



UvA-DARE (Digital Academic Repository)

Discourse-based lexical anticipation : the nature and contextual basis of predictions in language comprehension

Otten, M.

Publication date
2008

[Link to publication](#)

Citation for published version (APA):

Otten, M. (2008). *Discourse-based lexical anticipation : the nature and contextual basis of predictions in language comprehension*.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.



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 1A¹

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
<p>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.</p> <p><i>Sylvie and Joanna really feel like dancing and flirting tonight. Therefore they go to a stove [disco] where they also make very nice cocktails.</i></p>	<p>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.</p> <p><i>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</i></p>

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.

Pretest: Cloze test. 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.

Pretest: 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 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 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 $\pm 100 \mu\text{V}$, or which featured a linear drift of more than $\pm 40 \mu\text{V}$ 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

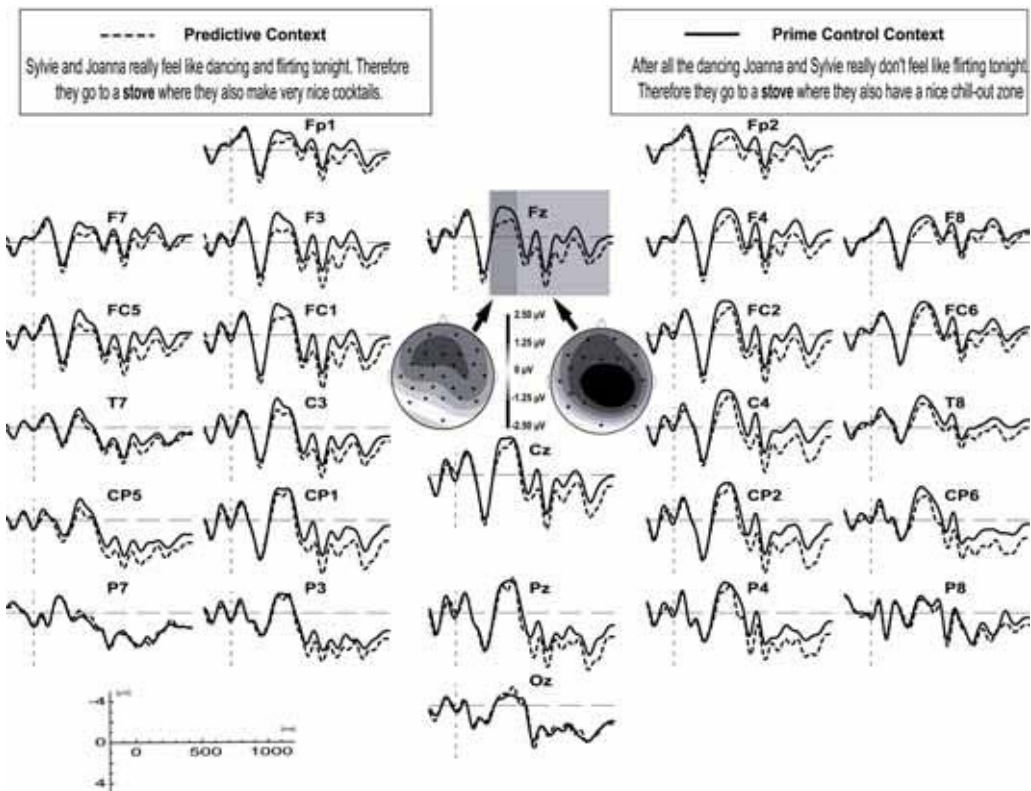


Figure 2.1 Grand average ERPs elicited by anomalous nouns (right panel) in the predictive (dotted line) and the prime control context (solid line). The two windows of analysis (300-500 ms and 500-1200 ms) are highlighted. The scalp distributions corresponding to the effect of discourse type (predictive 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

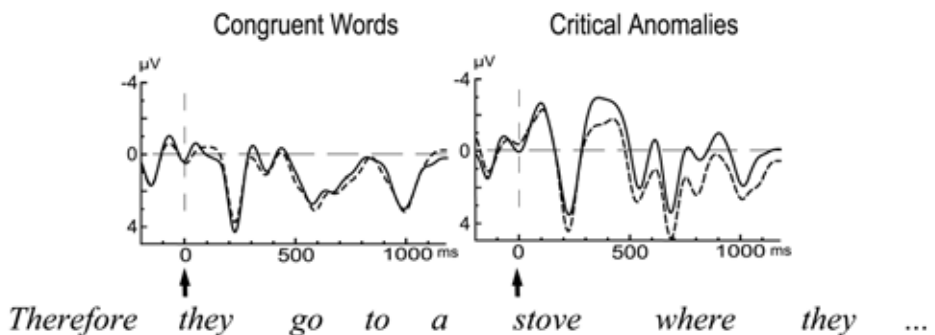


Figure 2.2 Grand average ERPs elicited by anomalous nouns (right panel) and by congruent words preceding the anomaly (left panel) in the predictive (dotted line) and the prime control context (solid line) for electrode Fz. The differential effect observed at the anomalies is not present in the words 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 suggest 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
<p>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.</p> <p><i>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.</i></p>	<p>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.</p> <p><i>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.</i></p>

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.

