources for low WMC readers could play an important part in predictive processes during language comprehension comes from the study of predictive inferences. Predictive inferences are optional, elaborative inferences about predictable events, for example when people assume that an actress who has fallen from a 14 story building is probably dead (McKoon & Ratcliff, 1986; O'Brien, Shank, Myers, & Rayner, 1988). Several experiments have provided evidence that high WMC readers process information that is in line with a predictive inference more easily than prediction-inconsistent information, whereas low WMC readers show no advantage for consistent information over inconsistent information (Calvo, 2001; Estevez & Calvo, 2000; St George, Mannes, & Hoffman, 1997). These results suggest the possibility that low WMC individuals have less resources that they can allocate to language comprehension, and are thus less able than high WMC individuals to make specific lexical predictions when they are processing linguistic input.

The second factor that might influence on-line lexical prediction in low WMC participants is suppression. There is much experimental evidence that people with high WMC are better in suppressing uninformative but related information, in the linguistic domain (M. A. Gernsbacher & M. Faust, 1991; Gernsbacher & Faust, 1995; M. A. Gernsbacher & M. E. Faust, 1991) as well as other aspects of cognitive functioning (Rosen & Engle, 1998). Consequently, it is possible that automatic activation plays a larger role in the process of prediction for low WMC individuals than for high WMC individuals. Predictive contexts usually contain one ore more words that are related to the predictable word. Previous studies from our lab with participants sampled randomly from the general college-population have shown that the effect of specific lexical prediction that emerges in a predictive context is not present in a prime control context (Otten et al., 2007; Otten & Van Berkum, in press), thus strongly suggesting that in the average participant prediction is based on the message of the discourse. However, low WMC individuals have been shown to be less able to suppress the automatic activation by primes (Engle, 2002). Therefore, it is possible that when we selectively examine low working

Table 4.1 Example of stimulus materials	
1. Predictive Discourse	
Predictable determiner & noun	**Unpredictable determiner & noun**
De actrice had een prachtige jurk aan, maar ze vond haar hals nog wat sober. Ze pakte ***de*** *verfijnde maar toch opvallende ketting die haar stylist had uitgezocht.* The actress wore a beautiful dress, but she thought her neck was a little plain. She picked up **the**$_{com}$ delicate yet striking necklace that had been selected by her stylist.	*De actrice had een prachtige jurk aan, maar ze vond haar hals nog wat sober. Ze pakte* ***het*** *verfijnde maar toch opvallende collier dat haar stylist had uitgezocht.* The actress wore a beautiful dress, but she thought her neck was a little plain. She picked up **the**$_{\mathbf{neut}}$ delicate yet striking collar that had been selected by her stylist.
2. Prime Control Discourse	
Predictable determiner & noun	**Unpredictable determiner & noun**
De actrice vond dat haar hals goed uitkwam in de sobere jurk. Ze pakte ***de*** *verfijnde maar toch opvallende ketting die haar stylist had uitgezocht.* The actress thought her neck looked beautiful in the plain dress. She picked up **the**$_{com}$ delicate yet striking *necklace* that had been selected by her stylist.	*De actrice vond dat haar hals goed uitkwam in de sobere jurk. Ze pakte* ***het*** *verfijnde maar toch opvallende collier dat haar stylist had uitgezocht.* The actress thought her neck looked beautiful in the plain dress. She picked up **the**$_{\mathbf{neut}}$ delicate yet striking collar that had been selected by her stylist.

memory participants we will see that these readers base their specific lexical predictions not on the message of the discourse, but on the primes that are present in the discourse. In this case, both low WMC and high WMC readers will make specific lexical predictions, but these predictions will differ in their contextual basis.

In the present experiment we explore whether WMC indeed affects specific lexical prediction, either by influencing the overall ability to create predictions or by influencing the contextual basis of the predictions. To test the actual effects of prediction, i.e. the *pre*-activation of a predictable word, we focussed not on the predicted word itself (*ketting* [*necklace*] in the example in Table 4.1), but on the definite article (definite determiner) that precedes it. In

Dutch, definite determiners vary with the arbitrary, lexically memorized gender of the noun they precede. Nouns of common gender are preceded by the common gender definite determiner *de* [the_{com}], whereas nouns of neuter gender are preceded by the neuter gender determiner *het* [the_{neut}].

Looking at the predictive story 1 in Table 4.1, if listeners strongly anticipate a specific noun like *ketting* (a common gender noun) a determiner that indicates neuter gender (*het*) will come as an 'unpleasant' surprise compared to the prediction-consistent determiner (*de*). Several experiments have shown that such prediction-inconsistent pronominal gender information evokes a differential event related potential (ERP) effect, even though, with the noun still to be presented, the gender information is at that point in the sentence fully unproblematic. Such a differential ERP effect shows that the readers have predicted the noun that is to follow, as well as its gender. Effects of specific lexical prediction have been observed for gender-inflected adjectives in Dutch (Otten et al., 2007; Otten & Van Berkum, in press; Van Berkum et al., 2005), gender-specific determiners in Spanish (Wicha, Bates et al., 2003; Wicha, Moreno et al., 2003; Wicha et al., 2004) and the a/an distinction in English (DeLong et al., 2005). To test whether the overall ability to anticipate specific words varies with WMC, we compared the ERPs elicited by predicted and unpredicted determiners for individuals with low and high WMC. If both groups generally make specific predictions about upcoming words, both should show the same difference between the ERPs evoked by predictable and unpredictable determiners. If, however, a low WMC indeed hampers or even precludes on-line prediction, the low WMC individuals should show a diminished effect compared to the high WMC group, or the effect could even completely disappear for individuals with low WMC.

As the experimental evidence we summarized above indicates, low WMC individuals are also less able to suppress information and more susceptible to interference. Consequently, low WMC individuals could rely more on automatic activation to make predictions than high WMC individuals. A predictive context, like our Example 1 in Table 4.1, usually contains one or more words that are somewhat to strongly related to the most predictable continuation of that story (i.e. "hals (neck)" -"ketting (necklace)"). The higher level of activation for the predictable word could thus rely on automatic activation through related words in the preceding discourse. To test the effect of automatic priming, we have created for each predictive story a non-predictive story that contained the same prime words as the original predictive story. As is clear from Example 2 (Table 4.1) the previously predictable word ("necklace") is not expected anymore based on the message of the discourse. However, if the pre-activation of the predictable word "necklace" simply depends on the presence of related words like "neck" and "dress" in the discourse, then the less predictable word "collar" may still evoke a different ERP effect in the prime control context. We know from previous studies (Otten et al., 2007; Otten & Van Berkum, in press) that for the generally high WMC college-population prediction is based on the message of the discourse. However, if low WMC individuals are indeed more susceptible to automatic activation, and thus are more inclined to base their predictions on related context words, then they might show an effect of prediction in the predictive as well as in the prime control context. Since high WMC individuals are hypothesized to be better in inhibiting unrelated intrusions, this should not be the case for the high WMC group.

Methods

Participants

38 right-handed native speakers of Dutch participated in the experiment. 19 Participants with a high WMC and 19 participants with a low WMC were selected from our subject pool. They were paid 25 euro or awarded course credit. 1 subject was excluded from the analyses because of technical problems, and a further 6 participants were excluded because more than 50% of the critical trials were deleted due to artifacts (see below for details). The remaining 31 participants (21 female participants) were on average 20 years old (range 18-25 years).

Reading Span Task

Participants were invited to participate based on their score on the Reading Span task originally designed by Daneman and Carpenter (Daneman & Carpenter, 1980). A computerized Dutch version of the Reading Span Task (Van den Noort, Bosch, & Hugdahl, 2005) was used to measure verbal working memory performance. This new version consists of five sets of 20 sentences, matched for sentence-length (number of syllables) and matched for the number of letters, number of syllables and frequency of the final word. The sentences were presented in different set sizes (2, 3, 4, 5 or 6 sentences) in random order. Participants read aloud the sentences from a computer screen. When a subject had finished a sentence, he immediately pressed the space bar triggering the onset of the next sentence. If the subject could not finish the sentence in 6.5 seconds, the next sentence was automatically presented. When a subject had completed all the sentences of a set, a recall-cue was presented and he/ she had to recall the final words of the sentences from that set. The experimenter registered and scored the responses of the subject. Participants were instructed to read for comprehension with a normal pace (though encouraged to read faster if they were not able to read the sentences in 6.5 s). Reading Span score was computed as the total number of final words that were correctly recalled. Participants were selected for the high WMC group if they recalled 75 or more words correctly, and for the low WMC group if their score was below 65 words. The 17 high WMC participants had an average score of 83 words (range 77 – 92, average age 20,8). The 14 low WMC participants had an average score of 56 words (range 49 – 64, average age 20,6).

Materials

The critical stimuli were 160 mini-stories, that consisted of a context sentence followed by a target sentence. For each item we created a predictive context sentence, that was predictive at a message level, as well as a prime control context sentence, that contained the same prime words but was *not* predictive at the message level. The following target sentence either contained the predicted word or an unexpected but still completely coherent alternative. We assessed the predictiveness of the constraining and prime control stories in a pencil-and-paper sentence completion test. In this so called cloze test, we presented participants with the items which were truncated at the position of the critical noun, after an indefinite gender-neutral determiner[3]. In predictive stories truncated after this indefinite determiner, the expected critical word had an average cloze value of 74% (sd = 14%), and the unexpected critical word had an average cloze value of 3% (sd = 6%). In non-predictive prime control stories these same two sets of critical words had average cloze values of 18% (sd = 15%) and 3% (sd = 7%) respectively.

[3] These cloze values were originally collected for another experiment (see Otten & Van Berkum, 2007), in which the critical prediction-inconsistent manipulation lay in the gender-related inflection of the adjectives that preceded the noun. The stories in this cloze test were truncated after the singular *indefinite* determiner "een", which is always identical for common gender and neuter gender nouns (contrary to the singular definite determiners "de" and "het"). The indefinite determiner does not provide the readers with additional information about the gender of the upcoming noun, and, because all our nouns were count nouns, its singularity also does not provide discriminative information. Therefore, we assume that these cloze values provide a reasonable estimate of message-level predictability at the point just before the *definite* determiners in the present experiment.

In order to help interpret the ERPs on later nouns, we also conducted a second cloze test in which we evaluated the expectancies after the participants had read an indefinite determiner *and* the intervening gender-inflected (and content-bearing) adjectives. In predictive stories truncated in this way just before the noun, the expected critical word had an average cloze value of 73% (sd = 44%), and the unexpected critical word had an average cloze value of 22% (sd = 42%). In non-predictive prime control stories these same two sets of critical words had average cloze values of 47% (sd = 45%) and 15% (sd = 35%) respectively. As can be seen, only the average cloze value for expected words in predictable stories remained virtually unchanged, whereas all other cloze values increased, presumably due to the combined information provided by content-bearing adjectives and gender inflections.

In this experiment, we tested the pre-activation of the predictable nouns with the preceding gender-marked definite determiner. In Dutch definite determiners can be of a common gender ("de") or of a neuter gender ("het"). As such, a definite determiner can be consistent or inconsistent relative to the gender of the predicted noun, but at the time that listener read these determiners neither poses an overt violation. Furthermore, to avoid grammatical violations later in the sentence, prediction-inconsistent determiners were always followed by a coherent but much less expected alternative noun, with a gender that matched the determiner.

The definite determiner preceding the target noun was always followed by three to five words before the critical noun was presented. The intervening words were the same in all four conditions. The expected or unexpected noun was never sentence-final, but was always followed by at least three more words. The first three words following the target noun were the same for all four conditions. In 98 out of the 160 items the expected nouns were of common gender, which results in 61 % of the definite determiners being the_{com} ('de'). Unexpected nouns were slightly longer than expected nouns (7.4 versus 6.1 characters). Furthermore, the unexpected nouns were less frequent than the expected nouns, with an average of 33 occurrences in 1 million written words (sd = 53) for the expected nouns, versus 24 occurrences (sd = 96) for unexpected nouns (word form frequencies taken from the Celex-database). A list with all critical items (in Dutch) can be obtained from the first author.

The 160 items of the present experiment (40 for each of the four conditions shown in Table 4.1) were rotated so that three more lists of stimuli were created. Each of the four lists contained all 160 experimental stimuli, 80 stories in the constraining context version and 80 with a prime control context. 40 of the 80 constraining items and 40 of the 80 prime control items contained the expected noun (and therefore the expected determiner) while the remaining 40 had an unexpected noun (and an unexpected determiner) at the target position. Each participant was shown one of these four lists of stimuli, so that one participant saw all the stimuli, but never in more than one condition.

Procedure and EEG recording

The 160 stories were shown to the subject in blocks of 40 with breaks between the blocks. Participants were asked to read for comprehension and were not required to perform any other task. The electroencephalogram (EEG) was recorded from 30 electrode sites, mounted in an elastic cap, each referenced to the left mastoid. Blinks and vertical eye-movements were registered by placing an electrode under the left eye, also referenced to the left mastoid. Electrode impedance was kept below 5 kOhms during the experiment. The EEG was amplified with BrainAmps amplifiers (BrainProducts, München), and-pass filtered at 0.03 Hz-100 Hz and sampled with a frequency of 500 Hz.

The stimuli were presented in black 36 point courier new font on a light grey background on a fast TFT display (Benq Q7C4) positioned approximately 80 cm away from the subject. Before each trial, a fixation cross was shown in the centre of the screen for 2.5 s. Participants were instructed to avoid blinks and eye-movement when the words were presented on screen, and were encouraged to blink when the fixation cross was shown. To signal the start of each trial to the subject, a beep was presented 1 s before the onset of the first word. The stories were then presented word for word. To make this presentation more natural, we used a Variable Serial Visual Presentation (VSVP) procedure in which the presentation duration of each non-critical word varied with its length and position in the sentence (Otten & Van Berkum, 2007). For the materials at hand, the average presentation time for all words (including critical words) was 326 ms. Critical determiners and nouns and the three words between these target words were presented with a fixed duration of 376 ms, based on the average critical word length across all stories. All words had the same ISI of 106 ms.

The electroencephalogram (EEG) was recorded from 30 electrode sites (FP1, FP2, F9, F7, F3, Fz, F4, F8, F10, FT9, FC5, FC2, FC6, FC1, FT10, T7, C3, Cz, C4, T8, CP5, CP1, Cp2, Cp6, P7, P3, Pz, P4, P8 and Oz), mounted in an elastic cap, each referenced to the left mastoid. The EEG signal was re-referenced off-line to the average of right and left mastoids. Blinks and eye movements were removed from the data using a procedure based on Independent Component Analysis (ICA) as described by Jung et al (Jung, Makeig, Humphries et al., 2000; Jung, Makeig, Westerfield et al., 2000). The data were then segmented in epochs from 500 ms before critical word onset until 1500 ms after critical word onset, for both the determiners and the nouns that followed. After baseline-correcting the signals by subtracting mean amplitude in the 150 ms preceding critical word onset, we eliminated segments in which the signal exceeded ±75 μV, or which featured a linear drift of more than ±50 μV, beginning before the onset of the critical word. We excluded 6 participants from subsequent analysis because they lost more than half of the critical trials as a result of this procedure (the average rejection rate ranged from 63% to 50% for these participants). For the other 31 participants, the main rejection rate was 11% (range 27% - 1%). For each subject and condition the remaining epochs were then averaged.

Analyses

To assess not only the effects consistency and context type, but also the possible interaction with electrode position the ERPs elicited by determiners and nouns were evaluated in an ANOVA with Consistency (prediction consistent/ prediction-inconsistent), Context (predictive/ prime control), Hemisphere (left/ right) and Anteriority (anterior/ posterior) as within participants factors and WMC (high/ low) as a between participants factor. This analysis thus involved four quadrants: (1) left-anterior, comprising FP1, F3, F7, F9, FC1, FC5 and FT9; (2) right-anterior, comprising FP2, F4, F8, F10, FC2, FC6 and FT10; (3) left-posterior, comprising C3, T7, CP1, CP5, P3 and P7; (4) right-posterior, comprising C4, T8, CP2, CP6, P4 and P8. Effects on the midline electrodes (Fz, Cz, Pz and Oz) were assessed in a separate ANOVA crossing the factors Context, Consistency and Electrode position with WMC. F tests with more than one degree of freedom in the numerator were adjusted by means of the Greenhouse-Geisser or Huynh–Feldt correction where appropriate. Uncorrected degrees of freedom and corrected P-values are reported.

Results

Determiners: effects of prediction and WMC

Figure 4.1 shows the ERPs evoked by expected and unexpected determiners in a predictive and prime control context, averaged over all participants. Unexpected determiners elicit a more negative ERP between 200 and 600 ms over right-frontal electrodes, followed by another, more centrally distributed

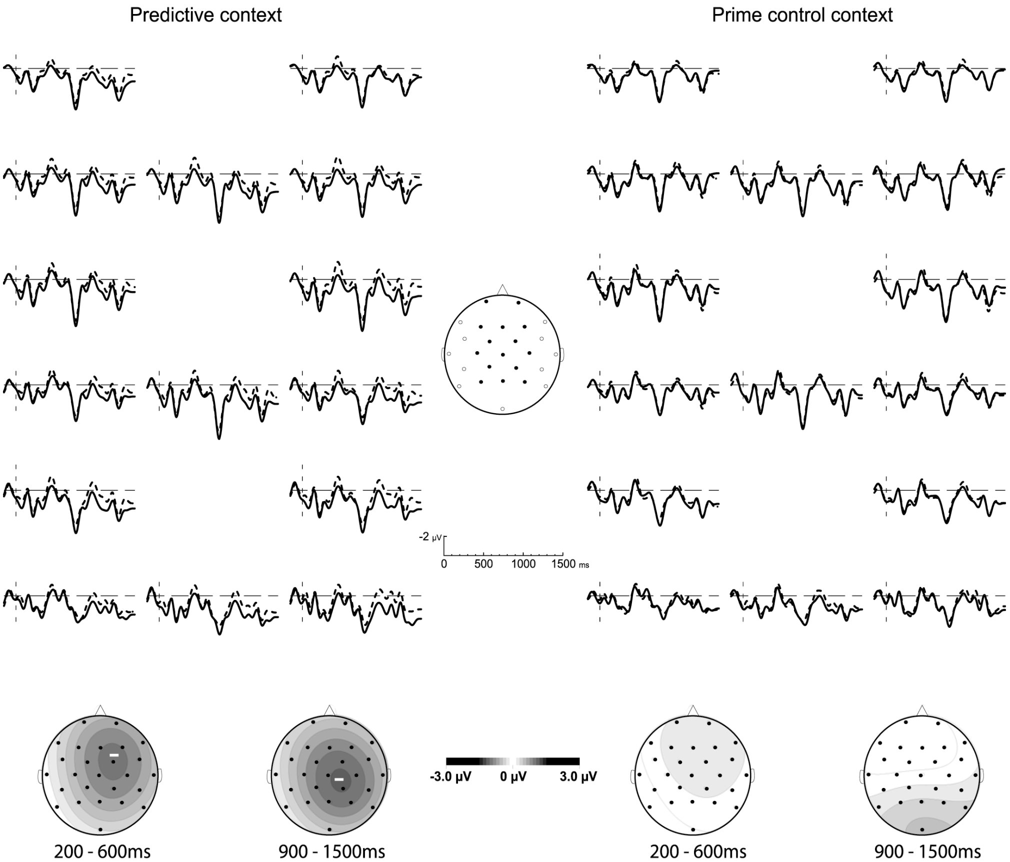


Figure 4.1 *ERPs elicited by determiners with a prediction-inconsistent gender (dotted line) and prediction-consistent gender (solid line) for both high and low WMC readers. The left-hand panel shows the ERPs for the determiners in the highly constraining predictive discourse, the left-hand panel shows the ERPs for the prime control context. The scalp distributions corresponding to the effect of prediction-consistency (prediction-inconsistent – prediction-consistent) are depicted for the two time intervals that were analysed.*

negative shift between 900 and 1500. These effects show in the predictive context, but not in the prime control context. Figures 4.2 and 4.3 show the same ERPs separated for respectively low WMC readers and high WMC readers. Low WMC readers show the same biphasic ERP pattern as seen in the grand average across all participants: an early negative shift after the onset of the determiner with an anterior maximum between 200 and 600 ms, followed, from about 900 ms onwards, by a more centrally distributed sustained negative shift. In contrast, high WMC participants only display an early and slightly right-lateralized anterior negativity, from about 200 until 600 ms after

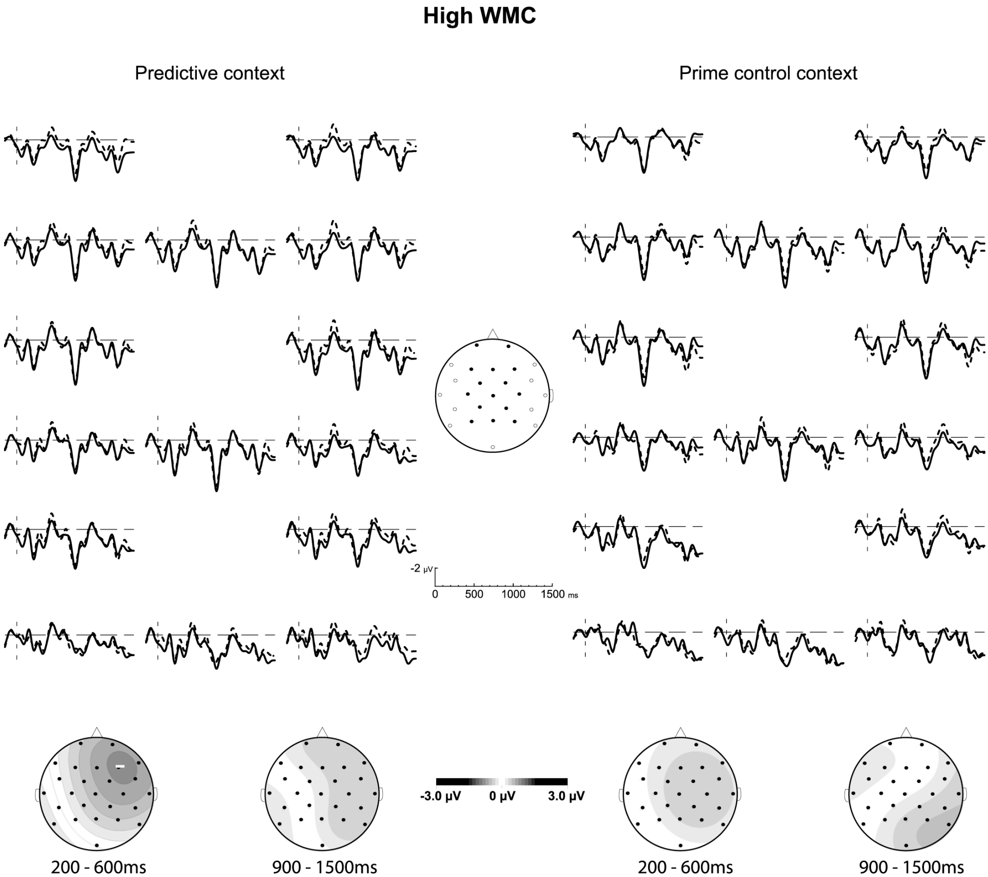


__Figure 4.3__ ERPs elicited by determiners with a prediction-inconsistent gender (dotted line) and prediction-consistent gender (solid line) in the predictive (left-hand panel) and prime control discourse (right-hand panel) for high WMC readers only.

the onset of the determiner. Based on the visual inspection of the data we have analysed the waveforms between 200 and 600 ms and between 900 and 1500 ms.

