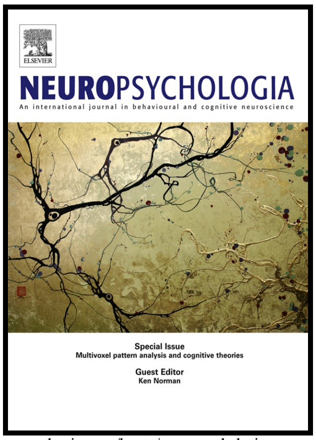
Author's Accepted Manuscript



www.elsevier.com/locate/neuropsychologia

How Context Features Modulate the Involvement of the Working Memory System during Discourse Comprehension

Xiaohong Yang, Xiuping Zhang, Yufang Yang, Nan Lin

PII: S0028-3932(18)30010-1
DOI: https://doi.org/10.1016/j.neuropsychologia.2018.01.010
Reference: NSY6639

To appear in: *Neuropsychologia*

Received date: 14 June 2017
Revised date: 11 January 2018
Accepted date: 11 January 2018

Cite this article as: Xiaohong Yang, Xiuping Zhang, Yufang Yang and Nan Lin, How Context Features Modulate the Involvement of the Working Memory System during Discourse Comprehension, *Neuropsychologia,* https://doi.org/10.1016/j.neuropsychologia.2018.01.010

How Context Features Modulate the Involvement of the Working Memory System during Discourse Comprehension

Xiaohong Yang[1,2] Xiuping Zhang[1,2] Yufang Yang[1,2] Nan Lin[1,2]*

[1]CAS Key Laboratory of Behavioral Science, Institute of Psychology, Beijing 100101, China

[2] Department of Psychology，University of Chinese Academy of Sciences, Beijing 100049, China

Corresponding author:

Nan Lin

CAS Key Laboratory of Behavioral Science, Institute of Psychology

Address :Lincui Road 16, Chaoyang District, Institute of Psychology, Chinese Academy of Sciences, Beijing, 100101, China

Tel: (86) 10 6787 6719

E-mail address: linn@psych.ac.cn

Abstract

Using functional magnetic resonance imaging (fMRI), we investigated the effects of context features on the involvement of the working memory (WM) system during discourse comprehension. During the fMRI scan, participants were asked to read two-sentence discourses in which the topic of the second sentence was either maintained, or was shifted from, the topic of the first. Changes in the level of coherence between the two sentences as well as context length were also investigated across discourse items. The WM system was identified with a verbal N-back task. Analysis of the reading comprehension task revealed that within the WM system, stronger activation in the left inferior frontal gyrus corresponded with increased bridging coherence demands between sentences, while greater activation in the left inferior and middle frontal gyri, bilateral superior frontal gyri, and bilateral inferior parietal lobules corresponded with increased context length. Topic variation showed no effect on activation of the WM system. These results provide new insights into understanding how different levels of context features modulate activation of the subcomponents of the WM system and indicate a role for the left inferior frontal gyrus as a core component of the WM system supporting discourse processing.

Key words: working memory; topic variation; coherence; length; discourse comprehension

Introduction

Discourse comprehension plays a core role in human communication. Successful discourse comprehension requires not only the involvement of the language system but also support from other cognitive systems. For some time, researchers have emphasized an important role of the working memory (WM) system in discourse comprehension. Theoretical psychologists have proposed working memory as a prerequisite memory store for discourse processing (Graesser et al., 1997; Van den Broeket al., 2005). The underlying assumption is that working memory provides both the working space for the analysis of linguistic inputs and also temporal storage for the resulting discourse representations.

Numerous empirical studies have also demonstrated the importance of working memory for discourse comprehension. The majority of these studies adopted a capacity perspective of working memory (Just & Carpenter, 1992), often by dividing participants into high and low working memory capacity groups (e.g., Palladino et al., 2001; Virtue et al., 2008). High working memory groups were consistently reported to perform better than low working memory groups, either in their greater sensitivity to discourse manipulations or with higher accuracy rates (e.g., Palladino et al., 2001; Virtue et al., 2008). At the neural level, Virtue et al. (2008) found that high working memory participants showed greater neural activity than low working memory participants in the right superior temporal gyrus and right inferior frontal gyrus when they generated inferences during story comprehension.

These previous studies have revealed how limitations of working memory resources can affect discourse comprehension. However, an over-reliance on the capacity perspective has stunted progress toward more knowledge about the contribution of working memory in discourse processing. Rather than considering working memory to be a static, and often insufficient, cognitive resource, most current neuropsychological theories suggest that working memory is best characterized as consisting of multiple subsystems, which are differentially involved in different aspects of high-level cognitive processing (Collette, et al., 2017; Collette & Van der Linden, 2002; Collette et al., 2007; D'Esposito et al., 1999; Kim et al., 2002; Manginelli et al., 2013; Rottschy et al., 2012; Shaywitz et al., 2001; Veltman et al., 2003). For example, a recent meta-analysis (Rottschy et al., 2012) of 189 fMRI studies of working memory showed that the WM system comprised a widespread bilateral frontoparietal network consisting of several regions sensitive to specific task components, among which the left inferior frontal gyrus was selectively active in verbal tasks while the ventral and dorsal premotor cortex were preferentially involved in memory for object identity and location, respectively. From the perspective of maintenance or manipulation of content, there were also reports demonstrating that multiple subsystems comprise the central executive component needed to manipulate or maintain information in the WM system (Cogan et al., 2017; Nyberg & Eriksson, 2015; Takahama et al., 2010). Working memory manipulations, such as the organization of working memory content into higher-level 'chunks' (Bor et al., 2003), updating object representation of dynamically moving objects (Takahama et al., 2010), or rule-related processing independent of stimulus identity (Cogan et al., 2017), have been consistently found to depend on frontal lobe activity. In contrast, working-memory maintenance was found to involve more widespread frontoparietal regions including the inferior-precentral sulcus, superior parietal lobule, and middle frontal gyrus (Habeck et al., 2005; Narayanan et al., 2005; Nyberg & Eriksson, 2015; Takahama et al., 2010).

Although this multiple-subsystem perspective of the WM system has proven to be quite useful for the understanding of human cognition (D'Esposito & Postle, 2015; Nyberg & Eriksson, 2015), to date, no study has related this functional segregation of the WM system to discourse processing. As a complex cognitive task, reading continuous discourse involves multiple processing levels: 1) lower-level reading components including phonology, morphology, word decoding, and vocabulary, and 2) high-level deeper comprehension components that include knowledge of discourse structures, construction of inferences, and use of background knowledge (Graesser & McNamara, 2011; Graesser et al., 2014). The processing demands of these levels are subjected to different discourse features including topic information, coherence, and length. The requirement for high-level processing can be increased when the topic changes or when there is less coherence between successive sentences, while the demand for low-level processing can increase when context length increases. Below, we consider these discourse features in more detail and describe how they relate to the different components of the WM system.

