available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

Does working memory capacity affect the ability to predict upcoming words in discourse?

Marte Otten[a,b,*], *Jos J.A. Van Berkum*[a,c,d]

[a]*Department of Psychology, University of Amsterdam, The Netherlands*
[b]*Department of Psychology, Harvard University, USA*
[c]*Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*
[d]*FC Donders Centre for Cognitive Neuroimaging, Nijmegen, The Netherlands*

ARTICLE INFO

Article history:
Accepted 15 July 2009
Available online 23 July 2009

Keywords:
EEG
ERP
Language
Prediction
Working memory capacity
Reading span

ABSTRACT

Prior research has indicated that readers and listeners can use information in the prior discourse to rapidly predict specific upcoming words, as the text is unfolding. Here we used event-related potentials to explore whether the ability to make rapid online predictions depends on a reader's working memory capacity (WMC). Readers with low WMC were hypothesized to differ from high WMC readers either in their overall capability to make predictions (because of their lack of cognitive resources). High and low WMC participants read highly constraining stories that supported the prediction of a specific noun, mixed with coherent but essentially unpredictive 'prime control' control stories that contained the same content words as the predictive stories. To test whether readers were anticipating upcoming words, critical nouns were preceded by a determiner whose gender agreed or disagreed with the gender of the expected noun. In predictive stories, both high and low WMC readers displayed an early negative deflection (300–600 ms) to unexpected determiners, which was not present in prime control stories. Only the low WMC participants displayed an additional later negativity (900–1500 ms) to unexpected determiners. This pattern of results suggests that WMC does not influence the ability to anticipate upcoming words per se, but does change the way in which readers deal with information that disconfirms the generated prediction.

1. Introduction

The average speaker produces between 150 and 190 words per minute (Marslen-Wilson, 1973; Riding and Vincent, 1980), which thus need to be processed at the same speed by the average listener. In our everyday life we have all experienced the seemingly effortless speed with which we can understand language. It has recently been suggested that one of the factors that makes our language comprehension system so effective and fast is our ability to anticipate what somebody might say next, and in what way (Levinson, 2000; Pickering and Garrod, 2007; Van Berkum, in press).

There is ample evidence that readers and listeners can anticipate grammatical (Lau et al., 2006) as well as conceptual structure (Federmeier, 2007; Kamide et al., 2003; Keefe and McDaniel, 1993) and that they can use both to predict specific upcoming words (DeLong et al., 2005; Van Berkum et al.,,2005; Wicha et al., 2004). In this paper we explore whether the ability

* *Corresponding author*. Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02141, USA. Fax: +1 617 384 9517.
E-mail address: motten@wjh.harvard.edu (M. Otten).

0006-8993/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2009.07.042

to anticipate upcoming language is related to the ability to temporarily store and manipulate information, also known as the working memory capacity (WMC) of a reader. Using event-related potentials (ERPs) we ask whether individuals with a less extensive WMC are equally capable of predicting specific upcoming words as high WMC readers.

Differences in WMC are often ascribed to differences in the availability of a central computational resource – the mental workspace – needed for language comprehension and all other non-automatic cognitive processes (Daneman and Carpenter, 1980, 1983; Just and Carpenter, 2002; Just et al., 1996). If the processes involved in the anticipation of upcoming language recruit the abovementioned central computational resource, the reduced ability to temporarily store and manipulate information would lead low WMC readers to be less able – possibly even unable – to adequately anticipate upcoming words in text, relative to high WMC readers. An indication that specific linguistic anticipations could be resource-demanding comes from the study of predictive inferences. Predictive inferences are optional, elaborative inferences about predictable events. For example, when people read that an actress has fallen from a 14 story building, they usually infer she has died (McKoon and Ratcliff, 1986; O'Brien et al., 1988). Several studies (Calvo, 2001; Estevez and Calvo, 2000; St George et al., 1997) have shown that whereas high WMC readers are able to take advantage of these predictive inferences, low WMC readers are not. Because a predictive inference is essentially an extrapolation of the conceptual situation model laid out in the discourse so far, it will usually be a necessary ingredient of discourse-based lexical anticipation (see Otten and Van Berkum, 2008 for a discussion). As such, the predictive inference literature suggests that low WMC individuals will be less likely to anticipate specific upcoming words.

To explore the influence of WMC on specific lexical prediction we use an experimental paradigm (Otten et al., 2007) that allows us to probe the prediction of upcoming nouns in unfolding text. To test for the pre-activation of a predictable word, we focus not on the predicted word itself but on the 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}].

Story 1 in Table 1 illustrates the way the constraining stories were set up: the critical noun *necklace* in the second sentence is expected because of the constraints set in the broader discourse (i.e. the first sentence) of the story, not just on the local sentence. If listeners strongly anticipate a specific noun like *necklace*, a common gender noun in Dutch, the determiner that indicates neuter gender (*het*) will come as an 'unpleasant' surprise compared to the prediction-consistent determiner (*de*). Several experiments have shown that adjectives and determiners whose morphological (or phonological) form does not agree with features of the expected head noun evoke a differential ERP effect, even though at this point in the sentence (with a noun yet to be presented) these prenominal elements are grammatically fully correct. For example, when participants hear a determiner with an unpredicted gender (Otten and Van Berkum, 2007) their ERPs show an enhanced frontal negative component between 300 and 600 ms compared to the determiner with a predictable gender. Effects like these show that the reader must 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 and Van Berkum, 2008; Van Berkum et al., 2005), gender-marked determiners in Spanish (Wicha, Bates et al., 2003; Wicha, Moreno et al., 2003; Wicha et al., 2004) and the a/an phonological in English (DeLong et al., 2005).

To test whether the overall ability to anticipate specific words varies with WMC, we examined the neural response to prediction-inconsistent determiners for individuals with low and high WMC. If both groups anticipate specific upcoming words to an equal extent, both should show the same ERP response to prediction-inconsistent determiners, relative to prediction-consistent determiners. If, however, readers with a low WMC are less able to anticipate specific upcoming words as the discourse unfolds, then these individuals should be less

Table 1 – One example of the stimulus materials used in this experiment in its four different versions.

