

RICE UNIVERSITY

**The role of working memory in interference resolution during
Chinese sentence comprehension: Evidence from event-related
potentials (ERPs)**

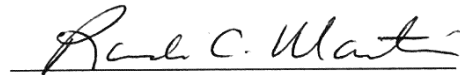
by

Yingying Tan

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

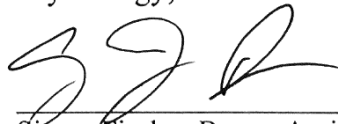
APPROVED, THESIS COMMITTEE



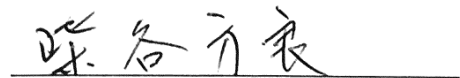
Randi C. Martin, Elma Schneider
Professor, Chair
Psychology



David Lane, Associate Professor
Psychology, Statistics and Management



Simon Fischer-Baum, Assistant Professor
Psychology



Matt Shibatani, Deedee McMurtry Professor
in Humanities
Linguistics

HOUSTON, TEXAS
[DECEMBER 2015]

ABSTRACT

The role of working memory in interference resolution during Chinese sentence comprehension: Evidence from event-related potentials (ERPs)

by

Yingying Tan

Interference during sentence comprehension occurs when readers use semantic and syntactic cues to retrieve earlier sentence information to integrate with later information and intervening material partially matches these cues, resulting in more parsing difficulty. This thesis collected event-related brain potentials (ERPs) while participants processed Chinese sentences with semantic and syntactic interference to address two main questions: 1) When and how do semantic and syntactic interference effects interact with each other? 2) What is the role of working memory (WM) mechanisms in interference resolution? Semantic and syntactic interference were examined during processing of the main verb (e.g., “*complain*”) that required the retrieval of a human subject noun. The degree of semantic interference was manipulated through varying the semantic plausibility of a distracting noun (e.g., human vs. non-human), and syntactic interference was manipulated through varying the distracting noun’s grammatical role (e.g., subject vs. object). Individual differences measures were collected on aspects of WM, executive function, and verbal knowledge. The ERP results at the critical verb showed negative anterior effects between 300 – 500 ms for both syntactic and semantic interference. Syntactic interference also induced a P600 effect and semantic interference also induced a late left anterior negativity. I interpret the early anterior negativities as reflecting a first stage process of detecting the semantic or syntactic interference, and the late ERPs as

reflecting a second stage of reanalysis during sentence processing. Importantly, the current results demonstrated that semantic processing plays an immediate and important role in Chinese, because the semantic interference effect was observed as early as the syntactic interference effect even when the distracting noun's syntactic features strongly eliminated it from the distractor set. In contrast, semantic interference has been shown to be delayed and even blocked in previous English studies. Additionally, the present study supports a role of attentional control underlying sentence comprehension. Subjects with better resistance to proactive interference had less difficulty in syntactic interference resolution, as indexed by a reduced mean amplitude of the P600 effect. This result is consistent with the argument that attentional control helps subjects to recover from interference during later controlled aspects of sentence processing.

Acknowledgements

First, I would like to express my sincerest gratitude and appreciation to my supervisor, Dr. Randi Martin, who has supported me with her patience, encouragement, and vast knowledge throughout my years at Rice. Randi has not only spent a significant amount of time helping me to sharpen my research skills, but also challenging me to think about the broader theoretic issues in psycholinguistic studies. I have been extremely fortunate to have Randi as my advisor. From Randi, I learned how to be a precise, rigorous, and responsible scientist, and conduct beautiful research.

I would like to thank my committee – Dr. Simon Fischer-Baum, Dr. David Lane, and Dr. Matt Shibatani, as well as the alternate member Dr. Tatiana Schnur, for their invaluable comments and insightful questions. I want to especially thank Simon, who has discussed everything from data analyzing to teaching strategies to future research directions with me. He has been a great source of inspirations. I would also thank David for providing me with statistical advice all these years, and Matt for providing knowledge from the linguistics perspective.

Many thanks are due to Dr. Xiaolin Zhou, my undergraduate advisor, for being a constant source of encouragements and help. He has generously offered me the opportunity to collect ERP data in his lab at Peking University. This work would not have been possible without the support from him.

I would like to acknowledge both past and present lab members of the Brain and Language Lab for their generous help both within and outside of my research life. Thanks to Dr. Cris Hamilton, Dr. Corinne Pittigrew, Dr. Yi Glaser, Azli Hassan, Heather Dial,

Hao Yan, Qiuhai Yue, and Curtiss Chapman. They have provided such a warm and supportive lab environment. Additionally, I would like to thank past and present lab members of the Center for Brain and Cognitive Sciences, Dr. Zheng Ye, Dr. Xiaoming Jiang, Dr. Yingyi Luo, Dr. Xi Chen, and Yunyan Duan, who helped me with the ERP data analysis. I learned a lot from our discussions over the years.

Last but not the least, I would like to thank my dear friends, Yi Da, Jiexin Cao, Xuebei Yang, Jennifer Chen, Anna Cragin, Fangbo Xu, Minyan Zhu, Xiaozhu Lin, Debshila Basu Mallick, and many many others, who made my life in Houston richer, better, and amazing. I would also thank the sweet secretaries of the Psychology Department, Lanita Martin and Jennifer Gucwa, who helped with every aspect of my life at Rice and made the department so heartwarming. I am going to miss everyone.

To my parents, who have cultivated my interest in science, I owe my special thanks to them for unending, unconditional love and support. Their supporting helps me go through the most difficult times. I have fractured my leg the third day after I went back to China for the data collecting. It was an extremely tough and frustrating time. In addition to the heavy working load my mother had, she had to taking care of my daily life, drive me to the lab every morning, and picked my up in the evening. For my father, although I think he may never fully understand the topic I am working on, he always showed great interests in my research and provided valuable insights as a physicist. Finally, I would like to thank my boyfriend Yiyang, who has been consistently supporting me with his incredible patience, wisdom, and love in all times.

In addition, I would like to thank Rice University Graduate Fellowship for supporting through most of my graduate school, the SSRI dissertation research improvement grant, Lodieska Stockbridge Vaughn fellowship, and Gertrude Maurin Fund for supporting my dissertation study.

Table of Contents

1	Introduction.....	11
1.1	Interference in sentence comprehension	11
1.1.1	<i>Cue-based parsing approach.....</i>	<i>14</i>
1.2	Cognitive mechanisms supporting interference resolution in sentence processing	36
1.2.1	<i>Capacity-based accounts of sentence processing.....</i>	<i>36</i>
1.2.2	<i>Non-capacity accounts of sentence comprehension</i>	<i>41</i>
1.2.3	<i>WM capacity, executive control, reading experience and their relation to interference resolution in sentence comprehension</i>	<i>49</i>
1.2.4	<i>Evidence from cross-linguistic studies about the memory mechanisms underlying interference resolution: Chinese sentence processing.....</i>	<i>54</i>
1.2.5	<i>Evidence from ERP studies of relative clause processing.....</i>	<i>74</i>
1.3	Exploring individual differences in sentence processing efficiency	80
1.4	Motivation for the present study	89
1.4.1	<i>Some special properties of Mandarin Chinese</i>	<i>90</i>
1.4.2	<i>Examining interference resolution using the ERP techniques.....</i>	<i>93</i>
1.4.3	<i>Predictions</i>	<i>94</i>
2	Experiment	98
2.1	Subjects	98
2.2	Materials, design, and procedure	98
2.2.1	<i>Sentence materials</i>	<i>98</i>
2.3	Procedures	102
2.3.1	<i>Sentence comprehension task</i>	<i>102</i>
2.3.2	<i>Cognitive capacity tests</i>	<i>105</i>
2.4	EEG recording and data analysis	113

3	Results	118
3.1	Comprehension question accuracy.....	118
3.2	Event-related potentials (ERPs)	119
3.2.1	<i>Critical verb</i>	120
3.2.2	<i>Spillover region</i>	130
3.2.3	<i>Exploratory analysis: sentence final region</i>	136
3.2.4	<i>ERP effects at the RC head noun and adverbial phrase</i>	143
3.2.5	<i>Summary of ERP effects during sentence processing</i>	155
3.3	Relations between individual differences measures and interference effect in ERPs	158
3.3.1	<i>Relations between the individual differences measures</i>	158
3.3.2	<i>Compute composite scores</i>	165
3.3.3	<i>Simple correlations between individual differences measures and semantic and syntactic interference resolution</i>	167
3.3.4	<i>Multiple regression approach</i>	172
3.3.5	<i>Relations between composite measures and RC processing</i>	175
3.4	Summary of major results	176
4	Discussion	180
4.1	Interference resolution during sentence comprehension.....	181
4.1.1	<i>Critical verb</i>	181
4.1.2	<i>Spillover region</i>	195
4.2	Time course of semantic and syntactic interference	200
4.3	Relations of semantic and syntactic interference effects to WM capacity.....	210
4.4	Chinese relative clause processing.....	221
4.5	Limitations of the study and future directions	227
4.6	Conclusion.....	231

List of Tables

Table 1 Predictions from different accounts about the relations between individual differences measures and interference resolution processes during sentence processing.	96
Table 2 Meaningfulness rating of the experimental sentences	102
Table 3 Descriptive data of mean accuracy (proportion correct) in sentence comprehension task..	119
Table 4 Results of repeated ANOVAs for the mean amplitude of the critical verb region.	125
Table 5 Results of repeated ANOVAs for the mean amplitude of the spillover region.	131
Table 6 Results of repeated ANOVAs for the mean amplitude of the sentence final region.	139
Table 7 Results of repeated ANOVAs for the mean amplitude of the RC head noun.	145
Table 8 Results of repeated ANOVAs for the mean amplitude of the adverbial phrase.	149
Table 9 Summary of the time course and scalp distribution of each reliable ERP component on the critical verb, spillover region, and sentence final position.	157
Table 10 Descriptive data and reliability estimates for all the individual differences measurements	159
Table 11 Full correlation matrix for the individual differences measures.....	161
Table 12 Full correlation matrix of the correlation tests between factors	166
Table 13 Correlations between interference effect size on the critical verb and spillover region (or SRC-advantage effect on the head noun) and composite score of individual difference measures.....	168
Table 14 Multiple regressions on the interference effect and individual differences measures	173

List of Figures

Figure 1 Grand average ERPs for each experimental condition time-locked to the onset of critical verb from -200 to 800 ms, at all electrodes included in the statistical analyses (negativity is plotted up).	127
Figure 2 Grand average ERPs for each experimental condition time-locked to the onset of critical verb from -200 to 800 ms, at midline electrode position Fz, FCz, CZ, CPz, and Pz (negativity is plotted up).....	128
Figure 3 Grand average ERPs for semantic (averaged across low and high syntactic interference conditions) and syntactic (averaged across low and high semantic interference conditions) interference effects time-locked to the onset of critical verb from -200 to 800 ms, at electrode position F5 and F6 (negativity is plotted up).	129
Figure 4 Grand average ERPs for each experimental condition time-locked to the onset of the first word following the critical verb (i.e., spillover region) from -200 to 800 ms, at all electrodes included in the statistical analyses (negativity is plotted up).	133
Figure 5 Grand average ERPs for each experimental condition time-locked to the onset of the first word following the critical verb (i.e., spillover region) from -200 to 800 ms, at electrode position Fz, FCz, CZ, and CPz (negativity is plotted up). As there was no interaction of interference effects with electrodes, a bar graph averaged across the five midline electrodes was presented in the bottom right corner for the mean amplitude between 250 – 800 ms. The error bars representing corrected standard error of the mean (Cousineau, 2005).....	134

Figure 6 Grand average ERPs for each experimental condition time-locked to the onset of the final word (i.e., the second word after the critical verb) from -200 to 800 ms, at all electrodes included in the statistical analyses (negativity is plotted up).	140
Figure 7 Grand average ERPs for syntactic interference effect (averaged across semantic interference conditions) time-locked to the onset of the final word from -200 to 800 ms, at electrode position F5, F6, and FCz (negativity is plotted up).	141
Figure 8 Grand average ERPs for each experimental condition time-locked to the onset of the head noun of the RC (e.g., " <i>resident</i> " in the Example 13) from -200 to 800 ms, at all electrodes included in the statistical analyses (negativity is plotted up).	146
Figure 9 Grand average ERPs for syntactic interference effect (averaged across semantic interference conditions) and semantic interference effect (averaged across syntactic interference conditions) time-locked to the onset of the RC head noun from -200 to 800 ms, at electrode position FC3, FC4, Fz, P3, and P4 (negativity is plotted up).	147
Figure 10 Grand average ERPs for each experimental condition time-locked to the onset of the adverbial phrase before the critical verb (i.e., " <i>everyday</i> " in the Example 13) from -200 to 800 ms, at all electrodes included in the statistical analyses (negativity is plotted up). .	150
Figure 11 Grand average ERPs for syntactic interference effect (averaged across semantic interference conditions) time-locked to the onset of the adverbial phrase before critical verb from -200 to 800 ms, at electrode position Fz, FC3, FC4, P3, and P4 (negativity is plotted up).....	151
Figure 12 Scatter plots of the significant correlations between the mean amplitude of the ERP effects and individual differences measures. Fig. a shows the relation of recent negatives effect and syntactic interference-elicited P600 in mean amplitude (μ V) in the critical verb region (e.g., " <i>complain</i> " in Example 13), collapsed across semantic interference conditions. Fig. b shows in the low semantic interference.....	171

The role of working memory in interference resolution during Chinese sentence comprehension: Evidence from event-related potentials (ERPs)

1 Introduction

1.1 Interference in sentence comprehension

Sentence comprehension is a highly incremental process in which the semantic and syntactic features of each word are accessed immediately as the word is perceived and these features are used in integrating current information with earlier information (Tanenhaus & Truswell, 1995). During sentence processing, it is often the case that later parts of a sentence have to be integrated with earlier parts that can be separated by an indefinite amount of materials. The ubiquitous presence of nonadjacent syntactic dependencies - that is, the one-to-one correspondence between two syntactically bound elements (e.g., subject - verb) at some distance from each other- indicates that some type of memory representation is needed for successful sentence comprehension. A number of studies have revealed that the memory mechanism underlying language processing operates according to similar principles as the memory system that subserves other memory-dependent tasks (Fedorenko, Gibson, & Rohde, 2006, 2007; Just & Carpenter, 1992; Van Dyke & McElree, 2006). In the literature on memory, numerous studies have

shown that both decay (Bartek, Lewis, Vasishth, & Smith, 2011; Brown, 1958; Conrad & Hille, 1958) and interference (Baddeley & Hitch, 1977; Watkins & Watkins, 1975) are primary determinants of forgetting. Therefore, one might expect that decay and interference within a sentence would also impair the ability to make appropriate linkages between earlier and later information in a sentence. However, most of the previous studies have focused on how the decay of the memory traces with the passage of time or with greater distance (in terms of more intervening elements) between the dependencies increases sentence processing difficulty (Gibson, 1998; Grodner & Gibson, 2005), although there are still disputes on whether decay plays any role in sentence comprehension (Konieczny, 2000; Vasishth & Lewis, 2006). On the other hand, the role of interference as a contributor to sentence comprehension has only recently gained attention (Gordon, Hendrick, & Johnson, 2001, 2004; Gordon, Hendrick, Johnson, & Lee, 2006; Van Dyke, 2007; Van Dyke & McElree, 2006, 2011).