In a continuous discourse, a sentence can either shift or continue the topic of the previous sentence. This has been acknowledged as a general organizing principle of discourse in both linguistic and psycholinguistic studies (Horne et al., 2001; Hyönä, 1994, 1995; Hyönä & Lorch, 2004; Kuppevelt, 1995; Oberlander, 2004; Yang et al., 2013). The notion that a text's topic structure is indeed mentally represented by comprehenders has been supported by numerous psychological studies. It has been consistently shown that topic shifts can bring about increased processing demands during discourse comprehension (Anderson et al.,1983; Binder & Morris, 1995; McKoon et al., 1993; O'Brien et al., 1986; Yang et al., 2013). For instance, readers were found to spend more time reading topic-shifted sentences than topic-maintained sentences (Hyönä, 1995). Studies using event-related potentials (ERPs) during discourse comprehension have found an enhanced late positivity for the processing of topic shifts compared with topic maintenance (Hung & Schumacher, 2014; Hung & Schumacher, 2012; Yang et al., 2015), which is likely to reflect the discourse updating process triggered by topic shifts. However, there is still a lack of evidence about whether the processing of topic shifts is supported by some or all subcomponents of the WM system. The structure building framework of discourse comprehension (Gernsbacher, 1997) assumes that topic shifts trigger an updating process: When encountering topic shifts, readers tend to suppress the current discourse representation and build a new substructure for the new topic in working memory. As such, it is likely that the processing of topic shifts should increase the demands for the manipulation of the contents of the WM system.

Another discourse feature that may affect the involvement of working memory is the semantic coherence between sentences (e.g., Helder et al., 2017). The construction of a coherent text representation is essential to text comprehension (e.g., Graesser et al., 1994; Kintsch & Van Dijk, 1978; McKoon & Ratcliff, 1992; van den Broek & Helder, 2017; Van den Broek et al., 2005; Zwaan & Radvansky, 1998). In past decades, numerous studies have investigated coherence processing with the underlying assumption that in order to process less coherent discourse, more inferences are required and thus more working memory resources must be employed (e.g., Virtue et al., 2006; Virtue et al., 2008). Lack of coherence could occur when the semantic connections between individual sentences are left implicit. For instance, it was shown to be more demanding for comprehenders to understand "*The next day they had bruises"* in the context of "The boys were having an argument. They became more and more angry. The next day they had bruises." as opposed to "The boys were having an argument. They began hitting each other. The next day they had bruises." (Kuperberg et al., 2006).[1] Lack of coherence could also occur as a result of referentially problematic expressions, such as in cases where two antecedents are equally plausible for an anaphor (Nieuwland et al., 2007a; Nieuwland & Van Berkum, 2008) or when a repeated name is used instead of a pronoun to refer back to an antecedent (Almor et al., 2007). In these situations, readers need to make anaphoric inferences to restore referential coherence. While manipulations of coherence varied across studies, lack of coherence was reliably associated with processing cost, including longer reading times (Albrecht & O'Brien, 1993; Myers et al., 1994) and enhanced ERP components such as the N400 and P600 (Kuperberg et al., 2011; Nieuwland & Van Berkum, 2008). In fMRI studies, coherence processing was reported to involve a wide range of brain regions, including the lateral prefrontal/temporal/parietal cortices, left precentral gyrus, medial prefrontal cortices, and posterior midline regions, which has been considered evidence of increased memory load associated with establishing coherence for the discourses (Egidi & Caramazza, 2014; Kuperberg et al., 2006; Mo et al., 2006).

An important question that has not been clearly addressed is whether the impact of topic shift on working memory is independent from that of semantic coherence. Although it has been widely assumed that a change in topic always lowers the semantic coherence between sentences (Dijkstra et al., 2004; Foltz et al., 1998), most previous studies on topic variation were not controlled for the

---

[1]In previous fMRI studies, two kinds of "coherence" effects have been described. One is obtained by manipulating coherence in natural discourse as in Kuperberg et al. (2006). The other is obtained by comparing connected sentences to strings of random sentences as in Yarkoni et al. (2008). It should be noted that the optimal approach to obtain the "coherence" effect depends on the research question to be addressed. If the scope of the study is to investigate how coherence difference affects processing demands during discourse processing, the first approach should be used. However, if the scope is to examine discourse-level coherence building as compared to sentence-level comprehension, then the second approach should be used. In the current study, the first approach was used.

effect of semantic coherence (Binder & Morris, 1995; Hung & Schumacher, 2012; Hyönä, 1994, 1995; Yang et al., 2013). Therefore, the observation of a topic shift effect may actually reflect the confounding effect of semantic coherence. Importantly, corpus analysis has indicated that in natural discourses, topic shifts can be introduced with less of a coherence gap than exists in less coherent text, which means that a text with topic shift is not always less coherent than a text without topic shift (Foltz et al., 1998). Therefore, it is possible to tease apart the effect of topic shift from that of semantic coherence on working memory by carefully and simultaneously manipulating the two factors.

In addition to topic and coherence variation, context length has also been suggested to affect the involvement of working memory. Processing any discourse requires integrating each upcoming word with its prior discourse context (Hagoort, 2013). It has been shown that compared with short discourse context, long discourse context can increase processing demands when readers attempted to integrate upcoming words with prior discourse context (Yang et al., 2015). As opposed to the effect of topic shifts or coherence, which both increase working memory load primarily by increasing the demand for content manipulation, increasing context length increases working memory demands primarily by requiring more effort for the storage of a larger number of items (Hammer et al., 2008; Yang et al., 2015).

Taken together, the field of discourse processing has dissected several context features of discourse, including topic structure, coherence, and length, that can potentially modulate working memory processing demands when individuals read or listen to connected discourse. These three features together can serve as an ideal linkage for assessing the relationship between discourse comprehension and the involvement of the working memory system. They constitute basic features of natural discourse (Foltz et al., 1998; Graesser & McNamara, 2011; Graesser et al., 2014; Hyönä, 1994, 1995; Oberlander, 2004). More importantly, as they belong to different levels of discourse processing, according to previous studies (Gernsbacher, 1997; Graesser et al., 1994; Hammer et al., 2008; McKoon & Ratcliff, 1992; Van den Broek et al., 2005; Zwaan & Radvansky, 1998), they are likely to affect different subcomponents of the WM system, with topic structure and coherence largely affecting WM manipulation and length largely affecting WM maintenance. However, this remains to be tested experimentally. Therefore, in the present study, we have investigated precisely how these context features modulate the activation of the WM system, using fMRI. Because previous findings have suggested that manipulation of working memory content is associated with increased prefrontal activity (Bor et al., 2003; Cogan et al., 2017; Nyberg & Eriksson, 2015) and that within the WM system, the left inferior frontal gyrus is selectively active in verbal working memory tasks (Rottschy et al., 2012), we expected that topic variation and coherence difference would increase activation in the frontal components of the WM system, particularly in the left inferior frontal gyrus. Moreover, given prior reports that working memory maintenance involves not only frontal regions, but also distributed parietal regions (Habeck et al., 2005; Narayanan et al., 2005; Nyberg & Eriksson, 2015), we expected that increasing length would modulate a more distributed frontoparietal network of the WM system.

2 Methods

2.1 Participants

Twenty-four undergraduate students (12 female, 12 male) participated in the current study. The mean age was 22.1 years (SD = 2.2). All participants were right-handed native Chinese speakers and had normal or corrected-to-normal vision. The Review Board of the Institute of Psychology, Chinese Academic of Sciences, approved this study, and written informed consent was obtained from each participant.

2.2 Design and Materials

During the experiment, the participants were asked to perform two tasks: an N-back localizer task and a reading comprehension task.

The verbal N-back task is a frequently-used paradigm to address neural activation patterns associated with verbal working memory functions (Owen et al., 2005; Rottschy et al., 2012). The task includes a two-back condition, which places high working memory demand on the participants, and a zero-back baseline condition, which does not require information manipulation within the WM system. Based on previous meta-analyses (Owen et al., 2005; Rottschy et al., 2012), we expected that relative to the zero-back baseline condition, the two-back condition would activate a bilateral frontoparietal working memory network.

