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

Otten, M.

Publication date
2008

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.*	*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**$_{com}$ delicate yet striking necklace that had been selected by her stylist.	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.*	*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**$_{com}$ delicate yet striking *necklace* that had been selected by her stylist.	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

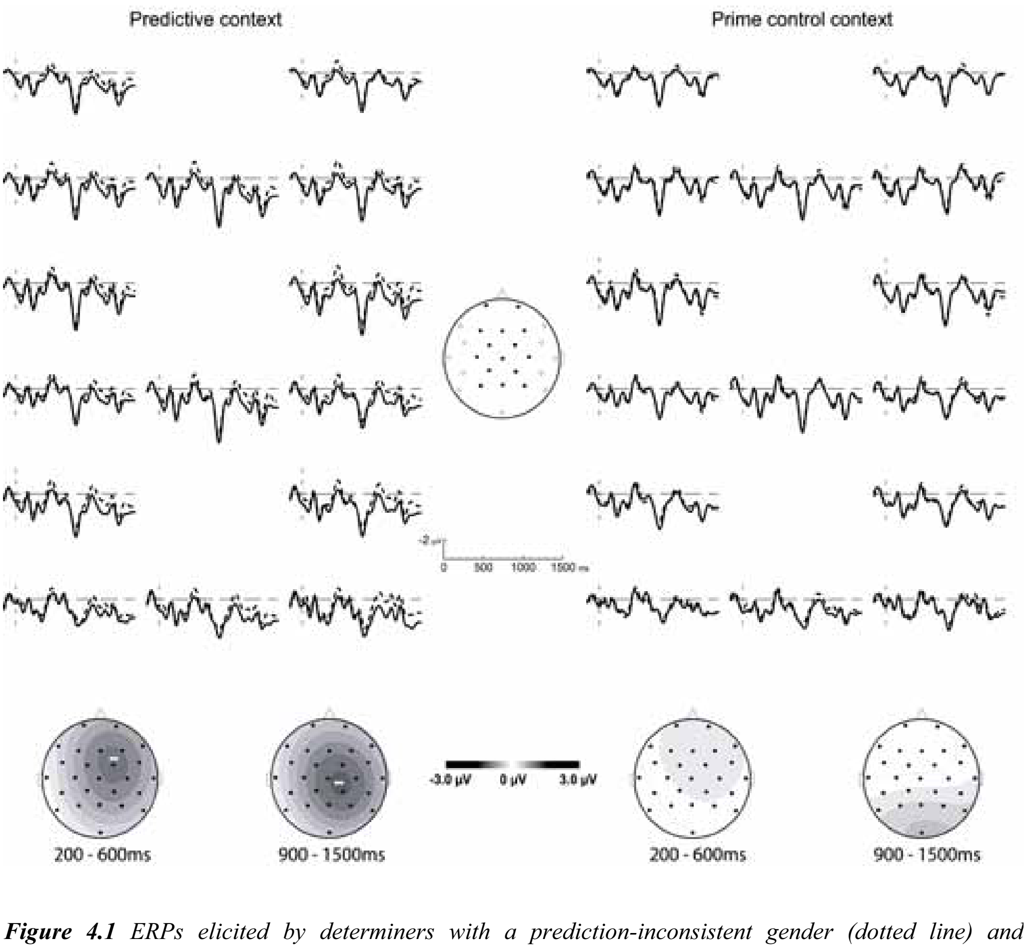

Figure 4.1 *ERPs elicited by determiners with a prediction-inconsistent gender (dotted line) and prediction-consistent gender (solid line) for both high and low WMC readers. The left-hand panel shows the ERPs for the determiners in the highly constraining predictive discourse, the left-hand panel shows 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