Between 200 and 600 ms the ERP elicited by unexpected determiners is significantly more negative than the one evoked by expected determiners ($F(1,29) = 4.70$, $p = .04$). This effect of expectancy interacts with context type ($F(1,29) = 4.28$, $p = .05$), and subsequent post-hoc test show that the effect of expectancy is only present in the predictive context ($F(1,30) = 4.44$, $p=.04$) and

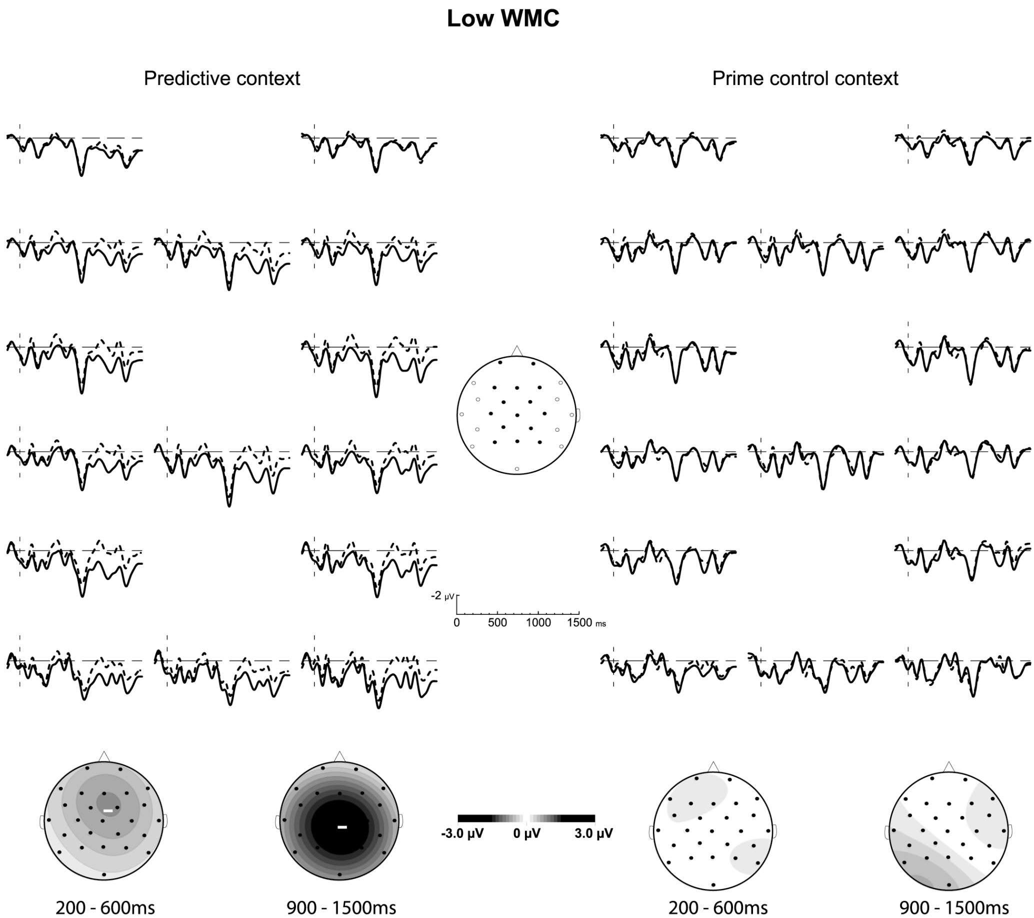


Figure 4.2 *ERPs elicited by determiners with a prediction-inconsistent gender (dotted line) and prediction-consistent gender (solid line) in the predictive (left-hand panel) and prime control discourse (right-hand panel) for low WMC readers only.*

not in the prime control context ($F(1,30) = .55$, $p = .46$). The absence of a significant interaction with WMC (Expectancy * Context * WMC: $F(1,29) = .03$, $p = .610$) confirms that this early negativity is present in both low WMC readers and high WMC readers. The distribution of the early effect of expectancy however does differ between the two WMC groups, since the negativity is widespread in the low WMC group and more right lateralised in the high WMC group, resulting in a significant interaction between Expectancy, Hemisphere and WMC ($F(1,29) = 5.13$, $p =.03$).

Also in the late time interval (900 to 1500 ms) unexpected determiners evoke a more negative inflection compared to the expected determiner ($F(1,29) = 7.62$, $p = .01$). The effect of determiner expectancy is differentially modified by context depending on WMC group (Expectancy * Context * WMC: $F(1,29) = 5.37$, $p= .03$). Separate follow-up ANOVAs for the two WMC groups show that the late negativity is only present for the participants with a low WMC (Expectancy: $F(1,29) = 4.89$, $p = .04$; Expectancy * Context Type: $F(1,29) = 5.26$, $p = .04$), and not for the participants with a high WMC (Expectancy: $F(1,29) = 2.51$, $p = .133$; Expectancy * Context Type: $F(1,29) = .01$, $p = .94$). Within the low WMC group the negativity is more pronounced over posterior than anterior electrodes, which is reflected by a significant interaction with the midline ($F(1,29) = 4.35$, $p = .03$) and the posterior/anterior factor ($F(1,29) = 5.85$, $p = .03$). Further ad-hoc tests show that within the low WMC group the late negativity is only present in the predictive context ($F(1,13) = 10.82$, $p = .006$) and not in the prime control context ($F(1,13) = .033$, $p = .86$).

Nouns: N400, P600 and WMC

Figure 4.4 shows the ERPs elicited by expected and unexpected noun, in the predictive and the prime control context. Unexpected words evoke a larger N400 followed by a long lasting positive shift in the predictive context. In the prime control context, where both nouns have comparable message-based levels of expectedness, unexpected nouns still elicit a larger N400 than expected nouns. The positive shift, however, is significantly decreased in the prime control context.

The larger N400 for unexpected nouns is reflected by a significant main effect of expectedness between 300 and 500 ms ($F(1,29) = 5.76$, $p =.02$), which is largest over posterior electrodes (Expectancy * PosteriorAnterior: $F(1,29) = 27,22$; $p = .00$). The type of context in which the (un)expected noun is presented does not significantly alter the N400 effect ($F(1,29) = .66$, $p = .42$). Both low and high WMC subjects show this N400, in both context types (Expectancy * WMC: $F(1,29) = .07$, $p = .80$; Expectancy * Context * WMC:

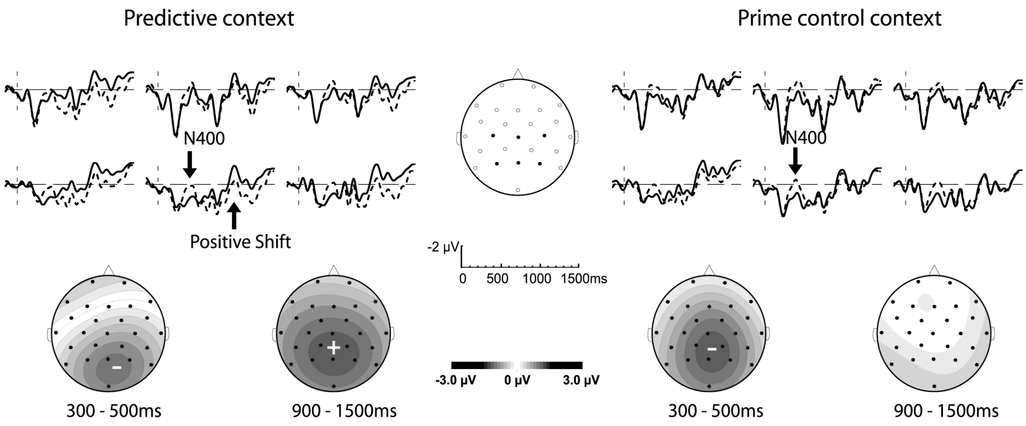


***Figure 4.4** ERPs elicited by unexpected (dotted line) and expected (solid line) nouns in the predictive (left-hand panel) and prime control discourse (right-hand panel) for high and low WMC readers.*

F(1,29) = .04, p = .84). Although the effect is present for both groups of readers, there is a difference in the laterality of the effect between low and high WMC readers, as shown by a significant interaction between expectancy, hemisphere and WMC (F(1,29) = 7,19, p = 0.01)

The results furthermore reveal a significant positive ERP effect of noun-expectedness from 900 ms after stimulus onset (F(1,29) = 9.25, p =.005). This positive shift is largest over posterior electrodes (Expectancy * PosteriorAnterior: F(1,29) = 7.29, p = .011) and interacts with context type (F(1,29) =4.08, p=.05). Post hoc tests show that the effect of expectancy is only significant in the predictive context (F(1,29) = 9.13, p = .005), not in the prime control context (F(1,29) = .76, p = .389). The late positive effect is present in both WMC groups (Expectancy * WMC: F(1,29) = .38, p = .55; Expectancy * Context * WMC: F(1,29) = 2.8, p = .11). Although Figure 4.5 suggest that there is a difference in strength and scalp-distribution of this late positive effect between the low and high WMC group, this is not backed up by any significant interaction (Expectancy * Hemisphere * WMC: F(1,29) = .48, p = .50; Expectancy * PosteriorAnterior * WMC: F(1,29) = .28, p = .60; Expectancy * Hemisphere * PosteriorAnterior * WMC: F(1,29) = .88, p = .36).

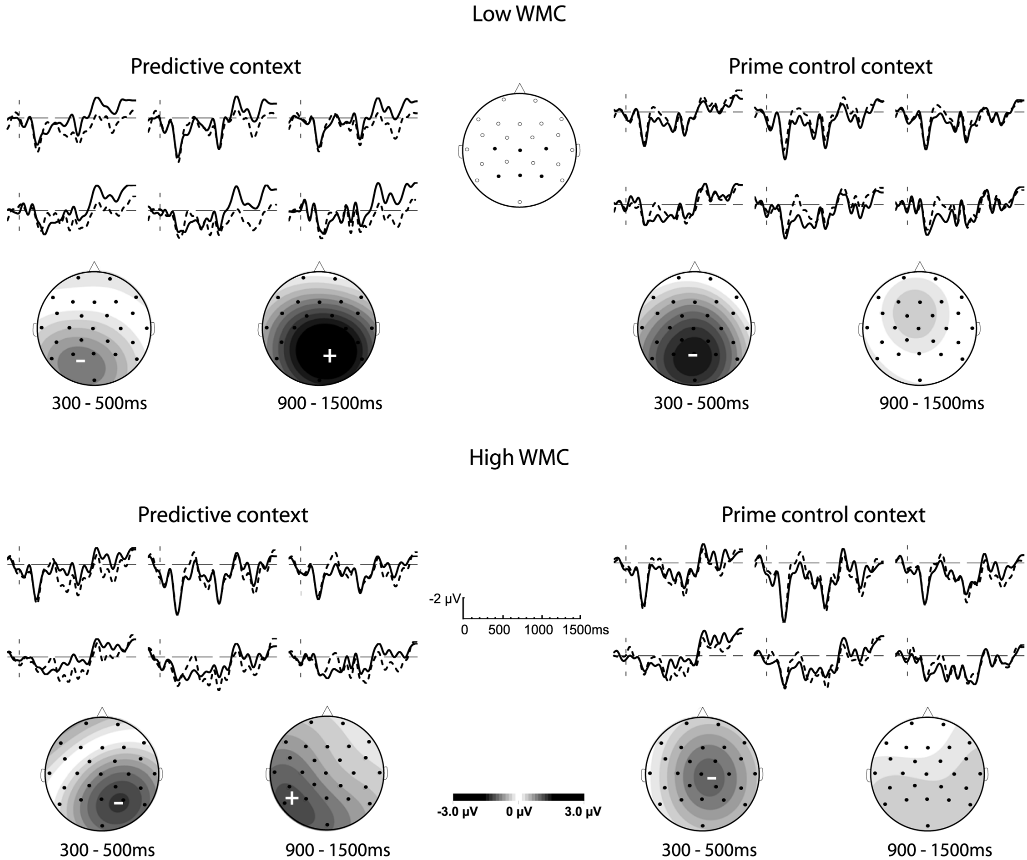


Figure 4.5 *ERPs elicited by unexpected (dotted line) and expected (solid line) nouns in the predictive (left-hand panel) and prime control discourse (right-hand panel) separated for high and low WMC readers.*

Discussion

Definite articles with a gender that was inconsistent with the gender of a discourse-predictable noun elicited a differential ERP compared to their prediction-consistent counterparts. Unexpected determiners evoke an anterior negative deflection between 200 and 600 ms, which is followed by a more central negative deflection between 900 and 1500 ms. Because the critical article and the later noun were always separated by at least three words (i.e., at least 1800 ms separated the onset of the critical determiner from the onset of the critical noun), these effects can not be attributed to the (un)expectedness of the noun that followed the determiner. Thus, the only difference between prediction-inconsistent and consistent articles was whether or not they agreed with the grammatical gender of the discourse-predictable noun. The differential ERP effects that accompany the unexpected articles therefore provide clear evidence for the fact that readers anticipate *specific* upcoming words, pre-activating the specific semantic as well as syntactic properties of the words. The early expectancy effect (between 200 and 600 ms) is present in the predictive context for both high WMC and low WMC groups. This shows that both low WMC and high WMC individuals have made a specific lexical prediction at the moment that the inconsistently inflected adjective was shown.

No such determiner-dependent effects emerged in the prime control context. Prime control stories contained the same prime words as predictive stories, but their message was not as constraining. The absence of a difference between expected and unexpected determiners shows that specific lexical predictions are based on the (predictive) message of the discourse, and not on the presence of related primes. Furthermore, both low and high WMC participants show no activation for unexpected determiners in the prime control context. This suggests that the contextual basis for predictive inferences is identical for both types of readers. Both groups use the actual message of the context to create specific lexical predictions.

The present findings confirm earlier observations that lexical predictive processes can not be traced back solely to priming (Otten et al., 2007; Otten & Van Berkum, in press). This does not mean, however, that message-based anticipation is the only process involved in lexical prediction. The stories used in this experiment were not specifically designed to contain strong primes, and the absence of a differential effect in prime control stories therefore does not provide compelling evidence against additional word-based priming in text comprehension. Furthermore, it seems highly likely that the concurrent syntactic analysis of the unfolding sentence also contributes to the anticipation of an upcoming noun (see Otten & Van Berkum, 2007; Van Berkum et al., 2005 for discussion). The only conclusion that can safely be drawn from the current pattern of results is that the effects observed in predictive stories in both WMC groups reflect the true message-dependent prediction of upcoming words.

There is a second reason why the absence of an effect in prime control stories must be interpreted with caution. The discourse-based prediction effect that we test for with the current experimental paradigm hinges on syntactic gender agreement between the determiner and an anticipated (but as yet to be presented) noun. This means that, given a determiner-dependent effect in predictive stories, we can infer not only that people were anticipating specific nouns, but also that they engaged in a form of anticipatory parsing, checking the agreement between the overt determiner and an anticipated noun (see Van Berkum et al., 2005 for discussion). As a consequence, though, the absence of an effect can reflect the absence of *either* causal links: it might be that predictions were not made here, but it might also be that people only take *message*-based predictions into account when they syntactically parse the determiner, and simply ignore whatever words are suggested by lower-level mechanisms. It seems difficult, perhaps even impossible, to disentangle these two possibilities. But the critical inference remains unaffected: the discourse-based prediction of upcoming words observed here and in similar experiments cannot be reduced to simple word-word (or scenario-mediated) priming mechanisms.

Predictions and working memory capacity

We did not observe a reliable difference between high and low WMC readers in their ability to make specific lexical predictions, or the contextual basis of these predictions. But the results do show a noticeable difference between the two WMC groups. When confronted with a prediction-inconsistent determiner, high and low WMC readers show an early effect (before 600 ms) of expectancy. Low WMC readers also show a late effect (after 900 ms). This suggests that low and high WMC individuals differ in the way they deal with information that affirms or disconfirms their predictions.

This additional electrophysiological response to the unexpected determiner displayed by the low WMC group could stem from the larger difficulties they have in suppressing their original prediction. When a reader strongly expects to read *schilderij*, while the preceding determiner has another gender ("de"), then this could call for adjustments to the prediction. The fact that the ERP response to these unexpected determiners involves a second, later ERP effect for the low WMC readers suggests that these adjustments are more demanding for the low than for the high WMC readers. This explanation is in line with the other evidence that shows that high WMC readers can more easily suppress information compared to low WMC readers (Engle, 2002; Whitney, Arnett, Driver, & Budd, 2001).

The separation in time and the difference in scalp distribution between the early and the late effect indicates that these two components could also reflect two different processes. This would suggest that reading an unexpected determiners invokes an additional process for the low WMC group, which is reflected by the late effect. In this light, the early effect most likely reflects the detection of the inconsistency, which is independent of WMC. The later effect could reflect the attempts to reintegrate this inconsistency with the prediction.

But why would only the low WMC readers attempt this, if the high WMC readers also have detected the inconsistency? An explanation might lie in the fact that low WMC readers are less able to maintain an overview of previous events, due to either a lack of resources (Daneman & Carpenter, 1980) or difficulty in maintaining attention (Engle, 2002). Previous events do not only include the preceding discourse of the present trial, but also of the previous trials. Some of these preceding trials will also have included inconsistencies. It is possible that high WMC readers are more able than the low WMC group to build a complete memory-representation of the preceding trials. If a high WMC reader detects a somewhat unexpected word, their (implicit) knowledge that encountering discrepancies in these stories is 'no big deal' might reduce the chances of an extensive repair process. The low WMC readers, on the other hand, will be relatively blank with regards to the preceding trials when they encounter an inconsistency. This could lead to more thorough processing of each individual unexpected word (see Brumback, Low, Gratton, & Fabiani, 2005 for a comparable argumentation). An interesting consequence of this interpretation is that the detection of prediction-inconsistent information is automatic, but that the consequences of such a detection are more controlled.

Our results show no evidence that the assumed lack of resources for the low WMC group decreases their ability to make specific lexical predictions. In this respect, our data contrasts with the results from predictive inference literature (Calvo, 2001; Estevez & Calvo, 2000; St George et al., 1997), which show that low WMC individuals generally do not make predictive inferences, whereas high WMC individuals do generate these elaborative inferences. Predictive inferences are assumptions about how a situation will develop, just like specific lexical predictions, but they are much more general and conceptual than specific lexical predictions. As such, they can form the basis of specific lexical predictions, but they are by no means identical. In a predictive inference study that was more comparable to the present design, Linderholm (2002) found that low WMC readers as well as high WMC readers showed a difference in reading times when processing information that was expected or unexpected. In this experiment, the inferences that were tested concerned specific words, and not broad concepts, and as such they were more comparable to the specific lexical predictions studied here than to the standard predictive inference. This seems to suggest that the observed differences high and low WMC readers (Calvo, 2001; Estevez & Calvo, 2000; St George et al., 1997) are task dependent, such that differences only arise in tasks where reactions to probe words are measured, and not in situations where more natural measures of prediction are involved.

Processing the expected and unexpected nouns

Unexpected nouns evoked a larger N400 followed by a relatively long-lasting positive shift in the predictive context, compared to expected nouns. In the prime control condition the N400 effect was also present, whereas the late positive shift disappeared.