For the reading comprehension task, we asked participants to read short discourses of two sentences each. In these discourses, the second sentence either shifted or maintained the topic of the first sentence. This topic manipulation allowed us to investigate the requirements of topic variation. The topic structure for each discourse adhered to the following guidelines. We kept the second sentence (i.e., the target sentence) identical across the topic-shifted and topic-maintained conditions to avoid potential confounding effects of syntactic and orthographic differences. For the topic-maintained condition, a name always appeared in the first sentence and a pronoun referring to the named individual was present in the second sentence. For the topic-shifted condition, the first sentence introduced a topic (a person, thing, event, or setting) that did not correspond to the pronoun introduced in the second sentence, although the topics of both sentences could be mentally organized within the same scene. The topic was defined as the information that appears in the initial position of the sentence and represents what the following information of the sentence is about (Li & Thompson, 1976; Li, 2004). Besides topic manipulation, while keeping coherence and length matched overall between the topic-shifted and topic-maintained conditions, we varied these two dimensions across individual items for each condition, which allowed us to investigate their effects with a parametric modulation approach (Altmann et al., 2012; Wang & Quadflieg, 2015).

We created 72 sets of short discourses. Each discourse consisted of two sentences, and for each discourse set, two conditions were developed, each with a different topic structure, although with the second sentence (the target sentence) always kept unchanged. The topic of the first sentence was varied such that the topic of the second sentence either continued or shifted the topic (topic-shifted vs. topic-maintained). An example is shown in Table 1. In the topic-shifted condition, the topic of the first sentence is "*The hotel*", whereas the topic of the second sentence is "*he*". However, in the topic-maintained condition, the topic of the first sentence is "*Chen Xi*" and the second sentence maintains this topic with the pronoun "*he*", which refers back to "*Chen Xi*".

Table 1 Example stimuli used in the present study

Condition	first sentence	second sentence
topic-maintained	陈西住进了一个很差的酒店，	他好几个晚上没睡好了。
	Chen Xi stayed in a very poor hotel.	He hasn't slept well for several nights.
topic-shifted	这个酒店的住宿条件非常差，	他好几个晚上没睡好了。
	The hotel has very poor accommodations.	He hasn't slept well for several nights.

Note: English translations are presented below each Chinese sentence.

Coherence was measured using coherence scores obtained from a rating pre-test in which 16 undergraduate students who did not take part in the fMRI experiment were asked to assess the coherence of the discourses. They were instructed to estimate how closely the two sentences of each discourse were connected to each other (Rinck & Weber, 2003) on a 5-point scale (1 indicating no connection at all and 5 indicating a very close connection). Two presentation lists were created so that a given discourse appeared only once in each list (i.e., either the topic-maintained or topic-shifted condition). The rating results showed that the coherence of the discourses was similar between the two topic conditions (mean ± SD = 3.99 ± 0.41 and 4.06 ± 0.45 for the topic-shifted and topic-maintained conditions, respectively; $t(15) < 1$). Thus, coherence between the two topic conditions should not confound the effect of topic structure while still allowing for variation across the trials.

Length was measured by character count of each sentence in the discourses. The length of the discourses was also matched between the two conditions on the basis of character count for each discourse (mean ± SD =23.36 ± 2.64; 23.39 ± 2.65; for the topic-shifted and topic-maintained conditions, respectively; $t(71) = 1.42$, $p = 0.16$) but varied across different discourse items. The length of the first sentence ranged from 8 to 15 Chinese characters (mean ± SD =11.75 ± 1.70). The length of the second sentence ranged from 8 to 16 Chinese characters (mean ± SD = 11.64 ± 1.79).

Using a Latin square design, the 72 sets of discourses were separated into two lists so that one version of each item appeared on each list. Each list was pseudorandomized to vary the order of the conditions within each list. To counterbalance the order of the conditions in each list, two more versions were created by reversing the presentation order of the conditions in each list. Thus, altogether four lists were created for the 72 sets of discourses.

2.3 Tasks and procedures

The N-back localizer task was performed first. Five blocks of the 2-back condition were alternated with five blocks of the 0-back condition. Before the presentation of each block, an instruction screen was presented for 5 sec, telling the participants whether the following block was a 2-back block or a 0-back block. In each block, 15 series of alphabetic letters were presented visually. Each letter was presented for 500 ms with an inter-stimulus interval of 2,500 ms. During the presentation of the letters, the participants were asked to judge whether the presented stimulus was the same as the one presented 2 trials previously (for the 2-back block) or whether a specific letter (e.g., ‘N’) had appeared (for the 0-back block). After the presentation of all blocks, an ending screen appeared for 5 sec indicating that the task was over. The N-back task was performed within a single run lasting 8 min 35 sec, with the first 10 sec of the run being a fixation.

After the N-back localizer task, the participants were asked to perform the discourse comprehension task. Each participant was assigned one of the four lists of stimuli. The experiment consisted of three runs lasting 8 min 16 sec each. The first 10 sec of each run was a fixation. Each run included 24 trials, 12 for each of the two conditions. In each trial, the first sentence of the discourse appeared in the middle of the screen for 4 sec, followed by a 2 to 8 sec Inter-stimulus Interval (ISI1). Then the second sentence appeared for 4 sec, followed by an ISI of 2 to 8 sec (ISI2). For a quarter of the trials, a comprehension question about the content of the discourses was then presented for 3 sec and the participants were told to make a Yes or No judgment by pressing a button with either their right index or middle finger[2]. These questions were followed by another ISI of 2–8 sec (ISI3).

For both tasks, stimuli were presented using the E-Prime 2.0 software, with black background and white color. The participants were asked to lie down while viewing the stimuli through a mirror mounted on the head coil. Before the formal experiments, each participant completed practice training sessions outside the scanner room for each task with additional stimuli.

2.4 Data acquisition and analyses

Magnetic resonance imaging (MRI) data were collected with a GE Discovery MR750 3 T scanner. Blood oxygen level-dependent signal data were obtained in 42 slices with no gap. The parameters of the sequence were: TR = 2000 ms, TE = 30 ms, flip angle = 70°, slice thickness = 3 mm, matrix size = 64 × 64, FOV = 192 mm×192 mm. T1-weighted structural images were collected in 176 sagittal slices with 1.0 mm isotropic voxels.

[2] These questions were designed to ensure that participants perform the task carefully. We did not delete trials according to the accuracy of probing questions in any data analysis because only a quarter of the experimental trials were followed by probing questions and subjects could give a correct response by chance in the Yes or No judgment task.

Preprocessing of the MRI data was performed using the Data Processing Assistant for Resting-State fMRI (DPARSF, http://www.restfmri.net), which is based on SPM8 software (http://www.fil.ion.ucl.ac.uk/spm) and the Resting-State fMRI Data Analysis Toolkit (REST, http://www.restfmri.net). We discarded the first five volumes of each run for steady-state magnetization. For the remaining data, we performed slice timing correction (except for the N-back task, which was a block design), head motion correction, spatial normalization, and spatial smoothing. During normalization, the structural image of each participant was first co-registered to the mean functional image and then segmented using the unified segmentation VBM module implemented in DPARSF. The resulting parameters were then used to normalize the functional images of each participant onto the Montreal Neurological Institute (MNI) space. Spatial smoothing was performed using an isotropic Gaussian filter with a kernel of 6 mm, at full-width half-maximum.

Statistical analyses were conducted with a two-level mixed-effects model. For the first level, we applied a general linear model (GLM) to explore single participant effects for each condition. In the first level analysis of the N-back task, the two conditions ("2-back" and "0-back") were set as two covariates of interest and each block was modeled with a boxcar function lasting 45 sec from the onset of stimulus presentation to the end of stimulus presentation and were convolved with a standard hemodynamic response function.