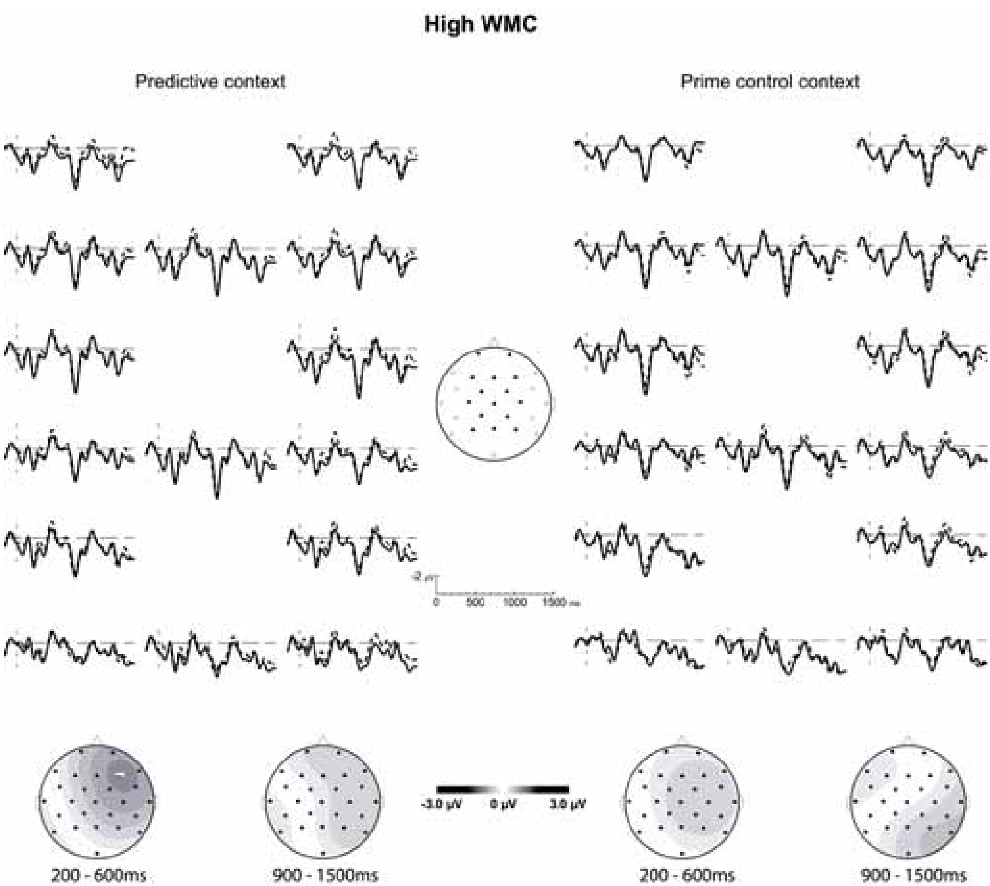

***Figure 4.3** ERPs elicited by determiners with a prediction-inconsistent gender (dotted line) and prediction-consistent gender (solid line) in the predictive (left-hand panel) and prime control discourse (right-hand panel) for high WMC readers only.*

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

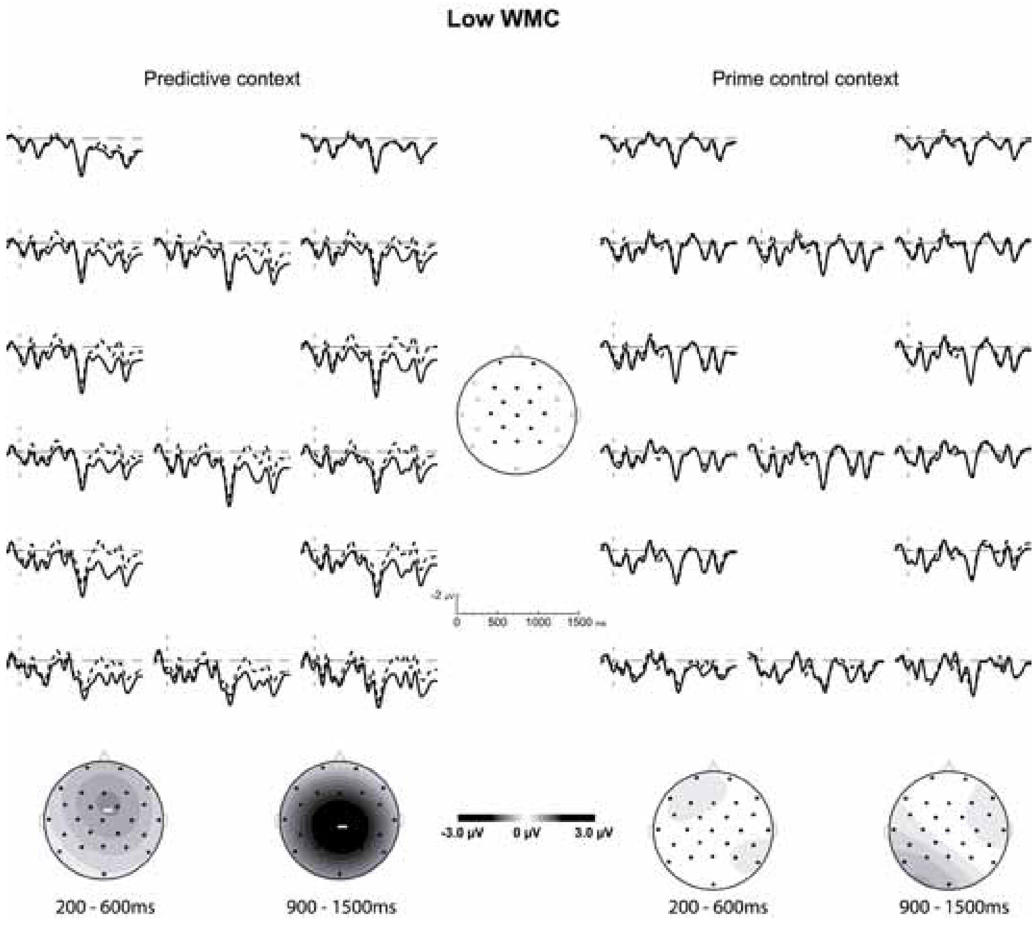

Figure 4.2 *ERPs elicited by determiners with a prediction-inconsistent gender (dotted line) and prediction-consistent gender (solid line) in the predictive (left-hand panel) and prime control discourse (right-hand panel) for low WMC readers only.*