1. Predictive discourse	
Prediction-consistent determiner	Prediction-inconsistent determiner
De actrice had een prachtige jurk aan, maar ze vond haar hals nog wat sober. Ze pakte de verfijnde maar toch opvallende ketting die haar stylist had uitgezocht. The actress wore a beautiful dress, but she thought her neck was a little plain. She picked up the$_{com}$ delicate yet striking necklace that had been selected by her stylist.	*De actrice had een prachtige jurk aan, maar ze vond haar hals nog wat sober. Ze pakte het verfijnde maar toch opvallende collier dat haar stylist had uitgezocht.* The actress wore a beautiful dress, but she thought her neck was a little plain. She picked up the$_{neut}$ delicate yet striking collar that had been selected by her stylist.
2. Prime control discourse	
Prediction-consistent determiner	Prediction-consistent determiner
De actrice vond dat haar hals goed uitkwam in de sobere jurk. Ze pakte de verfijnde maar toch opvallende ketting die haar stylist had uitgezocht. The actress thought her neck looked beautiful in the plain dress. She picked up the$_{com}$ delicate yet striking necklace that had been selected by her stylist.	*De actrice vond dat haar hals goed uitkwam in de sobere jurk. Ze pakte het verfijnde maar toch opvallende collier dat haar stylist had uitgezocht.* The actress thought her neck looked beautiful in the plain dress. She picked up the$_{neut}$ delicate yet striking collar that had been selected by her stylist.

responsive to prediction-inconsistent determiners than high WMC readers, and as such display a smaller – or possibly absent – ERP effect to these determiners.

Predictive stories such as our example 1 in Table 1 often contain one or more words that are mildly or strongly related to the most predictable continuation of the story at hand (i.e. "nek *(neck)*" to "ketting *(necklace)*"). This opens up the possibility that any observed pre-activation of a specific upcoming word is not based on the precise message of the story so far, but instead relies on the presence of one or more prime words. In two studies with the same materials and with participants randomly sampled from the general college population (Otten et al., 2007; Otten and Van Berkum, 2008) we specifically controlled for this possibility, and there we obtained no evidence for a word-word priming confound. To make sure that this also holds for the specific WMC subgroups tested here, the current experiment again included a set of control materials dedicated to ruling out a word-word prime confound.

As illustrated in example 2 in Table 1, these so-called prime control stories contained the same content words as their constraining counterparts, but their arrangement was changed such that the precise message no longer strongly supported the prediction of a single specific upcoming word (e.g., no strong message-level bias towards "ketting *(necklace)*"). The logic was that, whereas any determiner-elicited ERP effects observed in *both* prime control and predictive stories might be the result of low-level word–word priming, ERP effects *uniquely* elicited by prediction-inconsistent determiners in predictive stories must hinge on precise message-based anticipation. This logic not only allows us to test for uniquely message-dependent anticipation effects in each WMC group, but to also detect differences – if any – in the degree to which readers with low and high reading span can use a precise message-level representation to predict upcoming words.

2. Results

Fig. 1 shows the ERPs evoked by expected and unexpected determiners in a predictive and prime control context, averaged over all participants. In predictive stories (left), unexpected determiners elicited a negative deflection in the ERP between 200 and 600 ms ($F(1,29)=5.14$, $p=.03$), which is strongest over the right hemisphere ($F(1,29)=4.56$, $p=.04$). The absence of an interaction with the anterior–posterior factor ($F(1,29)=.158$, $p=.69$), suggests that unlike the N400 this effect of expectancy does not have a posterior maximum. This early negativity is followed by another negative shift between 900 and 1500 ms ($F(1,29)=7.62$, $p=.01$). As the scalp distribution in Fig. 1 suggests, this late effect is not lateralized as the early negativity, but more centrally distributed, which is confirmed by an absence of interaction with hemisphere ($F(1,29)=1.47$,

