Reading comprehension in elementary school children: cognitive studies of the reader, the text, and the task

Karlsson, A.K.J.

Chapter 5

The Role of Working Memory in Inference Generation during Reading and Listening Comprehension

Ganushchak, L, Y., Karlsson, J., & van den Broek P. (in preparation). The role of working memory in inference generation during reading and listening comprehension.

The Role of Working Memory in Inference Generation during Reading and Listening Comprehension

Abstract

We investigated the role of working memory (WM) during reading and listening comprehension. The same participants read (eye-tracking) and listened (EEG) to short stories. A target word had either high predictability or low predictability, depending on the prior discourse context. Readers and listeners showed a typical predictability effect, with longer total gaze durations, more regressions, and more pronounced N400 amplitudes for low predictable targets words than high predictable target words. This effect was modulated by WM during listening but not during reading. In the listening task, low-WM participants but not high-WM participants showed an effect of predictability. We conclude that working memory may be more taxed during listening than during reading comprehension. Making predictions from local context alone and inhibition difficulties are possible mechanisms discussed as underlying the predictability effect in low-WM participants during listening comprehension.

Keywords: reading comprehension, listening comprehension, working memory, N400, predictive inferencing.

5.1 Introduction

The ability to comprehend written and spoken language is essential for an individual's ability to function in a modern information society. It is important to understand not only the meaning of isolated bits of information, but also the relations between them. During reading and listening, the successful comprehender connects encountered events, persons, facts, and objects so that the text appears to be coherent (e.g., van den Broek, Risden, Fletcher, & Thurlow, 1996). Working memory (WM), a system responsible for active maintenance of information during ongoing language processing (Conway et al., 2005), has been implicated as a major factor in the success in creating coherence and as a source of individual differences (e.g., Calvo, 2001; Daneman & Carpenter, 1980; Otten & van Berkum, 2009; Whitney, Ritchie, & Clark, 1991). For instance, Just and Carpenter (1992) suggested that an individual's working memory capacity provides a central constraint on general language comprehension. However, it is possible that written and spoken information make different demands on WM. For instance, listening requires that a comprehender holds each preceding word in working memory, whereas reading provides memory support in the form of written words that the comprehender may re-read. Consequently, the role of working memory may differ for these two comprehension modalities. Here we investigate how individual differences in working memory affect reading and listening comprehension.

5.1.1 Working Memory

Working memory is a multi-component system responsible for active maintenance of information in the face of ongoing processing and/or distractions (Conway et al., 2005). Arguably, the most influential model of WM is the one proposed by Baddeley (for a review see Baddeley, 2012). According to this model, an attentional control mechanism (i.e., the central executive) oversees three auxiliary domain-specific subsystems responsible for the temporary maintenance of verbal (the phonological loop), visual (the visuospatial sketchpad), and long-term memory information (Baddeley, 2012). Verbal information is maintained for short periods by the phonological loop whose contents decay over time unless it is maintained by a process of subvocal articulation (Baddeley, 2007) or via the focusing of attention onto different components of long-term memory (LTM; Cowan, 1995; Oberauer, 2003). Working-memory processes involved in maintaining information relevant to the task are inhibition, switching, and updating (although the exact relation amongst them are debated; Diamond, 2013; Ecker, Lewandowsky, Oberauer, Chee, 2010). The extent to which information maintenance depends on

domain-specific skills versus domain-general executive attention varies as a function of individual ability, task context, and interaction between ability and context (Conway et al., 2005).

5.1.2 Inferences during Language Comprehension

Central to successful comprehension is the construction of a coherent mental representation of the provided information, a so-called mental model or a situation model (e.g., Graesser et al., 1997; Kintsch, 1988; Rapp, van den Broek, MacMaster, Kendeou, & Espin, 2007; Van den Broek, 2010). The processes of inference generation (i.e., extracting implicit information from the text) are important for the construction of such a model. Readers and listeners use discourse information and their world knowledge to make predictive inferences about upcoming events (e.g., Cook, Limber, & O'Brien, 2001; McKoon & Ratcliff, 1986; Weingartner, Guzmán, Levine, & Klin, 2003), but also to make predictions on the word level (e.g., activate aspects of word meaning, grammar, and form) prior to the onset of a predicted word (e.g., Delong, Urbach, & Kutas, 2005; Otten, Nieuwland, & van Berkum, 2007). It is believed that prediction plays an important role in effective use of language (e.g., Altmann & Mirković, 2009), and that predictions help readers and listeners to anticipate the forthcoming information and therefore facilitate subsequent processing (e.g., Calvo, 2001; Federmeier & Kutas, 1999a, 1999b; Fincher-Kiefer, 1995; Keefe & McDaniel, 1993; Klin, Murray, Levine, & Guzmán, 1999; Schwanenflugel & LaCount, 1988; Weingartner et al., 2003; Whitney, Ritchie, & Crane, 1992). Making predictions, however, depends on the text being constraining enough (Cook et al.., 2001; Klin, Guzmán, & Levine, 1999; van den Broek, 1990) as well as the comprehenders' characteristics, such as WM capacity (e.g., Linderholm, 2002; Virtue, van den Broek, & Linderholm, 2006).