As with interference in the memory domain, interference effects during sentence processing occur when there are similar items in memory. Interference effects can be either proactive or retroactive (Van Dyke, 2007; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2011). Retroactive interference (RI) results from later information interfering with earlier information and causes forgetting of older material. For example, as shown in sentence 1a, the distractor “*witness*”, which is interpolated between the dependency of the main verb “*compromised*” and its target subject “*attorney*”, would cause retroactive interference, when attempting to integrate “*attorney*” and “*compromised*” to form a grammatically dependent pair.

Example 1. Retroactive and proactive interference.

1a. Retroactive interference

The *attorney* who had rejected the witness yesterday compromised.

1b. Proactive interference

The *attorney* who arrived yesterday realized that the witness compromised.

Proactive interference (PI) results from previous information interfering with later information and causes difficulty in retaining new memories. As shown in sentence 1b, the distractor “*attorney*”, which is processed before the target subject “*witness*”, could cause PI in linking the complement clause verb “*compromised*” with its subject “*witness*”. PI could be induced with either sentence-internal words as shown in Example 1b (Van Dyke, 2007; Van Dyke & McElree, 2011), or by requiring subjects to maintain extra items (e.g., a list of words) presented for later recall before reading a sentence (Gordon et al., 2004; Gordon, Hendrick, & Levine, 2002). For the latter condition, for example, subjects might be required to maintain a word such as “*witness*” before reading a sentence such as “*The attorney who arrived yesterday compromised*”. Studies using the extraneous memory load manipulation showed that subjects’ decrease in sentence comprehension performance was not caused by a general effect of the secondary task (i.e., maintaining a list of words), but by the interference effects engendered from memorizing words sharing properties with the critical words in the sentences (e.g., both “*witness*” and “*attorney*” are role names, and are semantically plausible as the subject of “*compromise*”) (Gordon et al., 2002; Van Dyke & McElree, 2006). Generally, subjects show lower comprehension accuracy and longer reading time for sentences with high interference.

1.1.1 Cue-based parsing approach

Several studies have investigated the underlying cause of interference effects during sentence comprehension. Two proposals have been put forward. One is that interference arises from the confusability of the nouns in memory (Gordon et al., 2001, 2004; Gordon et al., 2006; Gordon et al., 2002). With respect to this position, Gordon and colleagues have conducted a series of studies investigating interference effects during sentence processing by manipulating the referential characteristics (i.e., common noun, proper names, pronoun) of noun phrases (NPs). They manipulated the referential characteristics of both sentence-internal NPs (Gordon et al., 2001, 2004), such as “*The banker that praised the barber/Joe/you/everyone climbed the mountain*”, and sentence-external NPs (Gordon et al., 2006; Gordon et al., 2002), for instance, as shown in Example 2, subjects were instructed to memorize a list of words such as “*Joel-Greg-Andy*” prior to reading sentences.

Example 2. (Gordon et al., 2002)

Memory-load set: Joel-Greg-Andy

2a. It was [Tony] that [Joey] liked before the argument began.

2b. It was [the dancer] that [the fireman] liked before the argument began.

In the matched condition, the NPs were matched in referential type (e.g., in Example 2a, both NPs were proper names, *Joel - Tony*), while the NPs were from different referential categories in the unmatched condition (e.g., *Joel - dancer*). The general finding from Gordon et al.’s studies (2001, 2002, 2004) is that reading performance suffered more in the matched condition, in which participants showed higher error rates to comprehension questions and longer online reading times. Therefore,

Gordon and colleagues attributed these results to an increase in similarity-based interference: when the two NPs had similar memory representations; such similarity would cause interference in retrieving target information. These results supported the claim that interference is an important determinant of sentence comprehension difficulty.

However, Lewis, McElree, and Van Dyke (Lewis, 1996; Lewis, Vasishth, & Van Dyke, 2006; Van Dyke, 2007; Van Dyke & McElree, 2006, 2011) have proposed a different approach, namely the cue-based parsing approach, which instead focuses on the match between the retrieval cues generated by a word and the features of the preceding noun phrases (NPs) that might be potential targets for integration with the verb, rather than the similarity of the nouns per se. According to the cue-based parsing approach, sentence parsing is accomplished through an efficient series of guided memory retrievals. These researchers assume that during sentence comprehension, the number of items that can be actively maintained in the focus of attention is extremely limited (only one or two items). Items outside the focus of attention must be retrieved into the focus of attention for processing. Thus, the processing of each word generates cues that are used to access preceding information to integrate with the incoming information. Retrieval cues are a subset of the features of the item to be retrieved, and they are derived from the incoming word, context, and grammatical knowledge. Previous studies have supported the view that the retrieval process works through a cue-driven, direct-access mechanism, and is not affected by the number of intervening words between the non-adjacent constructions (Lewis et al., 2006; Van Dyke, 2007; Van Dyke & McElree, 2006, 2011). Reliable retrieval cues are associated with a single, unique item in memory. However, interference occurs as a consequence of insufficiently distinct cues being available at retrieval – that

is, when the retrieval cues partially match the features of non-target information in the sentence. For example, for sentence 1a “*The attorney who had rejected the witness yesterday compromised*”, in order to understand the verb phrase “*compromised*”, readers need to find its grammatical subject. Although the “*attorney*” is the subject of this action of this action, interference arises from the intervening NP “*witness*” because it partially satisfies the retrieval cues generated by the verb (i.e., animate NP) even though “*witness*” has already been linked as the object/experiencer of “*rejected*.”

1.1.1.1 Capacity limits and the focus of attention

One of the core assumptions about cue-based parsing approach is that only an extremely limited number of items can be maintained in a highly accessible state. With respect to memory retrieval, McElree (2006), along with several recent memory researchers (Cowan, 2000; Lewis, 1996; Oberauer, 2002), have made a distinction between information stored within and outside the focus of attention. Items stored in the focus of attention afford privileged access and thus do not require retrieval, while the items outside of attention must be retrieved into the focus of attention for processing. However, there are differences among these researchers regarding the storage limits of the focus of attention. Cowan (2000) argued for a tripartite model of memory, including focal attention, activated portion of long-term memory (LTM) outside of focal attention (i.e., working memory, WM), and LTM. After summarizing a wide variety of data on short-term memory (STM) capacity limits, Cowan suggested that people could only maintain three to four chunks of information in the focus of attention regardless of presentation modality (Cowan, 2000; Cowan et al., 2005; Saults & Cowan, 2007). However, Lewis (1996) claimed that sentence processing requires the maintenance of no

more than two chunks, and many sentence processing phenomena could be explained as interference effects in an extremely limited WM system. Lewis and colleague (Lewis, 1996; Lewis & Vasishth, 2005) have successfully applied a theory taking this assumption to explain several core sentence processing phenomena, such as difficulty in processing center embedded structures (e.g., “*The salmon that the man that the dog chased smoked fell*”). Nonetheless, recently, McElree (2006) argued for a bipartite model of memory with two different representational states for this information: focal attention and LTM. McElree made a clear distinction between information stored within the focus of attention and information passively stored in LTM, but not a further distinction between WM and LTM. He studied the capacity of the focus of attention through the analysis of speed-accuracy trade-offs (SATs)¹ in recognition memory. McElree (2006) found that speed of access and rate of retrieval of information showed a dichotomous pattern across serial positions – both were faster for a probe matching the last item in a list than for all other serial positions and there were no differences in speed or rate among the other serial positions. Thus, McElree estimated the capacity of the focus of attention to be only one chunk². In many tasks, such as sentence comprehension, the information maintained in focal attention will typically be the last item or chunk of information the reader just processed (McElree, 2006; McElree, Foraker, & Dyer, 2003). Despite the differences in these models about the capacity limits of the focus of attention, these researchers agree

¹ In the SAT paradigm, participants are presented with a stimulus (e.g., a sentence), followed by a probe at a variable lag. They were required to make a two-alternative choice within a short time. The response accuracy is recorded as a function of response speed. The asymptotic d' of their performance is considered to reflect the availability of an item, and the intercept d' of their performance is considered to reflect the accessibility of an item.

² Some researchers have argued that McElree’s claim must allow at least two items – the last item and the probe - being stored in the focus of the attention to account for the list-final effect (Caplan et al., 2013).

on the notion that sentence processing is supported by a sharply limited memory system and the retrieval of information outside the focus of attention is content addressable.

1.1.1.2 The locus of interference effects

As in other studies on memory, the cue-based parsing approach carves up memory processes underlying sentence processing into stages of encoding, storage, and retrieval. Even now, there are still debates about whether interference occurs during encoding and storage or during the retrieval process. If interference occurs during encoding, then a slowdown in reading times should be observed as soon as the similar item is encountered in the pre-retrieval region (e.g., in processing “*Joey*” in “*It was Tony that Joey liked*” in Example 2). On the other hand, if interference occurs during the retrieval process, when the dependency between “*liked*” and “*Joey*” is resolved, then a slowdown in reading times should be observed at the verb but not in the pre-retrieval region. Although Gordon et al. (2002) did observe longer reading times in the pre-retrieval region, Van Dyke and McElree (2006) pointed out that Gordon et al. did not control for either word frequency or length between different conditions. For example, Gordon et al. directly compared the reading time between proper names such as “*Tony*” and roles such as “*the dancer*” as in Example 2. Thus, the source of the longer reading times was ambiguous. In order to clearly distinguish between an encoding effect and a retrieval effect, Van Dyke and McElree directly manipulated the nature of the retrieval cues rather than the similarity of referential properties shared by different memory representations. They extended the memory load paradigm used by Gordon and colleagues thorough manipulating both memory load (Memorize a list of words vs. No load) and the degree of interference (Interference vs. No interference). In the interference

condition, the nouns in the memorized list were plausible objects of the sentential verb, whereas they were implausible objects of the sentential verb in the no interference condition. For example, all the words in memory list are *fixable* in sentence 3a, whereas none of these words are possible objects of the verb “*sailed*” in sentence 3b.

Example 3 (Van Dyke & McElree, 2006).

Memory list: table-sink-truck

3a. [Interference] It was the boat that the guy who lived by the sea [**fixed**] in two sunny days.

3b. [No interference] It was the boat that the guy who lived by the sea [**sailed**] in two sunny days.

Van Dyke and McElree (2006) found that reading times on the relative clause verb increased significantly in the interference condition as compared to the no interference condition. More importantly, since the encoding conditions were constant in this experiment, Van Dyke and McElree argued that such results indicated that the interference effect in sentence comprehension occurs at the retrieval stage, which specifically arises when the retrieval cues generated by the verb partially overlap with the distractor NPs in some required features (e.g., “*fixable*” as a retrieval cue generated by the verb “*fix*”). They explained the similarity-based interference effect in Gordon et al.’s study (2002, 2004) as due to the fact that NP from the same types (i.e., proper names or NPs referring to a role) shared more referential properties and thus, reduced the retrieval cues’ relative ability to distinguish between them. For instance, as shown in Example 2, although the retrieval cues at the verb “*liked*” remain the same across conditions and may not specify what type of NP it was looking for, that higher similarity of referential characteristics between distractors and targets in the matched condition diminished the retrieval cues’ relative ability to distinguish between them. However, Van Dyke and McElree admitted that it was not very clear what kind of retrieval cues were used to

distinguish between the words, since the grammatical type of the NPs (e.g., pronoun, proper names, etc.) should not be weighed as crucial retrieval cues here, and further investigation is needed³. Therefore, although the effect of interference on retrieval has been confirmed in several later studies (Van Dyke, 2007; Van Dyke & McElree, 2011; Van Dyke, Johns, & Kukona, 2014), it could not fully rule out the possibility that interference might additionally arise from the encoding process and further investigation is required.

1.1.1.3 Different types of interference

As discussed in an earlier section, there are two classical types of interference: proactive interference and retroactive interference, both of which have been found to affect sentence processing. Interference during sentence processing can also be divided into semantic interference and syntactic interference (Van Dyke, 2007; Van Dyke & McElree, 2006, 2011). Semantic interference occurs when distractors partially match semantic retrieval cues whereas syntactic interference happens when distractors partially match syntactic retrieval cues. Van Dyke (2007) manipulated both semantic and syntactic interference and found that syntactic interference effect occurs earlier than semantic interference effect. As shown in Example 3, while processing the main verb phrase “*was complaining*”, readers need to retrieve its grammatical subject “*resident*”. Compared to sentence 4a, there is more syntactic interference in sentence 4b because the intervening

³ Van Dyke and McElree (2006) have pointed out that “fully specifying all the cues that might drive memory retrieval is beyond the scope of our current understanding, as it requires enumerating and investigating all forms of information that might be computed during sentence processing. However, general claims about cue-based retrieval in parsing do not depend on any particular assumed set of features. Rather the relevance of a cue-based approach is motivated by demonstrations that comprehenders have available some set of features that capture the distinctions necessary for identifying various grammatical dependencies within the sentence (page 164)”. As a result, we assumed that the grammatical type of NPs (e.g., pronoun, proper names, etc.) was not weighted as crucial retrieval cues here, and further hypothesized that the retrieval cues are the same across condition.