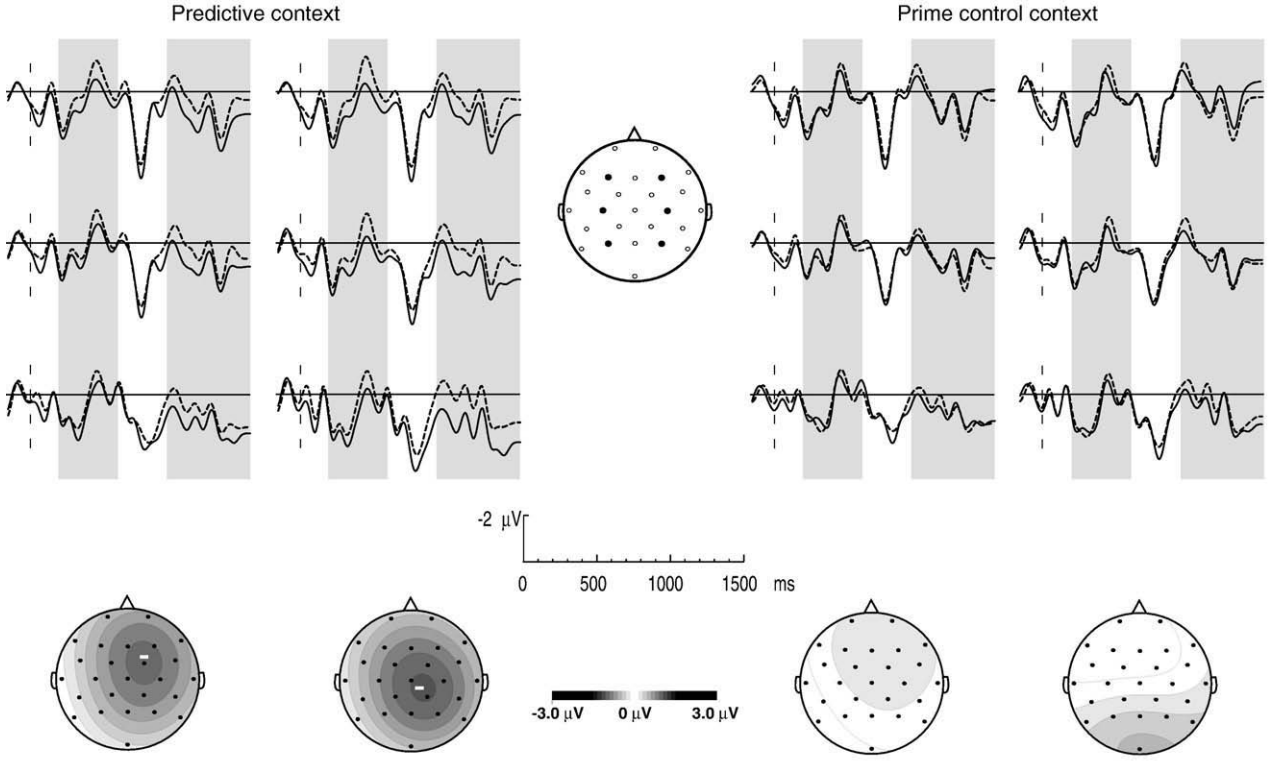

Fig. 1 – Effect of prediction-inconsistency collapsed over WMC groups: ERPs elicited by determiners with a prediction-inconsistent gender (dotted line) and prediction-consistent gender (solid line) for high and low WMC readers combined. The left-hand panel shows the ERPs for the determiners in the highly constraining predictive discourse, the right-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 analyzed.

$p = .24$). A trending interaction with anterior–posterior suggests a slight central-posterior distribution ($F(1,29) = 3.78$, $p = .06$).

The presence of these differential effects suggests that our participants indeed anticipated specific upcoming words. Furthermore, the right panel of Fig. 1 shows no evidence for any predictions in the prime control stories, suggesting that for the average participant, the observed prediction effects cannot be reduced to automatic activation by content words in the discourse. Thus, in line with our earlier work, the current findings suggest that the average participant uses a message-level representation of the prior discourse to anticipate specific upcoming words.

Figs. 2 and 3 show the ERPs evoked by expected and unpredicted determiners for low and high WMC readers separately. Both low and high WMC readers display the early negative shift in the predictive stories, and not in the prime control stories (Expectancy*Context: $F(1,29) = 4.28$, $p = .05$). This suggests that for the materials at hand, *both* groups generally make online lexical predictions. Furthermore, *both* groups use a precise message-level representation of the prior discourse to anticipate upcoming words. The scalp distributions in Figs. 2 and 3 show that the effect of expectancy is strongly right lateralized in the low WMC group, whereas in the high WMC group the effect is more centrally distributed (Expectancy*Hemisphere*WMC: $F(1,29) = 4.65$, $p = .04$).

In contrast to the early effect of expectancy, the late negativity is only displayed by low WMC participants and only in the predictive stories (Expectancy*Context*WMC: $F(1,29) = 5.37$, $p = .03$), indicating that WMC does influence the way in which readers process information that is not in line with their expectations.

3. Discussion

In constraining stories supporting the prediction of a specific noun, both high and low WMC readers show an early ERP effect to prediction-inconsistent determiners relative to consistent determiners. The fact that the unexpected gender of the determiner elicits this differential response, before any overtly unexpected noun has been shown, clearly shows that the readers have predicted and pre-activated a specific noun, as well as its lexical gender. Contrary to our expectations, our results show that even readers with a relatively low WMC routinely use their knowledge of the prior discourse to anticipate specific upcoming words as a story is unfolding.

In addition, in the less constraining stories that control for the presence of individual content words, the previously unpredicted determiner now evokes no differential processing. In the prime control condition, the expectancy effect is absent in both the low WMC and the high WMC group. This suggests that the linguistic anticipation observed here critically relies on a true message-level representation of the discourse, and is not rooted in simpler word–word activation.