The fact that the N400 effect is comparable for the predictive and the prime control context could be taken to suggest that the N400 does not reflect message-level expectancy or integration, but rather integrative or predictive processes related to word-based priming. However, two observations go against this interpretation. One is that results from the second cloze test (see methods) indicate that, at this later point in the sentence just before the noun, the cloze differences between unexpected and expected words have become much more similar in the predictive and prime control contexts (51% and 32%) than they were around the earlier determiner (71% vs 15%), presumably as the result of additional information provided by gender marking and content-bearing adjectives ("a delicate$_{com/neu}$ yet striking$_{com/neu}$ …"). Thus, on the assumption that cloze values primarily reflect message-level constraint, large differences in the two N400 effects were not to be expected.[4] The second observation is that in an experiment where we did have a strong manipulation of cloze differences just before the critical noun (71% difference in predictive stories, 15% in prime control stories(Otten & Van Berkum, 2007), a large part of the N400 effect elicited by unexpected words in predictable stories disappeared in prime control stories. In all, these findings suggest that the discourse-based N400 effect cannot be solely attributed to processes reflecting automatic activation by an unordered set of prime words.

The late positivity that is elicited by the unexpected nouns (e.g. "collar" instead of "necklace") in the predictive context is not present in the prime control context. Within the domain of language processing, late positive components have been related to (a) syntax-based analysis or reanalysis (Friederici, 1997; Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992), (b) the system checking up on its perception of the input (Kolk, Chwilla, van Herten, & Oor, 2003; Van Herten et al., 2005), (c) a conflict between various levels of linguistic analysis (syntactic, semantic, etc.) provided for by the input (Kuperberg, (2007), and (d) the processing of improbable events (Coulson et al., 1998). As for the first option, the unexpected nouns in the

[4] In fact, when we zoom in on the N400 components in all four conditions plotted together at a canonical electrode like Pz (not shown here), N400 amplitude seems to inversely track actual mean cloze values fairly well.

present experiment are not incongruous or problematic at any level, and they are therefore not very likely to induce re-analysis of earlier syntactic assignments. However, the second and third alternatives could perhaps explain the current positivity. Note that although there is no tension between the unfolding syntactic and semantic representations, there is a tension between what the system predicted and what it actually got. To the extent that a strongly predicted word is already entering the combinatorial analysis of what is being read (which is in fact precisely what the agreement-dependent determiner effects suggest), the incoming word form causes friction between form-based analysis on the one hand and the (extrapolated) syntactic/semantic analyses on the other (see also Vissers, Chwilla, & Kolk, 2006 for a comparable argument). Of course, this is also an improbable event, so we cannot rule out the more generic fourth account. We note that all but the first account must explain why standard semantic anomalies or cloze value manipulations often do not elicit a late positivity, and also that the options examined here need not exhaust possibilities. A single convincing interpretation of the current late positivity must thus await further research.

Unlike the determiner-induced prediction effect, the N400 and the late positivity are present for the high and low WMC readers. It is interesting to note that in this experiment the ERP response to *implicit* prediction-inconsistent information (prediction effect) varies with WMC, while the ERP response to the *overt* violation of a prediction (N400, late positivity) does not vary with WMC (see Nieuwland & Van Berkum, 2006a for a similar observation). However, several studies show that the N400 as well as the late positive components can be present or absent depending on the WMC of the reader (Brumback et al., 2005; St George et al., 1997; Van Petten et al., 1997). It is thus not always the case that overt violations are processed independent of WMC. It is possible that the nature of our overt prediction-violations can explain these different observations. The unexpected nouns were always coherent, even though they were not exactly in line with the predictions that the readers had created. This relatively high level of consistency for the unexpected nouns might have induced less 'deep' processing than other, completely inconsistent nouns would have induced. It is possible that we would have observed differences in how high and low WMC readers process unexpected nouns, if these nouns would have invoked more thorough attempts at repair.

Conclusion

Our results show that people can use a predictive discourse to anticipate with which word(s) a story will continue. With natural stories of the type examined here, these specific lexical predictions are based on the actual message of the discourse, and not on the primes that are present in the discourse. A diminished WMC does not influence our capability to make such specific lexical predictions nor the contextual basis of these predictions. However, the ERPs do show that the way that readers process prediction-inconsistent information is influenced by WMC. Low WMC readers show an additional, later ERP effect which is not present in high WMC readers. It is not clear whether this additional effect for the low WMC groups follows from their inability to suppress the initial prediction, or from their reduced memory for the overall context in which the stories are presented.

Chapter 5

Does appearance matter? A self-paced reading study exploring whether linguistic predictions extend to the visual form.

Sometimes a sentence or story is so constraining that we already know what our conversation partner is going to say, or how the paragraph will end, before we have even heard or read the entire utterance. In this study we have explored whether these linguistic predictions are limited to semantic and lexical features, or whether they also include the visual form of the expected word. In a self-paced reading experiment, participants were shown sentences that contained expected, unexpected or neutral nouns. These nouns could be presented in the standard font (same font as the other words in the story) or in a deviant font. We argued that, if people actually predict the visual form of an expected word, then an expected word in a deviant would mismatch this word-image prediction. We thus expected that the slowing effect of reading a deviant font word would be larger for expected words in a predictive context than for the neutral context. As expected, we observed that readers slow down more when they encountered a deviant typeface noun in a constraining discourse than in a neutral discourse. This suggests that readers make predictions about the actual visual appearance of the upcoming word. However, unexpected nouns showed a similar font-effect as the expected nouns, which could be taken to suggest that other factors than word-image prediction play a role. Thus, even though the findings indicate that linguistic predictions could have a visual component, this experiment does not provide the definitive answer.

Introduction

Electrophysiological studies suggest that people can use the linguistic context to make an informed guess about the word that will follow (DeLong et al., 2005; Otten et al., 2007; Otten & Van Berkum, in press, Van Berkum, 2005 #452; Wicha, Bates et al., 2003; Wicha, Moreno et al., 2003; Wicha et al., 2004). Levelt, Roelofs and Meyer (1999) suggest that the lexical representation of each word consists of three different information-stores: the lexical concept of the word, the lemma (which contains syntactic information about the word), and the lexeme/ morpheme (which specifies the phonological and the orthographic contents of the word). From the above-mentioned ERP studies we know that linguistic predictions include not only information about the meaning of the upcoming word, but also information about the syntactic properties of the word (in this case the grammatical gender Otten et al., in press; Van Berkum et al., 2005; Wicha et al., 2004). This shows that, when we predict a word, the lemma that is related to the anticipated word is also pre-activated. A recent study by DeLong, Urbach and Kutas (2005) suggests that linguistic predictions can also include the phonological information included in the lexeme. In this study Delong et al. used the fact that in English articles vary depending on the initial sound of the following word ('a' before words beginning with a consonant and 'an' before words beginning with a vowel). They observed a differential ERP to articles that were inconsistent with the initial sound of the predicted word. From this DeLong and colleagues conclude that linguistic predictions extend to the lexeme.

In this experiment we have explored whether linguistic predictions can go even beyond the lexeme, predicting exactly what the visual form of the predicted word will be, on paper or on the screen. To make predictions about the visual form of the word, first the orthographic structure, i.e. the letters of which it is composed, needs to be retrieved from memory. However, since the actual visual manifestation of these letters is not static (i.e. compare the visual form of "a" and "A"), pre-activation of orthographic information alone is not enough to make predictions about the visual form of a word. For this, readers need to combine memorised orthographic information with knowledge about the standard font in which the letters are presented. In this experiment, we have thus tested whether people are capable to translate their specific lexical prediction into a visual prediction, by combining orthographic information with dynamic, context-dependent information about the standard typeface.

Research on visual perception indicates that predictions can influence visual processing. Objects are more easily recognised when they are seen within, or shortly after, a consistent setting (i.e. a pan in a kitchen setting

Biederman, Mezzanotte, & Rabinowitz, 1982; Davenport & Potter, 2004; Palmer, 1975). For simple visual stimuli, there is evidence that predictive processes actually modify what we perceive (Jolij & Lamme, submitted). Taken together, these findings could be taken to suggest that anticipatory processes can influence visual perception (cf. Bar, 2004). When we combine these findings with the evidence that readers and listeners make highly specific predictions about upcoming words, this suggests that predictions about upcoming linguistic content could also include very specific predictions about the visual word-image.

We have tested this hypothesis in a self-paced reading experiment, which allows us to monitor the time it takes a reader to process each word in a sentence. We created stories that induced specific lexical predictions, such as Example story 1 in Table 5.1. In this story, the word *apotheek* (*pharmacy*) is highly expected to appear. Based on previous experiments (Otten et al., 2007; Otten & Van Berkum, in press) we know that readers use such a predictive context to pre-activate the supported word at the lexical level (lexical prediction). By combining knowledge about the orthography of the predicted word (i.e. the representation of the word in terms of its successive graphemes, i.e. "a" followed by "p" followed by "o" etc), and knowledge about the specific form of the letters in this context (i.e. the standard font) the reader can create a very detailed prediction about the visual word-image of this lexical prediction, i.e. apotheek.

If a reader indeed creates a word-image prediction, then seeing the expected word in an unexpected font ("apotheek") will mismatch the predicted visual word-image. Such a word-image mismatch will then lead to slower reading times. Of course, any differences in reading time between the word-image prediction match "apotheek" and the word-image prediction mismatch "apotheek" can not be solely attributed to the word-image mismatch, because besides a specific word-image mismatch, these nouns also have an unexpected and deviating font relative to the preceding context. To measure the more general effect that such an unexpected font has on the reading time of a word we also created non-predictive stories (Example 2 in Table 5.1). In these stories, the discourse did not support a lexical prediction. Any differences in reading time between "apotheek" and "apotheek" will thus be elicited by the difference in physical deviance for the noun in the standard and the deviating font.

Table 5.1. An example of the items used in experiment 1A, in the original Dutch version and an approximate English translation. The critical word is underlined and shown in the standard font and the deviating font.

Example 1: Expected Noun

Na het bezoek aan de streekdokter moest de boerin nog medicijnen ophalen. Ze ging daarvoor naar een apotheek/ apotheek in de grote stad.

After visiting the local doctor the farmer's wife still had to pick up the medication. For that, she went to the pharmacy/ pharmacy in the city.

Example 2: Neutral Noun

Piet fietste naar een apotheek/ apotheek om zijn medicijnen op te halen. Omdat hij daar toch was nam hij direct de pil voor zijn vriendin mee.

Piet cycled over to the pharmacy/ pharmacy to pick up his medication. On behalf of his girlfriend he also collected a prescription for the pill.

Example 3: Unexpected Noun

Na het bezoek aan de streekdokter moest de boerin nog medicijnen ophalen. Ze ging daarvoor naar een ziekenhuis/ ziekenhuis in de grote stad.

After visiting the local doctor the farmer's wife still had to pick up the medication. For that, she went to the hospital/ hospital in the city.

In the predictive story "apotheek" is presented in a deviating font *and* violates the word-image prediction, whereas in the neutral story, "apotheek" only has a deviating font. The critical comparison to test for the presence of a specific word-image prediction is thus between the predictive and the neutral stories. Both neutral and unexpected nouns presented in a deviant font will show *font-costs*, i.e. both types of words will be read slower when presented in a deviating font than in the standard font. However, if readers use a constraining discourse to predict the word-image, then these costs will be higher for the expected words than for the neutral words. The font-costs for neutral deviant font words reflect only the increase in reading times that is the consequence of simply reading a word in a font that differs from the standard font. In a constraining context, however, a reader can use the context to create a word-image prediction. If the expected word is then presented in a deviant font, the perceived word-image will diverge from the predicted word-image. In this case, the font-costs for the expected deviant font word will reflect the basic 'deviant font'-costs with the added 'unexpected word-image'-costs, and will thus be higher than for the neutral deviant font words. If readers do not make predictions about the specific word-image, based on the constraining context, then the font-costs for the neutral and expected words will be comparable.

To explore the interaction between meaning mismatches and the font manipulation, we also included constraining stories in which the expected word was replaced by a coherent but unexpected noun (see Example 3, Table 5.1). If a reader has a strong prediction for a specific target word (*apotheek*), but reads another, less expected noun ("`ziekenhuis`"), then both the lexical prediction and the word-image prediction will be violated at the same time. This double violation exists when the unexpected noun is written in the standard *and* in the deviant font. Unexpected nouns will be read slower than the expected nouns when presented in standard font, because unexpected nouns violate the prediction at a lexical and visual level. There will also be font-costs for the unexpected noun when it is presented in a deviant font, because of the general slowing that is the consequence of reading words in an unexpected font. If the font-costs for the unexpected word indeed only represent the basic 'deviant font'-costs, the font-costs for the unexpected words should be comparable to the neutral words.

Methods

Participants

57 students of the University of Amsterdam participated for course credit. 1 participant was excluded from the analysis because of technical problems during the experimental session, 7 others were excluded because of low scores on questions about the content of the sentences. The 49 remaining participants (42 females; mean age 20 years, range 18-33) were all native speakers of Dutch.

Materials

The critical materials for this experiment consisted of 180 two-sentence stories. Each story was created to be constraining, so that people predicted one specific word (cloze values: range from 73% to 100%, mean = 86%, sd = 8%). If the story actually contained the predicted word, the noun matched the prediction of the reader at the level of meaning (independent of the actual font that word is presented in). The predictive stories could also contain another noun than the predicted noun, which was coherent but less expected (cloze range 0% to 20%, mean = 1% ,sd = 3%). These unexpected nouns mismatch the prediction at the lexical level and at the level of the visual word-image (independent of the actual font that word is presented in). For each expected target word we also designed a neutral context story. In these neutral stories the critical noun always appeared very early in the first sentence of the story, such that the reader had not formed any specific prediction. These neutral nouns are never inconsistent with a lexical or word-image prediction, since the context does not give rise to any prediction. In all experimental conditions the critical noun could be shown in the same font as the rest of the story (Courier New), or in a deviating font (Schule 1995, resembling handwriting as it is taught in primary school). For the expected words, presented in a predictive context, the deviant font creates an additional word-image mismatch, which is not present in the neutral and unexpected deviant font words.

We constructed 60 filler items that resembled the structure of the critical items, thus also consisting of two sentences. One word in the first sentence of a filler item was always shown in the deviating Schule font. This deviating word was never a noun, or the first or last word. Each filler item was accompanied by a question about that story. The filler items were randomly mixed with the 180 experimental items. Of the experimental items, 60 items contained an expected noun, 60 items contained an unexpected noun and 60 contained a neutral noun. For each of these three types of stimuli the critical noun was either shown in the standard courier font or in the deviating Schule font, thus resulting in 30 items for each of the six conditions (expected standard font, expected deviant font, unexpected standard font, unexpected deviant font, neutral expected font, neutral deviant font). By rotating the conditions in this list, 6 more lists of stimuli were created. Each of the 6 lists contained all 180 experimental stimuli, and 60 identical filler items. Each participant was shown one of these six lists of stimuli, so that one participant saw all the stimuli, but never in more than one condition.

Procedure

We presented the stories in a standard noncumulative moving-window self-paced reading paradigm. Subjects read through each story word by word, with each button press disclosing the next word while replacing all other letters in the story with hyphens. As they pressed their way through a story, subjects could see its overall sentential and formatting layout (including punctuation) as well as the position of the currently visible word therein. Subjects were asked to process each story for comprehension and to adapt their speed to this. Simple yes–no comprehension questions were asked after the filler items, which made up 25% of the stories, to keep the participants focused on the content of each story. A reading session consisted of five trial blocks separated by a short break, and took approximately 60 minutes on average.

Analysis

Participants that answered more than 10% of the questions incorrectly were excluded from the analysis. Based on this criterion 7 participants (mean percentage correct: 66%, sd = 40%) were excluded. The remaining 49 participants answered 94% of the questions correctly (sd = 2.9%). Two items were not shown correctly during the experiment, and were thus excluded from the analysis, which as a result included 178 items.

We analyzed reading times in two regions of each story, namely a baseline-region that consisted of the three words that preceded the critical noun, and an experimental region that included the critical noun and the four words that followed this noun. Before analysis, outlying reading times were eliminated. Reaction times that deviated more than 2 standard deviations from the mean reading time for the subject in that condition *and* from the mean reading time for the item in that condition were removed. As a result 1.5% of the data, evenly distributed across the 8 X 6 cells of the design (ranging from .7% to 2.2%), was excluded from the analysis.

For each of the 8 wordpositions in the critical regions, we computed mean reading times per subject and per item for each of six conditions. For each word position we examined the resulting reading times in a by-subjects and a by-items two-way ANOVA with a factor Predictability (three levels: Expected Noun, Neutral Noun and Unexpected Noun) and a factor Font Consistency (two levels: Standard Font and Deviant Font). Conservative degrees of freedom were employed when violations of sphericity were found(epsilon <1). The Huynh–Feldt correction was used for smaller violations of sphericity (values between 0.75 and 1.00) while the Greenhouse–Geisser correction was employed for more severe violations of sphericity (epsilon <0.75). Significant results that concerned the factor Predictability where followed by planned comparisons to establish which levels of this factor differed.

Results

Figure 5.1 shows the reading times for all words in the baseline and critical region (for the exact reading times see Table 5.2). The bars in Figure 5.2 represent the corresponding font-costs for the critical region only. As expected, it is clearly visible that all deviant font nouns (grey lines at position 0) elicit longer reading times. This effect persists in the words that follow the critical nouns even though these words are all presented in standard font. Significant main effects of Font Consistency at the critical noun and the two words that follow (CW, CW 1 and CW 2, see Table 5.3) corroborate this observation. At the position of the critical word itself, the effect of a deviant font is significantly modulated by the predictability of the critical noun (Table 5.3, significant interaction between Predictability and Font consistency).

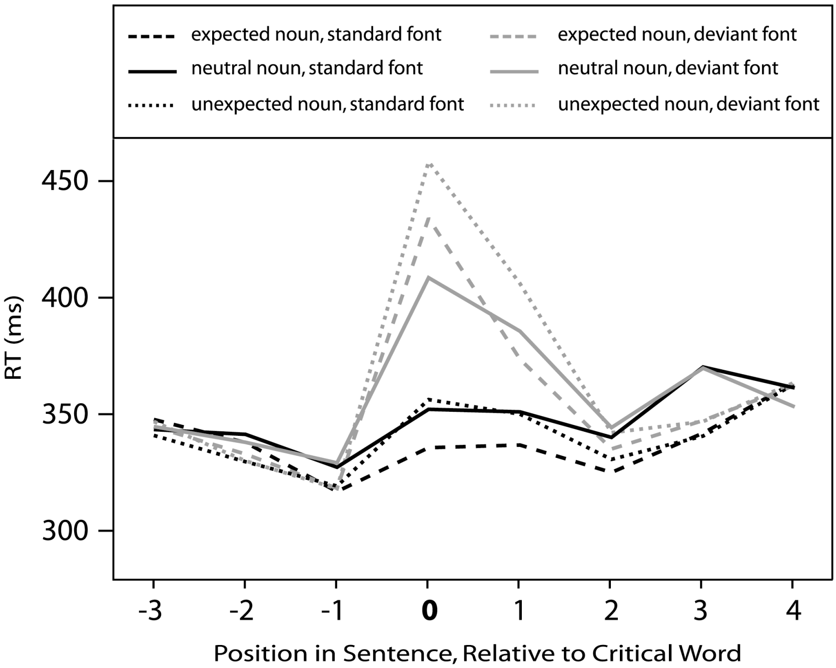


Figure 5.1 *Average reading times for all six experimental conditions, plotted for the 8 consecutive wordpositions that make up the baseline and critical region. The critical noun itself is referred to as position 0.*

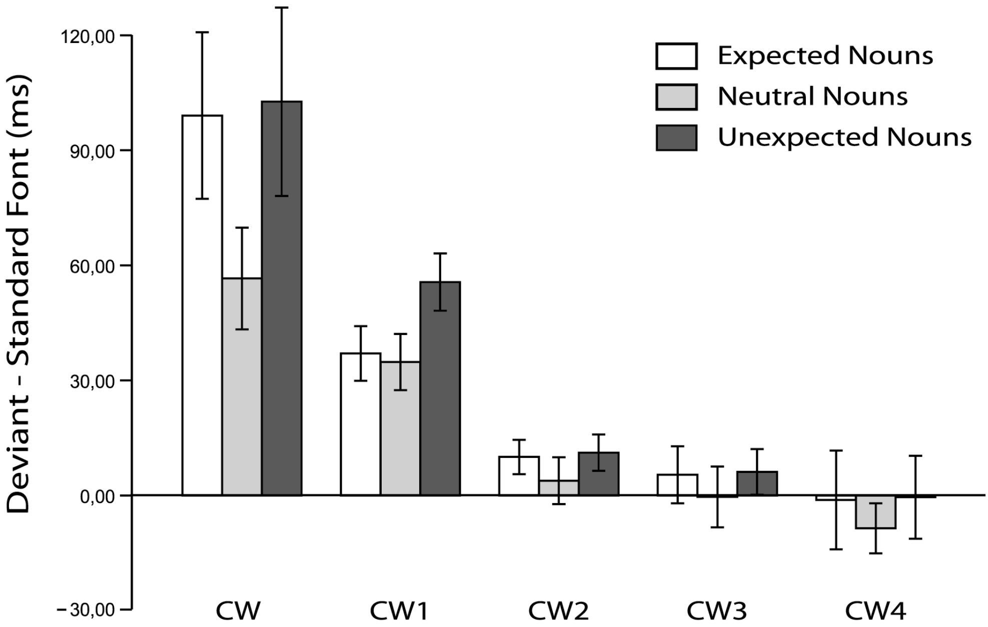


Figure 5.2 *The effect of Font Deviance (the difference in reading-time between nouns in a deviant and a standard font), depicted for the 3 different types of nouns. Error bars indicate the Standard Error.*

Table 5.2. Mean reading times for the words in the Baseline and Critical region.