NP “*warehouse/neighbor*” plays a subject role in the relative clause while it is a prepositional object in sentence 4a. Semantic interference can be seen by the contrast between using “*neighbor*” vs. “*warehouse*” in these two examples. There is more semantic interference in sentences with “*neighbor*” in the relative clause in that only “*neighbor*” fits the semantic retrieval cues of the verb phrase “*was complaining*” (i.e., a *warehouse* cannot complain).

Example 4. (Van Dyke, 2007)

4a. Low syntactic interference

The worker was surprised that the resident who was living near the dangerous [warehouse/neighbor] was complaining about the investigation.

4b. High Syntactic interference

The worker was surprised that the resident who said that the [warehouse/neighbor] was dangerous was complaining about the investigation.

There were two main findings from this experiment. First, Van Dyke observed both semantic and syntactic interference effects in terms of lower accuracy to probe questions and longer reading times in high interference conditions in the eye-movements data. Semantic interference occurred even when the syntactic properties were inappropriate (that is, the interfering noun could not serve as the subject of the verb). Second, the time course of these two interference effects was different: the syntactic interference effect occurred earlier than the semantic interference effect. Results from her study showed that the syntactic interference effect was obtained at the main VP (e.g., “*was complaining*” as in Example 4) for all eye-tracking measures (first pass, regression path, total reading time, and proportion of regressions back), whereas the semantic

interference effect was only observed later in the final region of the sentence (e.g., “*investigation*” as in Example 4) for the regression path measure. The finding that syntactically unavailable but semantically appropriate nouns (e.g., “*neighbor*”) showed interference is crucial to understanding the relationship between language processing and more general memory mechanisms. Van Dyke suggested that these findings present a challenge for grammar-driven parsers (Fodor & Frazier, 1980) which assume that syntactic processing has priority and thus would not predict semantic interference effects elicited by words that have already been integrated into the existing parse tree.

However, in a later study, Van Dyke and McElree (2011) failed to observe a semantic interference effect when the semantic distractor occurred as a direct object to the relative clause verb, instead of occurring as a prepositional object in Van Dyke (2007). For instance, for the sentence “*The attorney who the judge realized had rejected the [motion/witness] in the case compromised*”, no semantic interference was observed between the two conditions. Van Dyke et al. explained such differences by referring to the hierarchical distinctions between core arguments and other modifying adjuncts in English. They suggested that the syntactic features of the distractor that showed up in a core argument (e.g., direct object of relative clause as in Van Dyke & McElree, 2011) were strong enough to eliminate that noun from the distractor set, while the syntactic features that showed up in modifying adjuncts (e.g., prepositional object of embedded clause as in Van Dyke, 2007) were not strong enough to block that noun from the

distractor set⁴. Further studies will be required to fully address how and when different types of retrieval cues are bound together.

1.1.1.4 Neural basis of interference resolution

So far, few studies have directly looked into the neural basis of interference resolution during sentence processing. The general implication drawn from some relevant studies is that the left inferior frontal gyrus (LIFG) plays a critical in semantic selection (Kan & Thompson-Schill, 2004a, 2004b; Vuong & Martin, 2011), syntactic selection (Chen, West, Waters, & Caplan, 2006; Novick, Trueswell, & Thompson-Schill, 2005), or a general role in selecting among competing information in all domains (Hagoort, 2005; Rodd, Longe, Randall, & Tyler, 2010). In a recent functional magnetic resonance imaging (fMRI) study, Glaser et al. (2013) used materials with semantic and syntactic interference manipulations similar to those used in Van Dyke (2007) and investigated the neural basis involved in resolving syntactic and semantic interference. Glaser et al. (2013) found that interference resolution during sentence comprehension activated LIFG, with dissociable regions within LIFG supporting semantic or syntactic interference resolution. A region in the LIFG (BA 45) was involved in the resolution of both semantic and syntactic interference, while BA 47 was only involved in the resolution of semantic interference and BA 44 was only involved in the resolution of syntactic interference (Glaser, Martin, Van Dyke, Hamilton, & Tan, 2013). Moreover, similar to Van Dyke's

⁴ Van Dyke and McElree suggested that such differences might be explained by the fact that language processing mechanisms may weigh different types of information differently in certain contexts. The direct object is a core argument in the sentence, which is a required component and can be suppressed or added in different ways (e.g., passivization), while the prepositional object is in a modifying adjunct, which is an optional or structurally dispensable phrase. The direct object plays a more prominent role at the interface between syntactic and semantic processing than do modifying adjuncts, because the direct object specifies the thematic role that the NP plays (i.e., who-did-what-to-whom).

finding (2007), syntactic interference was observed earlier than semantic interference in this study - syntactic interference occurred during sentence reading but semantic interference only occurred during the answering of comprehension question. However, restricted by the relative low temporal resolution of the fMRI technique, these results could not distinguish online sentence processing occurring right on the critical verb, where interference effects were predicted to occur, from effects occurring in other sentence regions. The time course pattern implicated in this study needs be followed up by studies employing better temporal resolution technique, such as electroencephalography and magnetoencephalography.

1.1.1.4.1 ERP components related to language processing

In addition to fMRI studies, the event-related potential (ERP) technique, which was used in the current experiment, is a powerful tool for investigating sentence comprehension processes. It measures electrical activity in the brain through electrodes placed on the scalp across different brain regions. The high temporal resolution (on the order of milliseconds) and well-established linguistic effects in ERP allow the investigation of sentence comprehension moment-by-moment as each word is processed. For sentence processing, there are distinctive ERP effects related to semantic processing, syntactic processing, and general binding processes. The processing of semantic information is most often found to have influence on the N400 component, which is a negative-going wave between roughly 250 to 550 ms. The N400 component typically distributed over the centro-parietal scalp with a peak around 400 ms after word onset (Brown, Brown, & Hagoort, 1993; Kutas & Hillyard, 1980). It is often elicited by difficulty in semantic integration and the processing of thematic information; that is, the

more effortful the integration or processing, the larger its amplitude (Friederici & Frisch, 2000; Gunter, Friederici, & Schriefers, 2000; Weckerly & Kutas, 1999). The processing of syntactic information is most often found to have influence on two ERP components: an early left anterior negativity (LAN) and a late centro-parietal positivity (P600). The LAN effect will be discussed later within the family of anterior negativities. The P600 component is a positive-going wave between roughly 500 to 1000 ms, which peaks around 600 ms after word onset. It is most commonly associated with syntactic violations in processing ungrammatical sentences (e.g., violations of word category, subject-verb number agreement, reflexive-antecedent number agreement), though recent studies have demonstrated its role as reflecting general syntactic integration difficulty in processing grammatical sentences without any outright syntactic violation (Fiebach, Schlesewsky, Lohmann, von Cramon, & Friederici, 2005; Hagoort, 2003; Kaan, Harris, Gibson, & Holcomb, 2000; van de Meerendonk, Kolk, Vissers, & Chwilla, 2010). For example, Kaan et al. (2007) observed a P600 on the embedded verb (e.g., “*imitated*” in the following example) for processing long-distance syntactic dependencies, such as “*Emily wondered who the performer in the concert had imitated [GAP] ...*”, relative to a control sentence such as “*Emily wondered whether the performer in the concert had imitated ...*”, though both sentences were grammatical and contained no ambiguity. Recently, P600 components were also associated with some semantic/thematic violation when a semantically attractive interpretation is available (Kim & Osterhout, 2005; Kuperberg, 2007). For example, Kim and Osterhout (2005) reported a P600 effect for the sentence “*The hearty meal was *devouring*” when a semantically attractive interpretation can be achieved by assuming that the verb “*devour*” should be in a –ED form but not the –ING

form. Therefore, researchers suggested that at least under certain conditions, a P600 could be observed when semantic but not syntactic processing dominates sentence analysis, though the question about under what circumstance semantic processing dominates required future studies.

Additionally, several anterior negativities (ANs), such as a LAN effect, a sustained left anterior negativity, and a bilateral anterior negativity (or *Nref* effect as named by Van Berkum et al., 1999), have been associated with integration operations during language processing, which is very relevant to the goals of the current study. For these anterior negative shifts, although they vary in their time course (e.g., short lived vs. sustained) and topographical distribution (left anterior vs. bilateral), it has been argued that at least part of the neural generators of these ERP components are overlapping (Fiebach et al., 2002). One of the ANs that has been established for various language is the LAN effect, which is an anterior negativity going from 250 to 600 ms. It is usually left-lateralized, though in some conditions the distribution is bilateral (Hagoort, Wassenaar, & Brown, 2003). LAN has the same latency range as the N400, but with a clearly more frontal distribution (Hagoort, 2003). It has been associated with morphosyntactic processing, in which the syntactic integration depends upon the subject-verb agreement as signaled by number (e.g., number agreement errors in subject-verb agreement), person or gender information (e.g., gender agreement errors in Spanish or German) (Friederici & Weissenborn, 2007; Gunter et al., 2000; Krott, Baayen, & Hagoort, 2006; T. F. Münte, Heinze, & Mangun, 1993; Neville, Nicol, Barss, Forster, & Garrett, 1991; Osterhout & Mobley, 1995). It is worth noting that both LAN and P600 have been associated with morphosyntactic violations, with the results depending on the

language and the specific kinds of violations that were investigated (see Friederici & Weissenborn, 2007 and Van Berkum et al., 2007 for a review). Researchers suggest that the LAN effect is associated with the first-pass, automatic detection of a morphosyntactic mismatch, whereas the P600 effect reflects a more controlled semantic or syntactic repair or revision process. This argument is supported by the findings that LAN has been found for morphosyntactic mismatches of both real and pseudo words whereas the P600 has only been found for real words (Müntz, Matzke, & Johannes, 1997). Moreover, there is a double dissociation in these ERP effects in neuropsychological patients (Friederici, von Cramon, & Kotz, 1999). Friederici et al. (1999) reported that patients with left frontal cortical lesions, who have difficulty in syntactic processing, showed a P600 effect (thorough reduced) but failed to show the LAN, while patients with impaired basal ganglia functions showed both LAN and P600 to syntactic violations. This results support the dissociation of LAN and P600 effects, as LAN might reflect the early automatic first-pass process and P600 might reflect the late controlled second-pass processes for syntactic information.

ANs have also been related to an increased WM load in both ambiguous and unambiguous conditions during sentence processing, when the processing difficulties are not necessarily syntactic in nature (Felser, Clahsen, & Münte, 2003; Fiebach, Schlesewsky, & Friederici, 2001, 2002; King & Kutas, 1995; Kluender & Kutas, 1993; Matzke, Mai, Nager, Rüsseler, & Münte, 2002; Münte, Schiltz, & Kutas, 1998). For example, a bilateral sustained AN (i.e., Nref effect) that resembles the LAN effect has been associated with difficulty in establishing reference, when there was no (morpho)syntactic violation. Such AN effect usually emerges at about 300 ms after the

referring noun or pronoun onset and is particularly sustained (Van Berkum, Brown, & Hagoort, 1999; Van Berkum, Brown, Hagoort, & Zwitterlood, 2003; Van Berkum, Koornneef, Otten, & Nieuwland, 2007). For instance, Van Berkum et al. (2004) observed a Nref effect at the pronoun “*he*” of the sentence “*David shot at John as he jumped over the fence*” in which there was more than one possible antecedent, relative to the sentence “*David shot at Linda ...*”, indicating a difficulty in pronoun binding. Given its special topographical distribution and time course, Van Berkum et al. suggested that the AN reflects a processing difficulty different from those associated with semantic (e.g., N400) and syntactic processes (e.g., P600), and might itself reflect an increased WM demand involved in tracking all these processes.

In addition, AN has been observed in a number of studies investigating the establishment of long distance linguistic dependency (e.g., filler-gap) when there was no (morpho)syntactic violation or ambiguity (Fiebach et al., 2001, 2002; King & Kutas, 1995; Kluender & Kutas, 1993; Kluender et al., 1998). For example, Kluender and Kutas (1993) observed a LAN effect in the matrix clause when comparing *wh*-questions (e.g., “*Who did you decide that you should sing something for [GAP] ...?*”) to yes/no-questions (e.g., “*Couldn’t you decide if you should sing something for ...?*”), which was independent of the grammaticality of the eliciting condition. As a result, they suggested that the LAN effect reflected high WM load in the *wh*-questions, which required holding the filler “*who*” until a syntactic dependency could be established between “*who*” and the verb phrase “*searching for*”. In a later study, King and Kutas (1995) found that a sustained left AN on the main verb following an object-relative clause as compared to the verb following a subject-relative clause. More importantly, the negativity was modulated

by individuals' WM capacity as high reading-span subjects showed a more widespread negativity than low span subjects. Similar findings have been reported by Fiebach and colleagues in a German study (Fiebach et al., 2001, 2002). Fiebach et al. (2002) found that object *wh*-questions elicited a LAN effect as compared to subject *wh*-questions about 400 ms after the onset of the first prepositional phrase following the *wh*-pronoun (as shown in Example 5, e.g., "*Tuesday afternoon*"). This negativity became more widely distributed in the second prepositional phrase (i.e., significant in bilateral anterior and posterior regions on the phrase "*after the accident*"). Moreover, the negativity was dependent upon the length of the filler-gap distance in that it was stronger in the long distance condition relative to short distance condition (i.e., two prepositional phrase vs. one prepositional phrase), and was modulated by subjects' WM capacity as measured by reading span - participants with lower WM span showed a stronger and more widely distributed anterior negativity. Therefore, Fiebach et al. proposed that the sustained negativity reflects the increased WM load caused by maintenance of more complex syntactic prediction in the object *wh*-questions than for subject *wh*-questions.