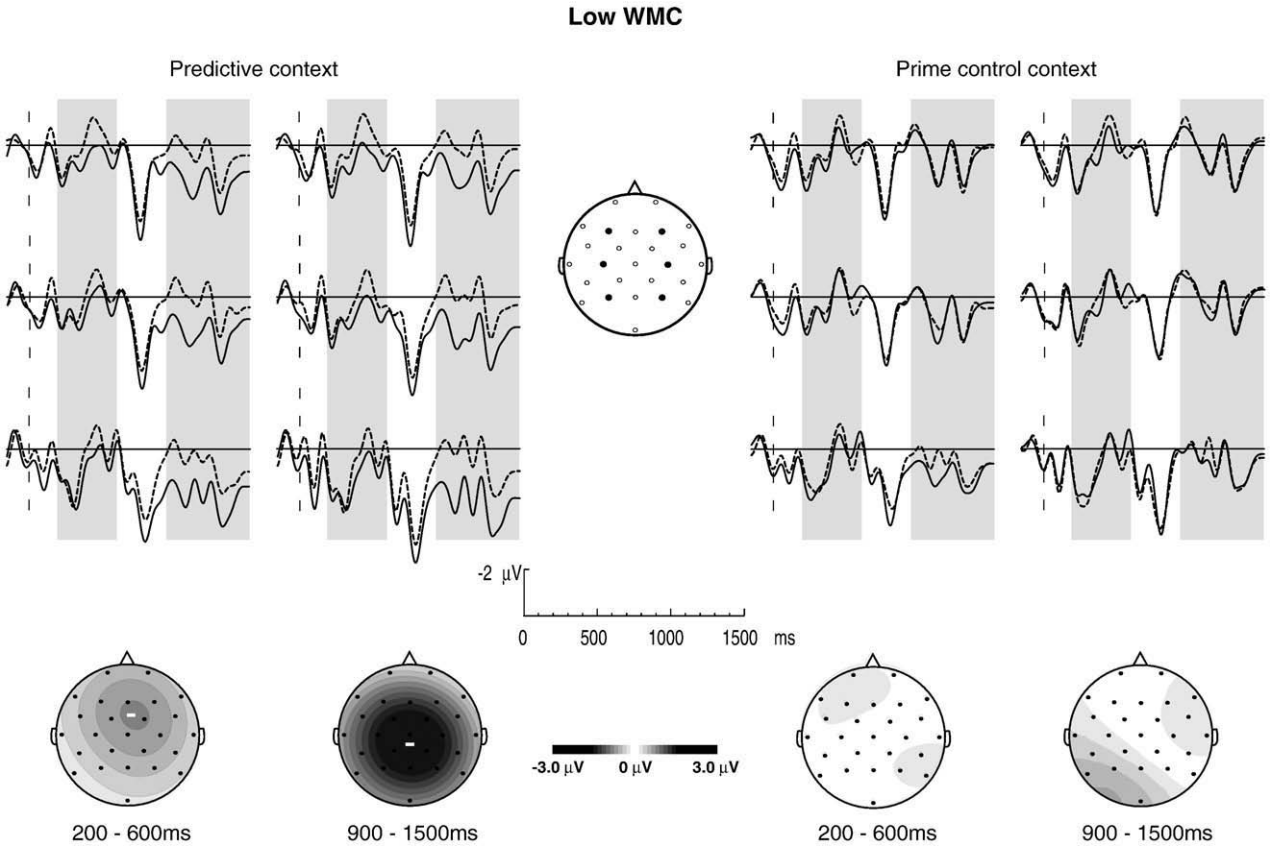

Fig. 2 – Effect of prediction-inconsistency for low WMC readers: ERPs elicited by determiners with a prediction-inconsistent gender (dotted line) and prediction-consistent gender (solid line) in the predictive (left-hand panel) and prime control discourse (right-hand panel) for low WMC readers only.

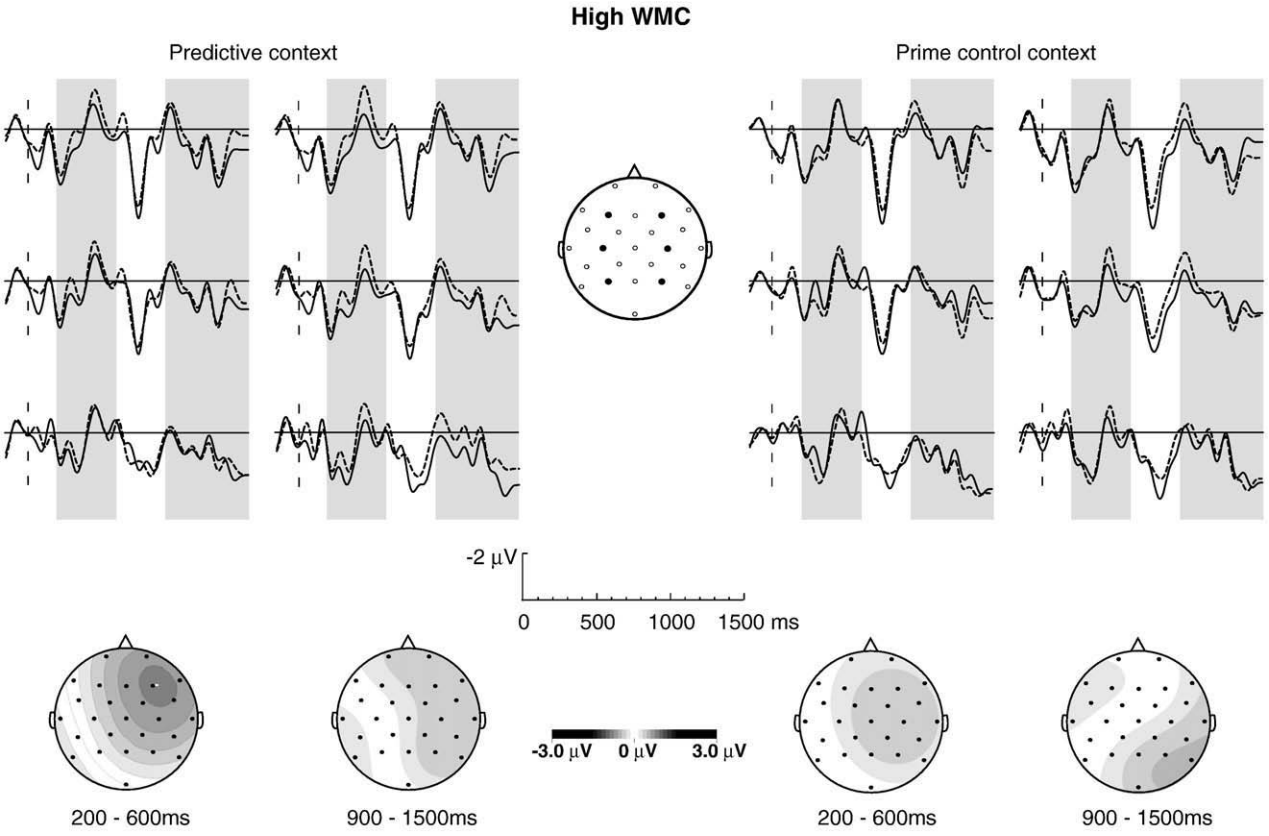

Fig. 3 – Effect of prediction-inconsistency for high WMC readers: ERPs elicited by determiners with a prediction-inconsistent gender (dotted line) and prediction-consistent gender (solid line) in the predictive (left-hand panel) and prime control discourse (right-hand panel) for high WMC readers only.

The ERPs also reveal a noticeable difference between the two WMC groups. Whereas both groups show evidence of discourse-based anticipation, only low WMC readers show the additional late effect of prediction-inconsistency, emerging at about 1 s after the mismatching determiner (but well before the earliest nouns are shown). This suggests that although both low and high WMC readers in the WMC range tested here create online lexical predictions, low and high WMC readers differ in their processing of information that is not in line with these predictions.