5.1.3 Comprehension and Working Memory

According to the causal inference process model (van den Broek, Fletcher, & Risden, 1993) and the capacity constrained comprehension theory (Just & Carpenter, 1992), (predictive) inferences are made when new information initiates a spreading of activation to associated concepts from background knowledge stored in long-term memory (LTM). Associated concepts are transferred from LTM into WM provided that there is enough activation, or cognitive resources, for further processing. Predictive inferences are made when explicit and associated concepts co-occur in WM and connections are made between them. Thus, enough cognitive resources are needed to keep the meaning of previous words active as the discourse proceeds and, at the same time, combine these with associated concepts to generate the predictive inferences. Accordingly, the more working memory capacity that

readers have available, the more likely it is that they will be able to expand on language processing and draw predictive inferences (Calvo, 2004; Just & Carpenter, 1992; Singer & Ritchot, 1996). For example, when reading short paragraphs requiring inference making for successful comprehension, high-WM participants were said to be more successful in generating predictive inferences than low-WM participants, as reflected by their N400 amplitudes (St George et al., 1997). The ERP component N400 is typically related to lexical and semantic processing (for a review, see Kutas & Federmeier, 2011). However, other research finds the N400 effect in both high- and low- WM participants (Otten & van Berkum, 2009), suggesting that both groups successfully make predictions. Instead, there may be qualitative differences in inference generation between comprehenders with different WM capacity. For instance, high- and low-WM participants appear to use different sources to generate an inference. High-WM participants rely both on discourse context and background knowledge to generate inferences from a larger pool of activated concepts, whereas low-WM participants tend to rely more on local discourse context (Boudewyn, Long, & Swab, 2013). Furthermore, low-WM but not high-WM participants may have difficulties in processing faulty predictions, shown by larger negativity at later processing (Otten & van Berkum, 2009). Low-WM participants have more problems with inhibiting unwanted information, compared to high-WM participants (e.g., Engle, 2002; Gernsbacher & Robertson, 1999), possibly explaining difficulties in processing faulty predictions.

5.1.4 Readings versus Listening Comprehension

Although it is clear that WM affects language comprehension, it is possible that this effect differs between spoken and written language. Previous studies have shown that listening and reading comprehension are two closely related skills (Booth, Perfetti, & MacWhinney, 1999; Booth, Perfetti, MacWhinney, & Hunt, 2000; Just & Carpenter, 1987). It is believed that higher-order cognitive processes (e.g., inference making) of text comprehension are modality independent and therefore the same in reading and listening comprehension (e.g., Booth et al., 2002). However, neuroimaging studies suggest that comprehension in these two modalities involves both similar and different underlying cognitive processes even for higher-level comprehension processes. Besides overlapping brain regions active during both reading and listening, there are also distinct brain regions that are only active during listening or reading comprehension (e.g., Buchweitz, Mason, Tomitch, & Just, 2009; Michael, Keller, Carpenter, & Just, 2001). Typically, listening comprehension results in more overall activation across the whole brain, whereas brain activation associated with reading comprehension is more lateralized in the left hemisphere (Buchweitz et al., 2009; Constable et al., 2004; Jobard

e al., 2007; Michael et al., 2001). One of the explanations of this pattern is that listening comprehension is sequential in nature and therefore more demanding than reading comprehension, where the reader can re-read words. Typically, more demanding tasks result in more widespread brain activation than easier tasks, for example due to recruitment of areas important for executive control or information maintenance (e.g., Buchweitz et al., 2009; Constable et al., 2004; Michael et al., 2001). If listening poses higher cognitive demands on the comprehender than reading, then it is possible that the role of WM is more important in listening than in reading comprehension. The primary focus of the current study is to determine whether individual differences in WM affect prediction equally in the two comprehension modalities, reading and listening.