		CW-3	CW-2	CW-1	**CW**	CW 1	CW 2	CW 3	CW 4
Expected Noun	mean	347	337	316	335	336	324	341	363
Standard Font	*SD*	*149*	*144*	*105*	*166*	*130*	*117*	*184*	*225*
Expected Noun	mean	344	332	317	433	373	334	346	361
Deviant Font	*SD*	*130*	*129*	*112*	*430*	*172*	*122*	*185*	*229*
Neutral Noun	mean	343	341	327	352	350	340	370	361
Standard Font	*SD*	*138*	*142*	*109*	*185*	*151*	*139*	*214*	*173*
Neutral Noun	mean	344	337	329	408	385	344	369	353
Deviant Font	*SD*	*152*	*144*	*131*	*325*	*175*	*153*	*263*	*168*
Unexpected Noun	mean	340	329	318	355	349	330	340	362
Standard Font	*SD*	*137*	*120*	*108*	*200*	*129*	*115*	*143*	*214*
Unexpected Noun	mean	346	329	318	458	406	342	346	363
Deviant Font	*SD*	*171*	*122*	*109*	*451*	*189*	*121*	*151*	*229*

Table 5.3. Overview of the ANOVAs for the critical region and the baseline region. Reported here are the results of a by subjects analysis (F_1, ε_1 and p_1) and a by items analysis (F_2, ε_2 and p_2).

		df	**CW –3**	**CW -2**	**CW -1**	**CW**
Predictability of Noun	F_1	*2,96*	.08	1.59	4.83	3.44
	F_2	*2,354*	.18	1.23	9.15	4.97
	ε_1		.82	-	.82	.77
	ε_2		-	.88	.95	.97
	p_1		.89	.20	.01	.05
	p_2		.83	.29	<.001	<.01
Font Consistency	F_1	*1,48*	.07	2.55	01	25.75
	F_2	*1,177*	.04	1.16	.02	71.23
	p_1		.79	.12	.95	<.001
	p_2		.85	.28	.89	<.001
Predictability * Font	F_1	*2,96*	.55	.30	.08	3.39
	F_2	*2,354*	.82	.27	.01	3.19
	ε_1		-	-	-	-
	ε_2		.94	-	.97	.95
	p_1		.58	.75	.93	.04
	p_2		.44	.76	.99	.05
		df	**CW 1**	**CW 2**	**CW 3**	**CW 4**
Predictability of Noun	F_1	*2,96*	9.89	4.50	11.14	.47
	F_2	*2,354*	11.60	3.97	10.00	.76
	ε_1		-	-	-	-
	ε_2		-	.87	.70	.62
	p_1		<.001	.01	<.001	.63
	p_2		<.001	.03	<.001	.41
Font Consistency	F_1	*1,48*	74.84	10.80	.87	.30
	F_2	*1,177*	179.6	11.49	1.53	.01
	p_1		<.001	<.01	.35	.58
	p_2		<.001	.001	.21	.92
Predictability * Font	F_1	*2,96*	2.98	.51	24	.19
	F_2	*2,354*	3.93	.84	.18	.64
	ε_1[a]		.92	-	-	.89
	ε_2[a]		-	.98	.88	.86
	p_1		.06	.60	.79	.83
	p_2		.02	.43	.81	.50

Table 5.4. Planned comparisons for the interaction between Predictability and Font Consistency				
		df	**CW**	**CW 1**
Predictability (Expected vs. Neutral)	F_1	*1,48*	4.42	.08
* Font Consistency	F_2	*1,177*	4.25	.01
	p_1		.04	.79
	p_2		.04	.98
Predictability (Unexpected vs. Neutral)	F_1	*1,48*	5.62	5.56
* Font Consistency	F_2	*1,177*	5.67	6.25
	p_1		.02	.02
	p_2		.02	.01
Predictability (Expected vs. Unexpected)	F_1	*1,48*	.04	2.92
* Font Consistency	F_2	*1,177*	.33	5.46
	p_1		.85	.09
	p_2		.56	.02

Planned comparisons (Table 5.4) show that the increase in reading times that accompanies a font mismatch is much larger for the expected than for the neutral nouns. Surprisingly, the increase in reading times for a deviating noun is also larger for unexpected than for neutral words.

The word that follows the critical noun (CW 1, which is always presented in standard font) still shows an increase in reading times as a result of the deviating font (Table 5.3, main effect of font consistency). At this position, the increase in reading times is roughly equal for the expected and the neutral noun, whereas the unexpected nouns show a larger increase than the nouns in the other two conditions (Table 5.4). Figure 5.2 clearly illustrates this difference between the critical word (CW) and the word that follows (CW 1) with regard to the interaction between noun predictability and font consistency.

The results in Table 5.3 also show a main effect for the noun predictability, which lasts from 1 word before the critical noun until 3 words after the critical noun. Planned comparisons (Table 5.5) show that the effect at CW -1 follows from a difference between neutral stories on the one hand, and predictive stories (with expected and unexpected words) on the other hand, with neutral trials resulting in slower reading times. These slower reading times for neutral stories are also visible at CW 2 and CW 3. At CW (the critical

Table 5.5. Planned comparisons for the main effect of Predictability							
		df	**CW -1**	**CW**	**CW 1**	**CW 2**	**CW 3**
Expected vs. Neutral	F_1	*1,48*	6.04	2.58	6.36	7.73	12.73
	F_2	*1,177*	14.14	.06	7.61	6.18	12.78
	p_1		.18	.61	.02	<.01	.001
	p_2		<.001	.82	<.01	.01	<.001
Unexpected vs. Neutral	F_1	*1,48*	5.51	3.87	3.72	1.86	19.05
	F_2	*1,177*	9.55	6.40	3.92	.43	9.32
	p_1		.02	.06	.06	.18	<.001
	p_2		<.01	.01	.05	.51	<.01
Expected vs. Unexpected	F_1	*1,48*	.08	6.49	19.33	3.44	.024
	F_2	*1,177*	.87	7.03	24.90	6.58	1.37
	p_1		.78	.01	<.001	.07	.88
	p_2		.35	<.01	<.001	.01	.24

noun) and CW 1, expected words are read at a faster pace than unexpected and neutral words.

Discussion

Our results show that for expected words the font-costs (i.e. the difference in reading times between words in a standard and a deviating font) are higher than for neutral words. This pattern of results suggests that readers make very specific predictions about the visual manifestation of highly predictable words. It thus seems that when readers make specific linguistic predictions, they also pre-activate the orthographic information associated with the predicted word. Perhaps even more interestingly, these results also suggest that readers are able to quickly integrate the activated orthographic information with the relevant contextual information about the standard font, to reach a specific visual prediction.

The fact that people are able to use contextual information to anticipate upcoming words not only at the lemma level (Otten et al., 2007; Otten & Van Berkum, in press; Van Berkum et al., 2005; Wicha, Bates et al., 2003; Wicha, Moreno et al., 2003; Wicha et al., 2004) but also at the level of the lexeme (DeLong et al., 2005, this experiment) is consistent with an interactive connectionist model of language perception as postulated by McClelland and Rummelhart (McClelland & Rumelhart, 1981). In this model activation of a word at the lexical level will automatically activate the related letters, and this activation consequently extends to the visual features that make up these letters. Although in the original model activation at the word level comes from real visual (or auditory) input, this model also acknowledges the role of higher level, top down input.

The difference in font-costs at the expected and neutral nouns thus suggests that linguistic predictions could extend to the visual features of the expected word. Our results, however, show one surprising finding. Unexpected words show almost identical font-costs as the expected words. However, according to our initial hypotheses, the font-costs for the unexpected words should have been comparable to the neutral words. This unexpected finding raises the possibility that the font-costs for the expected and unexpected words both have the same origin, which does not lie in visual linguistic predictions. In constraining stories (which contained the expected and unexpected words) the target noun was presented in the second sentence, whereas in the neutral stories the target word was always presented very early in the first sentence. The neutral stories thus differ from the predictive stories in the position of the critical noun. This leads to an overall difference in reading-times between the neutral and predictive stories: in the critical region, the predictive stories were read slower than neutral stories.

But can these overall differences in reading speed due to critical word position explain the larger font-costs for expected and unexpected words? Reading speed is higher for the constraining context than for the neutral context. One could thus reason that spillover-effects are more likely for deviant-font words that are presented in a constraining context: readers are pressing the button at a relatively fast rate, which increases the chances of a delayed response to the deviant noun (i.e. an increase in reading times that only becomes clear at one of the words that follows the critical word). If anything, however, a ‘fixed fast rate’ confound should lead to *smaller* font-costs for critical nouns that are presented later in the story, not *larger* font-costs.

Furthermore, the words that follow the critical nouns show that the processing of deviant expected and deviant unexpected words is not identical. The font-costs for unexpected words remain higher than for neutral words also after the critical word has been presented, whereas for the expected word this effect equals out immediately after reading the critical word. If the position of the critical word within the story is the only determiner of the difference between predictive and neutral stories, then the effects of expected and unexpected words should be the same, because these two words have exactly the same position in exactly the same story.

Taken together, it does not seem likely that the difference in word position is in itself responsible for the differences in font-costs between expected and unexpected nouns on the one hand and neutral nouns on the other hand. This would suggest that the difference in font-costs between neutral and expected words can indeed be attributed to the presence of a visual word-image prediction in the latter. But how then, if not by contextual factors, can we explain the additional processing costs that are visible for the unexpected deviant font words? Perhaps this answer lies in the extra emphasis that is indicated by a deviant font (McAteer, 1992). In contrast to the neutral and expected nouns, unexpected nouns were not consistent with expectations. Sanford and colleagues (2006) observed that when a word was foregrounding by a deviating typeface inconsistencies were noted more often than in sentences that did not contain such visual emphasis. It is thus possible that the extra emphasis on the mismatch, due to the conspicuous font, has made the meaning mismatch more striking, which in turn has led to the increased times for the unexpected nouns in comparison with the neutral nouns.

Conclusion

The results are consistent with the idea that people make predictions about the visual form of the expected word. The current experiment, however, does not allow us to conclude with certainty that the effects observed for neutral and expected words can be traced back exclusively to predictive processes. We are thus currently replicating this experiment with a neutral condition that resembles the predictive condition in length and critical word position. This design will allow us to determine whether the differences between the font effect for expected and neutral words can be attributed to the difference in critical word position, or whether this difference is due to the visual component of the linguistic prediction.

Chapter 6

What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse dependent N400 effect.

A discourse context provides a reader with a great deal of information that can provide constraints for further language processing, at several different levels. In this experiment we used event-related potentials (ERPs) to explore whether discourse-generated contextual constraints are based on the precise message of the discourse or, more 'loosely', on the scenario suggested by one or more content words in the text. Participants read constraining stories whose precise message rendered a particular word highly predictable ("The manager thought that the board of directors should assemble to discuss the issue. He planned a…meeting") as well as non-constraining control stories that were only biasing in virtue of the scenario suggested by some of the words ("The manager thought that the board of directors need not assemble to discuss the issue. He planned a…"). Coherent words that were inconsistent with the message-level expectation raised in a constraining discourse, for example "session" instead of "meeting", elicited a classic centroparietal N400 effect. However, when the same words were only inconsistent with the scenario loosely suggested by earlier words in the text, they elicited a different negativity around 400 ms, with a more anterior, left-lateralized maximum. The fact that the discourse-dependent N400 effect cannot be reduced to scenario-mediated priming reveals that it reflects the rapid use of precise message-level constraints in comprehension. At the same time, the left-lateralized negativity in non-constraining stories suggests that, at least in the absence of strong message-level constraints, scenario-mediated priming does also rapidly affect comprehension.

Introduction

Over the last two decades, event related potentials (ERPs) have considerably advanced our understanding of the cognitive processes underlying language comprehension. The N400, an ERP component that is particularly sensitive to semantic processing, has played a major role in this. The N400 has been used to study the comprehension of written, spoken and signed language (see Kutas, Van Petten and Kluender (2006) for review), as well as the breakdown of language or of related cognitive functions in for example aphasia (Hagoort, Brown, & Swaab, 1996; Swaab, Brown, & Hagoort, 1997), schizophrenia (Sitnikova, Salisbury, Kuperberg, & Holcomb, 2002), and Alzheimer's disease (Ford et al., 1996). The N400 component was discovered by Kutas and Hillyard (Kutas & Hillyard, 1980), who found that a sentence-final word that was incongruent with the preceding linguistic context evoked a larger negativity than a congruent ending. Later experiments showed that the N400 was more than a semantic anomaly detector. In particular, relative to highly expected words, semantically coherent but unexpected words also evoked a larger N400, albeit not as large as anomalies (Kutas & Hillyard, 1984). Based on these and other findings, Kutas and colleagues (2006) have recently argued that the amplitude of the N400 reflects the degree to which "context aids in the interpretation of a potentially meaningful stimulus". In line with this, language researchers generally agree that the word-elicited N400 indexes how well the meaning of a word fits the constraints set by the context (Brown, van Berkum, & Hagoort, 2000; Chwilla, Brown, & Hagoort, 1995; Coulson & Federmeier, in press; Friederici, 1995; Hagoort et al., 2004; Osterhout & Holcomb, 1995; Van Berkum, Brown et al., 2003).

But what are those constraints? Following up on the pioneering N400 research of St. George, Mannes and Hoffman (1994), ERP experiments with text-level manipulations suggest that the N400 is not only sensitive to constraints provided by a single word prime or an unfolding single sentence, but is also highly sensitive to what the wider discourse is about (Brown et al., 2000; Federmeier & Kutas, 1999b; Nieuwland & Van Berkum, 2006b; St George et al., 1997; Van Berkum, Brown et al., 2003; Van Berkum et al., 2005). In stories such as (1), for example, the discourse-supported word "movie" elicited a much smaller N400 than the discourse-inappropriate word "book" (Van Berkum et al., 1999; 2003). Such discourse-dependent N400 effects have been interpreted as evidence that the language comprehension system immediately evaluates the current word against a precise message-level representation of what has been said so far.

*(1) David and Sabrina had been thinking about what they were going to do that evening, and eventually decided to go to the cinema. They hoped the **movie/ book** would be fun, but it turned out to be quite boring.*

However, an important alternative interpretation has as yet not been ruled out. Models of text comprehension and memory suggest that in addition to contributing to a precise message-level representation of the discourse, the words in a text can also provide semantic constraints in a much less precise way, via the activation of related information stored in long term memory (Kintsch, 1988; McKoon & Ratcliff, 1992; Sanford, 1990). In the above story, for instance, the mere presence of the word "cinema" could activate a going-to-the-movies scenario, which includes seeing a film. The attenuation of the N400 in (1) might thus also come about because the word "movie" is relevant to the scenario suggested by one or several words in the preceding text.

The difference between a message-level and scenario-mediated account for discourse-dependent N400 effects may not be obvious at first. After all, our understanding of what has been said so far, the precise message, will in part depend on our default knowledge about what things tend to go together in the world (as captured in scenarios, scripts, etcetera). However, consider what would happen if we change the precise message of the discourse, as in (2). Although the going-to-the-movies-scenario is still implied by the words in the context, the actual message of the story does not really support either "movie" or "book".

*(2) David and Sabrina had been thinking about what they were going to do that evening, and eventually decided not to go to the cinema. They hoped the **movie/ book** would be out on dvd soon, and went to the pub.*

In the ERP study reported below, we try to disentangle the effects of message- and scenario-level constraints by exploiting the possibility to change the message of a story while leaving the scenario-relevant words in that story intact. The goal is to examine whether the discourse-dependent N400 effect hinges on constraints provided by the exact message-level representation of the prior text (as assumed in Van Berkum et al., 1999; 2003), or whether it can perhaps be accounted for – entirely or in part – by scenario-mediated lexical priming.

Behavioral experiments have shown that this scenario-mediated priming does play a role in comprehension, and can do so even when the scenario-generated information is irrelevant to, or at odds with, the actual message (Duffy et al., 1989; Garrod & Terras, 2000; O'Seaghdha, 1997). Garrod and Terras (2000), for example, showed that the word "pen" is initially just as effectively integrated when presented in a sentence following "The teacher wrote a letter" as it is after the sentence "The teacher wrote the exercise on the blackboard". Only in regression path analysis and second pass reading times a significant difference was observed between the appropriate and inappropriate contextual message. This indicates that participants did not at first notice the message-level incongruence of "pen", presumably because "pen" is strongly associated with the default scenario activated by the verb ("to write"). Results like these imply that contextually activated scenarios can prime scenario-related concepts, leading to facilitated processing of these concepts and the words that denote them. Furthermore, they show that scenario-related facilitation is in some cases initially stronger than the support provided by the actual message of the discourse

To account for such results, models of text comprehension (Kintsch, 1988; 1998; Sanford & Garrod, 1981; 1998) usually include an initial stage in which all potentially relevant information is retrieved in a way that is highly sensitive to the set of words and concepts in the text, regardless of the precise message of the text. For example, the scenario-mapping and focus theory of Sanford and Garrod (1981; 1998) proposes that the word currently read or heard (e.g. "movie") is initially evaluated in terms of a 'quick & dirty' match to prior words in the text and the scenario suggested by those words, before it is mapped more carefully onto the precise message conveyed by that text. Kintsch (1988; 1998) has made similar proposals. The dominant model for the relatively shallow initial retrieval process involved in these accounts is the resonance model (Cook et al., 1998; Myers & O' Brien, 1998; Myers et al., 1994). According to this model, individual concepts from the linguistic input send out a signal to long term memory. Concepts in memory then resonate as a function of their relatedness to the input, based on the overlap between the semantic and contextual features of the concepts involved. Eventually, those concepts that have the highest level of activation enter working memory. The resonance process is assumed to be fast-acting and autonomous (or "dumb" (Myers & O' Brien, 1998)). As a result, activated information in long-term memory can be irrelevant to the specific meaning conveyed by the complete discourse, and may in fact even be incongruent with that message.

In a recent ERP experiment Hoeks, Stowe and Doedens (2004) directly compared the effects of message-level constraint and scenario fit on the N400 by contrasting the ERP evoked by a scenario-related verb in a highly constraining sentence like "The javelin was by the athletes thrown" (approximate translation from Dutch) and in a much less constraining sentence like "The javelin has the athletes thrown". In spite of the difference in message, both sentences induce the same athletics-scenario due to the presence of related words ("javelin", "athletes") in the context. Surprisingly, the N400 to scenario-related verbs was independent of whether the actual message of the context supported or prohibited this verb. This result suggests that the N400 is sensitive to the effects of scenario fit, and that these scenario effects can under some conditions fully determine the amplitude of this component, at least in the early stages of comprehension.

If the N400 is indeed as sensitive to scenario fit as the results of Hoeks et al. (2004) suggest, this has important consequences for the functional interpretation of the N400, but also for the interpretation of experiments in which the N400 is used as a measure of message-level semantic integration or contextual facilitation. More specifically, if the N400 elicited by words in text is (also) dependent on scenario-based constraints, then differential N400 effects cannot be automatically taken as an indication that the words at hand are differentially integrated with a message-level representation of the text. As already discussed, this ambiguity affects the interpretation of discourse-dependent N400 effects, and we focus our experiment on this type of effect. However, consistent with the account given for discourse-dependent N400 effects in Van Berkum et al. (1999; 2003), the ambiguity also affects the interpretation of sentence-dependent N400 effects, such as for those elicited by, e.g., "He mailed the letter without a stamp" or "She locked the valuables in the safe" (Kutas & Hillyard, 1984). As with longer texts, it is not clear whether attenuated N400 effects in sentences like these reflect support from the precise message, or from the general scenario suggested by, say, "mailed" and "letter".

In our experiment, we examined whether the discourse-induced N400 effect results from a mismatch between the incoming word and the specific constraints set by the precise message of the discourse, or whether this N400 effect reflects scenario-based fit. We roughly followed the logic embodied in examples (1) and (2), avoiding semantic anomalies, since semantically congruent words that differ in sentence- or discourse-based predictability can also elicit a large differential N400 effect (DeLong et al., 2005; Hagoort & Brown, 1994; Kutas & Hillyard, 1984; Van Berkum et al., 2005).

Table 6.1 shows the two types of stories used in the experiment. In the so-called message- and scenario-biased (or MS-biased) story, the text leading up to the critical word has a highly constraining message, such that when people are asked to complete the story in a cloze test they predominantly converge on "meeting", and very rarely come up with the equally congruent word "session". This predictability critically hinges on the prior discourse

Table 6.1. An approximate English translation of one of the stories shown to the participants, in all four conditions. The critical nouns are printed in boldface.