Example 5. (Fiebach et al., 2002)

a. Subject *wh*-question

Thomas fragt sich, wer am Dienstag nachmittag nach dem Unfall den Doktor verständigt hat.

"Thomas asks himself, who_{ACC} on Tuesday afternoon after the accident the_{NOM} doctor called has."

b. Long object *wh*-question

Thomas fragt sich, wen am Dienstag nachmittag nach dem Unfall der Doktor verständigt hat.

“Thomas asks himself, who_{NOM} on Tuesday afternoon after the accident the_{ACC} doctor called has.”

*Note: “ACC” refers to “accusative”, and “NOM” refers to “nominative”.

It is worth noting that Fiebach et al.’s results support the notion of a separable WM capacity for maintenance and integration, as they observed different ERP components for the two processes. In their experiment, the object *wh*-questions elicited LAN effects was only observed during the maintenance period (i.e., on the prepositional phrases), but changed into a local positivity (400 – 700 ms) over central and parietal midline electrodes on the noun phrase where integration/retrieval happened (e.g., “*doctor*” in Example 5). However, Fiebach et al. did not observe any object *wh*-questions elicited differences relative to subject *wh*-questions at the main verb.

Importantly, the findings from all these studies, which claim a relationship between the AN family of effects and increased WM demands in sentence processing, pose a challenge to the cue-based parsing approach. The cue-based parsing approach argues that WM capacity per se is not critical for sentence processing because only an extremely limited number of items (e.g., 1- 2 items) can be activately maintained, and processing difficulty should mainly derive from interference. However, the cue-based parsing approach would have difficulty in interpreting the processing difference between subject and object *wh*-questions observed in Fiebach et al.’ study (2002), because the sentence contents were exactly the same during the maintenance period and there was no difference in the degree of either syntactic or semantic interference. Recently, Bartek et

al. (2011) have suggested that although the cue-based parsing model (Lewis & Vasishth, 2005) could successfully simulate a number of sentence processing effects, such as locality effects, it is very likely that interference, expectations (e.g., surprisal account, Hale, 201; production-distribution-comprehension theory, Gennari and MacDonald, 2009) and passive decay in STM interact together to determine sentence processing difficulty (Bartek et al., 2011). Our recent data (Tan et al., 2011, 2013), which will be discussed later, also revealed that although WM capacity *per se* may not be necessary for constructing linguistic dependencies via cue-based retrieval, both the ability to keep prior representation active and the attentional control component of general WM capacity are necessary to recover from interference when it occurs. Thus, I suggest that it is necessary to build in more factors into the cue-based parsing approach to explain for natural language processing.

1.1.1.4.2 ERP components related to interference resolution

So far, studies directly investigating interference effects during language processing using the ERP technique are sparse. Two ERP studies conducted by A. E. Martin and colleagues investigated morphosyntactic interference effects in Spanish (A. E. Martin, Nieuwland, & Carreiras, 2012, 2014). In the earlier study, A.E. Martin et al. (2012) manipulated agreement of grammatical gender between a determiner, an intervening NP, and an elided noun. Examples of sentences used in their study are shown below.

Example 6. (A. E. Martin et al., 2012)

a. Correct condition

Marta se compró la **camiseta** que estaba al lado de la **falda/vestido** y Miren cogió **otra** [...] para salir de fiesta.

“Marta bought the **t-shirt_{FEM}** that was next to the **skirt_{FEM} /dress_{MASC}** and Miren took **another_{FEM}** to go to the party.”

b. Incorrect condition

*Marta se compró la **camiseta** que estaba al lado de la **falda/vestido** y Miren cogió **otro** [...] para salir de fiesta.

“Marta bought the **t-shirt_{FEM}** that was next to the **skirt_{FEM} /dress_{MASC}** and Miren took **another_{MASC}** to go to the party.”

*Note: “FEM” refers to “feminine”, and “MASC” refers to “masculine”.

To understand the word “*another*”, subjects must retrieve its elided antecedent⁵ “*t-shirt*”, which is a feminine noun. The gender of the determiner agreed (e.g., sentence 6a) or disagreed (e.g., sentence 6b) with the gender of this antecedent (e.g., “*t-shirt*”). At the same time, the intervening NP (e.g., “*skirt*” or “*dress*”) inserted between this dependency (“*t-shirt*” - “*another*”) matched or mismatched the gender of the determiner, although it was syntactically unavailable as its antecedent. According to the cue-based parsing approach, a matching distractor (e.g., “*skirt*”) that partially satisfies the retrieval cues should cause more morphosyntactic interference than a mismatching distractor (e.g., “*dress*”), as indicated by an ERP effect with larger amplitude or wider scalp distribution. However, A. E. Martin et al. (2012) did not observe such a main effect of this interference manipulation. They only observed a central-anterior negative going wave (400 – 1000 ms) elicited by ungrammatical conditions, and this negativity was modulated

⁵ The words “*otra*” and “*otro*” are determiners, which allows nominal ellipsis. The determiners do not need to refer to the same instance of their antecedents. Therefore, it is not a pronoun or anaphor which refers back to an already given referent. In addition, these determiners have no grammatical gender of their own, but are inflected to agree with the gender of the elided noun.

by the gender of the local attractor NP – that is, in the grammatical sentences only, the gender mismatching distractor (e.g., “*dress*”) caused an enlarged negativity as compared to the gender matching distractor (e.g., “*skirt*”). They attributed the sustained negativity to subjects’ failure in retrieving an antecedent, while the interaction of grammatical and interference conditions indicated that the syntactically unavailable distractor was at least temporarily considered as target.

The ERP components found in A. E. Martin et al.’s study (2012) were neither consistent with the predictions from cue-based parsing approach, nor consistent with the robust finding that morphological disagreement should elicit a P600 effect (see Friederici, 2002 for a review). A potential problem of this study was that it was difficult to determine whether subjects comprehended the sentences correctly or not. Although the experimenter included comprehension questions (e.g., “*Did Miren take something to go to the party?*”), these questions did not specifically tap whether the appropriate antecedent had been retrieved. If subjects settled on the intervening noun in the distractor-matching condition some proportion of the time, they may have been unaware of doing so. In the distractor-mismatching condition, however, they might have detected the gender mismatch and thus shifted to favoring the appropriate noun, and this shifting process resulted in the ERP result.

As a result of these issues, in a later study, A. E. Martin and colleagues (2014) expanded on the early study (2012) by asking subjects to do a grammaticality judgment task after reading object-extracted relative clauses (instead of the subject-extracted relative clauses used in 2012), where the distractor was in a less prominent position (e.g., “*Rafaela lost the necklace_{MASC} that she always wore with the ring_{MASC/FEM} and Monica*

*recovered another_{MASC}/*another_{FEM} that had lost years before*”). They found an early and brief AN (100 – 400 ms) on the determiner associated with the grammaticality × interference interaction, with high morphosyntactic interference condition being more negative than the low morphosyntactic interference condition in the grammatical sentences only. In addition, they observed a P600 effect on the post-critical word indicating later syntactic revision. However, A. E. Martin et al. did not try to reconcile the different results among this study (2014), their earlier study (2012), and previous behavioral studies which found a robust syntactic interference effect (e.g., Van Dyke, 2007; Van Dyke & Lewis, 2013; Van Dyke & McElree, 2011). They suggested the differences in results were caused by task differences (e.g., irrelevant comprehension questions vs. grammaticality judgment), sentence structure differences (e.g., subject-extracted RC vs. object-extracted RC), and cross-linguistic differences (Spanish vs. English), although they admitted this explanation might overstate the differences between their paradigms and previous ones.

In addition, I suggest that the lack of a robust morphosyntactic interference effect in the expected direction and position in A. E. Martin et al.’s study (2012, 2014) might be attributed to the different type of matching or mismatching features they manipulated compared to most of the prior studies. Previous English studies have manipulated the grammatical role (e.g., subject vs. object, Van Dyke, 2007) or semantic role (e.g., human vs. non-human being, Van Dyke & McElree, 2011) of the distractor, whereas A.E. Martin et al. manipulated a morphological feature (i.e., grammatical gender). A number of studies have shown differences between syntactic processing and morphosyntactic processing, and even within morphosyntactic processing, further processing differences

have been reported (see Friederici, 2002, for a review). For example, researchers found that local agreement attraction only occurs for subject-verb agreement (e.g., “*the key to the cabinets was*”) but not for reflexive pronoun gender-agreement (Philips, Wagers, & Lau, 2010). Therefore, I suggest that although A. E. Martin et al.’ studies did provide some evidence in favor of a cue-based parsing approach in showing that a grammatically unavailable noun was apparently retrieved, further studies were needed to determine how differences in experimental paradigms and linguistic cues in different languages contribute to the interference effects.

Summary

To sum up, so far, a number of studies have examined how interference affects impair sentence processing by using behavioral measurements (Gordon et al., 2001, 2004; Gordon et al., 2002; Tan, Van Dyke, & Martin, 2011; Van Dyke, 2007; Van Dyke & McElree, 2011) and neuroimaging techniques (Glaser et al., 2013). Most of these studies have documented that individuals are less accurate, have slower reading times, and show greater brain activity while processing sentences with high interference. However, many questions are still under debate. These include: How are different retrieval cues combined during interference resolution? Are the semantic and syntactic retrieval cues combined in a weighted linear fashion? Is syntactic interference only caused by NPs in modifying adjuncts but not core arguments in other languages as well? Does the cue-based retrieval approach hold in other languages besides English with different syntactic structures and perhaps different weightings for semantic and syntactic information? Moreover, what is the role of WM mechanisms underlying interference resolution during sentence comprehension? Further investigation of these questions may

help us address some fundamental issues in psycholinguistic studies such as how semantics and syntax interact, and the nature of the WM mechanisms constraining sentence comprehension.

1.2 Cognitive mechanisms supporting interference resolution in sentence processing

A major question in cognitive science concerns the nature of the memory mechanisms. While many basic questions about WM remain controversial (Baddeley, 2012), in the field of psycholinguistic studies, a number of studies have shown a link between WM processes and language comprehension. The general implication drawn from these findings is that WM supports language processing. However, researchers have not come to a final conclusion about the nature of this relationship, such as what component(s) of general WM mechanism is most important, or whether there is a language-specific WM system. Recently, some sentence parsing models such as the experience-based parsing account and executive control-based account have even claimed that WM capacity per se is not important for online sentence processing. In the current study, I aimed to investigate the nature of the WM mechanisms underlying interference resolution in sentence comprehension. Thus, in the following section, I will briefly summarize some previous theories and studies that examined the role of WM in general language processing, and focus on two recent studies from our lab that investigated the role of WM in interference resolution during English sentence parsing.

1.2.1 Capacity-based accounts of sentence processing

Given that even simple expressions often contain non-adjacent constituents that need to be integrated, many researchers have assumed that WM capacity supports the

retention of sentence information during comprehension. There is controversy, however, regarding what kind of memory capacity is critical for sentence comprehension. Some researchers have claimed that there is a domain-general memory capacity common to all kinds of verbal tasks including language comprehension (Daneman & Carpenter, 1980; Daneman & Hannon, 2007; Fedorenko et al., 2006, 2007; Gordon et al., 2002; Just & Carpenter, 1992; J. King & Just, 1991). Most of these experiments used complex span tasks, such as reading span (Daneman & Carpenter, 1980) and operation span (Turner & Engle, 1989) to measure individuals' WM capacity. Different from simple span tasks, in which subjects are required to passively repeat back a list of random digits/letters/words, complex span tasks impose simultaneous processing (e.g., reading a sentence aloud in the reading span task, and solving a math equation in the operation task) and storage demands (e.g., actively maintain the to-be-remembered information). Considerable evidence has been put forward supporting the notion that WM capacity as assessed by complex span measures is a good predictor of people's performance in sentence comprehension, especially when the sentence is complex or there is an external memory load (e.g., maintaining a random sequence of digits) (Daneman & Carpenter, 1980; Daneman & Hannon, 2007; Fedorenko, Gibson, & Rohde, 2006, 2007; Just & Carpenter, 1992). However, some researchers questioned these findings as the WM-language relation occurs only under some task conditions and the interaction of sentence complexity and external memory load, which is strongly predicted by the general WM account, was absent in some of the later replications (Caplan & Waters, 1999, 2013; see later section for more discussion). Even assuming that the evidence is solid, it is not very clear what factor(s) as measured by the complex span tasks accounts for the relation

between general WM and language processing - whether it is the dynamic tradeoff between processing and storage tasks (Daneman & Carpenter, 1980; Just & Carpenter, 1992), attentional control (Engle, 2002; Kane & Engle, 2003; McVay & Kane, 2012), or storage limits (Cowan et al., 2005). Moreover, numerous studies have demonstrated that WM is not a single factor. Instead, it is composed of distinct processes, including capacity, attentional control, and secondary memory retrieval. All these processes independently and jointly contribute to the relation between WM and other higher cognition (e.g., general intelligence, language processing) (Shipstead, Lindsey, Marshall, & Engle, 2014; Unsworth, Fukuda, Awh, & Vogel, 2014). As a result, the nature of the WM-language relation remains unclear and further investigation is needed.