For the discourse comprehension task, separate models were conducted to explore different types of effects of interest. The first model was used to examine the effect of general discourse processing and the effect of topic structure. At the first level, the model included a regressor for the first sentence of the discourses, two regressors for the second sentence, in the topic-shifted and topic-maintained conditions, respectively, and a regressor for the comprehension questions. These regressors were time-locked to the onset of the respective sentence presentation, with their duration equivalent to their presentation length in the procedure. The effect of general discourse processing was represented as the average activation of all first and second sentences. The effect of topic structure was examined by comparing the topic-shifted condition with the topic-maintained condition.

To explore the effects of coherence and context length, we used the parametric modulation approach (Buchel et al., 1996; Buchel et al., 1998) in which the effect of interest is modeled as the interaction between the presence of a trial and a parametric variable associated with it. Several recent fMRI studies of language processing have suggested that the parametric modulation approach is very sensitive in capturing the effects of trial-specific continuous variables such as frequency, length, imageability, predictability, and affective lexical ratings (Graves et al., 2010; Schuster et al., 2016; Hsu et al., 2015). In our analysis, two models with distinct orders of parametric regressors were conducted. Both models included three constant regressors for the first sentence, the second sentence (with the two conditions of topic structure merged), and the comprehension question, respectively. These regressors were again time locked to the onset of sentence presentation with their duration equivalent to presentation length in the procedure. For the regressor of the second sentence (i.e., the target sentence), the coherence rating and length values were entered as trial-specific modulation parameters. We alternated the order of these in the two models. In both models, the length of the second sentence was entered as the first parameter to rule out its effect on the other parameters. In the model examining the effect of coherence, context length (the length of the first sentence) was entered before coherence rating while in the model examining the effect of context length, coherence rating was entered before the context length. By alternating the order of the modulation parameters, we serially orthogonalized the parameters so that parameters entered later only accounted for variance that was unaccounted for by parameters entered earlier (Mumford et al. 2015; Rao et al., 2011; Tong et al., 2014).

For the first-level models of both tasks, regressors for the six head motion parameters, as well as a high-pass filter (113 sec cutoff), were also included. After the estimation of model parameters, subject-specific statistical maps were generated and subsequently inputted into the second-level group analysis. For the localizer experiment, the contrast image of each subject ("2-back" vs. "0-back") was inputted into a one-sample t test. The false positive rate was controlled at $\alpha < 0.05$ using voxel-level FWE correction implemented in SPM8, combined with a cluster threshold of 10 voxels. The significant clusters showing the "2-back > 0-back" effect in the localizer experiment were then defined as regions of interest (ROIs).

For the discourse comprehension task, we conducted two sets of analyses. The first set of analyses was conducted to examine whether the WM system identified in the localizer task is involved in general discourse processing. The second set of analyses was conducted to examine how the discourse features modulate the activation of the WM system. Both sets of analyses were conducted at whole-brain as well as ROI levels. In the whole-brain analyses, we analyzed the effects of interest (general discourse processing, topic structure, coherence, and context length) and then examined whether the obtained clusters overlapped with the WM areas identified in the localizer task. The false positive rate was controlled at $\alpha < 0.05$ using cluster-level FWE correction implemented in SPM8 with the individual voxel threshold probability setting of $p < .001$. In the ROI analyses, the beta values for each regressor of interest and critical modulation parameters were extracted and averaged across all the voxels within each ROI for each participant. We then conducted paired t-tests to compare the beta values of the topic-shifted and topic-maintained conditions and conducted one-sample t-tests to examine the effects of general discourse processing, coherence, and context length across participants. The whole brain and ROI analyses were conducted to complement each other. Whole brain analysis of fMRI data always suffers from low statistical power due to the fact that it requires very strict multiple-comparison corrections (Eklund et al. 2016). One way to overcome this problem is to limit testing to specific ROIs (Poldrack, 2007). Therefore, the ROI analysis is statistically more sensitive than the whole brain analysis and our conclusions are mainly based on the results of the ROI analysis. The advantage of whole brain analysis is that it reveals not only activation within the ROIs but also activation outside the ROIs. Thus, we have also presented the results of the whole brain analysis to examine whether the discourse features modulate the activity of brain regions outside the WM system during discourse comprehension.

3 Results

3.1 The N-back localizer task

Behavioral Results

For the N-back task, the data collected from two participants whose correct response rates were below 70% (Chow et al., 2016; Mingtian et al., 2012) were discarded from the analyses of behavioral and fMRI data. Additionally, due to technical problems with the E-Prime software, the responses for one participant were not recorded successfully. Therefore, this participant was also excluded from further analysis. The remaining 21 participants had an average accuracy rate of 91%

(SD=4%). Paired t-tests revealed no significant differences in accuracy rates between the 2-back and 0-back condition (2-back: mean = 93.33%, *SD* = 2.63%; 0-back: mean = 89.60%, *SD* = 8.34%; $t(20) = 1.95$, $p > .01$). For the analysis of the RT data, those that were less than 200 ms or more than 3 SDs above each participant's mean were removed as outliers (Mogg et al., 2008). Paired t-tests for the RT data revealed that the 2-back condition resulted in significantly longer reaction times compared to the 0-back condition (2-back: mean = 886.72 ms, *SD* = 251.792 ms; 0-back: mean = 652.08 ms, *SD* = 13; $t(20) = -6.68$, $p < .001$). This suggested that increases in working memory demand resulted in additional processing demands.

fMRI results

As showed in Figure 1A and Table 2, compared with the 0-back condition, the 2-back condition evoked stronger activation in the left inferior and middle frontal gyri， bilateral superior frontal gyri, and bilateral inferior parietal lobules. Decreased activation for the 2-back condition was found in bilateral middle temporal gyri, right precentral gyrus, left precuneus/cuneus, right cingulate gyrus, left angular gyrus, right inferior parietal lobule, right middle occipital gyrus, bilateral insula, right parahippocampal gyrus, left hippocampus, and right cerebrum.