3.1. *The early prediction-dependent ERP effect*

In the present study, prediction-inconsistent determiners evoke a right frontal negativity between 200 and 600 ms. Previous studies from our own and other labs have also reported an early negative deflection in this latency range after the onset of the prediction-inconsistent information (determiners with prediction-inconsistent features in English (DeLong et al., 2005) and Spanish (Wicha, Bates et al., 2003; Wicha, Moreno et al., 2003) and adjectives with prediction-inconsistent inflections in Dutch (Otten et al., 2007)). The early negative deflection observed here is similar in both time course and polarity to the deflections reported in these earlier studies. The ERP effect in three of these studies has a slightly to strongly right-lateralized scalp distribution, comparable to the present effect (Otten et al., 2007; Wicha, Bates et al., 2003; Wicha, Moreno et al., 2003), similar to the right-lateralized negativity observed in this study. This scalp distribution does not resemble the scalp distribution of the standard N400 effect, and therefore we are reluctant to interpret the present effect as a canonical N400 effect. At the same time, though, the timing and polarity of the ERP effect and the critical involvement of high-level meaning are consistent with the idea that at least some of the neural generators that underlie the canonical N400 effect might also be activated when people hear something that (indirectly) mismatches their prediction.

Next to five studies revealing an early negativity, three other studies with very similar methodology reported rather different ERP effects to prediction-inconsistent information (i.e. an early positive shift (Van Berkum et al., 2005), a late positive shift (Wicha et al., 2004), and a late negative shift (Otten and Van Berkum, 2008)). A systematic inventarization across all studies shows that this variability cannot be accounted for by differences in language, stimulus modality, type of prediction probe, or differences in working memory capacity of participants. One possibility is that perhaps the broader context in which stimuli are presented (i.e. the type of filler that is used, the length of the experiment) matters more than commonly assumed, but we refrain from speculating about specific other factors that could critically influence the way people make predictions, or process prediction-inconsistent data. Ultimately, the variability in prediction-related ERP effects will need to be explained. However, the observation of a differential ERP effect by itself unequivocally reveals that people use the discourse to anticipate upcoming words.

As in our two earlier ERP experiments with a word–word prime control condition (Otten et al., 2007; Otten and Van Berkum, 2008), the current study reveals that the discourse-based lexical predictions observed in this gender-dependent paradigm depend on the precise message developed in the preceding discourse, and not on lower-level word–word priming. This finding converges with earlier evidence that a coherent discourse overrules automatic prime-based activation (Van Petten, 1993; see also Camblin et al., 2007; Hoeks et al., 2004; and Morris, 1994). We want to stress that the stories used in the present experiment were not deliberately designed to systematically vary lexical priming, or to include *strong* associative or semantic prime words. Therefore, the absence of a differential effect in prime control stories does not provide compelling evidence that additional word-based priming plays no role at all in the prediction of upcoming words. What it does show is that the anticipation observed in the present experiment cannot be completely reduced to word–word associative activation.

3.2. WMC and prediction

Our results reveal that when reading highly constraining stories, both low and high WMC readers predict specific upcoming words. Moreover, both types of readers arrive at this prediction not through a simple automatic word-based activation process, but by processing the complete meaning of the discourse. Even low WMC readers, who are typically believed to have less computational resources available (Daneman and Carpenter, 1980, 1983; Just and Carpenter, 2002; Just et al., 1996) can rapidly anticipate upcoming information in a fairly sophisticated way. In fact, at least with the materials tested here, there is no evidence in our data that they do so less, or less effectively, than their high WMC counterparts.

In Dutch, the mapping between determiners and nouns is not simply one-to-one, but depends on gender, number, and definiteness. Dutch determiner–noun agreement checking therefore requires some form of syntactic parsing. The implication is that in our paradigm, ERP effects to gender-mismatching prenominal determiners also reveal that these anticipated words can begin to participate in agreement-sensitive parsing operations (see Van Berkum et al., 2005, for a similar argument concerning gender-marked adjectives). Interestingly, and perhaps somewhat unexpectedly, our findings therefore also show that such anticipatory parsing is within the reach of both high and low WMC readers.

The observation that low and high WMC readers use a precise message-level representation of the prior discourse to anticipate upcoming words suggests that message-based anticipation is not necessarily a very resource-demanding affair, at least not to the extent that readers with lower WMC in the range tested here cannot handle it. Because we recruited university students for our study, the average WMC of our low WMC participants is most likely not the lowest reading span one might obtain when sampling from the general population. We cannot exclude that the discourse-based anticipation of specific words would break down in readers with yet lower WMC scores. However, most other studies, including those that report differences between low and high WMC readers recruit from a comparably restricted population, and our findings clearly contrast with the latter.

Particularly relevant here are studies that show that predictive inferences – elaborative inferences about predictable events – are only regularly created by people with a high WMC (Calvo, 2001; Estevez and Calvo, 2000; St George et al., 1997). It is very likely that the anticipation of specific upcoming words in our materials partly hinges on such inferences. Why, then, do the low WMC readers in our study anticipate upcoming words as easily as high span readers? Note that in the abovementioned predictive inference studies, either a blank time-interval or mildly related text separated the inference-inducing text from the probe. Thus, readers had to keep the predictive inferences active over time for it to be of any use. In a study that resembled our own with regard to the type and timing of the probe that was used to detect whether people had made a predictive inference (Linderholm, 2002) both low and high WMC readers were able to make and use predictive inferences. The overall pattern of results across studies could thus be taken to suggest that, whereas both low and high WMC comprehenders are in principle able to rapidly extrapolate the situation model (predictive inferences) and anticipate specific upcoming words, only high WMC readers are able to keep their predictions active for a longer period of time.

3.3. WMC and processing prediction-inconsistent information

The early ERP effect between 200 and 600 ms shows that high and low WMC readers make comparable discourse-based nominal predictions, and initially detect disconfirming gender information in the same way. However, the data also show a noticeable difference between the two types of readers: the ERPs of low WMC readers contain an additional neural response to prediction-inconsistent determiners. This later negativity between 900 and 1500 ms suggests that, beyond initial detection, information that is inconsistent with a prediction elicits some form of additional processing for the low but not for the high WMC readers.