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:

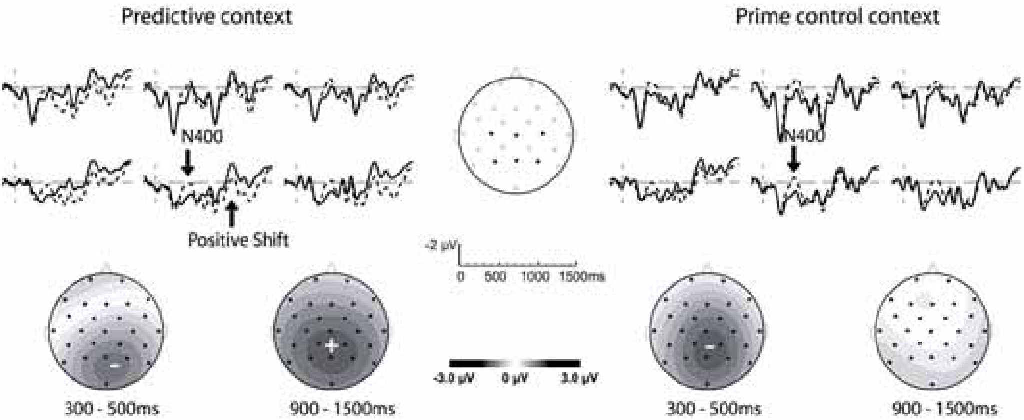

Figure 4.4 *ERPs elicited by unexpected (dotted line) and expected (solid line) nouns in the predictive (left-hand panel) and prime control discourse (right-hand panel) for high and low WMC readers.*

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).

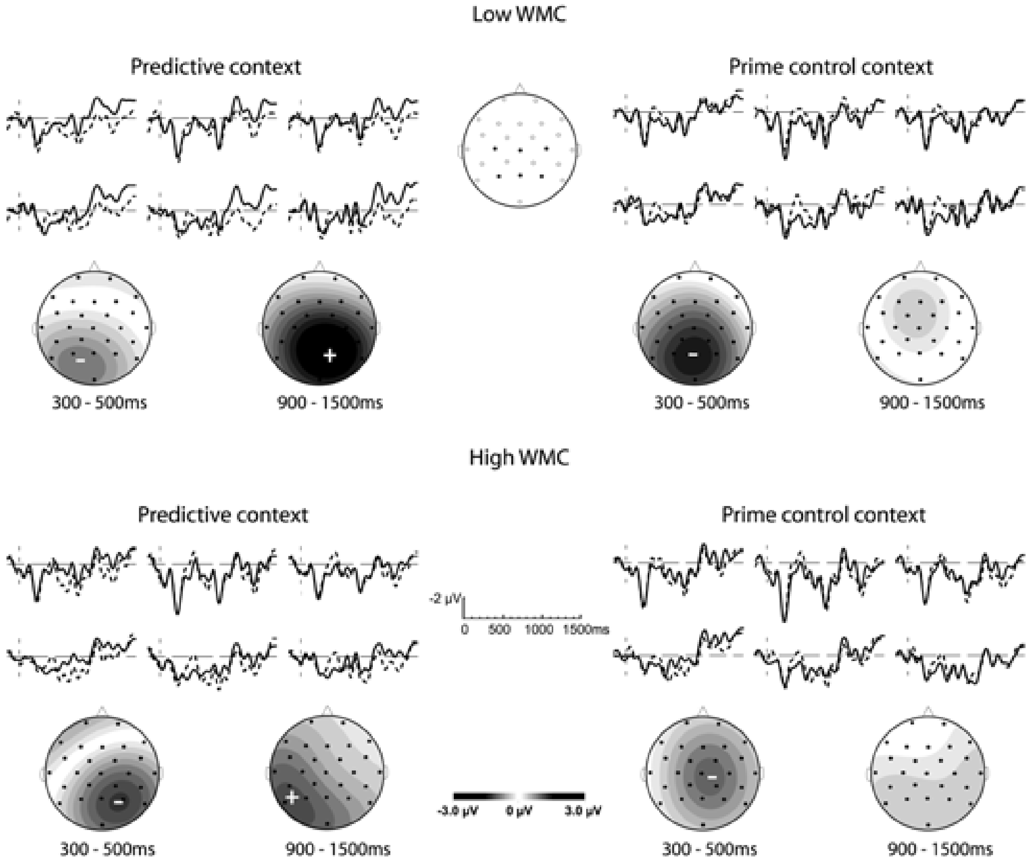

***Figure 4.5** ERPs elicited by unexpected (dotted line) and expected (solid line) nouns in the predictive (left-hand panel) and prime control discourse (right-hand panel) separated for high and low WMC readers.*

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.