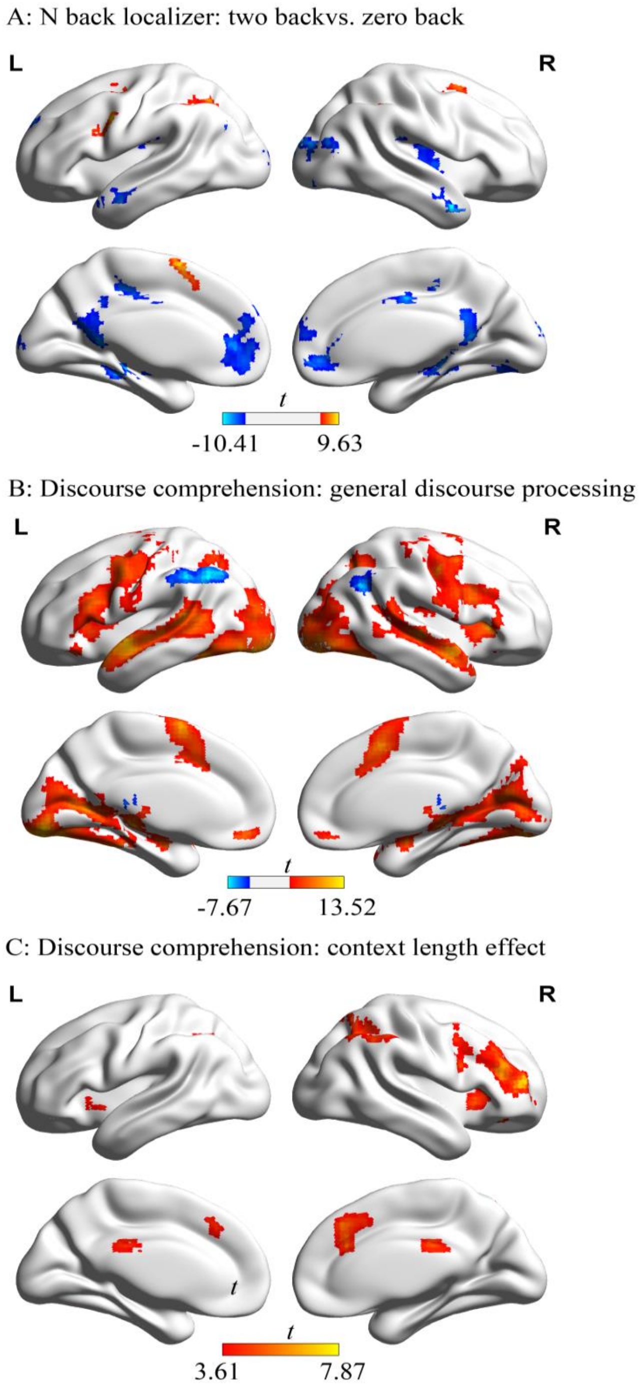


Figure 1 Results for the whole brain analyses of A) N-back localizer: two-back vs. zero-back, B) Discourse comprehension: general discourse processing, and C) Discourse comprehension: context length effect.

Table 2 Results of the N-back localizer.

Contrast	Anatomical region of the peak voxel	Cluster size (voxels)	MNI coordinates of peak voxel (x, y, z)			Peak t value
Two back > zero back						
	Left Inferior Parietal Lobule	146	-45	-42	42	8.54
	Left Inferior Frontal Gyrus	100	-48	3	36	9.40
	Left Superior Frontal Gyrus	74	-6	6	63	9.17
	Left Middle Frontal Gyrus	52	-30	0	63	8.37
	Right Superior Frontal Gyrus	46	24	6	60	8.04
	Right Inferior Parietal Lobule	40	42	-45	42	7.63
Two back < zero back						
	Left Precuneus	493	-12	-51	21	-9.55
	Left Anterior Cingulate	484	-9	39	6	-9.68
	Right Middle Occipital Gyrus	183	15	-99	12	-10.23
	Right Cingulate Gyrus	180	6	-9	39	-9.39
	Right Insula	131	36	-15	18	-8.28
	Right Parahippocampal Gyrus	100	30	-33	-12	-9.76
	Left Hippocampus	100	-30	-30	-12	-10.41
	Right Middle Occipital Gyrus	96	36	-72	-12	-9.56
	Left Middle Temporal Gyrus	73	-57	0	24	-8.79
	Right Middle Temporal Gyrus	50	54	3	-27	-10.40
	Left Cuneus	37	-15	-96	0	-8.02
	Left Insula	32	-33	-18	18	-7.48
	Right Inferior Parietal Lobule	21	45	-30	21	-7.78
	Right Precentral Gyrus	18	45	-18	45	-7.46
	Left Angular Gyrus	13	-45	-72	30	-7.35

Note. The false positive rate was controlled at $\alpha < 0.05$ using voxel-level FWE correction implemented in SPM8, combined with a cluster threshold of 10 voxels.

3.2 The discourse comprehension task

Behavioral Results

For the comprehension task, questions only appeared after a quarter of the trials and they were used to ensure that the participants attended to the stimuli while they were in the scanner. We found that five participants had an accuracy rate below 70%. Similar to the analysis of the N-back task, these participants were removed from the analyses of fMRI data. The remaining participants had an average accuracy rate of 86% (SD = 7%).

fMRI Results

Is the WM system involved in general discourse processing?

ROI analysis

The clusters showing the “2-back > 0-back” effect in the N-back task were defined as ROIs. As shown in Table 3, all the WM ROIs showed increased activation for general discourse processing. Thus, the results of the ROI analysis indicate that the entire WM system identified in the N-back task was involved in discourse comprehension.

Whole brain analysis

We conducted the whole brain analysis to comprehensively reveal the neural basis of general discourse processing. As shown in Figure 1B, general discourse processing evoked extensive and bilateral brain activation. These brain regions overlapped with the brain regions obtained in the N-back task in the left inferior frontal gyrus (94 voxels), left inferior parietal lobule (54 voxels), left superior frontal gyrus (71 voxels), left middle frontal gyrus (27 voxels), and right superior frontal gyrus (9 voxels). They also contained very extensive brain areas outside the WM system, indicating that discourse comprehension involved other cognitive systems besides the WM system.

How do the discourse features modulate the activation of the WM system?

ROI analysis

We found a significant effect of coherence in the left inferior frontal gyrus (as shown in Table 4), with increased activation for less coherent discourses. The left inferior and middle frontal gyri, bilateral superior frontal gyri, and bilateral inferior parietal lobule showed greater activation to increased context length. For the comparison between topic shift and topic maintenance, no significant effect was found.

Whole brain analysis

Whole brain analysis showed no significant clusters for either the topic effect or the coherence effect. With regard to the length effect, significant clusters were found in right inferior frontal gyrus, right superior parietal lobule, left medial superior frontal gyrus, left inferior parietal lobule, left superior temporal gyrus, left putamen, right cingulate gyrus, and right pallidum. All of these clusters showed stronger activation to increased context length (as shown in Table 5). We examined whether these brain regions and the WM regions identified in the localizer task overlap with each other. The overlap between these two sets of brain regions includes the bilateral inferior parietal lobule (left: 54 voxels; right: 4 voxel).

Table 3 Results for the ROI analyses of general discourse processing.

	Coherence			
Region	***Beta***	***SE***	***t***	***p***
LIFG	6.92	0.89	7.80	0.00***
LIPL	2.40	0.49	4.85	0.00***
RIPL	0.77	0.34	2.24	0.04*
LSFG	4.55	0.51	8.99	0.00***
LMFG	2.49	0.48	5.18	0.00***
RSFG	0.99	0.28	3.53	0.00***

Note: * p < 0.05, ** p < 0.01, *** p < 0.001. LIFG, Left Inferior Frontal Gyrus; LIPL, Left Inferior Parietal Lobule; RIPL, Right Inferior Parietal Lobule; LSFG, Left Superior Frontal Gyrus; LMFG, Left Middle Frontal Gyrus; RSFG, Right Superior Frontal Gyrus.

Table 4 Results for the ROI analyses of the discourse features.

	Coherence				Context length				*Topic: Shifted-maintained*			
Region	***Beta***	***SE***	***t***	***p***	***Beta***	***SE***	***t***	***p***	***Beta***	***SE***	***t***	***p***
LIFG	-0.38	0.14	2.71	0.01*	0.05	0.08	2.47	0.02*	0.08	0.15	0.52	0.61
LIPL	0.00	0.15	0.00	0.99	0.15	0.04	4.29	0.00***	0.05	0.11	0.49	0.62
RIPL	-0.12	0.14	0.87	0.40	0.16	0.05	3.30	0.00***	0.19	0.19	1.03	0.32
LSFG	-0.22	0.11	1.92	0.07	0.07	0.05	2.28	0.03*	0.09	0.12	0.70	0.49
LMFG	-0.16	0.13	1.24	0.23	0.12	0.05	2.80	0.01*	0.11	0.15	0.69	0.50
RSFG	-0.18	0.11	1.70	0.11	0.12	0.04	2.67	0.02*	0.22	0.16	1.33	0.20

Note: * p < 0.05, ** p < 0.01, *** p < 0.001. LIFG, Left Inferior Frontal Gyrus; LIPL, Left Inferior Parietal Lobule; RIPL, Right Inferior Parietal Lobule; LSFG, Left Superior Frontal Gyrus; LMFG, Left Middle Frontal Gyrus; RSFG, Right Superior Frontal Gyrus.