5.1.5 Current Study

We investigated the effects of individual differences in working memory on reading and listening comprehension using a task in which the ease of making predictive inferences was manipulated. Ease of making predictions was manipulated by varying expectancy of the target word. The same participants read and listened to short two-sentence stories. The first sentence provided the comprehender with a setting and context. In the second sentence we manipulated a target word to have either high or low predictability, depending on the discourse context provided by the first sentence. In example (1), the discourse context is set-up in such a way that *aquarium* rather than *bowl* is a more predictable word. In example (2) the reverse is true: *bowl* is more predictable than *aquarium*.

(1) "Peter vindt tropische vissen heel erg mooi. Thuis heeft hij veel verschillende soorten in een aquarium/kom zwemmen".
(Peter thinks that tropical fish are very beautiful. At home, he has a lot of different types swimming in an aquarium/bowl.)

(2) "Het jongetje was ontzettend blij met zijn nieuwe goudvis. Thuisgekomen deed hij de vis meteen in een kom/aquarium met schoon water."
(The boy was very happy with his new goldfish. At home he immediately put the fish in a bowl/ an aquarium with fresh water.)

To study cognitive processes during listening, we measured Event-Related Potentials (ERP) by means of electroencephalography (EEG). The ERP component of interest here is the N400. The N400 reflects the degree to which retrieval of semantic memory associated with a word is facilitated by the discourse context (for a review, see Kutas & Federmeier, 2011). Recently, it has been argued that

the N400 also reflects the ability to use background knowledge to predict upcoming words (e.g., Nieuwland, 2015). Overall, a reduced N400 amplitude is found on words that are semantically plausible, related, or predictable given the preceding language context compared to words that are semantically implausible, unrelated, or unpredictable (e.g., Federmeier & Kutas, 1999a, 1999b; Kutas & Hillyard, 1980; van Berkum, Hagoort, & Brown, 1999; van Petten, 1993; for a review, see Swaab, Ledoux, Camblin, & Boudewyn, 2012). Typically, the N400 effect observed in sentence processing shows that both readers and listeners immediately relate the incoming words to a semantic representation of the preceding language input (e.g., van Berkum et al., 1999).

Using ERPs during a listening task is suitable as the stimuli can be presented as natural language. However, when using ERPs in reading research, sentences are typically presented word-by-word, and thus pose a potentially higher cognitive load because participants need to remember each word and cannot look back. A word-by-word presentation may therefore overestimate the role of WM during reading. This does not make ERPs suitable to address this paper's research questions about reading processes because it would require changing those natural aspects of reading that make reading processes possibly different from listening processes. Thus, we chose two different methods to study discourse processing (and the role of WM) in the two modalities, each of which is eminently suited to record on-line processing in one of the modalities.

To study cognitive processes during reading we used an eye-tracking methodology, as it allows participants to read as naturally as possible. I.e. they could read at their own pace and preview as well as re-read parts of the text. For the eye-tracking data, a distinction between early and late processing was made. The index of early processing was (a) first gaze duration, i.e., the sum of all first-pass fixations durations on a target word prior to moving to another word. Indices of late processing (e.g., re-analysis) were (b) probability of regressions, i.e., backward eye movements from a region back to the target word, and (c) total gaze duration, i.e., the sum of all fixations on a target word. Prior research has shown that first gaze duration is sensitive to early comprehension processes, such as word recognition (for an overview see Clifton, Staub, & Rayner, 2007). Total gaze durations reflect late processing, such as re-analysis and discourse integration (e.g., Frisson & Pickering, 1999; Rayner, 1998; Sturt, 2007). For both first gaze and total gaze measures, longer durations are interpreted as an indication for more effortful integration processes (e.g., Rayner & Sereno, 1994). Typically, readers tend to look longer at and have more regressions (re-reading of earlier parts of the text) to words that have low semantic plausibility or low predictability than words that have high semantic plausibility or high predictability considering the context (e.g.,

Pickering & Traxler, 1998; Rayner, Warren, Juhasz, & Liversedge, 2004; Staub, Rayner, Pollatsek, Hyönä, & Majewski, 2007; Warren, McConnell, & Rayner, 2008). This effect is believed to reflect readers' difficulty to generate semantically coherent interpretations of the sentences.