The literature on working memory suggests several potential accounts for this additional late activity in low WMC readers. One is that it reflects the increased demands of adjusting and/or suppressing the original prediction. There is quite some evidence that low WMC readers have more difficulty with suppressing unwanted information (Engle, 2002), both in the linguistic domain (Gernsbacher and Faust, 1995; Gernsbacher and Faust, 1991; Gernsbacher and Robertson, 1999) and in other domains of cognitive functioning (Rosen and Engle, 1998). The late negativity that follows prediction-inconsistent determiners in low WMC readers could thus be related to their (larger) effort to adjust or suppress their initial prediction. Reprocessing of an earlier linguistic interpretation is often related to late positive shifts in the ERP (Kuperberg, 2007; Van Herten et al., 2005), but recent work on discourse model recomputation (Baggio et al., 2008) suggests that the suppression of an earlier discourse-model inference can show up in negative shifts as well.

A second possibility is that the late negativity observed for low WMC readers reflects an increased demand on working memory. Higher working memory load is generally associated with sustained negative shifts in the ERP, for simple memory tasks (Ruchkin et al., 1990) as well as taxing syntactic struc-

tures (Fiebach et al., 2001; Gunter et al., 1995; King and Kutas, 1995). Furthermore, in a study that directly manipulated linguistic memory load in low as well as high WMC participants (Vos et al., 2001), the load-related sustained negative shift was much more pronounced for low than for high WMC individuals. The reason why such a higher memory load might selectively arise in low WMC participants could be related to the fact that they are less able to suppress their original prediction: they may temporarily have to deal with two parallel analyses, one involving the anticipated noun, and one without it. In this account, the additional late negative shift observed for low WMC readers would reflect the consequences of a genuine increase in working memory load. However, it is also possible that upon encountering prediction-inconsistent information, both groups of readers face the same objective increase in memory load. In line with Vos et al. (2001) such an objectively comparable situation can also lead to an increased sustained negative shift in the ERP of low WMC individuals.

Our current findings do not allow us to unequivocally determine the functional generator(s) of the determiner-induced late negativity in low WMC readers. However, in the face of the earlier negativity, what we do know is that the late negative shift is unlikely to reflect differences in the degree to which low and high WMC readers anticipate upcoming words and/or detect a mismatching determiner.

3.4. *Conclusion*

The results of this experiment confirm that in a constraining discourse, readers can generate highly specific message-based predictions, as the text unfolds. Furthermore, strongly anticipated nouns enter into some form of anticipatory parsing, at least to the extent that determiner-noun agreement violations can be detected. Strikingly, these anticipatory abilities are not limited to readers with a high working memory capacity only. Our finding suggests that specific message-based predictions are generated and tested relatively automatically, without the necessity to allocate many additional mental resources. Interestingly, when confronted with prediction-inconsistent information, readers with a low working memory capacity show an additional ERP response that is not present in the high capacity group. This effect might be linked to increased working memory load while trying to resolve the inconsistency between the internally generated lexical prediction and the actually perceived sentence, or to a general inability to suppress the original prediction. Taken together, our findings suggest that differences in working memory capacity do not influence the ability to anticipate upcoming words per se, but do change the way in which readers deal with information that disconfirms the generated prediction.

4. Experimental procedures

4.1. *Participants*

Participants were 38 right-handed native speakers of Dutch, selected from our subject pool on the basis of their WMC (19 high, 19 low). They were paid 25 euro or awarded course credit. We excluded 1 participant from the analyses because of technical problems, and 6 participants because of >50% critical trial loss due to EEG artifacts (see below for details). The remaining 31 participants (21 female participants) were on average 20 years old (range 18–25 years).

4.2. *Reading Span Task*

Participant selection depended on their score on the Reading Span task originally designed by Daneman and Carpenter (1983). A computerized Dutch version of the Reading Span Task (Van den Noort et al., 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 and pressed the space bar triggering the onset of the next sentence after finishing a sentence. If the subject could not finish the sentence in 6.5 s, the next sentence was automatically presented. After reading all the sentences of a set, a recall-cue indicated that the participant had to recall the final words of each sentence in that set. Responses were registered by the experimenter. 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).

4.3. *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 content words but was *not* predictive at the message level. See the Appendix A for a sample of the stimulus materials that illustrates the different ways of adapting predictive stories so that they became less constraining (negation, changing the order and grammatical role of words and adding words). The subsequent target sentence either contained the predicted word or an unexpected but still completely coherent alternative. See Table 1 for an example story in both versions.

We assessed the predictiveness of the constraining and prime control stories in a pencil-and-paper sentence completion test (12 participants, 2 males, age range 19 to 26 years, with an average of 22.4). In this so-called cloze test, we presented participants with the items which were truncated before the critical determiner. In predictive stories, the expected critical word had an average cloze value of 70% (sd=20%), and the unexpected critical word had an average cloze value of 5% (sd=16%). In non-predictive prime control stories these same two sets of critical words had

average cloze values of 30% (sd=22%) and 4% (sd=10%) respectively.[1]

In this experiment, we tested the pre-activation of the predictable nouns using the gender-marked definite determiner as a probe. 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 with the gender of the predicted noun. However, at the time that a participant reads the determiner neither the expected nor the unpredicted determiner 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. By including these intervening words we made sure that any differential ERP effects observed at the unpredicted determiner cannot be attributed to the processing of the overtly incongruent noun that usually immediately follows that determiner, but can only be attributed to differential processing of the determiner. 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. A list with all critical items (in Dutch) can be obtained from the first author.

40 items for each of the four conditions shown in Table 1 were randomly intermixed to create the first list of items to be presented to the subject. The 160 items of this list were then rotated so that three more lists of stimuli were created. Each of the four lists contained all 160 experimental stimuli in random order, 80 stories in the predictive version and 80 with a prime control. Half of the stories in each set contained the unpredicted determiner, and the other half contained the expected determiner. 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.