Table 5 Results for the whole brain analysis of context length effect.

Anatomical region of the peak voxel	Cluster size (voxels)	MNI coordinates of peak voxel (x, y, z)			Peak t value
Right Inferior Frontal Gyrus	900	36	36	15	7.87
Right Superior Parietal Lobule	368	36	-60	54	6.51
Left Medial Superior Frontal Gyrus	227	0	30	33	7.57
Left Inferior Parietal Lobule	118	-30	-57	39	6.27
Right Pallidum	67	12	3	-3	5.35
Right Cingulate Gyrus	64	3	-27	27	5.03
Left Putamen	57	-24	9	12	5.40
Left Superior Temporal Gyrus	53	-54	9	-6	5.34

Note. The false positive rate was controlled using a height threshold of $p < .001$ with a spatial cluster extent threshold of $p<.05$ (- FWE-corrected).

4 **Discussion**

The present study aimed to provide a picture of how the WM system was modulated by the requirements of discourse processing. Our findings indicated that all subcomponents of the WM system were involved in general discourse processing. More importantly, our findings further revealed that subcomponents of the WM system were flexibly recruited to meet processing requirements associated with different discourse features: The left inferior frontal gyrus was recruited when more effort was needed to establish coherence; the left inferior frontal gyrus, left middle frontal gyrus, bilateral superior frontal gyri, and bilateral inferior parietal lobule responded to increasing context length in discourse context. These findings point to a functional dissociation within the WM system during discourse comprehension and suggest that internal discourse features play an important role in guiding the involvement of the subcomponents of the WM system.

We observed increased neural activity in the left inferior frontal gyrus when participants read less coherent discourses and when context length increased. When comprehending less coherent discourses relative to coherent discourses, participants are more likely to detect a coherence break and may make more semantic inferences, resulting in more manipulation of working memory content. In contrast, reading longer sentences compared to reading short sentences mainly increases the load for the maintenance of a larger number of words. Therefore, our results indicated that the left inferior frontal gyrus contributes to both manipulation and maintenance of information (which is likely to be semantic in nature) in working memory during discourse comprehension. This speculation accords nicely with many prior findings that reported inferior frontal gyrus activation in a variety of language processing tasks that involved manipulation and maintenance of working memory content. For instance, the left inferior frontal gyrus has been observed to show enhanced activation when individuals generate inferences to establish coherence (Kuperberg et al., 2006; Virtue et al., 2006; Virtue et al., 2008), when semantic information has to be retrieved from memory (Wagner et al., 2001), and when semantically unexpected constituents are present (Hagoort& van Berkum, 2007). These observations, together with our findings, highlight a central role of the inferior frontal gyrus as the core working memory component that subserves discourse comprehension.

In addition to the left inferior frontal gyrus, the effect of context length was also observed in the left middle frontal gyrus, bilateral superior frontal gyri, and bilateral inferior parietal lobule. The left middle frontal gyrus, bilateral superior frontal gyri, and bilateral inferior parietal lobule activations could simply reflect the additional effort required for the storage of discourse representation in working memory when context information increased, as has been found in previous studies (Habeck et al., 2005; Narayanan et al., 2005; Nyberg & Eriksson, 2015; Takahama et al., 2010). An alternative interpretation is that while the inferior parietal lobule serves as the storage buffer of the WM system (Kirschen et al., 2005), middle frontal gyrus and superior frontal gyrus activation could serve a general top-down function to the posterior cortical areas involved in memory storage (Nyberg & Eriksson, 2015). Thus, it could be the case that with more information being read, the situation changed into a more demanding one and more top-down control was required in order to allocate more attention to upcoming information, which then engaged the middle frontal gyrus and superior frontal gyrus.

The increased activation for context length in comparison with that for semantic coherence was consistent with the hypothesis that the frontal and posterior regions of the WM system serve different cognitive functions: While working memory manipulations depends on frontal activities, working memory maintenance involves more widespread frontal-parietal areas (Nyberg & Eriksson, 2015). Our results thus extend this functional segregation of the WM system to the domain of discourse processing. Importantly, the result that different pools of regions responded to different discourse features suggests that the contribution of working memory in discourse processing may be better understood in terms of functional dissociation between a network of regions rather than in terms of a specific association between one static resource and one higher-level cognitive process, as is prevalent in previous studies.

The whole brain analysis showed that context length also modulated the activation of brain regions outside the WM systems. These regions were mainly located in the lateral and medial frontal cortex and in the posterior parietal cortex, which have been found to play a critical role in inhibition control (Cole & Schneider, 2007). As context length is strongly related to the richness of contextual semantic information, we propose that the context length effect observed in the inhibition control system reflects the inhibition of the automatically activated semantic information that is inconsistent with the context.

No significant topic effect was found in either the ROI analysis or the whole brain analysis. Note that in previous studies that have found significant effects of topic shift (Binder & Morris, 1995; Hyönä, 1994, 1995; Yang et al., 2013), a topic shift was usually accompanied by a coherence break. In the present study, to dissociate the effect of topic structure from coherence difference, we have matched the coherence between the two topic conditions when constructing our stimuli. Therefore, the null effect of topic manipulation indicated that the effect of topic shift observed in previous studies may actually reflect the difference of coherence between the topic-shifted and topic-maintained conditions.

It is also worth noting that in the present study, the topic-shifted condition introduces a new, unknown referent into the discourse with a pronoun. This could result in referential processing difficulty, which in previous ERP studies has been reliably associated with a sustained, frontal negativity (the Nref, a brain signature that resembles that of increased working memory load) (Boudewyn et al., 2015; Nieuwland et al., 2007a; Nieuwland & Van Berkum, 2008; Van Berkum et al., 2007). Similar referential processing difficulty was also reported in a previous fMRI study, with referential ambiguity eliciting increased activations in medial and lateral parietal, medial frontal, and right superior frontal regions (Nieuwland et al., 2007b). However, in the present study, although we introduced unbound pronouns in the topic-shifted condition, no significant effect of topic manipulation was found. This again could have been because we have matched coherence between the topic-shifted and topic-maintained conditions. Thus, the lack of an effect of topic manipulations suggests that referential processing difficulty could be ruled out if the discourse containing referential problematic expression was perceived as coherent.

In conclusion, we observed that the involvement of the WM system was modulated by the internal features of a discourse and that context length difference exerted more influence on the working memory subcomponents than coherence difference. Our findings provide new insights into the links between subcomponents of the WM system and cognitive processes involved in discourse comprehension and suggest that the left inferior frontal gyrus may have a special role in coping with the changing demands of discourse processing.

Acknowledgments

This research was supported by Natural National Science Foundation of China (grant numbers: 31400968 and 61433018).

References

Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 19(5), 1061-1069.

Altmann, U., Bohrn, I. C., Lubrich, O., Menninghaus, W., & Jacobs, A. M. (2012). The power of emotional valence-from cognitive to affective processes in reading. *Frontiers in human neuroscience, 6*.