The current study employs the Mental Counters task to evaluate working memory (Huizinga, Dolan, & Van der Molen, 2006; Larson, Merritt, & Williams, 1988). In the studies reviewed above working memory was measured by the Reading Span task (Just & Carpenter, 1992) or by the Sentence Span task (Swanson, 1992). Both these tasks are complex span tasks that involve a tradeoff between maintenance and processing. Typically, participants are asked to read or listen to sets of unrelated sentences and to remember the final word of each sentence. Prior to recalling the final word of each sentence, participants are asked a question about one of the sentences. Thus, performance on such tasks is dependent on participants' comprehension ability. As our goal is to examine the relation between WM and listening and reading comprehension, a WM task independent of either modality is necessary. In the Mental Counters task participants are required to keep track of and update the score of visual counters in working memory, hence, performance on this task is independent from language comprehension ability.

We conducted the experiments to investigate (a) whether the role of WM differs during reading and listening comprehension, and (b) whether the inference generation process is comparable between the two modalities. If listening comprehension poses higher cognitive demands than reading comprehension, then a larger WM should be of more importance during listening than reading comprehension. Furthermore, if predictive inference processes are comparable between the two modalities, we expect to find correlations between the eye-tracking and ERP measures of inference making. Based on earlier findings discussed above, we predict that participants will have difficulty in successfully generating predictive inferences in the low predictability condition compared to the high predictability condition during both reading and listening comprehension. During reading, the relative difficulty of generating inferences will be reflected by longer gaze durations, and more regressions. During listening, relative difficulty in generating inferences will be reflected by a more pronounced N400. We also expect that inference processes will be affected by the WM of participants. On the one hand, if high-WM participants but not low-WM participants generate predictive inferences, then only the high WM comprehenders will show the predictability effect in form of longer gaze durations, more regressions, and a more pronounced N400. On the other hand, if both groups generate predictive inferences but only low-WM participants have difficulty in resolving faulty predictions, then only low-WM par-

ticipants will show the predictability effect as reflected by of longer gaze durations, more regressions, and a more pronounced N400, in the low-predictive condition compared to the high-predictive condition. If WM plays a greater role in listening than in reading comprehension, these differences would be more pronounced during the listening task than during the reading task. By using a within-subject design, we assess the effect of WM on comprehension processes in the two modalities in the same participants.

5.2 Methods

5.2.1 Participants

Forty-four students at Leiden University participated in the experiment (38 women; average age: 22 years, SD = 2.5 years). All participants were right-handed native Dutch speakers and had normal or corrected-to-normal vision and hearing. They gave written informed consent prior to participating in the study and received a small financial reward or course credits for their participation. Due to technical problems, the data of two participants had to be excluded from the analyses. Data of the remaining 42 participants (36 women; age: 22 years, SD = 2.5 years) was included into the analyses.

5.2.2 Materials and Design

5.2.2.1 Reading and Listening Task

The materials for the listening (ERP) and reading tasks (eye-tracking) were set up in the same way. For both tasks, target words were embedded in short stories consisting of two sentences: a context sentence followed by a target sentence. The second sentence contained the target word, which was a noun with an average length of 6.5 characters (SD = 2.6). The average length of the short stories was 24.3 words (SD = 4.9).

Short stories were designed in such a way that a target word was preceded by an indefinite article and always appeared towards the end of the second sentence (but was never the final word of the sentence). Depending on discourse context, the target word was either high or low in predictability. Each target word appeared in both conditions. The short stories in both high-/low- predictability conditions were semantically correct (e.g., see examples in Introduction). To pre-test the predictability of target words we used a cloze test, which is assumed to be an independent measure of critical word predictability (e.g, Nieuwland, 2015; Nieuwland, 2016). In the cloze task, 20 participants from a similar population as the main sample read the short stories up to the target word and were asked to complete the

sentences with the first word that came to their mind. The results of the cloze task showed that participants spontaneously used the high predictability target word (e.g., aquarium) with average cloze probability of 70% (SD = 5.3%).

After having read or listened to each story, participants were presented with a multiple-choice comprehension question with three answer possibilities. The purpose of the questions was to make sure that participants paid attention to the stories. The entire set of materials consisted of 336 short stories. Four stimulus lists were created to counterbalance target words across modality (i.e., reading and listening), with 4 practice items and 80 test sentence pairs and target words per list. Each target word occurred in the two conditions (i.e., high, and low predictability) on different lists, and each participant was exposed to each target word only once.

During the reading task, sentences were displayed 65 cm from the participants' eyes and appeared in Arial font 18 pt size. One degree of visual angle equaled approximately 3.0 characters. Each trial began with a fixation point presented about two character spaces to the left of the first character of the upcoming sentence. The fixation point also served as drift correction, to check whether the calibration was still acceptable. The eye tracker was recalibrated whenever the experimenter deemed necessary. The task was self-paced: after reading a sentence, the participant pressed the spacebar to move to the next sentence and, eventually, to the multiple-choice question. Participants were instructed to read for comprehension and answer the questions as accurately as possible.