On the other hand, several lines of research have highlighted the importance of specialized aspects of WM rather than a general capacity in supporting sentence processing. Neuropsychological studies have shown that brain damaged patients with very restricted STM span can nonetheless show excellent sentence comprehension (Butterworth, Campbell, & Howard, 1986; Caplan & Waters, 1999; Martin, 1987). In a review of their own studies and those from other labs, Caplan and Waters (1999) found that there was little relation between online syntactic comprehension and general WM capacity as measured by either simple or complex span tasks for either healthy individuals or brain-damaged patients, although a relation has been found between offline sentence processing (e.g., answering probe questions following the sentences) and standard memory tests. As a result, Caplan et al. claimed that the online syntactic processing is supported by a specialized verbal WM system which cannot be tapped by

standard span task, while offline syntactic processing is supported by a general verbal WM system shared with other verbal tasks.

A different specialized capacity approach is the multiple-components model put forward by Martin and colleagues (Hamilton, Martin, & Burton, 2009; Martin & He, 2004; Martin & Romani, 1994). In this multiple-components approach, Martin and colleagues argues that verbal working memory should be broken down into three separate capacities for the retention of phonological, semantic, and syntactic information (Martin & Romani, 1994). Phonological STM capacity is relevant to verbatim repetition and the learning new words, but is irrelevant to sentence comprehension. Semantic STM capacity is critical for maintaining several unintegrated word meanings during language comprehension and production, as well as for learning new meanings. Syntactic STM supports the maintenance of unintegrated syntactic structures. The supporting evidence for the dissociations of these different STM capacities has been obtained from both neuronanatomical and neuroimaging studies. In a series of neuropsychological studies, Martin and colleagues found that aphasic patients could be categorized into different groups based on their performance on tasks tapping phonological or semantic STM, for instance, on a rhyme probe task and a category probe task (Martin et al., 1994, 1999, 2004). In both tasks, subjects are presented with a spoken list of words followed by a probe word. On the rhyme probe task, subjects have to judge whether the probe word rhymes with any of the list items (e.g., *table-line-clock*, probe: *sign?*), while on the category probe task, they have to judge whether the probe word is in the same category as any of the list items (e.g., *snow-dress-apple*, probe: *jacket?*). The rhyme probe task has been used to measure individuals' phonological STM capacity, while the category probe

task has been used to measure individuals' semantic STM capacity. Although most aphasic patients have deficits in both tasks compared to age matched controls, they showed dissociation on these two tasks. Some patients scored significantly worse on the category probe task than on the rhyme probe task (e.g., patient M.L., A.B. as reported by Martin & He, 2004), while some patients showed the opposite pattern (e.g., patient E.A. as reported by Martin & Romani, 1994). In addition, although there is no appropriate test for syntactic STM capacity at this time, there were patients who had little deficits in performance on either category or rhyme probe task, but demonstrated a detrimental effect in a task that required active retention of syntactic information (e.g., patient M.W. as reported by Martin & Romani, 1994). Martin and colleagues have shown that patients with specific deficits in semantic STM had difficulty in a sentence anomaly judgment task that required holding the meaning of several individual words before integration (e.g., "*The rusty, old, red swimsuit was...*"), but not in a grammaticality judgment task that required maintaining syntactic information across intervening materials for later integration (e.g., "*Birds quite frequently *chirps in early morning*"). In contrast, patients with syntactic STM deficits showed the reverse pattern, and patients with phonological STM deficits performed normally in sentence anomaly judgment tasks but had severe deficits in sentence repetition (Martin & He, 2004; Martin & Romani, 1994; R.C. Martin, Shelton, & Yaffee, 1994). In some recent neuroimaging studies, researchers have demonstrated a dissociation in neural activity between semantic and phonological STM (Hamilton, Martin, & Burton, 2009; Martin, Wu, Freedman, Jackson, & Lesch, 2003; Shivde & Thompson-Schill, 2004), with the phonological STM tasks recruiting the left

inferior parietal region and the semantic STM manipulation recruiting left inferior frontal gyrus (LIFG) and middle frontal gyri.

1.2.2 Non-capacity accounts of sentence comprehension

However, the cue-based retrieval mechanism introduced in section 1.1.1 presents a strong challenge to the WM capacity-based accounts that have emphasized reduced capacity as the source of comprehension difficulty (Van Dyke & Johns, 2012; Van Dyke, Johns, & Kukona, 2014). Although the researchers who proposed the cue-based retrieval approach agree that sentence processing is supported by a memory system which operates according to similar principles as memory in other domains (Van Dyke & McElree, 2006), they argue that the number of items that can actively be maintained in the focus of attention is extremely limited (1 – 2 items) and should be within everyone's memory capacity. Thus, some researchers have called into question the capacity-based accounts to sentence processing and have suggested instead that the observed correlations between memory capacity and sentence comprehension performance actually derive from other factors, such as word knowledge, reading skills, general executive control ability, and so on, which affect both the capacity measures and the sentence processing measures (Acheson, Wells, & MacDonald, 2008; MacDonald & Christiansen, 2002; Novick, Trueswell, & Thompson-Schill, 2005; Perfetti, 2007; Real & Christiansen, 2007; Van Dyke & Johns, 2012; Wells, Christiansen, Race, Acheson, & MacDonald, 2009). In general, these researchers shift the emphasis away from addressing the role of memory storage capacity towards the role of other cognitive capacities. There are two main approaches under the non-capacity accounts: the retrieval-based account and the

experience-based account. Evidence for these two different positions is briefly summarized below.

1.2.2.1 Executive control-based account

The role that executive control ability, especially inhibition, plays in sentence comprehension has been less studied than WM capacity, though it has started gaining researchers' attention. In the behavioral studies, some researchers observed a link between executive control ability (specifically, verbal executive control ability) and many aspects of sentence comprehension (Mendelsohn, 2002; Vuong & Martin, 2013). For example, Mendelsohn (2002) found that subjects with better verbal inhibition ability (i.e., inhibit interference from irrelevant verbal information) had less difficulty in recovering from garden-path structures, in which readers' most likely initial interpretation will be shown to be wrong as the sentences unfolding over time. For example, while reading garden-path sentences such as "*Bill knew the truth was being kept from him*", it is initially unclear whether the "*truth*" is the direct object of the verb "*knew*" or the subject of a subordinate clause "*the truth was being kept from him*". Mendelsohn found that the participants, who made more errors due to a failure of inhibiting the previous rule in a verbal inhibition task, were more likely to accept the inappropriate interpretation of the sentences in a verification task following each sentence. In a later study, Vuong and Martin (2013) further investigated the link between executive function and online sentence processing (i.e., word-by-word reading times). They found that the longer it took subjects to resolve interference in the verbal Stroop task (i.e., $RT_{\text{inconsistent}} - RT_{\text{consistent}}$), the longer it took them to revise garden-path interpretation at the final region of the sentence. Thus, Vuong and Martin concluded that the ability to recover from

misinterpretations in garden-path sentences comprehension is predicted by individuals' verbal executive control ability. Interestingly, it should be noted that in both Mendelsohn (2002) and Vuong and Martin's (2013) study, non-verbal inhibition tasks, such as the antisaccade task and non-verbal Stroop task (in which subjects had to indicate the pointing direction of visually presented arrows and ignored their actual screen location) were not related to subjects' sentence processing performance. Therefore, these researchers argue for a role of verbal-specific executive control mechanism underlying syntactic ambiguity resolution during sentence comprehension.

In addition, there is a growing amount of supporting evidence for the executive control-based account from neuropsychological studies, in which researchers have found a relation between patients' linguistic deficits and a failure of executive control. Specifically, Novick, Trueswell, and Thompson-Schill (2005) have proposed a link between LIFG and executive cognitive control and conflict resolution during sentence comprehension. On the basis of reviewing previous neuropsychological evidence, they suggested that LIFG is involved in detecting and resolving incompatible representations in non-parsing tasks (e.g., verbal Stroop task), as well as in syntactically ambiguity resolution (e.g., recovering from a dispreferred parsing option in garden-path sentence processing). Novick and colleagues (2009) provided evidence for their hypothesis in a later case study of patient with restricted damage to LIFG (BA 44/45). This patient (I.G.), who showed an inflated error rate beyond the normal range of controls in tasks that measured resistance to proactive interference (e.g., recent-negatives task), had difficulty in language production and comprehension tasks when there was semantic, conceptual, or syntactic competition, such as in naming pictures with low name agreement (e.g.,

couch/sofa/loveseat) or in carrying out spoken instructions that required subjects to revise their early interpretation because of temporary ambiguity (e.g., “*Put the apple on the napkin into the box*”), and so on. Vuong and Martin (2011) have also reported that LIFG patients with attentional control deficits (e.g., as measured by verbal Stroop task) had difficulty in resolving lexical ambiguities during sentence processing (e.g., “*He drank the port*”). Additionally, studies on Parkinson’s disease (PD) patients have implied a correlation between patients’ deficits in sentence processing and their deficits in attentional control (Colman, Koerts, Stowe, Leenders, & Bastiaanse, 2011; Hochstadt, 2009; Hochstadt, Nakano, Lieberman, & Friedman, 2006). Hochstadt and colleagues (2006) have specifically linked PD patients’ deficits in processing center-embedded sentences (e.g., “*The king that is pulled by the cook is short*”) to their inability to inhibit irrelevant information (e.g., as measured by set-switching).

Moreover, neuroimaging evidence shows that there is an overlap between brain regions activated for sentence processing and brain regions activated for tasks required for attentional control ability (Mason, Just, Keller, & Carpenter, 2003; Novick, Trueswell, & Thompson-Schill, 2005; Thompson-Schill, Bedny, & Goldberg, 2005; Ye & Zhou, 2009a). Consistent with the neuropsychological evidence as summarized above, the results from a number of neuroimaging studies suggest that LIFG, which includes Broca’s area, is activated in both general cognitive control and demanding syntactic processing at sentence-level. Mason and colleagues (2003) found that reading dispreferred syntactic structures (e.g., “*The disgusted student threw the book on the battle/ground but picked it up moments later*”; “ground” is the preferred condition) or

syntactically ambiguous sentences (e.g., “*The experienced soldiers warned about the dangers conducted the midnight raid*”) gave rise to greater LIFG activation.

Nevertheless, it is important to point out that most of these studies summarized above as supporting evidence for an executive control-based account have only examined individuals’ capacity to process sentences that were syntactically complex or ambiguous or had syntactic violations. The cognitive capacity involved in a paradigm with violation or ambiguity might be different from that involved in resolving interference in the well-formed unambiguous sentences, such as the sentences used in the current study. What is more, most of these studies did not examine other factors, such as WM capacity or linguistic experience. In Vuong and Martin’s study (2013), verbal Stroop performance accounted for only 11-13% of the variance observed in garden-path revision, leaving considerable variance to be explained. In Hochstadt et al.’s study (2006), although PD patients’ executive control ability could best predict their performance in comprehending center-embedded sentences, verbal WM also significantly correlated with overall sentence comprehension accuracy. As a result, a more complete investigation regarding the role that executive control ability plays in interference resolution is required for future studies.

The general implication from these studies, in which a role of executive control was observed, for my current project was that it might be the attentional control component, rather than or together with the storage capacity of WM mechanisms, that supports language processing. It is noteworthy that as in the development of the contemporary WM models, many researchers have demonstrated that WM capacity as measured by complex span tasks does not only reflect differences in memory capacity per

se (e.g., the number of items or memory traces could be actively maintained), but also reflects differences in the ability to retrieve information from long-term memory and differences in attentional control (Shipstead et al., 2014; Unsworth et al., 2014).

Therefore, while WM capacity *per se* may not be necessary for online sentence comprehension, controlled attention could be necessary to overcome processing difficulty (e.g., recover from interference or recover from misinterpretation such as in garden-path sentences), and it could be this process that the general WM measures have indexed: those with lower WM capacity show a larger processing difficulty because they have more difficulty resolving it. Consequently, the executive control-based account could be reconciled with the general WM account, in a way that both agreed on that the controlled attention play a role in the relation between general WM capacity and higher cognitions.

1.2.2.2 Experience-based account

Another account, namely the experience-based account, claims that the distinction commonly drawn between language processing ability and verbal WM capacity is artificial because all these tasks are “simply different measures of language processing skills” (MacDonald & Christiansen, 2002). Researchers who support experience-based account argue that skilled sentence comprehension is actually affected by variation in subjects’ exposure to language and some biological differences that affect processing accuracy (e.g., differences in the precision of phonological representations)(MacDonald & Christiansen, 2002). There is supporting evidence for the experience-based account in that the ease of processing a sentence is predicted by participants’ performance in several tasks that are linguistic experience-dependent, such as: 1) receptive vocabulary test

(Perfetti, 2007; Van Dyke & Johns, 2012; Van Dyke, Johns, & Kukona, 2014); 2) rapid automatized naming (RAN) task, in which subjects are required to name some digits/letters/words as fast as possible and their performance is assumed to reflect their phonological decoding ability (Gordon et al., 2013; Kuperman & Van Dyke, 2011); and 3) author recognition test (ART) or magazine recognition test, which measures subjects' knowledge of authors and literature (Acheson, Wells, & MacDonald, 2008; Kuperman & Van Dyke, 2011; Moore & Gordon, 2014; Stanovich & West, 1989; Van Dyke et al., 2014). It should be noted that the experience-based account and retrieval-based account are not mutually exclusive since the retrieval-based account also suggests that retrieval success depends on the quality of lexical representations (Van Dyke et al., 2014), although it puts more emphasis on general executive control ability.