Anderson, A., Garrod, S. C., & Sanford, A. J. (1983). The accessibility of pronominal antecedents as a function of episode shifts in narrative text. *Quarterly Journal of Experimental Psychology, 35*(3), 427-440.

Almor, A., Smith, D. V., Bonilha, L., Fridriksson, J., & Rorden, C. (2007). What is in a name? Spatial brain circuits are used to track discourse references *NeuroReport*, 18(12), 1215-1219.

Binder, K. S., & Morris, R. K. (1995). Eye Movements and Lexical Ambiguity Resolution: Effects of Prior Encounter and Discourse Topic. [doi:]. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(5), 1186-1196.

Bor, D., Duncan, J., Wiseman, R. J., & Owen, A. M. (2003). Encoding Strategies Dissociate Prefrontal Activity from Working Memory Demand. *Neuron*, *37*(2), 361-367.

Boudewyn, M. A., Long, D. L., Traxler, M. J., Lesh, T. A., Dave, S., Mangun, G. R., . . . Swaab, T. Y. (2015). Sensitivity to referential ambiguity in discourse: The role of attention, working memory, and verbal ability. *Journal of Cognitive Neuroscience*, 27(12), 2309-2323.

Büchel, C., Holmes, A. P., Rees, G., & Friston, K. J. (1998).Characterizing stimulus–response functions using nonlinear regressors in parametric fMRI experiments. *Neuroimage*, 8(2), 140-148.

Büchel, C., Wise, R. J. S., Mummery, C. J., Poline, J. B., & Friston, K. J. (1996). Nonlinear regression in parametric activation studies. *Neuroimage*, 4(1), 60-66.

Chow, H. M., Jingling, L., & Tseng, C. H. (2016). Eye of origin guides attention away: An ocular singleton column impairs. *Journal of Vision, 16*(1), 1-17.

Cogan, G. B., Iyer, A., Melloni, L., Thesen, T., Friedman, D., Doyle, W., . . . Pesaran, B. (2017). Manipulating stored phonological input during verbal working memory. [Article]. *Nat*

Neurosci, *20*(2), 279-286.

Cole, M. W., & Schneider, W. (2007). The cognitive control network: Integrated cortical regions with dissociable functions. *Neuroimage*, 37(1), 343-360.

Collette, F., & Van der Linden, M. (2002). Brain imaging of the central executive component of working memory. *Neuroscience & Biobehavioral Reviews, 26*(2), 105-125.

Collette, F., Van der Linden, M., Laureys, S., Arigoni, F., Delfiore, G., Degueldre, C., . . . Salmon, E. (2007). Mapping the Updating Process: Common and Specific Brain Activations Across Different Versions of the Running Span Task. *Cortex, 43*(1), 146-158.

D'Esposito, M., Postle, B. R., Ballard, D., & Lease, J. (1999). Maintenance versus manipulation of information held in working memory: an event-related fMRI study. *Brain and Cognition*, 41(1), 66-86.

D'Esposito, M. and B. R. Postle (2015).The cognitive neuroscience of working memory. *Annual review of psychology*, **66**: 115-142.

Dijkstra, K., Bourgeois, M. S., Allen, R. S., & Burgio, L. D. (2004). Conversational coherence: Discourse analysis of older adults with and without dementia. *Journal of Neurolinguistics*, *17*(4), 263-283.

Egidi, G., & Caramazza, A. (2014). Mood-dependent integration in discourse comprehension: Happy and sad moods affect consistency processing via different brain networks. NeuroImage, 103(0), 20-32.

Eklund, A., Nichols, T. E., &Knutsson, H. (2016). Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, 201602413.

Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). The measurement of textual coherence with latent semantic analysis. *Discourse Processes*, *25*(2-3), 285-307.

Gernsbacher, M. A. (1997). Two decades of structure building. *Discourse Processes*, *23*(3), 265-304.

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

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual review of psychology, 48*(1), 163-189.

Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in cognitive science*, 3(2), 371-398.

Graesser, A. C., McNamara, D. S., Cai, Z., Conley, M., Li, H., & Pennebaker, J. (2014). Coh-Metrix measures text characteristics at multiple levels of language and discourse. *The Elementary School Journal*, 115(2), 210-229.

Graves, W. W., Desai, R., Humphries, C., Seidenberg, M. S., & Binder, J. R. (2009). Neural systems for reading aloud: a multiparametric approach. *Cerebral cortex*, 20(8), 1799-1815.

Habeck, C., Rakitin, B. C., Moeller, J., Scarmeas, N., Zarahn, E., Brown, T., & Stern, Y. (2005). An event-related fMRI study of the neural networks underlying the encoding, maintenance, and retrieval phase in a delayed-match-to-sample task. *Cognitive Brain Research*, *23*(2–3), 207-220.

Hagoort, P., & van Berkum, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481), 801-811.

Hagoort, P. (2013). MUC (memory, unification, control) and beyond. *Frontiers in Psychology*, 4: 416.

Hammer, A., Jansma, B. M., Lamers, M., & Münte, T. F. (2008). Interplay of meaning, syntax and working memory during pronoun resolution investigated by ERPs. *Brain research, 1230*, 177-191.

Helder, A., van den Broek, P., Karlsson, J., & Van Leijenhorst, L. (2017). Neural correlates of coherence-break detection during reading of narratives. *Scientific Studies of Reading*, 1-17.

Horne, M., Hansson, P., Bruce, G., Frid, J., & Filipsson, M. (2001). Cue words and the topic structure of spoken discourse: The case of Swedish men [`]but'. *Journal of Pragmatics, 33*(7), 1061-1081.

Hsu, C. T., Jacobs, A. M., Citron, F. M., & Conrad, M. (2015).The emotion potential of words and passages in reading Harry Potter–An fMRI study. *Brain and language*, 142, 96-114.

Hung, Y.-C., & Schumacher, P. B. (2014). Animacy matters: ERP evidence for the multi-dimensionality of topic-worthiness in Chinese. *Brain research, 1555*, 36-47.

Hung, Y. C., & Schumacher, P. B. (2012). Topicality matters: Position-specific demands on Chinese discourse processing. *Neuroscience Letters, 511*(2), 59-64.

Hyönä, J. (1994). Processing of topic shifts by adults and children. *Reading Research Quarterly, 29*, 76-90

Hyönä, J. (1995). An Eye Movement Analysis of Topic-Shift Effect During Repeated Reading. [doi:]. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(5), 1365-1373.

Hyönä, J., & Lorch, R. F. (2004). Effects of topic headings on text processing: evidence from adult readers' eye fixation patterns. *Learning and Instruction, 14*(2), 131-152.

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

Kim, J.-J., Kim, M. S., Lee, J. S., Lee, D. S., Lee, M. C., & Kwon, J. S. (2002). Dissociation of working memory processing associated with native and second languages: PET investigation. *NeuroImage, 15*(4), 879-891.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363 -394.

Kirschen, M. P., Chen, S. H. A., Schraedley-Desmond, P., & Desmond, J. E. (2005). Load- and practice-dependent increases in cerebro-cerebellar activation in verbal working memory: an fMRI study. *NeuroImage, 24*(2), 462-472.

Kuperberg, G. R., Lakshmanan, B. M., Caplan, D. N., & Holcomb, P. J. (2006). Making sense of discourse: An fMRI study of causal inferencing across sentences. *NeuroImage*, *33*(1), 343-361.

Kuperberg, G. R., Paczynski, M., & Ditman, T. (2011). Establishing causal coherence across sentences: An ERP study. *Journal of Cognitive Neuroscience*, 23(5), 1230-1246.

Kuppevelt, J. V. (1995). Discourse structure, topicality and questioning. *Journal of linguistics, 31*(01), 109-147.