During the listening task, short stories were presented auditorily through speakers. All short stories and comprehension questions were recorded by a male native Dutch speaker. Each trial started with the fixation point, which remained on the screen for the duration of the story. The short stories were on average 6141 ms long (SD = 1200 ms). After each short story, a comprehension question was presented auditorily simultaneously with the visual presentation of the answer options. Participants were instructed to listen to the stories for comprehension and answer the question as accurately as possible by pressing the corresponding keys on the button-box.

5.2.2.2 Working-Memory Task

The Mental Counters is a computerized working memory task (Huizinga et al., 2006; Larson et al., 1988). In the Mental Counters task, participants were asked to keep track of and update the score of counters in their working memory. There were two (block 1) or three (block 2) independent counters. The counters are horizontal lines, positioned to the left and to the right from the middle of the computer screen. For each trial, the starting score for each counter was 0. Above and below the lines, squares appeared in a random order. Participants were required to add

one point to the value of the counter, when a square appeared above it, and to subtract one point when it appeared below the counter. Before each trial began, participants received a criterion value to remember throughout the trial. Participants were instructed to press a button when any counter reached the given criterion value, before the next block appeared. The score for the task was the proportion of correct trials.

5.2.3 Procedure

Participants were tested individually. After signing the informed consent, participants started with the Mental Counters task. The Mental Counters began with a block of practice items, before moving on to the two test blocks. After completing this task, participants were asked to read and listen to the short stories. The order of the reading and listening tasks were counterbalanced across participants. Reading and Listening tasks started with four practice trials. The entire test session took approximately two hours.

5.2.4 Apparatus

Eye-movement recording. Eye movements were recorded with an Eyelink 1000 eye- tracker (SR Research Ltd.; 500 Hz sampling rate). Eye calibration was done at the beginning of the experiment, using a 9-point calibration procedure. The participant's head was kept immobile with the use of a chin and head rest. Viewing was binocular, but only the movements of the dominant eye were recorded.

Electrophysiological recoding. Electroencephalograms (EEG) were recorded using an EEG cap with 32 active electrodes (Ag/AgCl), mounted according to the extended International 10-20 system. The EEG was collected using BioSemi ActiView, and the EEG signal was digitized at a rate of 512 Hz with a band pass filter of DC-128 Hz. All electrodes were offline re-referenced to the two mastoids. Lateral eye movements were measured using a bipolar montage of two electrodes placed on the right and left external canthus. Vertical eye movements were measured using a bipolar montage of two electrodes placed above and below the eyes.

5.2.5 Analyses

5.2.5.1 Eye-tracking Data

In total, 1% of the data were lost due to eye blinks or technical problems. For each target word, we determined first gaze durations (the sum of all fixation durations on a target word prior to moving to another word), total gaze durations (the sum of all fixation duration on a target word), and regression rates (the percentage of backward eye movements to the target region from a succeeding region). In the

analyses of total gaze durations, only those trials were included where the target word was fixated during first-pass reading. The data were subjected to repeated measures ANOVA with Predictability (high- vs. low- predictability) as independent variable. WM group (high-WM vs. low-WM) was used as between-subject factor. The working memory groups were created based on the median split of mental counters scores.

5.2.5.2 Electrophysiological Data

Epochs from –200 ms to +800 ms were obtained relative to the onset of each target, including a 200 ms pre-stimulus baseline. To correct for ocular and non-ocular artefacts, epochs with amplitudes above or below 75 μV were rejected. The EEG signal was applied to a high-pass filter of 0.01 Hz/ 24 dB and a low-pass filter of 40 Hz/ 24 dB. Analyses were performed for a post-stimulus time window of 250 – 550 ms. For this time window, we calculated mean amplitude values per participant and per condition.

Amplitudes were submitted to repeated-measures ANOVAs with Predictability (high- vs. low- predictability), and Location (anterior, i.e. F7, F3, FC1, FC5, AF3, FC6. FC2, F4, F8, AF4, Fz vs. posterior, i.e. CP2, P4, PO4, O2, CP1, CP5, P3, PO3, O1, CP6, Pz) as independent variables. This led to a division of electrodes into two areas, which were used to investigate the distribution of the possible effects. WM group (high-WM vs. low-WM) was used as between-subject factor. The groups were the same as in the analysis of the eye-tracking data.