4.4. *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 electrodes (FP1, FP2, F9, F7, F3, Fz, F4, F8, F10, FT9, FC5, FC2, FC6, FC1, FT10, T7, C3, Cz, C4, T8, CP5, CP1, Cp2, Cp6, P7, P3, Pz, P4, P8 and Oz), mounted in an elastic cap, each referenced to the left mastoid. Blinks and vertical eye movements were registered by placing an electrode under the left eye, also referenced to the left mastoid. Electrode impedance was kept below 5 kΩ 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. Blinks and eye movements were removed from the data using a procedure based on Independent Component Analysis (ICA) as described by Jung and colleagues (Jung, Makeig, Humphries et al., 2000; Jung, Makeig, Westerfield et al., 2000). 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. The main rejection rate was 11% (range 1–27%). For each subject and condition, the remaining epochs were then averaged.

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 and 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.

Since the critical determiner was always separated from the noun by at least 3 words, the determiner and the noun were always separated in time by at least 1900 ms (the presentation time of the critical word plus the presentation times of the intervening words).

4.5. *Analyses*

The ERPs elicited by determiners were evaluated in an ANOVA with Consistency (prediction-consistent/prediction-inconsistent) and Context (predictive/prime control stories) as within participants factors, and WMC level (high/low) as a between participants factor. To assess the possible interaction with electrode position the ERPs were also assessed in an ANOVA crossing Consistency, Context and WMC with the factors Hemisphere (left/right) and Anteriority (anterior/posterior). 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. *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.

[1] Although the percentage of completions using the (message-predictable) critical noun in prime-control stories (30%) is considerably lower than the percentage of completions with this noun in the message-predictable stories (70%), it is somewhat higher than we had expected. Analysis of specific responses suggests that in prime control stories, the cloze test overestimates the message-level appropriateness of the critical noun, and that in this condition it actually picks up on scenario-based priming.

Appendix A

A sample of the materials used in this experiment. The critical determiner is printed in boldface. These stories exemplify the different ways in which predictive stories were changed into less predictive prime control stories (negation, changing the order and grammatical role of words and adding words)

Predictive discourse	Prime control discourse
Nadat hij uren naar het enorme lege doek had gekeken voelde de schilder inspiratie opkomen. Hij greep naar **de** *grote vanwege intensief gebruik sleetse kwast/***het** *grote vanwege intensief gebruik sleetse paletmes en smeet de verf op het doek.* After watching the empty canvas for several hours the painter felt a surge of inspiration. He reached for **the**$_{com}$ big, thoroughly worn brush/**the**$_{neu}$ big, thoroughly worn palette knife and threw the paint on the canvas.	*Nadat hij uren naar het enorme lege doek had gekeken had de schilder nog steeds geen inspiratie. Hij greep naar* **de** *grote vanwege intensief gebruik sleetse kwast/***het** *grote vanwege intensief gebruik sleetse paletmes en smeet deze door zijn atelier.* After watching the empty canvas for several hours the painter still didn't feel a surge of inspiration. He reached for **the**$_{com}$ big, thoroughly worn brush/**the**$_{neu}$ big, thoroughly worn palette knife and threw it through the studio.
Anne had eindelijk een rustig plekje gevonden waar ze kon studeren. Ze ging zitten en pakte **het** *dikke en behoorlijk beduimelde boek/***de** *dikke en behoorlijk beduimelde roman uit haar tas.* Anne had finally found a quiet place to study. She sat down and took **the**$_{neu}$ thick and pretty well-thumbed book/**the**$_{com}$ thick and pretty well-thumbed novel out of her bag.	*Na het studeren had Anne een rustig plekje in het park gevonden. Ze ging zitten en pakte* **het** *dikke en behoorlijk beduimelde boek/***de** *dikke en behoorlijk beduimelde roman uit haar tas.* After studying Anne had finally found a quiet place in the park. She sat down and took **the**$_{neu}$ thick and pretty well-thumbed book/**the**$_{com}$ thick and pretty well-thumbed novel out of her bag.
De leraar probeerde de verveelde pubers iets uit te leggen. Hij schreef op **het** *kleine en ietwat gammele schoolbord/***de** *kleine en ietwat gammele flip-over en praatte tegelijkertijd.* The teacher tried to explain something to the bored teenagers. He wrote on **the**$_{neu}$ small and somewhat wobbly blackboard/**the**$_{com}$ small and somewhat wobbly flip-over and talked at the same time.	*De puber probeerde de verveelde lerares iets uit te leggen over zijn thuissituatie. Hij schreef op* **het** *kleine en ietwat gammele schoolbord/* **de** *kleine en ietwat gammele flip-over en praatte tegelijkertijd.* The teenager tried to explain something to the bored teacher about his home situation. He wrote on **the**$_{neu}$ small and somewhat wobbly blackboard/**the**$_{com}$ small and somewhat wobbly flip-over and talked at the same time.

REFERENCES

Baggio, G., van Lambalgen, M., Hagoort, P., 2008. Computing and recomputing discourse models: an ERP study. J. Mem. Lang. 59 (1), 36–53.

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

Camblin, C.C., Gordon, P.C., Swaab, T.Y., 2007. The interplay of discourse congruence and lexical association during sentence processing: evidence from ERPs and eye tracking. J. Mem. Lang. 56 (1), 103–128.

Daneman, M., Carpenter, P.A., 1980. Individual differences in working memory and reading. J. Verbal Learn. Verbal Behav. 19 (4), 450–466.

Daneman, M., Carpenter, P.A., 1983. Individual differences in integrating information between and within sentences. J. Exper. Psychol.: Learn., Mem., Cogn. 9 (4), 561–584.

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

Engle, R.W., 2002. Working memory capacity as executive attention. Curr. Dir. Psychol. Sci. 11 (1), 19–23.

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

Federmeier, K.D., 2007. Thinking ahead: the role and roots of prediction in language comprehension. Psychophysiology 44 (4), 491–505.

Fiebach, C.J., Schlesewsky, M., Friederici, A.D., 2001. Syntactic working memory and the establishment of filler-gap dependencies: insights from ERPs and fMRI. J. Psycholinguist. Res. 30 (3), 321–338.

Gernsbacher, M.A., Faust, M., 1995. Skilled suppression. Interfer. Inhib. Cogn. 295–327.