1. Message & Scenario biased Discourse	
Bias-Consistent Noun	*Bias-Inconsistent Noun*
The manager thought that the board of directors should assemble to discuss the issue. He planned a **meeting** where the staff members involved would be present as well.	The manager thought that the board of directors should assemble to discuss the issue. He planned a **session** where the staff members involved would be present as well.
2. Scenario biased Discourse	
Bias-Consistent Noun	*Bias-Inconsistent Noun*
The manager thought that the board of directors need not assemble to discuss the issue. He planned a **meeting** where the staff members involved would be present and nobody else.	The manager thought that the board of directors need not assemble to discuss the issue. He planned a **session** where the staff members involved would be present and nobody else.

supplied by the preceding sentence (Van Berkum et al., 2005). Next, because the discourse-dependent attenuation of the N400 for a word like “meeting” might also be caused by its support from the meeting-scenario suggested by the individual words in the MS-biased story quoted in Table 6.1 like “manager”, “board”, “directors”, “assemble” and “discuss”, we created so-called scenario-biased (or S-biased) control stories. In these stories, we changed the precise message of the first sentence such that neither of the two critical words (nor any other word) was particularly predictable. At the same time, we made sure to preserve the scenario-relevant content words of their MS-biased counterparts (e.g., “manager”, “board”, “directors”, “assemble” and “discuss”), so that these S-biased control stories would remind people of the same scenario.

If the discourse-dependent N400 effect is solely a result of the mismatch between the incoming word and a precise message-level representation of the discourse, then the N400 effect elicited by MS-biased stories should not be elicited in S-biased stories. Alternatively, if the discourse-dependent N400 effect hinges solely on fit to the scenario induced by a set of content words, the N400 effect elicited by MS-biased stories should be comparable to the one elicited in S-biased stories. Of course, to the extent that the precise message conveyed by the discourse and the scenario induced by a set of words in that discourse both affect the N400, both types of stories should evoke an N400effect. However, in this case the N400 effect for S-biased stories, where only the scenario supports the bias-consistent word, should be smaller than the N400 effect observed in MS-biased stories, in which the bias-consistent word is supported by both the scenario and the message of the story.

Methods

Participants

36 right-handed native speakers of Dutch (27 female participants, mean age 21, range 18-26 years) took part in the experiment, as part of a course requirement. None had any neurological impairment, had experienced neurological trauma, or used neuroleptics. None of the participants had participated in the pre-tests conducted during the material construction phase.

Materials

The stimuli in this experiment were 160 mini-stories of two sentences, with the first sentence establishing the discourse context, followed by the local carrier sentence containing the critical word. For each item a *message- and scenario-biased* (or *MS-biased*) discourse as well as a *scenario-biased* (or *S-biased*) control discourse was created, both containing the same scenario-relevant words (see Table 6.1 and the Appendix 3 for examples of the stimulus materials). All stories were designed to suggest a specific message-predictable word right after the indefinite article in the target sentence (the second sentence) in the MS-biased condition, but not in the S-biased condition. In addition, S-biased stories suggested a scenario that favoured the bias-consistent critical word over the bias-inconsistent one. Across the MS- and S-biased conditions, we refer to the message/scenario-predictable words as *bias-consistent* words, and to the coherent but essentially unpredictable control word as *bias-inconsistent* words.

The level of *message-level* constraint for each of the two conditions of each critical story was determined in a pencil-and-paper cloze test, prior to the EEG-experiment. In this pretest, we showed the MS-biased and S-biased mini-stories up to (and thus not including) the critical word to 66 participants, and asked them to complete the story with the first thing that came to mind. The two versions of the items were divided over separate lists, so that a participant never saw an item in more than one context condition. For each item the cloze value (the proportion of participants who filled in the critical word) for the bias-consistent word and the bias-inconsistent control word was calculated, in both the MS-biased and the S-biased condition. Only those items were selected in which the cloze value for the *bias-consistent* word was higher than .50 in the MS-biased version of the story and lower than .30 in the S-biased version, with a difference between these two values of at least .25. In the resulting itemset, bias-consistent words had a mean cloze value of .77 (sd = .13) across all MS-biased stories and .18 (sd = .15) across all S-biased stories. The mean cloze value for corresponding bias-inconsistent control words was .04 in both the MS-biased stories (sd = .06) and the S-biased stories (sd = .07).

Bias-consistent and bias-inconsistent words were matched on average length and frequency: The mean length of the discourse-predictable and control word was respectively 6.2 (sd = 2.2) characters and 6.8 characters (sd = 2.5), and the mean frequency for discourse-predictable and control words was respectively 30.5 (sd = 52.6) and 30.3 (sd = 65.6) per 1 million, as stated in the Celex database. All words preceding the critical noun and the two words that followed the critical noun in the second sentence were identical for each of the four conditions, and as such did not differ in length or frequency. After the two identical words that followed the critical noun, the remainder of the story sometimes varied between MS- and S-biased stories, to avoid coherence breaks at the message level.

The 160 items (40 for each of the four conditions shown in Table 6.1) were pseudo-randomly mixed with 80 filler items addressing an unrelated issue (Van Berkum et al., 2007). By rotating the conditions in this list, three more lists of stimuli were created. Each of the four lists contained all 160 experimental stimuli, 80 stories in the MS-biased discourse version and 80 with a S-biased discourse. Half of the 80 MS-biased items and half of the 80 S-biased items contained the bias-consistent word, while the remaining 40 ended with a bias-inconsistent word. Each participant was shown one of these four lists of stimuli, so that one participant saw all stories, but never in more than one condition.

Procedure, EEG recording and Analysis

Each participant saw 240 stories, 160 of which were critical for the current issue. Participants were asked to read for comprehension and were not required to perform any other task. The electroencephalogram (EEG) was recorded with 30 electrodes (FP1, FP2, F9, F7, F3, Fz, F4, F8, F10, FT9, FC5, FC2, FC6, FC1, FT10, T7, C3, Cz, C4, T8, CP5, CP1, Cp2, Cp6, P7, P3, Pz, P4, P8 and Oz) mounted in an elastic cap, each referenced to the left mastoid. Blinks and vertical eye-movements were registered by placing an electrode under the left eye, initially referenced to the left mastoid, but later rereferenced to an electrode above the left eye (Fp1). Electrode impedance was kept below 5 kOhms during the experiment. The EEG was amplified, band-pass filtered at 0.03 Hz-100 Hz and sampled with a frequency of 500 Hz.

During the comprehension task the participants sat in a comfortable chair in a normally lit room. The stimuli were presented in black 36 point courier new font on a white background on a fast TFT display (Iiyama TXA 3834 MT) positioned approximately 80 cm away from the participant. Before each trial, a fixation cross was shown in the center of the screen for 2.5 s. Participants were asked to avoid blinks and eye movements when the words were presented on screen, and were encouraged to blink when the fixation cross was shown. To signal the start of each trial to the participant a beep sounded 1 s before the onset of the first word.

The stories were then presented word for word. To make the visual presentation more natural, words were presented using a Variable Serial Visual Presentation (VSVP) procedure, in which the presentation time of each non-critical word varied with its length. Non-critical word duration consisted of a standard offset of 187 ms plus and additional 27 ms per letter (with an upper bound of 10 letters for each word). In the present experiment, durations varied from 214 ms for a one-letter word to 450 ms for words consisting of ten or more letters. The interword interval was always 106 ms. The presentation of clause-final words preceding a comma was prolonged with an additional 200 ms. In addition, presentation time for sentence-final words was extended with an extra 293 ms, followed by a 1 s pause until the next sentence began. These various parameters were based on natural reading times (Haberlandt & Graesser, 1985; Legge et al., 1997), a subjective assessment of the naturalness of the resulting presentation, and technical constraints imposed by the video refresh rate. Note that to the extent that critical words, or words close to the critical word, differ in average length, the above procedure will induce unintended shifts in the ERP waveforms (particularly the exogenous deflections associated with visual word onset and offset). To avoid spurious ERP effects due to these shifts, words whose exogenous components fall in the critical EEG epoch (or baseline) should therefore be equated across condition on their presentation time. In the present study, the critical noun and the three words that followed were presented with a fixed duration of 346 ms, based on the average critical word length across all 240 stories in the experiment (6 characters). Participants did not notice the alternation between completely variable and semi-fixed word duration presentation within a single story.

The data were re-referenced off-line to the average of right and left mastoids. Blinks and eye movements were removed from the data using a procedure based on Independent Component Analysis (ICA) as described by Jung and colleagues. (Jung, Makeig, Humphries et al., 2000; Jung, Makeig, Westerfield et al., 2000). After that the data were segmented, timelocked to the onset of the critical word, from 500 ms before critical word onset until 1200 ms after critical word onset. Segments in which the signal exceeded (-)100 µV and those containing linear drift that was not related to the onset of the critical word and exceeded (-) 40 µV were eliminated off-line. Due to the presence of artifacts, for each of the four conditions about 7% of the trials were deleted. The remaining trials were normalized by subtracting the mean amplitude in a 200 ms pre-stimulus interval. For each participant the trials were then averaged for each of the four conditions, timelocked to the onset of the word.

Analyses of Variance (ANOVAs) were conducted, using mean amplitude values computed for each participant and condition in the 300-500 ms N400 window for each electrode. Univariate F tests with more than one degree of freedom in the numerator were adjusted by means of the Greenhouse-Geisser or Huynh–Feldt correction where appropriate. Uncorrected degrees of freedom and corrected *P*-values are reported in the Results. The results were evaluated in an overall ANOVA with the factors Context (MS-biased and S-biased) and Consistency (bias-consistent word and bias-inconsistent word). To evaluate differences in scalp distribution of the observed effect, an ANOVA with the factors and Context, Consistency and Electrode was conducted after the data were scaled by vector-length (McCarthy & Wood, 1985). Each participant's amplitude value was divided by the square root of the sum of the squared amplitudes over all electrodes for each condition of that participant (i.e. the vector-length for each condition). The nature of significant interactions with Electrode was assessed in a quadrant ANOVA crossing Context and Consistency with a Hemisphere (left/right) and Anteriority (anterior/posterior) factor. Four quadrants were defined: (1) left-anterior, comprising FP1, F3, F7, F9, FC1, FC5 and FT9; (2) right-anterior, comprising FP2, F4, F8, F10, FC2, FC6 and FT10; (3) left-posterior, comprising C3, T7, CP1, CP5, P3 and P7; (4) right-posterior, comprising C4, T8, CP2, CP6, P4 and P8.

Results

Figure 6.1 shows for each electrode the ERPs evoked by the bias-consistent (solid line) and bias-inconsistent (dotted line) word in the MS-biased condition. Figure 6.2 shows the ERPs for the same words in the S-biased condition. As can be seen in Figure 6.1, semantically coherent words that are not what would be expected given the highly constraining message of the discourse so far elicit a clear N400 effect, relative to coherent expected words. However, Figure 6.2 reveals that the same bias-inconsistent words also evoke a negativity in the N400 latency range when the discourse is only biasing through a related scenario, whilst the actual message is not constraining. Figure 6.3 shows the difference waves (ERP for inconsistent words – ERP for consistent words) for the MS-biased (solid line) and the S-biased discourse (dotted line), together with the scalp topographies of the bias inconsistency effect between 300 and 500 ms for the two context conditions. What can be

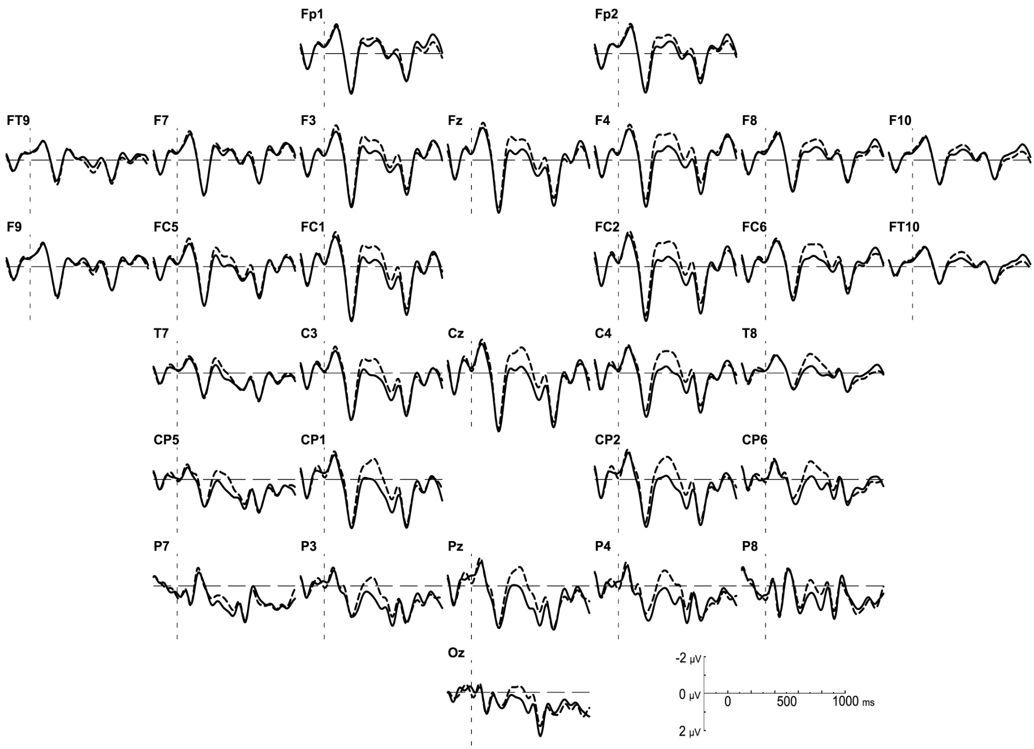


Figure 6.1 *Grand average ERPs elicited by the critical nouns in a MS biased discourse. The solid waveforms, represent the response to bias-consistent target nouns; the dotted waveforms represent responses to bias-inconsistent target nouns.*

seen here is that, whereas the bias inconsistency effect in MS-biased stories has a typical N400 distribution (a centro-parietal and slightly right-lateralized maximum), the effect in S-biased stories has a very different scalp distribution, with a more anterior, left-central maximum.

The overall Context x Consistency ANOVA on mean amplitudes in the 300-500 ms latency range reveals a significant main effect of Consistency ($F(1,35) = 11,7$; $p = .002$), with inconsistent words evoking a larger negativity than consistent words, and a main effect of Context ($F(1,35) = 5.4$, $p = .03$), which is a result of a larger negativity for words presented in a MS-biased discourse than a S-biased discourse, regardless of the nature of the word (consistent or inconsistent).

In spite of the difference in overall effect size visible in Figure 6.3, the Consistency x Context interaction in the overall ANOVA was not significant

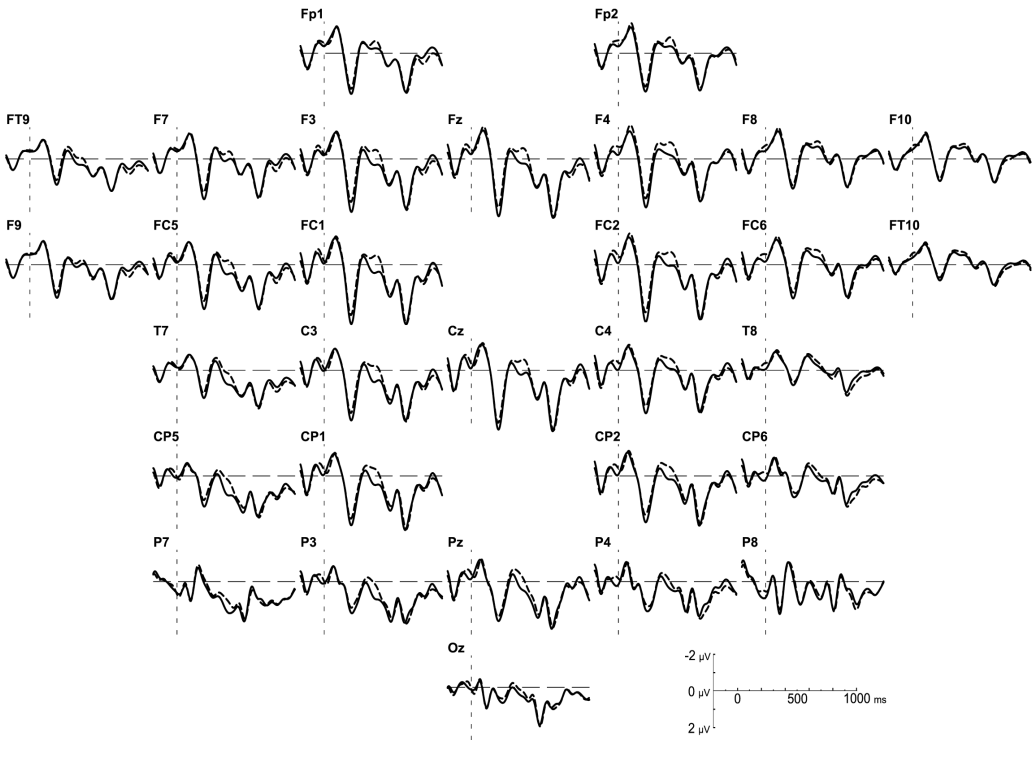


Figure 6.2 *Grand average ERPs elicited by the critical nouns in an S biased discourse. The solid waveforms, represent the response to bias-consistent target nouns; the dotted waveforms represent responses to bias-inconsistent target nouns.*

(F(1,35) = 1.6, p = .22). However, the difference in effect distribution was statistically corroborated, after scaling the data, by a reliable Consistency x Context x Electrode interaction (F(29,1015) = 2.4, p = .04) in the overall ANOVA, as well as a reliable Consistency x Context x Hemisphere interaction (F(1,35) = 8.0, p = .008) in the quadrant ANOVA. Follow-up tests on the unscaled data indicated that whereas the two consistency effects did not reliably differ over the left hemisphere (F(1,35) = 0.4, p = .54), they did differ

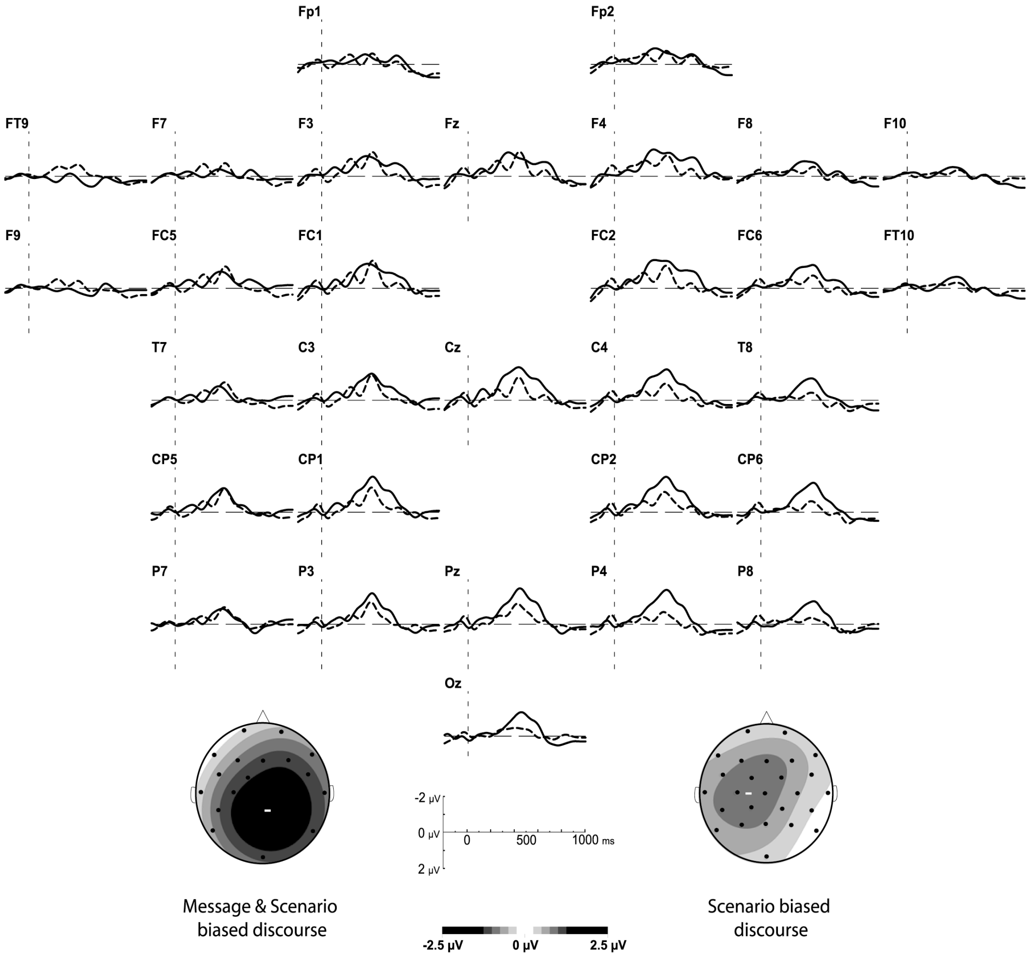


Figure 6.3 *Difference waves for Message & Scenario and Scenario biased Discourses. The difference waves depicted here result from subtracting the ERP elicited by bias-consistent targets from the ERP evoked by bias-inconsistent targets. Solid waveforms represent the difference wave for the MS biased discourse; dotted waveforms represent the difference wave for the Scenario biased discourse. Also shown are the scalp distributions for the effect of Bias-Consistency in each of the two conditions.*

over the right hemisphere ($F(1,35) = 4.6$, $p = .04$). Further simple main effects analyses revealed a reliable negativity over the left hemisphere in the MS-biased condition ($F(1,35) = 5.9$, $p = .020$) as well as the S-biased condition ($F(1,35) = 5.6$, $p = .023$), but a reliable negativity over the right hemisphere in the MS-biased condition only ($F(1,35) = 10.7$, $p = .002$), and not in the S-biased condition ($F(1,35) = 1.8$, $p = .19$).