5.2.5.3 Combining Eye-tracking and Electrophysiological Data

To investigate the relation between reading comprehension (eye-tracking) and listening comprehension (ERP data), we computed Pearson's two-tailed correlations between differences scores (low predictable – high predictable) for First Gaze Durations, Total Gaze Durations, Regression Rates, and the N400 effect.

5.3 Results

5.3.1 Reading Comprehension: Eye-Tracking Data

Participants answered 91.8% (SD = 0.03) of the comprehension questions correctly, showing that they read the sentences attentively.

For the first gaze durations there was no significant effect of Predictability ($F < 1$). Across both conditions, high-WM participants fixated for shorter duration (M = 240 ms; SD = 32 ms) than low-WM participants (M = 243 ms; SD = 36 ms). However, this difference was not significant ($F(1, 37) = 3.75$, $p = .06$). The interaction between Predictability and WM group was not significant ($F(1, 37) = 1.95$, $p = .17$).

Analysis with total gaze durations as dependent variable showed that all participants spent more time reading target words in the less-predictable condition (M = 471 ms; SD = 287 ms) than in the semantically predictable condition (M = 432 ms; SD = 589 ms; $F(1, 37) = 7.19$, $p = .01$). The main effect of WM group and the interaction between Predictability and WM group were not significant ($F < 1$ and $F(1, 37) = 2.22$, $p = 0.15$, respectively).

Participants were more likely to make regressions to the target in the less-predictable condition (M = .37; SD = .18) than in the high-predictable condition (M = .33; SD = .18; $F(1, 37) = 13.78$, $p = .001$). Across both conditions, there was a lower probability that high-WM participants would look back (M = .28; SD = .18) than that low-WM participants would do so (M = .43; SD = 14; $F(1, 37) = 6.79$, $p = .01$). The interaction between Predictability and WM group was not significant (both $F < 1$).

5.3.2 Listening Comprehension: Electrophysiological Data

Participants answered 90.2% (SD = 0.02) of the comprehension questions correctly, showing that they listened to the sentences attentively.

Analysis of the mean amplitudes showed a main effect of Predictability ($F(1, 37) = 10.77$, $p = .002$). This effect was qualified by a significant interaction between Predictability and WM group ($F(1, 37) = 5.63$, $p = .023$). To further investigate the significant interaction, separate ANOVA's were run for high-WM and low-WM participants. For high-WM participants, there was no significant effect of Predictability (low-predictable: M = -0.99μV; SE = 0.55μV; high-predictable: M = -0.61μV; SE = 0.38μV; $F < 1$, see Figure 5.1). By contrast, for low-WM participants there was a significant effect of Predictability ($F(1, 16) = 14.62$, $p = .001$, see Figure 5.2): N400 amplitudes were more negative for the low predictable condition (M = -1.61μV; SE = 0.41 μV) than for the high predictable one (M = 0.77μV; SE = 0.50 μV;).

Neither main effects of WM group and Location were significant nor were other interactions (Predictability x WM group x Location: $F(1, 37) = 1.62$, $p = .21$; all other effects: Fs < 1).