However, in most of these studies, in which a link between linguistic experience and language comprehension has been reported, researchers also found a relation between WM capacity and sentence processing. For example, in a large-scale experiment examining the relation between individuals' reading skills and eye-movements during general sentence processing, Kuperman and Van Dyke (2014) found that subjects' general WM capacity (as measured by reading span) was related to the early eye movement (as indexed by initial landing position, which taps into the decision about where to move the eyes and is typically made before the word is foveated), and their phonological STM capacity (as measured by digit span) was related to the late eye movement (as indexed by total fixation time)⁶. However, because subjects' linguistic

⁶ In eye-tracking measurement, the initial landing position is the landing position of the first fixation on the word, which measured in characters from the word's left boundary; the total fixation time is the summed duration of fixation that landed on the word. The mean regression path duration (Go-past time) is the

experience (as indexed by RAN and word identification performance) were more consistently related to both early and late eye-movements, Kuperman and Van Dyke concluded that word decoding skills (measured by word identification test), and phonological/orthographic processing skills (as measured by RAN) are better predictors for general sentence processing ability. In another study, Van Dyke et al. (2014) found that in addition to the link between receptive vocabulary and sentence comprehension, there was also a relation between subjects' general WM capacity and sentence processing efficiency, though such relation disappeared after controlling for subjects' fluid intelligence. I suggest that a potential problem of experience-based account based on these findings was that they did not take the multifaceted nature of WM mechanism into consideration. As shown by a number of studies, all the factors of WM mechanism (i.e., capacity, attentional control, and long-term retrieval) independently and collectively account for the correlation observed between WM and higher cognition abilities (Shipstead et al., 2014; Unsworth et al., 2014), such as fluid intelligence (Engle, Tuholski, Laughlin, & Conway, 1999). Therefore, by using a single WM task (i.e., reading span) or partialling out the common variance shared between WM and IQ, one may not be able to observe a robust relation between WM and language comprehension. This will be further discussed in later sections.

summed duration of all fixations on the word and words to its left after the first fixation on the word is made and before the eyes moved to the right of the word for the first time (Kuperman & Van Dyke, 2011)

1.2.3 WM capacity, executive control, reading experience and their relation to interference resolution in sentence comprehension

As summarized in section 1.1, interference is a primary determinant for sentence processing difficulty. Different types of interference impair sentence comprehension, with this interference occurring primarily during the retrieval process. Different accounts of the underlying mechanisms supporting sentence processing make different predictions about the relation between subjects' interference resolution ability and other cognitive abilities. All the memory capacity accounts would predict that individuals with smaller memory capacity would experience greater difficulty in interference resolution than individuals with high WM capacity, as low-span individuals generally have fewer resources being available for maintaining clear representations, especially when the processing is demanding (e.g., Daneman & Carpenter, 1980; Just & Carpenter, 1992), or less capable of keeping the prior items active (e.g., Martin & He, 2004; Tan et al., 2011), or worse performance in attentional control or secondary memory retrieval as compared to high-span individuals (Shipstead, Lindsey, Marshall, & Engle, 2014; Unsworth, Fukuda, Awh, & Vogel, 2014). However, these accounts differ in what kinds of memory component are critical. In a recent behavioral study, Tan, Martin, and Van Dyke (2011) examined retroactive interference in English sentence comprehension and the cognitive mechanisms underlying such processes through an individual differences approach, through relating subjects' interference resolution ability (as indexed by self-paced sentence reading times) to capacities on a set of individual difference measures as would be predicted by different accounts (i.e., semantic STM, phonological STM, general WM, executive function, and vocabulary tests). In this experiment, syntactic and semantic

properties of intervening noun phrases were manipulated to result in either high or low interference in a similar manner as in Van Dyke (2007) (see Example 7).⁷

Example 7 (Tan, Martin, & Van Dyke, 2011, modified from Van Dyke, 2007; with semantic manipulation in brackets)

a. Low Syntactic interference condition

The resident who was living near the dangerous [house/neighbor] last month has complained about the investigation.

b. High Syntactic interference condition

The resident who said that the [house/neighbor] was dangerous last month has complained about the investigation.

Results from mixed-effects model analyses including all the individual differences measures demonstrated that first, as predicted by the cue-based parsing approach, subjects showed higher error rates to comprehension questions and longer reading times in the high semantic or syntactic interference condition relative to the low interference conditions. There was no interaction between the semantic and syntactic interference. More importantly, as predicted by the multiple capacity account (Martin et al., 1999, 2004), participants' semantic STM capacity (as measured by the category probe task) predicted the magnitude of their semantic interference effects, but not syntactic interference effects; while phonological STM was not related to either semantic or syntactic interference. In addition, participants with better general WM capacity (as measured by the composite WM score calculated from reading span and operation span)

⁷ In Tan et al. (2011), in the low syntactic interference condition (e.g., sentence 7a), the intervening noun ("house/neighbor") was the object of a prepositional phrase, while in the high syntactic interference condition (e.g., sentence 7b), the intervening noun ("house/neighbor") was the subject of the sentential complement. In the high semantic interference condition, the NP "neighbor" was a semantically plausible subject for the verb phrase "had complained", whereas in the low semantic interference condition, the NP "house" was a semantically implausible subject for the verb.

had less difficulty in resolving online syntactic interference. Both WM -language relations remained significant after controlling for the quality of individuals' basic lexical representation (as measured by vocabulary test). In a later study with aphasic patients using similar sentence materials, Tan, Van Dyke and Martin (2013) provided convergent results by showing that there was a negative correlation between patients' semantic STM span and the magnitude of their semantic interference effect, and a negative correlation between patients' resistance to interference (i.e., as measured by Stroop and picture-word interference tasks) and the magnitude of their syntactic interference effect, even after controlling for patients' verbal knowledge deficits. Moreover, as in the early study with healthy subjects (Tan et al., 2011), the correlation between phonological STM and either semantic or syntactic interference effect was far from significance.

Overall, the results from these two experiments (Tan et al., 2011, 2013) are most consistent with the multiple capacities approach that semantic STM determines the ease of semantic processing during sentence comprehension, while phonological STM is not critical for online sentence comprehension (Martin et al., 1996, 1999, 2004). In addition, the results also provide some support to the general WM capacity account and executive control account as syntactic interference resolution efficiency was predicted by the WM composite or the executive control in these studies. However, in Tan et al.'s study (2011), it was not very clear what was the factor(s) of WM that remained after controlling for the variance shared with semantic/phonological STM, general control ability, and vocabulary in the mixed-effects models. As discussed earlier, recent memory studies have demonstrated that all the three components capacity of WM (i.e., attentional control, and secondary memory retrieval) make independent and joint contribution to the relation

between WM and higher cognitions (Shipstead et al., 2014; Unsworth et al., 2014). We suggested it might be certain attentional control (e.g., resistance to proactive interference) that modulates the relation of WM-syntactic interference. It is interesting that the resistance to interference as measured by verbal Stroop only predicted subjects' syntactic interference resolution for aphasic patients (Tan et al., 2011) but not for the young healthy subjects (Tan et al., 2013)⁸. Given the many differences between these two studies, such as differences in subject population, dependent measures (log RT vs. accuracy), and independent variables (self-paced reading vs. error rates), it is hard to explain the nature of the relation between general WM/EF and sentence comprehension based on our current evidence. Nonetheless, both studies demonstrated that there is a general link between attentional control and syntactic interference during sentence comprehension. Further study is needed to examine the correlations between different types of inhibition function, WM capacity, and language processing.

On the other hand, some researchers who embrace the experience-based accounts, have demonstrated that subjects' variation in the linguistic experience determines their interference resolution ability better than any variation in the WM capacity (Van Dyke & Johns, 2012; Van Dyke, Johns, & Kukona, 2014). Van Dyke et al. (2014) tested participants on the same self-paced reading task as used in Van Dyke and McElree's study (2006, as shown in Example 3), as well as an extensive battery of individual differences measures, including working memory, simple memory span, vocabulary knowledge and so on. The results showed that after removing variance shared with a

⁸ We would also expect to observe a link between general WM capacity and syntactic interference effect size in the neuropsychological study. However, our patients were not able to complete the complex memory span measures. Thus, we do not have a compatible measurement for patients' general WM capacity as in the patient study as in the healthy young subjects study.

measure of fluid intelligence (IQ), receptive vocabulary knowledge was the only significant predictor of sentence comprehension performance. Thus, Van Dyke et al. concluded that WM capacity is not an important predictor for language processing if considered general cognitive ability as indexed by IQ. Instead, only knowledge of word meanings, which could relate to the richness of semantic representations for words in the sentence, predicted comprehension performance. Individuals with poor quality lexical representation may not be able to discriminate between representations of target and distractors during retrieval.

However, I suggest that a potential problem with Van Dyke et al.'s conclusions is that they treated IQ as a "black box" with some unknown factors underlying differences in IQ, though a number of studies have reported a very high correlation between WM capacity as measured by complex span tasks and general intelligence (Engle, Tuholski, Laughlin, & Conway, 1999; Hambrick, Kane, & Engle, 2005; Kane et al., 2004; Unsworth, Redick, Heitz, Broadway, & Engle, 2009). Thus, by factoring out IQ, one may be also factoring out working memory capacity. Therefore, their results could not unambiguously rule out an influence of WM capacity on sentence processing efficiency. The results from Tan et al.'s study (2011) do reveal a role for specific WM capacities in sentence comprehension, even after controlling for vocabulary. Despite our findings of a role for WM capacity, we suggest that our findings could be reconciled with the cue-based parsing theory. We agree with Van Dyke et al. that it is necessary to put some emphasis on the content and quality of memory representation rather than merely focusing on how many items can be held in an active state in WM. During sentence processing, successful retrieval of earlier sentence constituents is also affected by the

strength of memory traces. However, we suggest that the quality of memory representation for earlier linguistic constituents, which are outside the focus of attention, is not just decided by individuals' vocabulary knowledge, but also affected by individuals' semantic and syntactic STM capacities. Participants with better semantic/syntactic STM have stronger memory traces (e.g., slower decay rates) and more distinguishable representations for the to-be-retrieved items. As a result, these participants will show higher efficiency and accuracy when they need to selectively retrieve earlier sentence information into focus of attention for processing.

1.2.4 Evidence from cross-linguistic studies about the memory mechanisms underlying interference resolution: Chinese sentence processing

In psycholinguistics studies, one important way to evaluate existing theories is to examine whether the predictions hold up in languages with different grammatical properties to ensure that the findings are not limited to specific properties of only one language. Regarding interference effects during sentence processing, there are two questions that have not been fully resolved by English studies. First, is the observed syntactic interference effect in previous studies confounded with sentence hierarchy differences, e.g., the intervening NP was in a sentential complement in the high syntactic interference condition but was in a propositional clause in the low syntactic interference condition? Second, how are syntactic and semantic retrieval cues combined during sentence processing, e.g., are they combined in a simple linear fashion? More important, for the purpose of better understanding the memory mechanisms underlying sentence processing, I suggest that it is worth examining the correlations between WM capacities

and semantic/syntactic interference resolution capacity, as observed in Tan et al.' study (2011, 2013), in a language with different properties, such as Chinese.

Regarding the first question, in most of the previous studies in English (Tan et al., 2011, 2013; Van Dyke, 2007; Van Dyke & McElree, 2011), it is hard to fully rule out the possibility that greater processing difficulty in the high relative to the low syntactic interference condition is due to having to process an additional embedded clause (e.g., “*who said that [the warehouse was dangerous]*” vs. “*who was living near the dangerous warehouse*” in Example 4, Van Dyke, 2007). Some researchers have suggested that there is little or no confounding problem caused by syntactic hierarchy differences. McElree and colleagues (2003, 2006, 2011) have argued for a direct, content-addressable retrieval during sentence processing and thus, a structured tree is only necessary to describe relations between words but the retrieval mechanism does not access items via the tree and the syntactic hierarchy should have little influence on the retrieval process. In addition, Van Dyke (2007) has also shown that the extra storage or processing cost associated with the additional embedded clause in the high syntactic interference condition, which is predicted by Gibson’ dependency locality theory (DLT; 1998, 2000), should have been resolved before the critical verb (see Van Dyke, 2007, p420 – 421, for a detailed discussion on these issues). However, both McElree et al. and Van Dyke’s explanations were based on the assumption that the cue-based parsing account (or the DLT as tested by Van Dyke, 2007) alone could explain all the phenomena of sentence processing. Nonetheless, some researchers have argued that there are certain aspects of sentence processing, such as the cost associated with maintaining syntactic predictions (e.g., Bartek et al., 2011; Fiebach, et al., 2002) or generating retrieval cues that could not

be captured by either the cue-based parsing approach or the DLT alone (e.g., Bartek et al., 2011; Fiebach, et al., 2002; see further discussion later). Therefore, it is still possible that at least part of the effect observed at the critical verb is a spillover effect associated with processing an additional embedded clause, which extends all the way past the final word of the RC structure.

One possible solution to the problem of confounding variables is to compare sentences containing subject relative clause (SRC) versus object relative clause (ORC). As shown in Example 8, the number of clauses (i.e., both only contain one RC), the syntactic hierarchical role of the intervening NP (i.e., both are core arguments), and even the content words are kept the same in sentences 8a & b. Through comparing individuals' reading time on the matrix verb (e.g., "*admitted*"), one could test the syntactic interference effect predicted by cue-based parsing approach, as the intervening NP "*senator*" plays a grammatical object role in the sentence 8a and a grammatical subject role in the sentence 8b. Thus, a syntactic interference effect should be observed on the matrix verb in sentence 8b while readers are integrating "*admitted*" with its real subject "*reporter*".

Example 8 (King & Just, 1991).

- a. SRC: The **reporter** [who [GAP] **attacked** the **senator**] **admitted** the error.
- b. ORC: The **reporter** [who the **senator** **attacked** [GAP]] **admitted** the error.

Previously, a number of studies have demonstrated that in the ORC sentences, individuals are less accurate to probe questions (e.g., "*Did the reporter admit the error?*") and slower in processing the RC regions and the matrix verb in the ORC

sentences (e.g., “*attacked*”), relative to SRC sentences (J. King & Just, 1991). However, interpretation of the comparison between SRC and ORC structures is difficult in English because a variety of reasons have been offered for the greater ease of processing SRC than ORC structures in English – including, for example, the longer distance between linguistic dependencies in the ORC structure (Gibson, 1998, 2000; Grodner & Gibson, 2005), the accessibility hierarchy (Keenan & Comrie, 1977), the Perspective Shift account (MacWhinney, 2005), or the experience-based explanation which attributes the contrast to the frequency differences between ORC and SRC structures (MacDonald & Christiansen, 2002). All these various explanations predict a relative ease of processing the SRC structure in the RC region (i.e., shorter RT), and such SRC-advantage might spillover to the matrix verb following the RC region (Gibson, 1998; Gordon, Hendrick, & Johnson, 2004; Gordon, Hendrick, & Levine, 2002; Grodner & Gibson, 2005; Hale, 2001; Lewis, Vasishth, & Van Dyke, 2006; MacDonald & Christiansen, 2002; Traxler, Morris, & Seely, 2002; Wells, Christiansen, Race, Acheson, & MacDonald, 2009). Thus, through using English materials, it is hard to distinguish between the cue-based parsing approach and other approaches to explaining the longer RT on the matrix verb.