Discussion

We examined whether word-elicited N400 effects within a coherent discourse are solely determined by constraints set by the precise message of the discourse or by the contextually activated scenario. In an MS-biased discourse, where both the message of the discourse *and* the invoked scenario supported a specific noun, critical words that were inconsistent with these biases indeed elicited a larger N400 than the completely consistent critical word. However, in the S-biased discourse, where the consistent word was only supported by the scenario suggested by one or more content words in the text, bias-inconsistent words also elicited an increased negativity around 400 ms, now with a more anterior, left-lateralized maximum. The fact that the discourse-dependent N400 effect cannot be reduced to scenario-mediated priming reveals that it reflects the rapid use of precise message-level constraints in comprehension. At the same time, the presence of residual left-dominant negativity in S-biased stories suggests that, at least in the absence of strong message-level constraints, scenario-mediated priming does also rapidly affect comprehension. We discuss both implications in turn, and then discuss their possible relation.

A message-based discourse-level N400 effect

In MS-biased stories designed to generate specific word expectations, coherent words that did not meet those expectations elicited a classic N400 effect, peaking at 400 ms, and with a centroparietal, slightly right-lateralized maximum. As suggested by Figure 6.3, this effect cannot be explained by scenario-mediated (or other word-based 'non-message') priming, for when the message-level constraint was removed while all potentially scenario-relevant words were still present (as in S-biased control stories), the same critical words no longer elicited a classic N400 effect. Although critical words in S-biased control stories also gave rise to a negativity, its very different scalp distribution implies that the N400 effect observed in MS-biased stories at least partly reflects other processes than those induced by scenario consistency. Since the two types of stories only systematically differ in the precise message and its degree of predictability, the N400 effect observed in MS-biased stories must reflect, at least in part, the impact of message-level constraints.

This result supports earlier claims (Brown et al., 2000; Nieuwland & Van Berkum, 2006b; St George et al., 1994; Van Berkum, Brown et al., 2003) that comprehenders immediately relate the meaning of every incoming word to a precise message-level representation of the wider discourse so far. That is, comprehension is incremental all the way up to the level that matters most: what the story or conversation is about. Note that earlier reports on discourse-dependent N400 effects invariably relied on semantic anomalies (the only earlier study in which discourse-dependent cloze differences also elicited an N400 effect was aimed at a different phenomenon, and did not cleanly manipulate N400-relevant cloze probabilities; Van Berkum et al., 2005). In the present study a discourse-dependent N400 was elicited by acceptable but unexpected words, which were thus not anomalous. This is entirely consistent with the fact that subtle cloze manipulations can also induce N400 effects in coherent single sentences (DeLong et al., 2005; Hagoort & Brown, 1994; Kutas & Hillyard, 1984).

The effect of scenario fit

As just discussed, the data show that the discourse dependent N400 effect cannot be reduced to mere differences in scenario fit: the actual message of the discourse is taken into account when evaluating new information. However, when the message of the discourse was not particularly constraining, as is the case in the S-biased discourse, bias-inconsistent words still evoked a differential negativity in the N400 time window. The difference in scalp distribution for the two effects after normalization shows that the negativity in the S-biased condition is not just a smaller version of the standard N400 effect present in the MS-biased discourse. As a result, the negativity in the S-biased condition cannot be uniquely attributed to the small remaining difference in cloze value between the scenario-consistent and the scenario-inconsistent word.

But how then should we interpret this unexpected effect related to scenario bias? The timing and polarity of the effect are in line with a modulation of the classic N400, but whereas the distribution of the latter is consistently described to have a centro-parietal maximum (Curran, Tucker, Kutas, & Posner, 1993; Johnson & Hamm, 2000; Kutas & Van Petten, 1994; Osterhout & Holcomb, 1995), the current effect has a more anterior left-lateralized maximum. This distribution resembles the scalp topography of the left anterior negativity or LAN (Coulson et al., 1998; Friederici, Hahne, &

Mecklinger, 1996; Gunter, Stowe, & Mulder, 1997 ; Osterhout & Mobley, 1995). LAN components, however, are related to morphosyntactic violations, whereas the critical manipulation in the present design is purely lexical-semantic. A more appealing explanation for this might therefore be that the observed scenario-related negativity recruits some (but clearly not all) of the neuronal generators that also underlie the classic N400 effect. This is in line with the observed similarities in timing and polarity for the MS- and S-related effects. Such an account would suggest that scenario-based priming lies at the basis of the classic N400 as well, but that it is not the only factor that determines the fit of incoming semantic information. [5]

Apart from this, the mere fact that scenario-based differences in support for particular words elicit a left-lateralized central negativity, instead of a classic N400 effect, may have an interesting additional implication for language research on (or with) the N400. Although researchers agree that the word-elicited N400 indexes how well the meaning of a word fits the constraints set by the context, there is disagreement over whether N400 context effects reflect contextual modulations of the ease of lexical information retrieval (Federmeier, Segal, Lombrozo, & Kutas, 2000), whether they instead directly reflect the compositional processes involved in making sense of language (i.e., semantic integration or 'unification', (Brown et al., 2000; Chwilla et al., 1995; Hagoort et al., 2004; Van Berkum, Brown et al., 2003)), or whether they perhaps necessarily reflect both because the two aspects cannot be meaningfully separated (Coulson & Federmeier, in press). It seems that only a perspective that includes message-level semantic integration as a critical component can account for our current findings. An account in which the word-elicited N400 purely indexes lexical retrieval (and contextual modulations thereof) cannot easily explain why, in our study, differences in scenario-based (or other word-based priming) support for a critical word do not modulate the classic, centroparietal N400.

One thing that is important to note is that it is difficult to separate scenario fit from simple "intra-lexical" word-word priming. In the example

---

[5] The fact that violations of message- and scenario-level constraints engage (at least partially) nonoverlapping neuronal systems might also be related to findings that the two hemispheres are differentially sensitive to the different levels of contextual constraint, as is apparent from divided visual field studies (Atchley, Burgess, & Keeney, 1999; Beeman, Friedman, Grafman, Perez, & et al., 1994; Chiarello, Liu, Quan, & Shears, 2000; Faust, 1998; Faust & Gernsbacher, 1996) and ERP research (Coulson, Federmeier, Van Petten, & Kutas, 2005).

story in Table 6.1, the scenario-biased context also contains words that in isolation would be a moderate to strong prime for the consistent word. Since it has often been shown that individually presented primes do not exert a facilitative effect over intervening words or longer SOA's (> 700 ms) (Masson, 1995; Neely, 1977, 1991; Ratcliff & McKoon, 1988, 1995), it seems unlikely that the N400 effect observed in natural language utterance (where related words are usually separated by time and intervening words) is dependent on strict single-word priming. Kutas and colleagues have shown that within a sentential context, incongruent words that are related to the contextually expected target also evoke an attenuated N400 (Federmeier & Kutas, 1999a, 1999b; Kutas & Hillyard, 1984). This was taken to imply that N400 amplitude serves as an index of relative priming by words present in context. However, taking into account the fuzzy boundary between word-word priming and scenario-based priming, this effect could just as well be ascribed to scenario-based priming. Previous research therefore does not support a conclusion that all effects of scenario-based priming are actually effects of simple lexical priming, or vice-versa. These two types of priming could very well be instantiations of the same underlying process. Both processes require that information in working memory activates dormant information in long term memory through previously established connections. In case of scenario priming, the information that generates the inference, and the knowledge that is in turn activated by that inference, is more elaborate than in word-word priming, but the underlying mechanisms of storage and retrieval seem closely related.

The role of background information in language comprehension

Our experiment suggests that both message- and scenario-based constraints determine the early integration of words within a discourse. This is consistent with a core assumption in several models of language comprehension, such as the Construction-Integration model (Kintsch, 1988; 1998), or the Scenario-Mapping and Focus theory (Sanford & Garrod, 1998). The assumption is that during comprehension, words are mapped onto prior discourse in two different ways: a relatively crude mapping onto potentially relevant scenarios and other static 'default' knowledge structures in long-term memory (e.g., via an automatic resonance process (Myers & O' Brien, 1998; Myers et al., 1994)), and a more precise mapping onto the actual meaning dynamically constructed for the discourse.

Our findings do not support the idea that scenario- and message-level information is (always) used in *consecutive stages* of integration, as suggested by Sanford and Garrod (1998). Instead, our results suggest that people can use scenario- and message-related constraints equally rapidly. This implication of our findings must be treated with some caution, for we do not know whether the N400 evoked in the MS-condition is actually a result of an addition of a message-based N400 and the scenario-based negativity observed in the S-biased condition, or a 'pure' N400 effect elicited solely by message-based constraints. Determining this would require 'message-only' biased discourses without scenario-based or other lower-level forms of contextual support (which appear very difficult, perhaps even impossible, to create). Thus, our data do not allow us to infer whether message- and scenario-level information can actually be used *simultaneously*, or whether strong message-level constraint simply overrides the scenario-based information. What we can infer is that strong message-level expectations can be brought to bear on processing as rapidly as scenario-based constraints in the absence of strong message-level expectations.

Other ERP studies that focused on the interaction between contextual constraint and scenario-based (or word-word) priming have found mixed results. Evidence for the prevalence of message-level congruity over lexical association (Van Petten et al., 1999) has been found, as well as additivity of these two factors (Van Petten, 1993). In addition, as discussed in the introduction, Hoeks et al (2004) found that scenario-related information can overrule the actual message of the discourse. In a recent review Ledoux et al. (2006) state that the exact interaction of message-level information and lexical association depends on the amount of constraint imposed by the message. In a constraining context, effects of lexical association are overridden by message-level effects, whereas in a less constraining discourse the effects of lexical association are present. The present results suggest that for scenario and message information this could also very well be the case. The data presented by Hoeks et al. (2004) furthermore suggest that when the scenario constraint is particularly high, the effect of message-level information is temporarily diminished. Recent N400 evidence that listeners momentarily fail to notice a coherence break when the anomalous word is scenario-relevant (Nieuwland & Van Berkum, 2005) also points in that direction. Taken together, the extant ERP data suggest that both types of information, scenario-based (or based on lexical association) and message-based, are taken into account when processing new information, and that the amount of influence each type of information exerts seems to depend on the relative amount of constraint it receives from the context.

Conclusion

Our first conclusion is that discourse-based N400 effects cannot completely be attributed to scenario fit (or other lower-level priming mechanisms). The actual message of the broader discourse plays a significant role in the early processing of incoming information. Second, message-level fit does not seem to be the only factor influencing early processing of words in coherent text. When the message of the discourse does not generate strong expectations, a clear differential effect of scenario fit shows up, as a left-dominant negativity around the same time as the discourse-based N400 effect. Thus, both the actual message of the context and the scenario suggested by the 'bag of words' in the context can affect early processing in the N400 latency range, as people read a coherent text. Third, the different scalp distributions of the ERP effects observed in the scenario-biased and the message- & scenario-biased condition suggest that different constellations of neural generators are involved when incoming information mismatches constraints based on message-level information as opposed to when the constraints are based on induced scenario's. It is clear, from our own experience, that readers nearly always extract the correct message from a discourse and our results show that message-dependent processing begins extremely rapidly. At the same time, however, basic scenarios play their automatic part in the early stages of language processing, irrespective of that message.

Chapter 7

General discussion and summary

Throughout the studies collected in this dissertation, we find that people are capable of predicting how a story will continue while they are hearing or reading that story. These predictions are highly specific, incorporating meaning (chapter 2), lexical features such as the grammatical gender of the upcoming word (chapter 2, 3 and 4), and possibly also the visual characteristics of the expected word in writing (chapter 5). Importantly, the results show that specific lexical predictions are not simply based on automatic word- or scenario-based activation. The experiments presented in chapter 2, 3 and 4 only showed evidence for the presence of specific lexical predictions when the actual meaning of the discourse supported such an inference. These predictions were not present for stories that contained the same words as the original predictive stories. This shows that people can anticipate upcoming words on the basis of what the discourse is really about.

Within the range of predictive processes examined here, limited cognitive or attentional resources (as indicated by the working memory capacity of the reader) do not seem to influence either the ability to make specific linguistic predictions or the contextual basis of these predictions (chapter 4). Readers with low and with high working memory capacity both make message-based predictions. Interestingly, readers with low working memory capacity show additional activation when confronted with information that is not in line with their prediction. This suggests that a lack of resources leads to increased processing of unexpected information.

Although the effect of prediction is consistently only present in predictive stories and not in prime control stories, the expected and unexpected nouns that follow the critical determiners and adjectives show a different pattern. The difference in N400 amplitude for expected and unexpected nouns that is present in the predictive context is also present in the prime control context. As we by now know, the message of the prime control stories did not support any specific predictions, and based on this there should thus not be a difference between the expectedness of the 'expected' and the 'unexpected' noun in a prime control story. The similarity of the effect in the predictive and the prime control story thus could be taken to suggest that this N400 effect does not reflect message-based expectancy, but scenario- or prime-based automatic activation. The nouns in chapter 2, 3 and 4 are, however, always preceded by gender-marked adjectives or determiners, which are likely to influence how a reader or listeners expects the story to continue. Indeed, in an additional cloze test which included gender-inflected adjectives, reported in chapter 2, there was a difference in cloze value between expected and unexpected words in the prime control condition, which was not present in the original cloze test. This indicates that the difference in N400 amplitude observed for the expected and unexpected nouns in the prime control condition does not necessarily indicate that N400 amplitude reflects automatic activation. In chapter 6, we used a 'cleaner' design in which the gender-marked articles or adjectives were removed from the stories, to explore the influence of scenario-mediated priming and message-level meaning on the amplitude of the N400. In the prime control context, where the cloze values were very much comparable for the previously expected and unexpected nouns, unexpected nouns still evoked a more negative ERP in the N400 time-window (300-500). However, this effect had a different distribution compared to the N400 effect observed in the predictive context. These results show that the discourse dependent N400 effect cannot simply be completely reduced to (scenario-mediated) priming. The N400 effect reflects the rapid use of precise message-level constraints in comprehension. However, these results also show that, at least in the absence of strong message-level constraints, scenario-mediated priming can rapidly affect comprehension.

The role of discourse-message and lexical association in predictive processes

The results reported in chapters 2, 3 and 4 show that the message of the discourse lies at the basis of the on-line lexical predictions that readers and listeners make when confronted with a constraining story. This does not mean, however, that lexical association can play no role in the predictive process. There is quite a lot of evidence that the close proximity of a strongly related prime influences the processing of a related word, even when the message of the discourse is actually unrelated (Camblin et al., 2007; Ditman, Holcomb, & Kuperberg, 2007). The stories used in chapter 2, 3 and 4, however, were not designed to specifically include strong primes, and if they did, the distance between the strong prime and the critical region (inflected adjective and (un)expected noun) was relatively large. It is thus impossible to evaluate the exact role of lexical association from the present results. What the present results unequivocally show, though, is that lexical association is definitely not the only contextual factor that plays a role in the development of linguistic predictions. What the text is actually about plays an important role as well.

Implications for current models of language comprehension

Several models of lexical activation in language comprehension state that simply reading words activates other, related words (Collins & Loftus, 1975; Cook et al., 1998; Myers & O' Brien, 1998; Myers et al., 1994). Such models could thus explain how a reader or a listener can pre-activate words that are likely to follow in that sentence or story. DeLong and colleagues (2005) indeed suggested that words or combinations of words that are present in the preceding sentence activate related words and world-knowledge, which leads to the pre-activation of specific words. However, as discussed before, the results of the experiments in chapter 2, 3 and 4 show that this cannot be the whole story.

So, what are the other options? We suggest that what underlies the specific prediction observed here is the result of convergent predictions being made at several levels of unfolding structure. It is well known that language comprehenders compute the syntactic and conceptual analysis of the incoming language *incrementally* and in parallel (see Jackendoff, 2002; Jackendoff, 2007 for an overall framework and; Vosse & Kempen, 2000 for an explicit computational model of the syntactic side of things). As a consequence, at any point in an unfolding sentence, readers and listeners have at their disposal a partial syntactic and conceptual analysis of the preceding sentence fragment. Each of these partial representations can by itself suggest what might come next. Although these predictions arise at different levels of representation, it is not difficult to see how they might come together and converge onto a specific word. As laid out by Jackendoff (2002; 2007; see also Kempen & Huijbers, 1983; Levelt, 1989), an individual lexical item consists of bits of orthographic, phonological, syntactic, and conceptual information, bundled together into a single multi-leveled structure. If people actually read or hear a word the associated fragments of syntactic and conceptual structure are activated via their orthography or phonology, and merged ('unified') with the syntactic and conceptual analyses constructed for the language input so far. However, within the same framework, the preceding syntactic and conceptual context can, if sufficiently constraining, also each *pre-activate* the relevant bits of structure, resulting in the prediction of the related lexical item. In this case, it is the convergent pull of syntactic and message-based conceptual constraints that activates a particular word, and not the orthographic or phonological input.

Electrophysiological consequences of disconfirming a lexical prediction

In three experiments that explored on-line lexical predictions in language processing (chapter 2, 3 and 4) we consistently found that the ERPs to prediction-consistent and prediction-inconsistent information differ in the predictive context. However, the nature of this ERP effect of prediction is not as consistent over experiments. When gender-inflected adjectives were used as prediction-probes (chapter 2, experiment 1B), the (visually presented) adjectives with an inconsistent inflection evoked a late negative deflection, from 900 to 1200 ms after stimulus onset with a right-frontal distribution, compared to consistent adjectives. In a replication of this experiment with spoken stimuli (chapter 3), the inconsistent adjectives also evoked a negative deflection with a right-frontal distribution, but much earlier, from 200-600 ms after the acoustic onset of the adjective. Finally, when gender-marked articles were used as the critical probes in written stories (chapter 4), prediction-inconsistent articles evoked a negative deflection between 200 and 600 ms over right frontal electrodes, which was followed, for readers with a low working memory capacity only, by a later more centrally distributed negativity between 900 and 1500 ms.

It is important to note that, irrespective of the variation in the nature of the prediction effects observed in the predictive context, we have found a consistent difference between predictive and prime control contexts: significant effects of prediction-mismatch are observed in predictive stories, whereas the prime control stories show no differential activation. This consistent pattern, that is related to our contextual predictiveness manipulation, suggests that the effects observed in the predictive condition are not simply 'false alarms'. If this were the case, then the same noisy processes that underlie the effects in the predictive context should have been present in the prime control context, leading to ERP 'effects' in both types of discourse. Therefore, the ERP effects observed in chapters 2, 3 and 4 can thus not simply be disposed of as noise.

Furthermore, there are important similarities between the prediction-related ERP effects described in this dissertation and previous experiments. The early negative deflection observed in chapters 3 and 4 show important similarities with other ERP effects found with the prediction-probe paradigm. Three previous studies have observed an early negative inflection with a widespread, central distribution, between 300 and 500 ms after the onset of the prediction-inconsistent information (in this case determiners with prediction-inconsistent features in English (DeLong et al., 2005) and Spanish (Wicha, Bates et al., 2003; Wicha, Moreno et al., 2003)). The time course and polarity of these effects seems comparable to the findings reported in chapter 3 and 4, although the scalp distribution differs slightly (visual inspection of the scalp distribution of the prediction effect in Wicha, Bates et al., 2003 however does seem to suggest a somewhat right-frontal distribution).

Additionally, the late negativity observed in chapter 2 is somewhat similar to the additional late negativity that is displayed by low WMC readers in chapter 4. This resemblance could be taken to suggest that these two effects are actually the same. Such a conclusion, however, immediately raises the question why the late negativity observed in chapter 2 is not preceded by the early negative component that is present for both low and high span readers in chapter 4. One important difference between the two experiments was the type of prediction-probe used (adjectives with gender-dependent inflections in chapter 2 and gender-marked determiners in chapter 4). Indeed, one could argue that an inflected adjective provides a somewhat more subtle clue about the gender of the upcoming noun than a gender marked determiner, since determiners are very common and very short. The unexpected gender of the determiner are thus likely to be perceived at a glance (i.e. compare "de" vs. "het") whereas the subtle difference in adjective inflection, which usually relies on the presence or absence of a single "-e" is less conspicuous (i.e. compare "klein" vs "kleine"). The fact that the inconsistency of the adjective inflection is less striking than the inconsistency of the determiner could have led to a delay in noticing the inconsistency or to a different, less profound response for the participants in chapter 2. However, if this would be the case, then we should have observed the same pattern of results in the experiment reported in chapter 3, where, just as in chapter 2, inflected adjectives were used as the critical probes. As stated above, the spoken adjectives with a prediction-inconsistent inflection from chapter 3 showed an early negative shift that is comparable to the early negative shift evoked by the prediction-inconsistent determiners in chapter 4. It could thus be that perhaps other experimental features, such as individual differences of participants, or specific features of the experimental design such as the type of fillers, influence the way in which a reader or a listener makes predictions, or processes prediction-inconsistent materials.

Taken together, we must conclude that we can not draw any final conclusion about the nature of the prediction effect. One of the core goals of future research should thus be to determine what factors underlie this variability. One important factor that deserves to be explored is the influence of strategic processes in prediction: can we as readers unconsciously 'decide' to rely more strongly on predictive processes, for example in a noisy environment (see Pickering & Garrod, 2007), or, on the contrary, diminish the influence of prediction, for example when everything that we read follows unexpected and seemingly incongruent paths? It could be that such broad contextual factors influence the presence or nature of the predictions that are made, and with that, perhaps can explain the variability in observed ERP effects.