Pre-test: Cloze test. 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

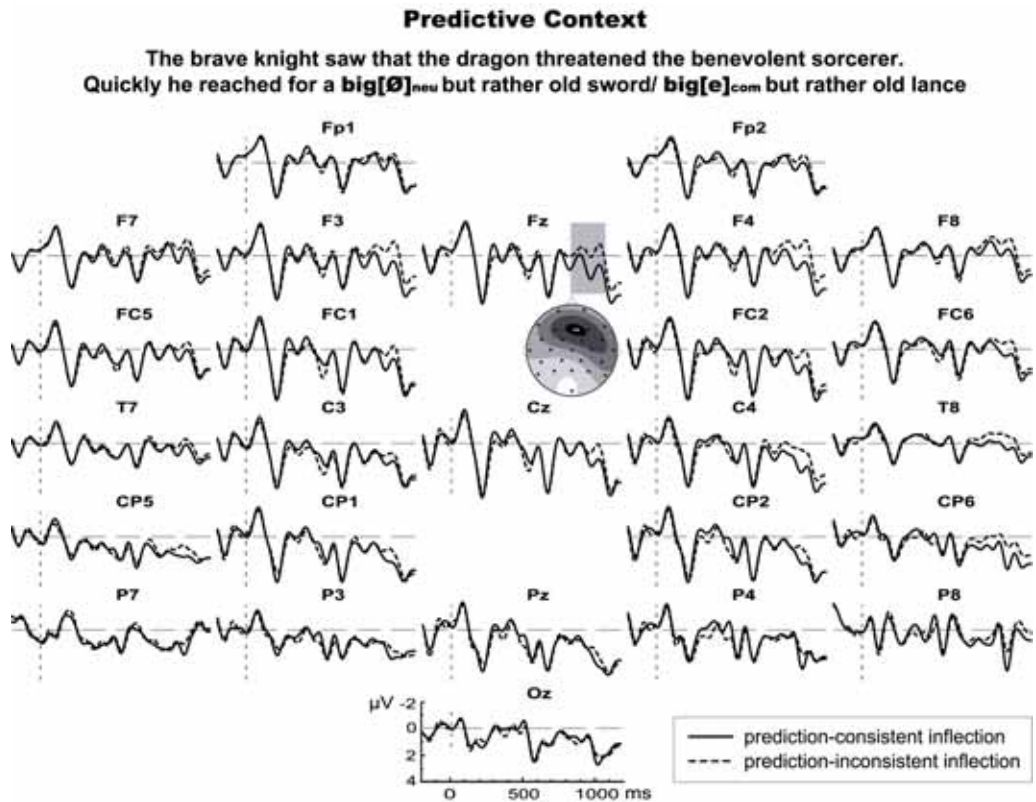


Figure 2.3 Grand average ERPs elicited by the prediction-consistent (solid line) and prediction-inconsistent (dotted line) adjectives in the predictive condition. The time window where prediction inconsistent adjectives show a significant difference in the predictive context (900 - 1200 ms) is highlighted.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