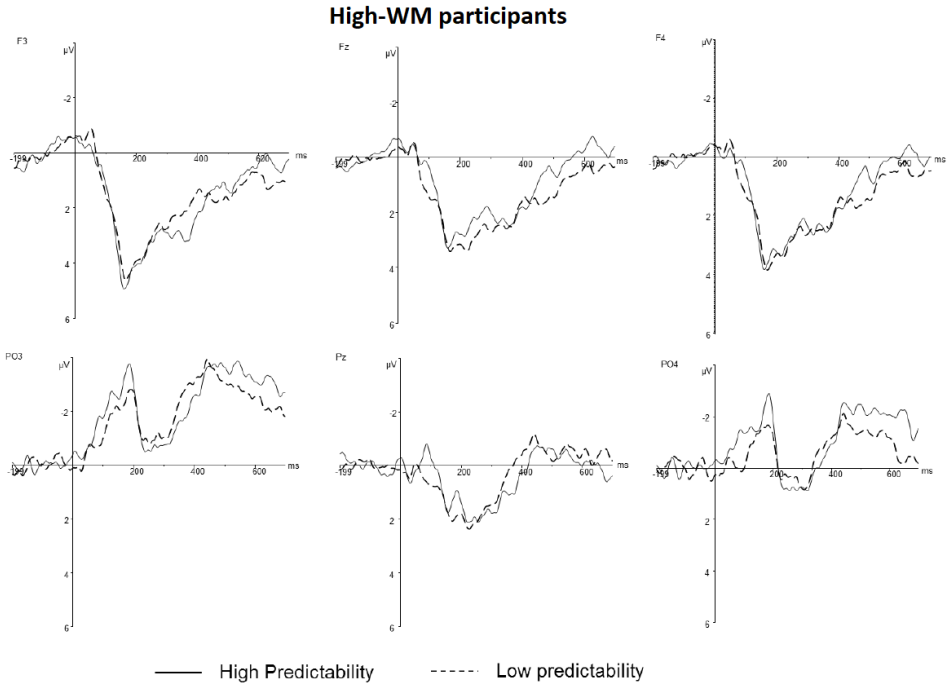

Figure 5.1. Averaged stimulus-locked ERP waveforms for high predictable (solid lines) versus low predictable (dashed lines) target words for high-WM participants. Amplitudes (μV; Y-axis) are plotted against time (ms; X-axis); zero represents onset of the target word.

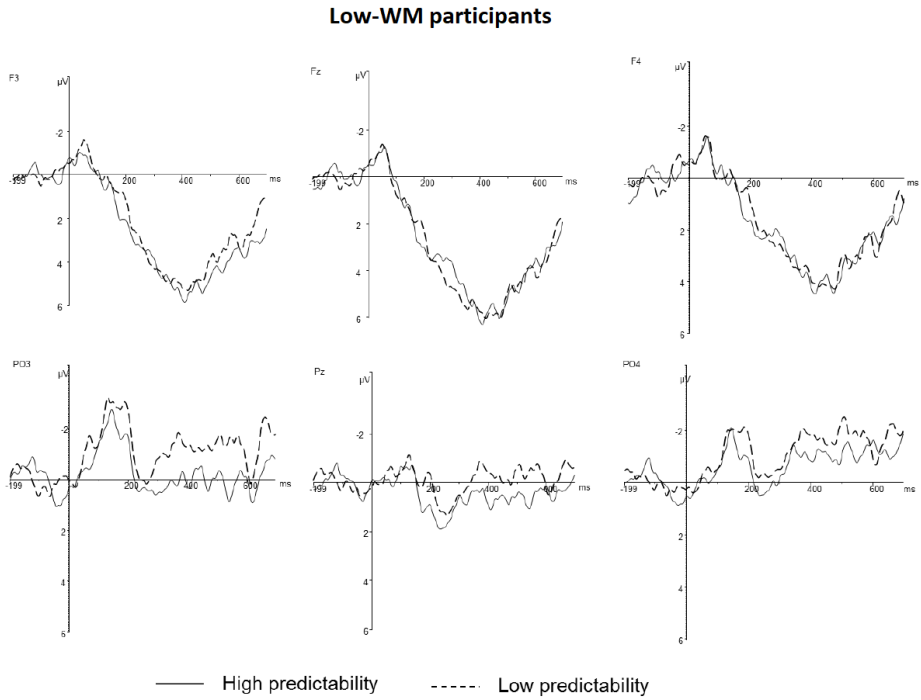

Figure 5.2. Averaged stimulus-locked ERP waveforms for high predictable (solid lines) versus low predictable (dashed lines) target words for low-WM participants. Amplitudes (μV; Y-axis) are plotted against time (ms; X-axis); zero represents onset of the target word.

5.3.3 Relation between Reading and Listening Comprehension

There was a significant positive correlation between first gaze durations and N400 (r = .36; p = .02). Participants who showed larger differences between experimental conditions during reading also had larger differences during listening comprehension. The correlations between total gaze durations and N400, as well as between regression rates and N400 were not significant (r = .15, p = 0.23; r = .12, p = .29, respectively).

5.4 Discussion

The aims of the present study were to investigate (a) whether the role of Working Memory (WM) differs during reading and listening comprehension, and (b) whether the inference generation process is comparable between the two modalities. Overall, for both the reading and listening tasks, we replicated the standard effect. All participants spent more time reading low-predictable words in late reading processes and having more pronounced N400 amplitudes than high-predictable words. However, only during listening this effect was modulated by the WM of participants; low-WM participants, but not high-WM participants, showed an effect of predictability. Furthermore, there was a positive moderate correlation between measures of predictive inference generation during reading and listening comprehension. Below we discuss these findings in more detail.