References

a

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. <u>Cognition, 73</u>(3), 247-264.

Altmann, G. T. M., van Nice, K. Y., Garnham, A., & Henstra, J. A. (1998). Late closure in context. <u>Journal of Memory and Language, 38</u>(4), 459-484.

Atchley, R. A., Burgess, C., & Keeney, M. (1999). The effect of time course and context on the facilitation of semantic features in the cerebral hemispheres. <u>Neuropsychology, 13</u>(3), 389-403.

b

Baddeley, A. (2003). Working Memory: Looking Back and Looking Forward. <u>Nature Reviews Neuroscience, 4</u>(10), 829-839.

Bar, M. (2004). Visual objects in context. <u>Nature Reviews Neuroscience, 5</u>(8), 617-629.

Beeman, M., Friedman, R. B., Grafman, J., Perez, E., & et al. (1994). Summation priming and coarse semantic coding in the right hemisphere. <u>Journal of Cognitive Neuroscience, 6</u>(1), 26-45.

Beeman, M. J., Bowden, E. M., & Gernbacher, M. A. (2000). Right and left hemisphere cooperation for drawing predictive and coherence inferences during normal story comprehension. <u>Brain and Language, 71</u>(2), 310-336.

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. <u>Electroencephalogr Clin Neurophysiol, 60</u>(4), 343-355.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. <u>Cognitive Psychology, 14</u>(2), 143-177.

Brown, C. M., van Berkum, J. J. A., & Hagoort, P. (2000). Discourse before gender: An event-related brain potential study on the interplay of semantic and syntactic information during spoken language understanding. <u>Journal of Psycholinguistic Research, 29</u>(1), 53-68.

Brumback, C. R., Low, K. A., Gratton, G., & Fabiani, M. (2005). Putting Things into Perspective: Individual Differences in Working-Memory Span and the Integration of Information. <u>Experimental Psychology, 52</u>(1), 21-30.

Budd, D., Whitney, P., & Turley, K. J. (1995). Individual differences in working memory strategies for reading expository text. <u>Memory and Cognition, 23</u>(6), 735-748.

C

Calvo, M. G. (2001). Working memory and inferences: Evidence from eye fixations during reading. Memory, 9(4-6), 365-381.

Calvo, M. G., & Castillo, M. D. (1996). Predictive Inferences Occur On-Line, but with Delay: Convergence of Naming and Reading Times. Discourse Processes, 22(1), 57-78.

Calvo, M. G., Castillo, M. D., & Estevez, A. (1999). On-Line Predictive Inferences in Reading: Processing Time during versus after the Priming Context. Memory and Cognition, 27(5), 834-843.

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(1), 103-128.

Campion, N. (2004). Predictive inferences are represented as hypothetical facts. Journal of Memory and Language, 50(2), 149-164.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. Behavioral and Brain Sciences, 22(1), 77-126.

Carroll, P., & Slowiaczek, M. L. (1986). Constraints on semantic priming in reading: A fixation time analysis. Memory and Cognition, 14(6), 509-522.

Chiarello, C., Liu, S., Quan, N., & Shears, C. (2000). Imageability and word recognition in the left and right visual fields: A signal detection analysis. Brain and Cognition, 43(1-3), 90-94.

Chomsky, N. (1957). Syntactic structures. Den Haag, The Netherlands: Mouton.

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

Clark, H. H. (1996). Using language. Cambridge [England]: Cambridge University Press.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. Psychological Review, 82(6), 407-428.

Cook, A. E., Halleran, J. G., & O'Brien, E. J. (1998). What is readily available during reading? A memory-based view of text processing. Discourse Processes, 26(2-3), 109-129.

Cook, A. E., Limber, J. E., & O'Brien, E. J. (2001). Situation-Based Context and the Availability of Predictive Inferences. Journal of Memory and Language, 44(2), 220-234.

Cook, A. E., & Myers, J. L. (2004). Processing discourse roles in scripted narratives: The influences of context and world knowledge. Journal of Memory and Language, 50(3), 268-288.

Coulson, S., & Federmeier, K. (in press). Words in context: ERPs and the lexical/postlexical distinction. Journal of psycholinguistic research.

Coulson, S., Federmeier, K. D., Van Petten, C., & Kutas, M. (2005). Right Hemisphere Sensitivity to Word- and Sentence-Level Context: Evidence From Event-Related Brain Potentials. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(1), 129-147.

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

Curran, T., Tucker, D. M., Kutas, M., & Posner, M. I. (1993). Topography of the N400: Brain electrical activity reflecting semantic expectancy. Electroencephalography and Clinical Neurophysiology: Evoked Potentials, 88(3), 188-209.

d

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior, 19(4), 450-466.

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9(4), 561-584.

Davenport, J. L., & Potter, M. C. (2004). Scene Consistency in Object and Background Perception. Psychological Science, 15(8), 559-564.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. Psychological Review, 93(3), 283-321.

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

Ditman, T., Holcomb, P. J., & Kuperberg, G. R. (2007). The contributions of lexico-semantic and discourse information to the resolution of ambiguous categorical anaphors. Language and Cognitive Processes, 22(6), 793-827.

Duffy, S. A., Henderson, J. M., & Morris, R. K. (1989). Semantic facilitation of lexical access during sentence processing. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15(5), 791-801.

e

Elman, J. L. (1990). Representation and structure in connectionist models. In G. T. M. Altmann (Ed.), Cognitive models of speech processing: Psycholinguistic and computational perspectives. (pp. 345 - 382). Cambridge, MA, US: The MIT Press.

Engle, R. W. (2002). Working memory capacity as executive attention. Current Directions in Psychological Science, 11(1), 19-23.

Estevez, A., & Calvo, M. G. (2000). Working memory capacity and time course of predictive inferences. Memory, 8(1), 51-61.

f

Faust, M. (1998). Obtaining evidence of language comprehension from sentence priming, Beeman, Mark (Ed); Chiarello, Christine (Ed). (1998). Right hemisphere language comprehension: Perspectives from cognitive neuroscience. (pp. 161 185). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers. xii, 408 pp.

Faust, M. E., & Gernsbacher, M. A. (1996). Cerebral mechanisms for suppression of inappropriate information during sentence comprehension. Brain and Language, 53(2), 234-259.

Federmeier, K. D., & Kutas, M. (1999a). Right words and left words: Electrophysiological evidence for hemispheric differences in meaning processing. Cognitive Brain Research, 8(3), 373-392.

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

Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: an ERP study. Psychophysiology, 39(2), 133-146.

Federmeier, K. D., Segal, J. B., Lombrozo, T., & Kutas, M. (2000). Brain responses to nouns, verbs and class-ambiguous words in context. Brain, 123(12), 2552-2566.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. Brain Research, 1146, 75-84.

Fincher Kiefer, R. (1993). The Role of Predictive Inferences in Situation Model Construction. Discourse Processes, 16(1-2), 99-124.

Fincher Kiefer, R. (1995). Relative Inhibition Following the Encoding of Bridging and Predictive Inferences. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(4), 981-995.

Fincher Kiefer, R. (1996). Encoding Differences between Bridging and Predictive Inferences. Discourse Processes, 22(3), 225-246.

Ford, J. M., Woodward, S. H., Sullivan, E. V., Isaacks, B. G., Tinklenberg, J. R., Yesavage, J. A., & Roth, W. T. (1996). N400 evidence of abnormal responses to speech in Alzheimer's disease. Electroencephalography and Clinical Neurophysiology, 99(3), 235-246.

Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. Brain and Language, 50(3), 259-281.

Friederici, A. D. (1997). Diagnosis and reanalysis: Two processing aspects of the brain may differentiate. In J. Fodor & F. Fereira (Eds.), Reanalysis in sentence processing. (pp. 177-200). Dordrecht: Kluwer.

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22(5), 1219-1248.

G

Garrett, M. (2000). Remarks on the architecture of language processing systems. In Y. Grodzinsky & L. P. Shapiro & D. Swinney (Eds.), Language and the brain: Representation and processing. (Vol. xxi, pp. 31-69). San Diego, CA, US: Academic Press.

Garrod, S., & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and resolution. Journal of Memory and Language, 42(4), 526-544.

Gernsbacher, M. A., & Faust, M. (1991). The role of suppression in sentence comprehension, Simpson, Greg B (Ed). (1991). Understanding word and sentence. (pp. 97 128). Oxford, England: North Holland. xiv, 400 pp.

Gernsbacher, M. A., & Faust, M. (1995). Skilled suppression, Dempster, Frank N (Ed); Brainerd, Charles J (Ed). (1995). Interference and inhibition in cognition. (pp. 295 327). San Diego, CA, US: Academic Press. xvii, 423 pp.

Gernsbacher, M. A., & Faust, M. E. (1991). The mechanism of suppression: A component of general comprehension skill. Journal of Experimental Psychology: Learning, Memory, and Cognition, 17(2), 245-262.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. Psychological Review, 101(3), 371-395.

Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. Psychophysiology, 34(6), 660-676.

h

Haberlandt, K. F., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. Journal of Experimental Psychology: General, 114(3), 357-374.

Hagoort, P., & Brown, C. (1994). Brain responses to lexical ambiguity resolution and parsing, Clifton, Charles Jr. (Ed); Frazier, Lyn (Ed); Rayner, Keith (Ed). (1994). Perspectives on sentence processing. (pp. 45 80). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc. xiv, 483 pp.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. Language and Cognitive Processes, 8(4), 439-483.

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: A Journal of Neurology, 119(2), 627-649.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of Word Meaning and World Knowledge in Language Comprehension. Science, 304(5669), 438-441.

Hess, D. J., Foss, D. J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. Journal of Experimental Psychology: General, 124(1), 62-82.

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

j

Jackendoff, R. (2002). Foundations of Language. New York: Oxford University Press.

Jackendoff, R. (2007). A Parallel Architecture perspective on language processing. Brain Research, 1146, 2-22

Johnson, B. W., & Hamm, J. P. (2000). High-density mapping in an N400 paradigm: evidence for bilateral temporal lobe generators. Clin Neurophysiol, 111(3), 532-545.

Jolij, J., & Lamme, V. A. F. (submitted). Top-down effects in conscious visual perception make you see what is not there: evidence from a magnetic stimulation-induced visual illusion.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. Psychophysiology, 37(2), 163-178.

Jung, T. P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. Clinical Neurophysiology, 111(10), 1745-1758.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. Psychological Review, 99(1), 122-149.

Just, M. A., Carpenter, P. A., & Keller, T. A. (1996). The capacity theory of comprehension: New frontiers of evidence and arguments. Psychological Review, 103(4), 773-780.

k

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. Journal of Memory and Language, 49(1), 133-156.

Keefe, D. E., & McDaniel, M. A. (1993). The Time Course and Durability of Predictive Inferences. Journal of Memory and Language, 32(4), 446-463.

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. Cognition, 14(2), 185-209.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. Journal of Memory and Language, 30(5), 580-602.

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

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

Kintsch, W., Patel, V. L., & Ericsson, K. A. (1999). The role of long-term working memory in text comprehension. Psychologia: An International Journal of Psychology in the Orient, 42(4), 186-198.

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

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

Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., & Caplan, D. (2003). Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. Journal of Cognitive Neuroscience, 15(2), 272-293.

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

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

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. Nature, 307(5947), 161-163.

Kutas, M., & Van Petten, C. K. (1994). Psycholinguistics electrified: Event-related brain potential investigations, Gernsbacher, Morton Ann (Ed). (1994). Handbook of psycholinguistics. (pp. 83 143). San Diego, CA, US: Academic Press. xxii, 1174 pp.

Kutas, M., Van Petten, C. K., & Kluender, R. (2006). Psycholinguistics Electrified II (1994-2005). In M. A. Gernsbacher, & Traxler, M. (Ed.), Handbook of Psycholinguistics, 2nd edition. (pp. 659-724). New York: Elsevier Press.

ℓ

Lau, E., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. Brain and Language, 98(1), 74-88.

Ledoux, K., Camblin, C. C., Swaab, T. Y., & Gordon, P. C. (2006). Reading Words in Discourse: The Modulation of Lexical Priming Effects by Message-Level Context. Behavioral and Cognitive Neuroscience Reviews, 5(3), 107-127.

Legge, G. E., Ahn, S. J., Klitz, T. S., & Luebker, A. (1997). Psychophysics of reading-XVI. The visual span in normal and low vision. Vision Research, 37(14), 1999-2010.

Levelt, W. J. M. (1989). Working models of perception: Five general issues. In B. A. G. Elsendoorn & H. Bouma (Eds.), Working models of human perception. (Vol. xiii, pp. 489-503). San Diego, CA, US: Academic Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. Behavioral and Brain Sciences, 22(1), 1-75.

Linderholm, T. (2002). Predictive inference generation as a function of working memory capacity and causal text constraints. Discourse Processes, 34(3), 259-280.

m

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. Cognitive Psychology, 24(1), 56-98.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. Psychological Review, 101(4), 676-703.

Masson, M. E. J. (1995). A distributed memory model of semantic priming. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(1), 3-23.

McAteer, E. (1992). Typeface emphasis and information focus in written language. Applied Cognitive Psychology, 6(4), 345-359.

McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. Electroencephalogr Clin Neurophysiol, 62(3), 203-208.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. Psychological Review, 88(5), 375-407.

McDaniel, M. A., Schmalhofer, F., & Keefe, D. E. (2001). What is minimal about predictive inferences? Psychonomic Bulletin and Review, 8(4), 840-846.

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12(1), 82-91.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. Psychological Review, 99(3), 440-466.

Meyer, D. E., & Schvaneveldt, R. W. (1976). Meaning, memory structure, and mental processes. Science, 192(4234), 27-33.

Morris, R. K. (1994). Lexical and message-level sentence context effects on fixation times in reading. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20(1), 92-103.

Munte, T. F., Heinze, H. J., Matzke, M., Wieringa, B. M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: no evidence for specificity of the syntactic positive shift. Neuropsychologia, 36(3), 217-226.

Murray, J. D., & Burke, K. A. (2003). Activation and encoding of predictive inferences: The role of reading skill. Discourse Processes, 35(2), 81-102.

Murray, J. D., Klin, C. M., & Myers, J. L. (1993). Forward inferences in narrative text. Journal of Memory and Language, 32(4), 464-473.

Myers, J. L., & O' Brien, E. J. (1998). Accessing the discourse representation during reading. Discourse Processes, 26(2-3), 131-157.

Myers, J. L., O'Brien, E. J., Albrecht, J. E., & Mason, R. A. (1994). Maintaining global coherence during reading. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20(4), 876-886.

N

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(3), 226-254.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories, Besner, Derek (Ed); Humphreys, Glyn W (Ed). (1991). Basic processes in reading: Visual word recognition. (pp. 264 336). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc 350 pp.

Nieuwland, M. S., & Van Berkum, J. J. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. Brain Res Cogn Brain Res, 24(3), 691-701.

Nieuwland, M. S., & Van Berkum, J. J. A. (2006a). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. Brain Research, 1118(1), 155-167.

Nieuwland, M. S., & Van Berkum, J. J. A. (2006b). When Peanuts Fall in Love: N400 Evidence for the Power of Discourse. Journal of Cognitive Neuroscience, 18(7), 1098-1111.

O

O'Brien, E. J., Shank, D. M., Myers, J. L., & Rayner, K. (1988). Elaborative inferences during reading: Do they occur on-line? Journal of Experimental Psychology: Learning, Memory, and Cognition, 14(3), 410-420.

O'Seaghdha, P. G. (1997). Conjoint and dissociable effects of syntactic and semantic context. Journal of Experimental Psychology: Learning, Memory, and Cognition, 23(4), 807-828.

Osterhout, L. (1994). Event-related brain potentials as tools for comprehending language comprehension, Clifton, Charles Jr. (Ed); Frazier, Lyn (Ed);

Rayner, Keith (Ed). (1994). Perspectives on sentence processing. (pp. 15 44). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc. xiv, 483 pp.

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

Osterhout, L., & Holcomb, P. J. (1995). Event related potentials and language comprehension, Rugg, Michael D (Ed); Coles, Michael G. H (Ed). (1995). Electrophysiology of mind: Event related brain potentials and cognition. (pp. 171 215). New York, NY, US: Oxford University Press. xii, 220 pp.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. Journal of Memory and Language, 34(6), 739-773.

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

Otten, M., & Van Berkum, J. J. A. (2007). What makes a discourse constraining? A comparison between the effects of discourse message and priming on the N400. Brain Research, 1153, 166-177.

Otten, M., & Van Berkum, J. J. A. (in press). Discourse-based word anticipation during language processing: Prediction or priming? Discourse Processes.

p

Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. Memory and Cognition, 3(5), 519-526.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? Trends in Cognitive Sciences, 11(3), 105-110.

r

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. Psychological Review, 95(3), 385-408.

Ratcliff, R., & McKoon, G. (1995). Sequential effects in lexical decision: Tests of compound-cue retrieval theory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(5), 1380-1388.

Rosen, V. M., & Engle, R. W. (1998). Working memory capacity and suppression. Journal of Memory and Language, 39(3), 418-436.

S

Sanford, A. J. (1990). On the nature of text-driven inference, Balota, David A (Ed); Flores d'Arcais, Giovanni B (Ed); Rayner, Keith (Ed). (1990). Comprehension processes in reading. (pp. 515 535). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc. xix, 656 pp.

Sanford, A. J., & Garrod, S. C. (1981). Understanding written language. New York: Wiley.

Sanford, A. J., & Garrod, S. C. (1998). The role of scenario mapping in text comprehension. Discourse Processes, 26(2-3), 159-190.

Sanford, A. J. S., Sanford, A. J., Molle, J., & Emmott, C. (2006). Shallow Processing and Attention Capture in Written and Spoken Discourse. Discourse Processes, 42(2), 109-130.

Schmalhofer, F., McDaniel, M. A., & Keefe, D. (2002). A unified model for predictive and bridging inferences. Discourse Processes, 33(2), 105-132.

Singer, M., Andrusiak, P., Reisdorf, P., & Black, N. L. (1992). Individual differences in bridging inference processes. Memory and Cognition, 20(5), 539-548.

Sitnikova, T., Salisbury, D. F., Kuperberg, G., & Holcomb, P. J. (2002). Electrophysiological insights into language processing in schizophrenia. Psychophysiology, 39(6), 851-860.

St George, M., Mannes, S., & Hoffman, J. E. (1994). Global semantic expectancy and language comprehension. Journal of Cognitive Neuroscience, 6(1), 70-83.

St George, M., Mannes, S., & Hoffman, J. E. (1997). Individual differences in inference generation: An ERP analysis. Journal of Cognitive Neuroscience, 9(6), 776-787.

Swaab, T., Brown, C., & Hagoort, P. (1997). Spoken sentence comprehension in aphasia: Event-related potential evidence for a lexical integration deficit. Journal of Cognitive Neuroscience, 9(1), 39-66.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. Journal of Verbal Learning and Verbal Behavior, 18(6), 645-659.

T

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension, Miller, Joanne L (Ed); Eimas, Peter D (Ed). (1995). Speech, language, and communication. (pp. 217 262). San Diego, CA, US: Academic Press. xviii, 415 pp.

Traxler, M. J., & Foss, D. J. (2000). Effects of sentence constraint on priming in natural language comprehension. Journal of Experimental Psychology: Learning, Memory, and Cognition, 26(5), 1266-1282.

Traxler, M. J., Foss, D. J., Seely, R. E., Kaup, B., & Morris, R. K. (2000). Priming in sentence processing: Intralexical spreading activation, schemas, and situation models. Journal of Psycholinguistic Research, 29(6), 581-595.

v

Van Berkum, J. J. A. (1996). The psycholinguistics of grammatical gender: Studies in language comprehension and production. Unpublished Doctoral Dissertation, Nijmegen University Press, Nijmegen, The Netherlands.

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

Van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999b). When does gender constrain parsing? Evidence from ERPs. Journal of Psycholinguistic Research, 28(5), 555-571.

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

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

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

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

Van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. Cognitive Brain Research, 17(3), 701-718.

Van den Noort, M. W. M. L., Bosch, M. P. C., & Hugdahl, K. (2005). Understanding the role of working memory in second-language acquisition. Paper presented at the KogWis05 The German Cognitive Science Conference.

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

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

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(2), 394-417.

Van Petten, C., Weckerly, J., McIsaac, H. K., & Kutas, M. (1997). Working memory capacity dissociates lexical and sentential context effects. Psychological Science, 8(3), 238-242.

Vissers, C. T. W. M., Chwilla, D. J., & Kolk, H. H. J. (2006). Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. Brain Research, 1106(1), 150-163.

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

w

Whitney, P., Arnett, P. A., Driver, A., & Budd, D. (2001). Measuring central executive functioning: What's in a reading span? Brain and Cognition, 45(1), 1-14.

Wicha, N. Y. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: human brain potentials to gender expectation and agreement in Spanish spoken sentences. Neuroscience Letters, 346(3), 165-168.

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2003). Expecting gender: An event related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in Spanish. Cortex, 39(3), 483-508.

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. Journal of Cognitive Neuroscience, 16(7), 1272-1288.

z

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

Appendices

APPENDIX 1. A sample of the materials used in experiment 1A, chapter 2 (approximate English translation). The critical anomalous word is printed in boldface italics, the bracketed word is the expected noun implicated by the context (and not shown to the subjects).