However, different from most languages⁹, Chinese has an advantage in that it is the first language in which some evidence has shown that ORC structures are easier to process than SRC structures, as demonstrated by shorter RT in the ORC regions (i.e., RC verb and NP) (Chen, Ning, Bi, & Dunlap, 2008; Gibson & Wu, 2013; Hsiao & Gibson, 2003). This finding is of particular relevance to the aims of my current experiment, as the cue-based parsing approach would predict an SRC-advantage on the matrix verb, even if

⁹ A number of cross-linguistic studies have shown that this SRC advantage also exists in German (Mecklinger, Schriefers, Steinhauer, & Friederici, 1995), Japanese (Ishizuka, Nakatani, & Gibson, 2006), Dutch (Mak, Vonk, & Schriefers, 2002) and many other languages (see a review in Lin & Bever, 2006).

there were an ORC-advantage in the preceding RC regions, because the distractor NP in the ORC structure plays a subject role and thus should cause greater syntactic interference due to a better match with the syntactic retrieval cues generated on the critical verb. Therefore, if there were an ORC-advantage in processing the RC structure as reported in these studies, but one still observed an SRC-advantage on the critical verb, then such result would provide a strong support to the cue-based parsing approach. Actually, although most of the previous Chinese studies have focused on the processing differences in the RC region and the head noun, several studies did observe an SRC-advantage on the critical verb (Jäger et al., 2015; Hsiao, 2003). As a result, in the current project, I employed Chinese sentences with SRC or ORC structures in order to dissociate syntactic interference effect from the potential confounding variable of sentence hierarchy differences. In the next section, I will first summarize the previous studies on Chinese RC processing.

1.2.4.1 The debate about the SRC- vs. ORC-advantage

Mandarin Chinese is a head-initial language with the canonical word order of SVO (subject-verb-object). RC structure in Chinese differs from that in English in several ways. First, in Chinese the RC comes before its head noun, while in English the RC always follow its head noun. Second, Chinese has a RC marker “*de*”, which roughly corresponds to “*that*” or “*who*” in English, but “*de*” always comes after the relative clause (e.g., “[*Eat apple*] *de* boy ...”; English - *The boy who [eats the apple]*). Chinese relative clause processing has been shown to be an interesting test case for evaluating WM capacity-based versus experience-based accounts of sentence processing, which is

one of the main purposes of the current project. The long and ongoing debate about how to explain the SRC-advantage in English, and whether there is an ORC-advantage in Chinese RC processing is briefly summarized in this section, as it affected the choice of materials in my current project.

1.2.4.1.1 SRC-advantage in English RC processing

In previous English studies, some researchers have claimed that the processing advantage of SRCs could be best accounted for by differing WM demands for the two sentence types (Gibson, 1998; Just & Carpenter, 1992; Warren & Gibson, 2002), as the distance between the linguistic dependencies (i.e., gap-filler) is greater in the ORCs relative to SRCs¹⁰. The greater distance results in greater WM demand (e.g., storage cost) in maintaining and retrieving incomplete dependencies¹¹ (Gibson, 1998). In addition to the storage cost, Gibson and colleagues (1998, 2000, Warren & Gibson, 2002) have claimed that the integration cost, which is associated with integrating incoming elements into currently existing sentence structure(s), is larger in ORCs because there are more

¹⁰ According to Gibson (1998), distance between the linguistic dependencies is quantified by the number of new referents (nouns and verbs). A discourse referent is “an entity that has a spatio-temporal location so that it can later be referred to with an anaphoric expression, such as a pronoun for NPs, or tense on a verb for events (Webber, 1988)”.

¹¹ During sentence processing, there is an incomplete dependency after processing the first element of the two relevant sentence elements (syntactic dependency) and before processing the second element of the dependency. For example, in sentence “*The boy laughed*”, the processing of NP “*boy*” is dependent on the following verb “*laughed*”. Thus there is an incomplete dependency immediately after processing the NP “*the boy*”. According to Gibson, the more predictions the readers need to maintain before completing the dependency, the more processing difficulties. For example, as shown in Example 8, after processing the embedded subject “*the senator*”, subjects have to maintain three incomplete dependencies in sentence b with an ORC: one between the main clause subject “*the reporter*” and its verb “*admitted*”; a second between the embedded subject “*the senator*” and its verb; a third one between “*who*” and its predicted GAP position. In contrast, they only need to maintain at most two incomplete dependencies in sentence 8a when processing the word “*who*”: one between main subject “*the reporter*” and its verb “*admitted*”; a second one between “*who*” and its predicted GAP position. As a result, there is a relative ease of processing SRC sentences.

new discourse referents inserted between the dependencies.¹² For example, in sentence 8a, the attachment of the empty-category (i.e., GAP) in subject-position of the RC to “*reporter*” crosses zero new discourse referents while in sentence 8b, the attachment of the gap as object of the embedded verb “*attacked*” to “*reporter*” needs to cross two new discourse referents – the object referent “*senator*” and the verb referent “*attacked*”. As a result, the activation level of “*senator*” decays more in the ORC structure due to more discourse referents being processed and fewer resources being available for maintaining syntactic representations.

In contrast to the explanations based on WM demands, researchers who support the experience-based account have attributed the SRC processing advantage in English to the fact that SRC structures are much more frequent than corresponding ORC structures during reading in English (SRCs: 86% vs. ORCs 13%; MacDonald & Christiansen, 2002). Supporting evidence has come from studies in English in which researchers directly manipulated subjects’ reading experience through training them on certain sentence structures (Wells et al., 2009), or manipulated comprehenders’ frequency-based assumptions about the relation between noun animacy and sentence structure (Traxler et al., 2002). For example, Wells et al. (2009) trained subjects on reading relative clause with SRC or ORC structures. They found that increased relative clause experience improved reading speeds for ORCs more than for SRCs, while no improvement was observed for a control group who was trained on other types of complex sentences. Wells et al. argued that the interaction of RC types and degree of improvement is consistent with the predictions from the experience-based account that less regular sentence types

¹² Gibson (1998) pointed out that although processing all words (e.g., “*the*”, “*and*”, etc.) probably causes increment of integration cost, most of the integration cost should be caused by processing words indicating new discourse referents (i.e., verbs and nouns).

(i.e., ORC structure) depend more heavily on specific experience (frequency) of that exact structure. In addition, Traxler et al. (2002) provided data supporting frequency-based assumptions in a study manipulating the animacy of the NPs. They argued that the processing difficulty for ORCs is caused by the comprehenders' frequency-based assumptions that the sentential subject is the subject of the relative clause as well as the main clause. They predicted to see a reduced ORC-disadvantage when the sentential subject was inanimate as compared to the animate subject condition (e.g., "*The movie/actor that the director admired was funny*"), because inanimate entities are more typically objects than subjects and thus, it should be easier to revise the interpretation of an inanimate noun (e.g., "*movie*") as compared to an animate noun (e.g., "*actor*") to be the object rather than the subject of the relative clause since. As expected, Traxler et al. did find an interaction between animacy of the NPs and relative clause processing advantage, and they explained such results by appealing to a frequency-based account.

However, researchers who support WM-based accounts point out that some results cannot be easily explained by the experience-based account, such as the locus of difficulty in ORC processing (Gibson & Wu, 2013; Grodner & Gibson, 2005) and interference effects in a dual-task paradigm (Fedorenko, Gibson, & Rohde, 2006, 2007; Gordon, Hendrick, & Johnson, 2001; Gordon et al., 2004; Gordon, Hendrick, Johnson, & Lee, 2006; Gordon et al., 2002). Gibson and Wu (2013) pointed out that the experience-based account would predict processing difficulty with the less frequent structure (i.e., ORC) as soon as readers recognized this structure (e.g., in Example 8, right at the embedded subject "*the senator*" when reading the ORC structure "*who the senator attacked*"). However, the processing difficulty has been reported to occur most often on

the embedded verb “*attacked*” rather than the embedded subject. In addition, in experiments using a dual-task paradigm, researchers have found that individuals’ sentence processing performance was only impaired when the secondary task was verbal (e.g., remembering a set of words or solving arithmetic operations) but not spatial (e.g., spatial orientation) (Fedorenko et al., 2006, 2007), or when the words in the memory list shared similarities with to-be-retrieved items (Gordon et al., 2002, 2006; Van Dyke & McElree, 2006). This verbally specific, similarity-based interference is important to debates between experience- and capacity-based account, because the experience-based account would predict the same results regardless of types of secondary task or the characteristics of the memorized items, if only reading experience matters. As a result, Gibson and colleagues (2013) have suggested that WM-based accounts could provide a better explanation for robust SRC-advantage in English.¹³

1.2.4.1.2 ORC-advantage in Chinese RC processing

Recently, several studies investigating RC processing in Chinese have provided strong support for the WM-based account by showing that there was an advantage of processing ORC structure, in which the syntactic integration is more distant than in the SRC structure (Chen et al., 2008; Gibson & Wu, 2013; F. Hsiao & Gibson, 2003; Packard, Ye, & Zhou, 2011). Despite the fact that in Chinese, as in many other

¹³ Regarding the empirical evidence from cross-linguistic studies, in most languages, for both head-initial languages (e.g., English, Dutch, German, etc.) and head-final languages (e.g., Japanese, Korean, etc.), there is a SRC processing advantage as in English (Ozeki & Shirai, 2007; Traxler et al., 2002) – e.g., in German SRCs: 74% vs. ORCs 26% (Korthals & Christian, 2001) and in Japanese 87.4% of the sentences produced by native speakers with RCs that modifying animate nouns were SRC structures (Ozeki & Shirai, 2007), and so on. Since both WM-based accounts and non-capacity accounts can explain such an advantage, studying the SRC-advantage is not informative for deciding among these theories,

languages, SRCs are more frequent than ORCs (72.6% vs. 27.4%, Chen et al., 2010)¹⁴, Chinese is the first language in which a reversed RC processing effect with an ORC-advantage has been reported. In contrast to English, the distance between the filler-gap dependencies is larger in SRCs than in ORCs. For example, as shown in Example 9, the linear distance between the head noun “*official*” and its empty NP position in the RC (as marked by “GAP” in the example) is larger in the SRC sentence, while there is just a local integration in the ORC structure.

Example 9 (Hsiao & Gibson, 2003)

a. Subject-modifying SRC

	邀请	富豪	的	官员	图谋不轨
[[GAP	yaoqing	fuhao	de]	guanyuan]	xinhuaibugui
[[GAP	invite	tycoon	REL]	official]	have bad intentions
“The official who invited the tycoon had bad intentions.”					

b. Subject-modifying ORC

富豪	邀请		的	官员	图谋不轨
[[fuhao	yaoqing	GAP	de]	guanyuan]	xinhuaibugui
[[tycoon	invite	GAP	REL]	official]	have bad intentions
“The official who the tycoon invited had bad intentions.”					

¹⁴ The early corpus study carried out by Hsiao and Gibson (2003) based on 882 instances from the Chinese Treebank (version 3.0) showed that 42.5% of these instances were ORCs, while the remaining 57.5% instances were SRCs. Hsiao et al. have restricted their counts to active RCs (omitting passives) and argument relativization (omitting adjunct relativizations such as “*the reason why he left*”). In particular, they mentioned that they “also did not count simple phrases that lacked copula verbs that could be analyzed as reduced subject-extracted RCs, e.g., prepositional phrases such as ‘*The company in China*’ cf. ‘*The company that is in China*’, or adjectival phrases such as ‘*The big company*’ vs. ‘*The company that is big*’. We thought that the inclusion of such items could artificially increase the number of subject-extracted RCs”.

In a more recent corpus analysis, through searching through a different corpus containing more words (Sinica Corpus 3.0 developed in Taiwan), Chen et al. (2010) reported a consistent but slightly different frequency pattern as Hsiao et al., with SRCs (72.6%) much more frequent than ORCs (27.4%). Chen et al. (2010) examined 639 sentences in total. 164 out of these sentences had similar structures to the sentences included in Hsiao and Gibson’s corpus study, which only examined sentences with animate RC subject. Chen et al. pointed out that a potential problem of Hsiao and Gibson’s corpus analysis was that if restricted RC subject and object to animate only, all the six instances remained were SRC structure, though 13/16 instances were SRC structure in Chen et al.’s study. Overall, the SRC structure is more frequent than ORC structure in Chinese.

Therefore, examining Chinese RC processing provides a particularly important test case to help decide between WM-based and experience-based theories: WM-based accounts predict an ORC-advantage due to longer distance or more intervening words between the unintegrated dependencies in the SRCs, whereas experience-based accounts predict an SRC-advantage due to higher frequency of SRC structure. In a series of studies conducted by Gibson and colleagues, they confirmed the ORC-advantage in Chinese RC processing as predicted by WM-based account (Gibson & Wu, 2013; Hsiao & Gibson, 2003; Hsiao, 2003). Specifically, the relative clause marker “de” and head noun (e.g., “official” in Example 9) were processed significantly faster in sentences with ORCs than sentences with SRCs. In addition, there is also neuropsychological evidence showing that both Mandarin Chinese-speaking aphasic patients (Su, Lee, & Chung, 2007) and Cantonese-speaking aphasics (Law & Leung, 1998, 2000) have more difficulty in answering question for sentences with SRC structures relative to ORC structures, while English-speaking aphasic patients typically showed the opposite pattern (Caplan & Futter, 1986; Caramazza & Zurif, 1976; Hickok, Zurif, & Cansecogonzalez, 1993). As a result, Gibson et al. have suggested that the ORC-advantage in Chinese RC processing is most consistent with WM-based accounts.