Li, W. (2004). Topic chains in Chinese discourse. *Discourse Processes,* 37(1): 25-45.

Li, C., Thompson, S., 1976. Subject and topic: a new typology of language. In: Li, C. (Ed.), Subject and Topic. Academic Press, New York, pp. 457–489.

Manginelli, A. A., Baumgartner, F., & Pollmann, S. (2013). Dorsal and ventral working memory-related brain areas support distinct processes in contextual cueing. *NeuroImage, 67*(0), 363-374.

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

McKoon, G., Ward, G., Ratcliff, R., & Sproat, R. (1993). Morphosyntactic and pragmatic factors affecting the accessibility of discourse entities. *Journal of Memory and Language, 32*, 56-75.

Mingtian, Z., Shuqiao, Y., Xiongzhao, Z., Jinyao, Y., Xueling, Z., Xiang, W., . . . Wei, W. (2012). Elevated amygdala activity to negative faces in young adults with early onset major depressive disorder. *Psychiatry Research: Neuroimaging, 201*(2), 107-112.

Mo, L., Liu, H. L., Jin, H., Ng, Y. B., & Lin, C. D. (2006). Passive reactivation of background information from long-term memory during reading. NeuroReport, 17(18), 1887-1891.

Mogg, K., Holmes, A., Garner, M., & Bradley, B. P. (2008). Effects of threat cues on attentional shifting, disengagement and response slowing in anxious individuals. *Behaviour research and therapy, 46*(5), 656-667.

Mumford, J. A., Poline, J.-B., & Poldrack, R. A. (2015). Orthogonalization of regressors in fMRI models. *PloS one, 10*(4), e0126255.

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

Narayanan, N. S., Prabhakaran, V., Bunge, S. A., Christoff, K., Fine, E. M., & Gabrieli, J. D. E.

(2005). The Role of the Prefrontal Cortex in the Maintenance of Verbal Working Memory: An Event-Related fMRI Analysis. *Neuropsychology, 19*(2), 223-232.

Nieuwland, M. S., Otten, M., & Van Berkum, J. J. A. (2007a). Who are you talking about? Tracking discourse-level referential processing with event-related brain potentials. *Journal of Cognitive Neuroscience*, 19(2), 228-236.

Nieuwland, M. S., Petersson, K. M., & Van Berkum, J. J. A. (2007b). On sense and reference: Examining the functional neuroanatomy of referential processing. *NeuroImage*, 37(3), 993-1004.

Nieuwland, M. S., & Van Berkum, J. J. A. (2008). The interplay between semantic and referential aspects of anaphoric noun phrase resolution: Evidence from ERPs. *Brain and Language*, 106(2), 119-131

Nyberg, L., & Eriksson, J. (2015). Working Memory: Maintenance, Updating, and the Realization of Intentions. *Cold Spring Harb Perspect Biol*, *8*(2), a021816.

O'Brien, E. J., Duffy, S. A., & Myers, J. L. (1986). Anaphoric inference during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*(3), 346 -352.

Oberlander, J. (2004). On the reduction of discourse topic. *Theoretical Linguistics, 30*(2-3), 213-225.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N‐back working memory paradigm: A meta‐analysis of normative functional neuroimaging studies. *Human brain mapping*, *25*(1), 46-59.

Palladino, P., Cornoldi, C., De Beni, R., & Pazzaglia, F. (2001). Working memory and updating processes in reading comprehension. *Memory & cognition*, 29(2), 344-354.

Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Social cognitive and affective neuroscience*, 2(1), 67-70.

Rao, L.-L., Zhou, Y., Xu, L., Liang, Z.-Y., Jiang, T., & Li, S. (2011). Are risky choices actually guided by a compensatory process? New insights from FMRI. *PloS one, 6*(3), e14756.

Rinck, M. and U. Weber (2003). Who when where: An experimental test of the event-indexing model. *Memory & cognition*, 31(8): 1284-1292.

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., . . . Eickhoff, S. B. (2012). Modelling neural correlates of working memory: A coordinate-based meta-analysis. *NeuroImage, 60*(1), 830-846.

Rinck, M., & Weber, U. (2003). Who when where: An experimental test of the event-indexing model. *Memory & cognition*, 31(8), 1284-1292.

Sanders, T., & Maat, H. P. (2006). Cohesion and Coherence: Linguistic Approaches. *reading, 99*, 440-466.

Shaywitz, B. A., Shaywitz, S. E., Pugh, K. R., Fulbright, R. K., Skudlarski, P., Mencl, W. E., . . . Gore, J. C. (2001). The Functional Neural Architecture of Components of Attention in Language-Processing Tasks. *NeuroImage, 13*(4), 601-612.

Schuster, S., Hawelka, S., Hutzler, F., Kronbichler, M., &Richlan, F. (2016). Words in context: The effects of length, frequency, and predictability on brain responses during natural reading. *Cerebral Cortex*, 26(10), 3889-3904.

Takahama, S., Miyauchi, S., & Saiki, J. (2010). Neural basis for dynamic updating of object representation in visual working memory. *NeuroImage, 49*(4), 3394-3403.

Tong, X., Tong, X., & McBride-Chang, C. (2014). Tune in to the tone: Lexical tone identification is associated with vocabulary and word recognition abilities in young Chinese children. *Language and Speech*,58(4), 441-458.

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

van den Broek, P., & Helder, A. (2017). Cognitive Processes in Discourse Comprehension: Passive Processes, Reader-Initiated Processes, and Evolving Mental Representations. *Discourse Processes*, 1-13

Van den Broek, P., Rapp, D. N., & Kendeou, P. (2005). Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes, 39*(2-3), 299-316.

Veltman, D. J., Rombouts, S. A. R. B., & Dolan, R. J. (2003). Maintenance versus manipulation in verbal working memory revisited: an fMRI study. *NeuroImage*, *18*(2), 247-256.

Virtue, S., Haberman, J., Clancy, Z., Parrish, T., & Jung Beeman, M. (2006). Neural activity of inferences during story comprehension. *Brain research, 1084*(1), 104-114.

Virtue, S., Parrish, T., & Jung-Beeman, M. (2008). Inferences during story comprehension: Cortical recruitment affected by predictability of events and working memory capacity. *Journal of Cognitive Neuroscience, 20*(12), 2274-2284.

Wagner, A. D., Paré-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering Meaning: Left Prefrontal Cortex Guides Controlled Semantic Retrieval. *Neuron*, 31(2), 329-338.

Wang, Y., & Quadflieg, S. (2015). In our own image? Emotional and neural processing differences when observing human–human vs human–robot interactions. *Social cognitive and* *affective neuroscience*, 10(11), 1515-1524.

Yang, X., Chen, X., Chen, S., Xu, X., & Yang, Y. (2013). Topic Structure Affects Semantic Integration: Evidence from Event-Related Potentials. *PloS one, 8*(12), e79734.

Yang, X., Shuang, C., Chen, X., & Yang, Y. (2015). How Distance Affects Semantic Integration in Discourse: Evidence from Event-Related Potentials. *PloS one, 10*(11).

Yarkoni, T., Speer, N. K., & Zacks, J. M. (2008). Neural substrates of narrative comprehension and memory. *NeuroImage, 41*(4), 1408-1425.

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

Highlights

1 Discourse features guided the involvement of working memory subcomponents.

2 The left IFG was recruited for coherence processing.

3 The left IFG and MFG, bilateral SFG, and bilateral IPL responded to context length.

4 The left IFG played a central role in coping with discourse feature variations.