Predictive Discourse	Prime Control Discourse
The doctors found that the young woman had an aggressive but curable tumor. She got a **pear** *[chemo]* but died a few months later.	The doctors found that the young woman had an aggressive tumor which was not curable. She got a **pear** but died a few months later.
The rich man wanted to have a house designed especially for him. He approached an **animal** *[architect]* to make an estimate of the costs.	The rich man had had a house designed especially for him. He approached an **animal** to make a unique staircase.
The woman was very satisfied with the waiter's service. So she gave him a **maniac** *[tip]* on top of the bill to show her appreciation.	The woman was not very satisfied with the waiter's service. So she gave him a **maniac** on top of the bill to elevate his mood.
The strict Presbyterian family is going to have dinner together. They start with a **strike** *[prayer]* to thank god for the meal.	The strict Presbyterian family is going to play goose after dinner together. They start with a **strike** to thank god for their good life.

APPENDIX 2. A sample of the materials used in experiment 1B (chapter 1), and the experiments reported in chapter 2 and 3. The critical adjective is printed in boldface. These four stories exemplify the different ways in which predictive stories were changed into less predictive prime control stories.

Predictive Discourse	Prime Control Discourse
Nadat hij uren naar het enorme lege doek had gekeken voelde de schilder inspiratie opkomen. Hij greep naar een **grote** vanwege intensief gebruik sleetse kwast/ **groot** vanwege intensief gebruik sleets paletmes en smeet de verf op het doek.	Nadat hij uren naar het enorme lege doek had gekeken had de schilder nog steeds geen inspiratie. Hij greep naar een **grote** vanwege intensief gebruik sleetse kwast/ **groot** vanwege intensief gebruik sleets paletmes en smeet deze door zijn atelier.
Anne had eindelijk een rustig plekje gevonden waar ze kon studeren. Ze ging zitten en pakte een **dik** en behoorlijk beduimeld boek/ **dikke** en behoorlijk beduimelde roman uit haar tas.	Na het studeren had Anne een rustig plekje in het park gevonden. Ze ging zitten en pakte een **dik** en behoorlijk beduimeld boek/ **dikke** en behoorlijk beduimelde roman uit haar tas.
De misdadiger is opgepakt en veroordeeld en zit voor drie jaar in een gevangenis. Hij zit bijna altijd in een **verouderde** en daarom behoorlijk onprettige cel/ **verouderd** en daarom behoorlijk onprettig gevang maar komt binnenkort vrij.	De misdadiger heeft zijn leven gebeterd nadat hij was opgepakt en veroordeeld tot drie jaar gevangenis. Hij zit bijna altijd in een **verouderde** en daarom behoorlijk onprettige cel/ **verouderd** en daarom behoorlijk onprettig gevang maar komt binnenkort vrij.
Het kleine kind had het warm vanwege de hittegolf en liep te zeuren. Ze wilde een **koud** en liefst ook lekker ijsje/ **koude** en liefst ook lekkere ijslolly om af te koelen.	De moeder had het warm vanwege the hittegolf en vond dat haar kind liep te zeuren. Ze wilde een **koud** en liefst ook lekker ijsje/ **koude** en liefst ook lekkere ijslolly om af te koelen.

APPENDIX 3. A sample of materials used in chapter 4. The critical nouns are printed in boldface. The first bold-face noun is the expected noun, the second bold-face noun is the unexpected noun.

MS biased Discourse	S biased Discourse
De voetballers vierden de goede afloop van hun wedstrijd in het café. Ze namen allemaal een **biertje/ pilsje** voordat ze weer naar huis fietsten.	De voetballers hoopten straks in het café de goede afloop van hun wedstrijd te vieren. Ze namen allemaal een **biertje/ pilsje** voordat ze het veld op gingen.
De cardioloog wist dat zijn jonge patiëntje een transplantatie nodig had. Hij wachtte op een **hart/ donor** voor het kleine ventje.	De cardioloog had zijn jonge patiëntje geholpen met een transplantatie. Hij wachtte op een **hart/ donor** voor het kleine tweelingbroertje van het ventje.
De vakbond was verontwaardigd over de lage loonsverhoging. Ze organiseerden een **staking/ protest** vanwege de slechte omstandigheden en belachelijke betaling in de fabriek.	De vakbond accepteerde de lage loonsverhoging vanwege de verontwaardigde directeur. Ze organiseerden een **staking/ protest** vanwege de slechte omstandigheden in de fabriek.
Iedereen dacht dat Jesse het niet ver zou brengen als kok, maar hij zette door. Hij werkt nu bij een **restaurant/ zaak** waar vrij veel beroemde mensen komen en maakt kans op een ster.	Iedereen dacht dat Jesse het ver zou brengen als kok, maar hij zette niet door. Hij werkt nu bij een **restaurant/ zaak** waar vrij veel penoze komt, in de bediening.

Nederlandse samenvatting

Lexicale predictie tijdens taalverwerking

Gesproken of geschreven taal brengt een boodschap over van de schrijver of spreker naar de ontvanger, de lezer of luisteraar. Meteen als de ontvanger begint met lezen of luisteren, begint het extraheren van die boodschap. Dit begripsproces gaat continu door, terwijl de rest van de tekst binnen komt. Sommige teksten geven al vroeg een indicatie hoe ze naar alle waarschijnlijkheid verder zullen gaan. Een verhaaltje als "*De koene ridder zag dat de draak de goede tovenaar bedreigde. Hij pakte snel een …*" is behoorlijk voorspellend voor het woord *zwaard*. Als wij nu in staat zouden zijn om die informatie te gebruiken om het woord *zwaard* alvast te pre-activeren, zou dat heel nuttig zijn. Als de zin inderdaad verdergaat met het voorspelde woord, dan bespaart dat de ontvanger veel (taalbegrips-)werk. Aan de andere kant, als het verhaal juist anders verloopt dan de verwachting, bijvoorbeeld "*De koene ridder zag dat de draak de goede tovenaar bedreigde. Hij pakte snel een lans.*" dan geeft deze *mismatch* juist aan dat het nuttig is om extra aandacht geven aan het verwerken van deze onverwachte wending in de tekst. Maar wordt predictieve contextuele informatie ook daadwerkelijk gebruikt om deze behoorlijk specifieke, talige predicties te maken?

Op basis van ervaringen uit het dagelijks leven zou je zeggen van wel. Het gebeurt bijvoorbeeld regelmatig dat je de zinnen van je gesprekspartner kan afmaken bij een aarzeling of onderbreking. In zo'n geval is er echter wel wat meer tijd dan normaal om na te denken over wat die gesprekspartner gezegd zou hebben, juist omdat het gesprek stokt. Een belangrijke vraag is dus of we ook kunnen anticiperen als we tegelijkertijd nog druk bezig zijn met het vaststellen en verwerken van de echte zinsbetekenis. Deze vraag hebben we in dit proefschrift onderzocht (in hoofdstuk 2, 3 en 4) met behulp van een paradigma dat oorspronkelijk is ontworpen door Van Berkum en collega's (2005). Als een lezer inderdaad het woord *zwaard* verwacht als hij/ zij het verhaal "*De koene ridder zag dat de draak de goede tovenaar bedreigde. Hij pakte snel een …*" leest, dan zal op dat moment het adjectief "*scherpe*" als een onverwachte verrassing komen, terwijl het adjectief "*scherp*" wel overeenkomt met het woordgeslacht van het verwachte zelfstandig naamwoord *zwaard*. Om te testen of mensen inderdaad meer moeite hebben met een woord als "*scherpe*" versus "*scherp*", hebben wij mensen verhaaltjes met zulke woorden getoond, en tegelijkertijd hun elektrische hersenactiviteit geregistreerd met een electro-encefalogram (EEG). Dit EEG wordt opgenomen terwijl de deelnemer aan het experiment gewoon leest of luistert, er is geen extra (onnatuurlijke) taak nodig om bijvoorbeeld een reactietijd vast te stellen. Op basis van dit EEG kan worden vastgesteld welke elektrische activatie precies samenhangt met het aanbieden van een specifieke gebeurtenis (zoals het lezen van een predictie-consistent of een predictie-inconsistent adjectief). Deze gebeurtenis-gerelateerde hersenpotentialen worden *event-related potentials* of ERPs genoemd.

De experimenten die zijn verzameld in dit proefschrift laten zien dat adjectieven met een predictie-inconsistent woordgeslacht (zoals "*scherpe*" als het woord *zwaard* hoogverwacht is) stelselmatig leiden tot een ander ERP dan consistente adjectieven (zoals "*scherp*"). Dit bevestigt dat lezers en luisteraars in staat zijn om te voorspellen hoe een zin of verhaal zal verdergaan *terwijl* ze druk bezig zijn input te verwerken. Deze experimenten laten bovendien zien dat deze voorspellingen heel gedetailleerd zijn: mensen anticiperen naast de betekenis van een verwacht woord (hoofdstuk 2, experiment 1A) ook specifieke lexicale eigenschappen zoals het woordgeslacht van het verwachte woord (hoofdstuk 2 (experiment 1B), 3 en 4). Een *self-paced reading* experiment (hoofdstuk 5), waarin lezers een verhaaltje op hun eigen tempo woord voor woord lezen, door na ieder gelezen woord op een knop te drukken, geeft aan dat deze lexicale voorspellingen mogelijk nog preciezer zijn. De resultaten van dit experiment suggereren namelijk dat de lezers ook de visuele verschijningsvorm van het hoogverwachte woord kunnen voorspellen.

De contextuele basis van lexicale predicties

Lezers en luisteraars zijn dus in staat om te anticiperen hoe een zin of verhaal waarschijnlijk zal verder gaan. Het lijkt waarschijnlijk dat deze predicties zijn gebaseerd op de voorspellende boodschap van het voorafgaande verhaaltje. Het is echter ook mogelijk dat deze lexicale predicties voortkomen uit de automatische activatie van het verwachte woord op basis van gerelateerde woorden die in de eerdere tekst stonden. In het eerdergenoemde voorbeeld is het woord "*ridder*" bijvoorbeeld duidelijk gerelateerd aan *zwaard*. Volgens de *spreading activation* theorie (Collins & Loftus, 1975) worden voor ieder woord dat wordt geactiveerd in het lexicon (het mentale woordenboek) alle woorden die samenhangen met dat woord ook geactiveerd. Als lexicale predictie inderdaad gebaseerd is op dergelijke automatische activatie, dan zijn alleen de losse woorden waaruit een uiting bestaat van belang voor predictieve processen. In dit geval zou de eigenlijke boodschap of betekenis van het verhaal er dus niet toe doen.

Een centrale vraag in dit proefschrift betreft dan ook de contextuele basis van lexicale voorspellingen. Zijn deze predicties een bijproduct van relatief "domme" automatische activatie op basis van individuele, gerelateerde woorden in de context? Of zijn de voorspellingen juist gebaseerd op een meer complete representatie van de betekenis van de context? Dit hebben wij onderzocht te vergelijken in hoeverre mensen specifieke woorden voorspelden op basis van verhaaltjes (waarvan de boodschap voorspellend was zoals "*De koene ridder zag dat de draak de goede tovenaar bedreigde. Hij pakte snel een …*") en controle verhaaltjes (waarvan de boodschap minder of niet voorspellend was, maar die wel dezelfde woorden bevatten, zoals "*De goede tovenaar zag dat de draak de koene ridder bedreigde. Hij pakte snel een …*").

In hoofdstuk 2, 3 en 4 vinden wij alleen maar evidentie voor lexicale voorspellingen in de predictieve verhalen en niet in de controle verhalen. Dus, alleen als de betekenis van de zin de predictie ondersteunde was er een verschil in de ERPs voor predictie-consistente en predictie-inconsistente informatie. Deze resultaten laten duidelijk zien dat lexicale voorspellingen niet alleen maar gebaseerd zijn op automatische activatie op basis van gerelateerde woorden in de context, maar dat de boodschap van de context ook een belangrijke rol kan spelen.

De elektrofysiologische gevolgen van een predictie-schending

In de drie experimenten waarin lexicale predictie tijdens taalverwerking werd onderzocht (hoofdstuk 2, 3 en 4) hebben we stelselmatig gevonden dat predictie-consistente informatie een ander ERP oproept dan predictie-consistente informatie. Het geobserveerde ERP effect zelf was echter niet compleet vergelijkbaar over de verschillende experimenten. Als adjectieven met een woordgeslacht-afhankelijke inflectie ("groot" vs "grote") werden gebruikt als predictie *probe* (hoofdstuk 2, experiment 1B) dan lieten de (visueel gepresenteerde) inconsistente adjectieven een late, negatieve, rechts-frontale deviatie zien in het ERP, 900 tot 1200 ms nadat het adjectief werd gepresenteerd. Als dezelfde stimuli gesproken werden aangeboden (hoofdstuk 3), dan lieten de inconsistente adjectieven een veel vroegere negatieve, rechts-frontale deviatie zien, namelijk van 200 tot 600 ms na aanvang van het gesproken adjectief. Als laatste werden in hoofdstuk 4 de verschillen in bepaalde lidwoorden ("*de*" vs "*het*") gebruikt als predictie probe. In deze geschreven verhalen lieten de predictie-inconsistente adjectieven ook een negatieve, rechts-frontale deviatie zien tussen 200 tot 600 ms. Deze werd echter (alleen voor lezers met een lage werkgeheugen capaciteit) gevolgd door een late, meer centraal gedistribueerde negatieve deflectie tussen 900 en 1500 ms.

Op basis van de huidige gegevens niet mogelijk om vast te stellen welke factoren er precies voor zorgen dat het ERP effect van lexicale predictie zo variabel is. De factoren die wij hebben gevarieerd, namelijk het type predictie probe (adjectief vs. lidwoord), de modaliteit waarin de verhaaltjes worden aangeboden (gesproken vs. geschreven) en de werkgeheugencapaciteit van de deelnemers, bieden geen van allen een verklaring voor de waargenomen variatie in ERP effecten. Het lijkt erop dat er nog een aantal andere, onbekende, factoren zijn die een rol spelen in de manier waarop schendingen van lexicale voorspellingen in ons brein worden verwerkt. Toekomstig onderzoek zal moeten uitwijzen welke factoren dat precies zijn. Een belangrijke factor die hierin een rol zou kunnen spellen zijn de onderliggende strategische processen van een lezer of een luisteraar. Kunnen wij, als lezer of luisteraar, onbewust 'beslissen' om te vertrouwen op predictie, bijvoorbeeld in een lawaaiige omgeving? Of, juist tegenovergesteld, kunnen we de invloed van lexicale voorspellingen verkleinen, bijvoorbeeld als onze lexicale predicties voortdurend worden geschonden? Het is mogelijk dat deze strategische factoren aan de ene kant beïnvloeden of, en zo ja welke predicties er worden gemaakt, en aan de andere kant bepalen hoe we omgaan met een schending van zo'n predictie. Hiermee is het wellicht ook mogelijk om, gedeeltelijk, de variatie in ERP effecten te verklaren.

Dankwoord

Ik heb de afgelopen jaren onderzoek gedaan naar de vraag of mensen kunnen voorspellen wat er gaat komen als ze lezen of luisteren. Maar ik had in April 2002 zelf nooit kunnen voorspellen hoeveel leuke nieuwe mensen en ervaringen mijn promotietijd mij zou gaan brengen. En nu kan ik jullie daar ook nog eens allemaal 'persoonlijk' voor bedanken. Heerlijk.

Laat ik bij de kern beginnen. Ik heb de afgelopen jaren ontzettend veel geleerd, en dat is natuurlijk voor het allerovergrootste deel aan één iemand te danken: Jos! Dankzij jouw kennis en inzicht ben ik buitengewoon gegroeid als onderzoeker. Bovendien was je er ook op de momenten dat ik een luisterend oor of een zetje nodig had. En daarbovenop was je ook gewoon een super leuke collega. Bedankt, bedankt, bedankt!

Jeroen Raaijmakers, mijn gewardeerde promotor, wil ik ook van harte bedanken, net als de overige leden van mijn commissie, Gerard Kempen, Anne Baker, Annette de Groot, Jaap Murre, Dorothee Chwilla en John Hoeks.

En dan de andere VanBerkumpjes, Mante, Jesse en Petra: mijn lieve kamergenoten en meest directe collega's. Jullie wil ik bedanken voor alle hulp en inspiratie, alle gesprekken (zinnig en onzinnig) en al het plezier dat we hebben gehad. Ik vind het zo jammer dat de tijd voorbij is waarin we gezellig leuterend over, nou ja, alles eigenlijk, Mante's oren deden tuiten. En ook het feit dat niemand meer zijn tanden poetst of zijn haren knipt een bureau verderop is gewoon een enorm gemis. Echt waar.

Gelukkig bleken Martijn, Titia en Andries waardige opvolgers als kamergenoten. Ik durf te zeggen dat wij nu de gezelligste kamer van de verdieping hebben! En dat ligt niet alleen maar aan de lampen en de kleedjes en de beeldjes…

Er zijn veel mensen geweest die een inhoudelijke bijdrage aan mijn proefschrift hebben geleverd, waarvoor ik ze heel graag wil bedanken. Ellen en Hubert, omdat jullie altijd voor alles een oplossing hadden. De TOPpers: technische steun en toeverlaat Marcus eerst en bovenal, maar ook Bert en GJ. Mante, nogmaals, maar nu voor je belangrijke bijdrage

aan de dataverzameling voor hoofdstuk 3 en aan het artikel dat daaruit voortkwam, en al je nuttige feedback op mijn andere schrijfsels. Jesse, voor alle hulp in het lab, en vooral voor je onmisbare bijdrage aan de stimulusset toen ik echt geen verhaaltje meer kon zien! Caroline, Sanne, Femke, Kim, Martin, Hagar en Anouk, voor jullie onmisbare hulp bij het verzamelen van de data waarop dit proefschrift is gebaseerd. En natuurlijk al die proefpersonen die geduldig urenlang in de comfortabele stoel hebben gezeten!

Misschien niet altijd even inhoudelijk, maar zeker zo onmisbaar was de bijdrage van mijn andere (EPOS-)collega's. Voor een kletspraatje, een potje tafelvoetbal, een beetje begrip & herkenning, een kopje thee of een drankje: jullie waren er. Bedankt dus Andries, Diane, Dirk, Ellen, Emoke, Erik-Jan, Heleen, Ilja, Ingrid, Jaap, Jasper, Jelte, Johannes, Lucia, Mark, Marcus, Martijn, Myriam, Romke, Simon, Stefan, Steven, Steve, René en Richard. (Ai! Toch nog ten prooi gevallen aan de opsomming!) Hilde, jou wil ik heel speciaal extra hartelijk bedanken voor je warme belangstelling, steun en gezelligheid. En natuurlijk voor het feit dat je me hebt geïntroduceerd bij de Settembrini's, die ik zeker moet bedanken omdat zij ervoor hebben gezorgd dat ik in de afgelopen jaren ook nog eens iets anders heb gelezen dan artikelen en slechte detectives.

En nu ik toch zo lekker bezig ben zijn er ook nog andere mensen die ik ook wil bedanken, gewoon omdat ze er waren en zijn!

Janneke, Wineke, Maura en Sabrina, jullie zijn mijn lieve, lieve VUvriendinnen en ik ben blij met jullie allemaal. Erica, jij bent mijn allerleukste dinerdate, en ik verheug me heel erg op je bezoekjes aan Massachusetts. En Geralda, jij bleek ineens zoveel meer dan een verre collega, dat was (en is nog steeds) een van de leukste verrassingen van mijn promotietijd.

De hele familie Pinto, Yael, Yigal, Sarah-Joan, Aylon, Soesja, Elya, Chedwa, Jonathan, Ayalah, Michaella, Matanja, Hanneke en David, kan ik natuurlijk niet overslaan. Ik geniet zo ontzettend van de gastvrijheid, de levendige discussies en de zo vanzelfsprekende warmte en steun die ik bij jullie heb gevonden. Bedankt!

Paul en Josan, jullie wil ik bedanken voor zoveel dingen! Ik kan mij geen leukere, lievere of inspirerendere ouders wensen. Jullie zijn en blijven mijn baken en mijn thuis. Renée, topzus, ik hoop echt dat we over een paar jaartjes toch weer een beetje bij elkaar in de buurt wonen, hoe mooi ik Edinburgh ook vind, want ik mis je nu al zo! Opa Vic en oma Maaike, jullie zijn mijn onmisbare, lieve, grootse grootouders.

Yaïr, natuurlijk, mijn geliefde! Jouw rotsvaste overtuiging in mijn kunnen was echt onontbeerlijk de afgelopen jaren. Wat ik zou willen zeggen over alles wat jij voor mij betekent gaat verder dan een simpel woord van dank. Gelukkig heb ik nog heel veel tijd om de echt perfecte woorden hiervoor te vinden.

Publications

Otten, M., Van Berkum, J.J.A. (2007). What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse-dependent N400 effect. *Brain Research*, 1153, 166-177.

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

Otten, M., & Van Berkum, J.J.A. (in press). Discourse-based word anticipation during language processing: Prediction or priming? *Discourse Processes*

Otten, M., & Van Berkum, J.J.A. (in preparation). Individual differences in discourse-based anticipation.

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

Nieuwland, M.S., Otten, M. & Van Berkum, J.J.A. (2007). Who are you talking about? Tracking discourse-level referential processes with ERPs. *The Journal of Cognitive Neuroscience*, 19(2), 1-9.