To investigate the role of WM during reading and listening we recorded eye-movements during reading and EEG during listening in low- and high-WM participants, and found differences between the two language comprehension modalities. Predictive inference processes during reading were not modulated by individual differences in working memory: Participants spent more time reading low-predictable words than high-predictable words. This predictability effect emerged only for total gaze duration, not for first-pass gaze duration, and was due to readers re-fixating on the low-predictable words more often than on the high-predictable words. Because highly predictable words fit better with discourse context they require less overall processing time and, consequently, are easier to integrate into preceding context without regressions. These results are in line with previous findings showing that semantic integration affects late processes more than early processes (e.g., Rayner et al., 2004). Hence, during reading we only see the predictability effect in late processing, and this is independent of WM.

During listening, N400 amplitudes to low-predictable words were more negative than to high predictable ones. This predictability effect was modulated by individual differences in working memory. Low-WM participants, but not high-WM participants, showed more pronounced N400 amplitudes for the low-predictable words compared to the high predictable ones. Similar to our results, Otten and van Berkum (2009) showed a significant predictability effect for low-WM participants but not for high WM participants. Here we propose two possible explanations for the differences in the predictability effect between the two groups: There may be differences in how high- and low-WM participants make predictions, or there may be differences in how high- and low-WM participants process information that is not in line with the made prediction.

First, successful comprehension depends not only on information provided by the discourse context but also on the comprehenders ability to retrieve information

from semantic memory (i.e., background knowledge; e.g., Ericsson & Kintsch, 1995; Kintsch, 1998; Till, Mross, & Kintsch, 1988). The N400 reflects not only sensitivity to local context information (e.g., from preceding context) but also the ease with which one accesses information from one's background knowledge (for review see Kutas & Federmeier, 2000). High-WM participants are more efficient in retrieving and/or integrating information from background knowledge than low-WM participants. Consequently, high-WM participants have a wider network of words active in their lexicon during listening (e.g., both bowl and aquarium may be active) thereby decreasing the differences between levels of predictability induced by the local discourse context. Low-WM participants may rely more on recently activated information by the local context than on activation from background knowledge (Boudewyn et al., 2013). This in turn results in a more pronounced difference between levels of predictability induced by manipulating local discourse context. Hence, during prediction generation, high-WM participants use both local context and their background knowledge, whereas low-WM participants rely more on local context alone.

Second, it is possible that the observed differences between high-WM and low-WM participants do not reflect difficulty in making predictions per se but rather a greater effort to suppress their initial prediction about an upcoming target word which was not in line with the local context. Inhibition is an important predictor for successful listening comprehension (Kim & Phillips, 2014) and low-WM participants have more problems with inhibiting unwanted information than high-WM participants do (e.g., Engle, 2002; Gernsbacher & Robertson, 1999). Furthermore, an inability to inhibit irrelevant information may lead to increased N400 amplitudes compared to situations when inhibition was successful (e.g. Debruille, 2007; Debruille et al., 2008; Kutas & Federmeier, 2011). Hence, the N400 effect observed in the low-WM participants may reflect their difficulty suppressing their initial prediction.

To address our second aim, comparing inference generation processes between the comprehension modalities, we compared performance on the reading measures with performance on the listening measure. Results showed that participants who experienced predictive inference difficulties during reading also tended to experience inference difficulties during listening comprehension. The positive correlation was significant only between first gaze duration and N400, and not between total gaze duration or between regression rate and N400. This may be due to the fact that total gaze duration and regression rates reflect later reading processes associated with revisiting the target word, whereas first gaze fixation duration and N400 reflect early comprehension processes. These results lend some support to an amodal view of higher-order comprehension processes. But the correlation is

moderate (r = 0.36), leaving enough room for modality-specific differences in processing.

We propose that similar prediction generation mechanisms (e.g., use of local discourse context and background knowledge, and inhibition mechanisms) could play a role during reading as well as listening comprehension. However, a likely modality-specific difference between reading and listening comprehension concerns the demands they make on working memory. During reading there were no differences between high- and low-WM participants concerning the predictability effect. In contrasts, during listening only low- WM participants showed a predictability effect. Spoken language generally does not allow control over the rate of processing and revision of already heard parts of the story (e.g., Constable et al., 2004). Because listening comprehension poses greater demands on the comprehender's cognitive system, a larger WM is of more importance during listening than during reading.

To conclude, whereas most theories of language comprehension take an amodal approach, stating that high order processes are the same between the two modalities, few studies have compared higher-order cognitive processes during reading and listening. Our study expands on previous literature by taking a comparative approach between reading and listening. The results point to specific commonalities as well as differences between the two modalities. Similar underlying mechanisms may play a role during both reading and listening, but our results show that even on a higher-order processing level comprehension during reading and listening may differ. Specifically, we showed that WM plays a greater role in predictive inferencing during listening than reading.