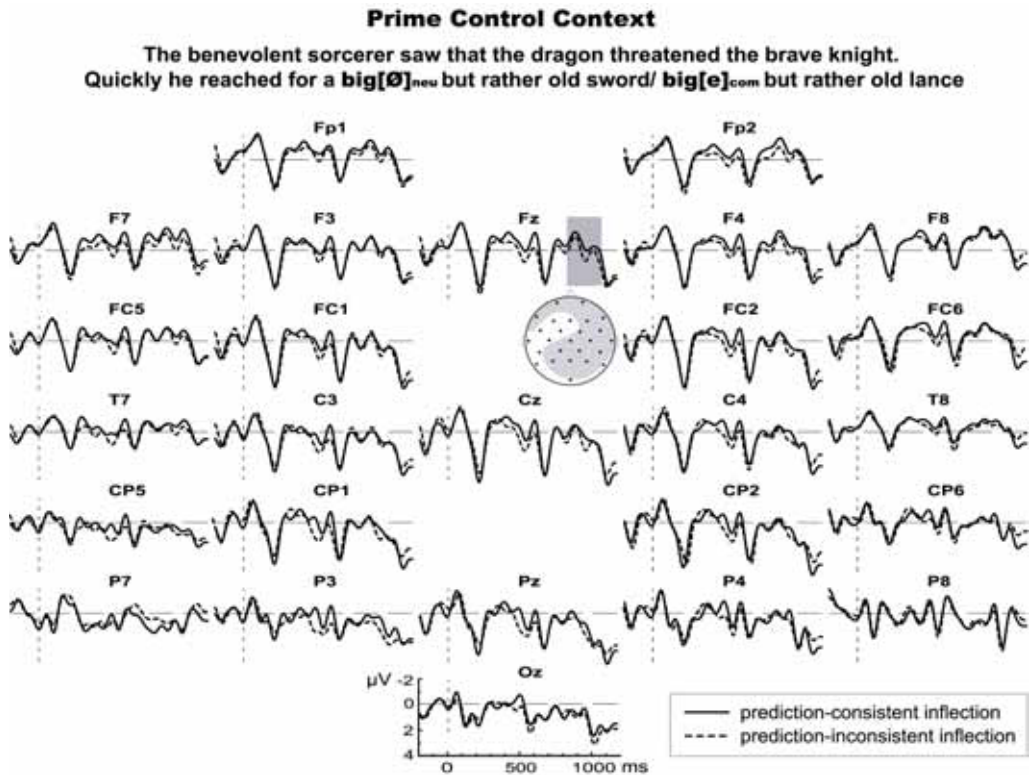


Figure 2.4 Grand average ERPs elicited by the prediction-consistent (solid line) and prediction-inconsistent (dotted line) adjectives in the prime control condition. The time window where prediction inconsistent adjectives show a significant difference in the predictive context (900 - 1200 ms) is highlighted. 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 ($F(1,23) =$

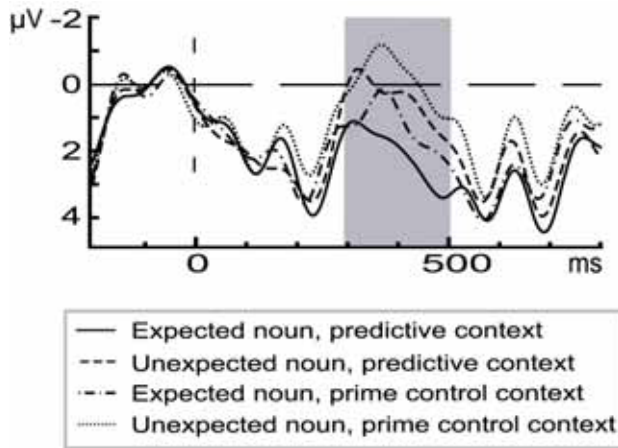


Figure 2.5 Grand average ERPs elicited by the expected and unexpected nouns in the predictive and prime control context for electrode Pz. Unexpected nouns evoke a larger N400 between 300 and 500 ms in both the predictive and the prime control context.

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

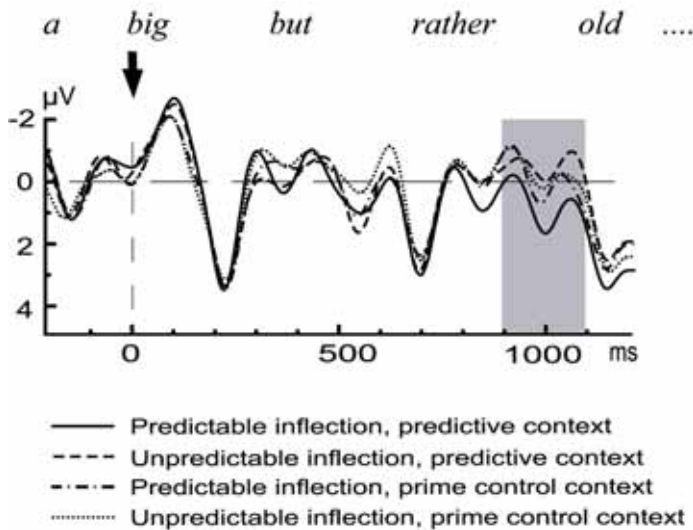


Figure 2.6 Grand average ERPs elicited by the prediction-consistent and prediction-inconsistent adjectives in the predictive and prime control context for electrode Fz. Prediction-inconsistent adjectives elicit a late positive deflection between 900-1100 ms, in the predictive context only.

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.