Gernsbacher, M.A., Faust, M.E., 1991. The mechanism of suppression: a component of general comprehension skill. J. Exper. Psychol.: Learn., Mem., Cogn. 17 (2), 245–262.

Gernsbacher, M.A., Robertson, R.R.W., 1999. The role of suppression in figurative language comprehension. J. Pragmat. 31 (12), 1619–1630.

Gunter, T.C., Jackson, J.L., Mulder, G., 1995. Language, memory, and aging: an electrophysiological exploration of the N400 during reading of memory-demanding sentences. Psychophysiology 32 (3), 215–229.

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

Jung, T.P., Makeig, S., Humphries, C., Lee, T.W., McKeown, M.J., Iragui, V., et al., 2000. Removing electroencephalographic artifacts by blind source separation. Psychophysiol. 37 (2), 163–178.

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

Just, M.A., Carpenter, P.A., 2002. A capacity theory of comprehension: individual differences in working memory. Cogn. Modeling 131–177.

Just, M.A., Carpenter, P.A., Keller, T.A., 1996. The capacity theory of comprehension: new frontiers of evidence and arguments. Psychol. Rev. 103 (4), 773–780.

Kamide, Y., Altmann, G.T.M., Haywood, S.L., 2003. The time-course of prediction in incremental sentence processing: evidence from anticipatory eye movements. J. Mem. Lang. 49 (1), 133–156.

Keefe, D.E., McDaniel, M.A., 1993. The time course and durability of predictive inferences. J. Mem. Lang. 32 (4), 446–463.

King, J.W., Kutas, M., 1995. Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. J. Cogn. Neurosci. 7 (3), 376.

Kuperberg, G.R., 2007. Neural mechanisms of language comprehension: challenges to syntax. Brain Res. 1146, 23–49.

Lau, E., Stroud, C., Plesch, S., Phillips, C., 2006. The role of structural prediction in rapid syntactic analysis. Brain Lang. 98 (1), 74–88.
Levinson, S.C., 2000. Presumptive Meanings: The Theory of Generalized Conversational Implicature. MIT Press.
Linderholm, T., 2002. Predictive inference generation as a function of working memory capacity and causal text constraints. Discourse Process. 34 (3), 259–280.
Marslen-Wilson, W., 1973. Linguistic structure and speech shadowing at very short latencies. Nature 244 (5417), 522–523.
McKoon, G., Ratcliff, R., 1986. Inferences about predictable events. J. Exper. Psychol.: Learn., Mem., Cogn. 12 (1), 82–91.
Morris, R.K., 1994. Lexical and message-level sentence context effects on fixation times in reading. J. Exper. Psychol.: Learn., Mem., Cogn. 20 (1), 92–103.
O'Brien , E.J., Shank, D.M., Myers, J.L., Rayner, K., 1988. Elaborative inferences during reading: do they occur on-line. J. Exper. Psychol.: Learn., Mem., Cogn. 14 (3), 410–420.
Otten, M., Nieuwland, M.S., Van Berkum, J.J.A., 2007. Great expectations: specific lexical anticipation influences the processing of spoken language. BMC Neurosci. 8 (89).
Otten, M., Van Berkum, J.J.A., 2007. What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse-dependent N400 effect. Brain Res. 1153, 166–177.
Otten, M., Van Berkum, J.J.A., 2008. Discourse-based lexical anticipation: prediction or priming? Discourse Process. 45 (6), 464–496.
Pickering, M.J., Garrod, S., 2007. Do people use language production to make predictions during comprehension? Trends Cogn. Sci. 11 (3), 105–110.
Riding, R.J., Vincent, D.J.T., 1980. Listening comprehension: the effects of sex, age, passage structure and speech rate. Educ. Rev. 32 (3), 259–266.
Rosen, V.M., Engle, R.W., 1998. Working memory capacity and suppression. J. Mem. Lang. 39 (3), 418–436.
Ruchkin, D.S., Johnson, R., Canoune, H., Ritter, W., 1990. Short-term memory storage and retention: an event-related brain potential study. Electroencephalogr. Clin. Neurophysiol. 76 (5), 419–439.
St George, M., Mannes, S., Hoffman, J.E., 1997. Individual differences in inference generation: an ERP analysis. J. Cogn. Neurosci. 9 (6), 776–787.
Van Berkum J.J.A. (in press). The neuropragmatics of 'simple' utterance comprehension: an ERP review. In U. Sauerland, K. Yatsushiro, (Eds.), Semantic and pragmatics: from experiment to theory: Palgrave.
Van Berkum, J.J.A., Brown, C.M., Zwitserlood, P., Kooijman, V., Hagoort, P., 2005. Anticipating upcoming words in discourse: evidence from ERPs and reading times. Learn., Mem. 31 (3), 443–467.
Van den Noort, M.W.M.L., Bosch, M.P.C., Hugdahl, K., 2005. Understanding the role of working memory in second-language acquisition. Paper presented at the KogWis05. The German Cognitive Science Conference.
Van Herten, M., Kolk, H.H.J., Chwilla, D.J., 2005. An ERP study of P600 effects elicited by semantic anomalies. Cogn. Brain Res. 22 (2), 241–255.
Van Petten, C., 1993. A comparison of lexical and sentence-level context effects in event-related potentials. Lang. Cogn. Process. 8 (4), 485–531.
Vos, S.H., Gunter, T.C., Kolk, H.H.J., Mulder, G., 2001. Working memory constraints on syntactic processing: an electrophysiological investigation. Psychophysiology 38 (01), 41–63.
Wicha, N.Y.Y., Bates, E.A., Moreno, E.M., Kutas, M., 2003. Potato not Pope: human brain potentials to gender expectation and agreement in Spanish spoken sentences. Neurosci. Lett. 346 (3), 165–168.
Wicha, N.Y.Y., Moreno, E.M., Kutas, M., 2003. Expecting gender: an event related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in Spanish. Cortex 39 (3), 483–508.
Wicha, N.Y.Y., Moreno, E.M., Kutas, M., 2004. Anticipating words and their gender: an event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. J. Cogn. Neurosci. 16 (7), 1272–1288.