However, results supporting the experience-based account have also been reported, demonstrating that there is an SRC-advantage in Chinese as in many other languages (Kuo & Vasishth, 2006; Lin & Bever, 2006; Jäger et al., 2015). Researchers favoring the experience-based account have argued that the ORC-advantage found in some Chinese studies is due to a potential temporary ambiguity in the ORC structures: the initial fragment of an ORC structure (NP + VP, e.g., “tycoon invite” in sentence 8b)

is misinterpreted as a main clause rather than a relative clause due to the fact that there is no relative clause marker at the beginning of the sentence. Thus, subjects are faster at the beginning of their reading of the relative clause because they treat the first part as a main clause but then they slow down in a later part when they find out it is a RC structure. Thus, subjects have to revise their initial main clause interpretation of an ORC structure at a later point. In a replication of Hsiao and Gibson's study (2003), Lin and Bever (2006) found a consistent SRC–advantage in Chinese with SRC structure being read significantly faster than the ORC structure on both the relativizer “*de*” and the head noun, where filler-gap integration happened¹⁵, although no RT differences were observed during processing of the RC. To explain the divergent results from Hsiao and Gibson's study (2003), Lin et al. (2006) argued that there was an ambiguity problem in Hsiao and Gibson's study. The embedded verbs used in their experiment did not match across relative clause condition on syntactic features - some of the verbs they used could take sentential complements in addition to nominal objects, while some could only sentential complements. Lin et al. suggested that this syntactic ambiguity affected subject's expectation about the incoming structure.

In response, Gibson and Wu (2013) argued that the Lin et al.'s study was problematic in that they collapsed across both subject-modifying RCs (i.e., the head noun of RC is the subject of the main clause) and object-modifying RCs (i.e., the head noun of RC is the object of the main clause). As shown in Example 9, Hsiao and Gibson (2003) had only used subject-modifying RCs in their experiment. Gibson et al. pointed out that

¹⁵ The syntactic category of the Chinese relative marker “*de*” is not very clear. Some researchers argue that it is a complementiser, corresponding to “that” in English, while some researchers argue that it is a general linker of a modifier to a head (see Gibson & Wu, 2013, for a discussion). However, no matter what kind of syntactic category of “*de*”, many Chinese studies have shown that filler-gap integration could happen even before the presentation of head noun (Gibson & Wu, 2013; Packard et al., 2011; Yang et al., 2012).

in the object-modifying RCs (e.g., sentence 9b), there was a temporary ambiguity on the ORC structure that required later re-analysis. It was very likely that the grammatical subject of the ORC structure (e.g., “*dean*”) was initially interpreted as the direct object of the main verb (e.g., “*bumped into*”) due to higher corpus statistics (Kuo & Vasishth, 2006), lower storage costs (Gibson, 1998), or the Minimal Attachment heuristic (Frazier, 1978, summarized by Gibson & Wu, 2013). Indeed, both Lin (2006) and Hsiao (2003) only found an SRC-advantage on the relativizer “*de*” and head noun for object-modifying RCs but not for subject-modifying RCs (Hsiao, 2003; Lin, 2006). Thus, there was an SRC preference after collapsing across subject- and object-modifying RCs. Moreover, in a later study, Gibson and Wu (2013) tried to eliminate this ambiguity by adding supportive contextual information that biased the comprehender towards a RC interpretation of the sentences (which would support the processing of both kinds of RCs), then subjects should be aware of an upcoming RC at an early point. Gibson and Wu replicated the ORC-advantage reported in the earlier study (Hsiao & Gibson, 2003). The ORC-advantage occurred at the head noun following the RC region but did persist to the later part of the sentence (i.e., there were three more words after the head noun). However, because I am most interested in the processing differences at the critical verb, it should be noted that a potential problems of Gibson and Wu’s study was that they did not control for the word phrases after the head noun. The regions following the head noun differed lexically and syntactically across the 16 sets of sentences they used - about half of the head noun were followed by the word “*is*”, while the other half were followed by a conjunction word, an adverbial phrase, and so on. Such variation might create confounds

making it difficult to observe any effect associated with RC processing in the regions following the head noun.

In a very recent study, Jäger et al. (2015) further tested the experience-based account and capacity-based account in Chinese RC processing, through eliminating the local ambiguities by adding structural cues (i.e., determiner, adverbial phrase) to generate a syntactic configuration in which a RC is highly predicted (as shown in Example 10). They manipulated both modification type (i.e., subject-modifying vs. object-modifying) and RC type (subject RC vs. object RC). The results from both self-paced reading and eye-tracking experiments found an SRC-advantage in the relative clause region (e.g., embedded verb + N), which disappeared in the relativizer *de* and the head noun. However, in the subject modifying sentences, such SRC-advantage showed up again on the matrix verb in the eye-tracking measures (e.g., “*know*” in Example 10 a & b) and on the first phrase following the main verb in the self-paced reading measures.¹⁶ Jäger et al. argued that this latter effect was hard to explain by either Gibson’s theory or the experience-based account, as both would predict no processing differences in any region after the head noun¹⁷. Therefore, they did not further discuss this SRC-advantage on the matrix verb. However, I suggest that the SRC-advantage on the matrix verb was actually quite consistent with the cue-based parsing approach, according to which a processing difficulty is predicted on the matrix verb following ORC sentences as there is higher syntactic interference. For example, the intervening NP “*boy*” in Example 10 was a grammatical object in the relative clause in 10a (SRC condition); while it was a

¹⁶ As shown in Example 10, the words following the head noun differed lexically and syntactically across modification type (subject- vs. object - modification). Thus, Jäger et al. analyzed the RC type differences within each modification type for the post-head regions,

¹⁷ Jäger et al. suggested that this might be explained as that there is a higher conditional probability to produce RC head noun in SRCs compared to ORCs within subject-modifying sentences.

grammatical subject in the relative clause in 10b (ORC condition). Therefore, there should be more processing difficulty on the matrix verb in sentence 10b, caused by higher syntactic interference.¹⁸ The finding of an SRC-advantage on the main verb in Jäger et al.’s study is of great relevance to my current study.

Example 10. (Jäger, Chen, Li, Lin, & Vasishth, 2015)

a. Subject-modifying SR															
那个			上个月	邀请了	男孩	几次	的		女孩	认识	王老师	因为	上过	她的	课。
Na-ge	[_{RC}	t _i	shanggeyue	yaoqing-le	nanhai	ji-ci	de	]	nūhai _i	renshi	Wang	yinwei	shang-guo	ta-de	ke
DET-CL			last month	invite-ASP	boy	several-CL	REL		girl	know	teacher Wang	because	attend-ASP	her	class
‘The girl who invited the boy several times last month knows teacher Wang because [she] has attended her class.’															
b. Subject-modifying OR															
那个			上个月	男孩	邀请了	几次	的		女孩	认识	王老师	因为	上过	她的	课。
Na-ge	[_{RC}	shanggeyue	nanhai	yaoqing-le	t _i	ji-ci	de	]	nūhai _i	renshi	Wang	yinwei	shang-guo	ta-de	ke
DET-CL		last month	boy	invite-ASP		several-CL	REL		girl	know	teacher Wang	because	attend-ASP	her	class
‘The girl who the boy invited several times last month knows teacher Wang because [she] has attended her class.’															
c. Object-modifying SR															
王老师	认识	那个			上个月	邀请了	男孩	几次	的		女孩	因为	教过	她的	课。
Wang	renshi	na-ge	[_{RC}	t _i	shanggeyue	yaoqing-le	nanhai	ji-ci	de	]	nūhai _i	yinwei	jiao-guo	ta-de	ke
laoshi	know	DET-CL			last month	invite-ASP	boy	several-CL	REL		girl	because	teach-ASP	her	class
teacher Wang															
‘Teacher Wang knows the girl who invited the boy several times last month because [she] has taught her class.’															
d. Object-modifying OR															
王老师	认识	那个		上个月	男孩	邀请了	几次	的		女孩	因为	教过	她的	课。	
Wang	renshi	na-ge	[_{RC}	shanggeyue	nanhai	yaoqing-le	t _i	ji-ci	de	]	nūhai _i	yinwei	jiao-guo	ta-de	ke
laoshi	know	DET-CL		last month	boy	invite-ASP		several-CL	REL		girl	because	teach-ASP	her	class
teacher Wang															
‘Teacher Wang knows the girl who the boy invited several times last month because [she] has taught her class.’															

To summarize, the results from Chinese RC processing studies are inconclusive.

Although many studies have reported an ORC-advantage in the filler-gap integration process and thus support the WM-based account of sentence comprehension, there is also

¹⁸ Moreover, in Jäger et al.’ study (2015), the results that no RC type differences was found in the post-head noun regions for object modifying sentences were also consistent with the cue-based parsing approach, as subjects only need to do a local integration (e.g., integrate “teach” with “her class”), while no distant integration is necessary. It should be noted that there was actually an ambiguity in the last region of the sentences presented in Example 10 as used by Jäger and colleagues. The pronoun “her” could refer to either the “teacher” or the “girl”, though the pragmatic cues bias the “teacher”. This might pose a problem to their analysis on the sentence final region, though which would not the focus of our current discussion.

evidence arguing for an SRC-advantage as predicted by the experience-based account. Most of the previous Chinese studies have mainly focused on the RC processing differences at the relativizer “*de*” and its following head noun, where distant integration happens, in order to test the predictions generated by WM-based vs. experience-based account. However, I am most interested in which RC type causes processing differences at the main verb, which could potentially help us test the predictions from the cue-based parsing approach. Among the previous studies, some found no RC type difference at the main verb (Gibson & Wu, 2013; Hsiao, 2006; Lin, 2006), some found an ORC-advantage (Chen et al., 2010; Packard et al., 2011), and some found an SRC-advantage after controlling for the temporary ambiguity problem (Jäger et al., 2015; Yang et al., 2010). However, so far, none of the previous studies has related the processing effect on the main verb to the syntactic interference effect which is relevant for testing predictions of the cue-based parsing approach. I hypothesize that no matter whether there is an ORC-advantage (predicted by the WM-based account) or an SRC-advantage (predicted by the experience-based account) on the relativizer “*de*” and the following head noun, there should be an ORC-disadvantage at the main verb caused by higher syntactic interference. While both the DLT account and the experience-based accounts predict little effect on the main verb, the cue-based parsing account predicts an SRC-advantage as there is higher syntactic interference in the ORC structure (see more discussion later). Thus, in the current project, I aimed to test the cue-based parsing approach through examining subjects’ performance on reading Chinese sentences with RC structures, in order to provide cross-linguistic evidence to the cue-based parsing approach, and further investigate the WM mechanisms underlying sentence comprehension.

1.2.4.2 Explaining the subject-object RC processing differences through the cue-based parsing approach

The studies summarized above mainly focused on whether the subject-object RC processing asymmetry in both English and Chinese could be explained by the WM load differences between the two RC structures. Instead, researchers who embrace the cue-based parsing proposed that such processing asymmetry could be explained by the differences in memory retrieval between SRC and ORC structures. Lewis and Vasishth (2005) have developed a computational model of the WM processes that subserve syntactic processing in the ACT-R (Adaptive Control of Thought-Rational) architecture, based on the cue-based parsing theory (Lewis & Vasishth, 2005). Their model construes parsing as an efficient series of memory retrievals, and interference arises at retrieval when the strength of association between a cue and a target reduced as a function of cue overlap. The model was successful in accounting quantitatively for processing differences in varied sentence types, including the basic contrast between SRCs and ORCs in English. In Lewis and Vasishth's study, the simulation results showed that both the embedded verb and the main verb were processed slower in the ORC condition in English. They suggested that the RT differences are due only to differences in WM retrieval times (i.e., extra retrieval cycles), which is modulated by activation fluctuation. Considering the ORC shown in Example 8 "*The reporter who the senator attacked [GAP] admitted the error*", for both the embedded verb "*attacked*" and the matrix verb "*admitted*", there is an additional retrieval associated with retrieving the relative pronoun "*who*" to fill the gap. However, in the corresponding SRC condition such as "*The*

reporter who [GAP] attacked the senator admitted the error”, the subject gap is filled at the relative pronoun “*who*”, thus there is no additional retrieval. Such results could be conceptually explained through the cue-based parsing approach as there is more syntactic interference at both the embedded and main verb in the ORC structure. At the embedded verb, in the SRC condition, there is only one NP (e.g., “*reporter*”) that could be retrieved. Thus, there is no proactive or retroactive interference. However, in the ORC condition, there is proactive syntactic interference from the head noun (e.g., “*reporter*”) while readers are retrieving “*senator*” as subject of the embedded verb. At the main verb, there are two NPs that could be retrieved, but the intervening NP in the ORC condition is a grammatical subject (e.g., “*senator*” in the clause “*who the senator sent to the editor*”) while it is an object in the SRC condition (e.g., “*senator*” in the clause “*who sent the senator to the editor*”). Therefore, there should be more syntactic interference at the matrix verb in the ORC condition.

Moreover, the simulation results from Lewis and Vasishth’s model demonstrated that there was an interaction between RC type (SRC vs. ORC) and verb type (embedded vs. matrix). Generally, RTs were longer in the ORC as compared to the SRC condition, but this SRC-advantage was smaller at the main verb than at the embedded verb (27 vs. 105; Note that the unit for the model output is different from real RT data). The interaction of RC type and verb type is consistent with the results from previous behavioral studies (Gordon et al., 2001; Grodner & Gibson, 2005; Just & Carpenter, 1992). Some researchers have suggested that the processing differences on the main verb are just a spillover effect from the embedded verb (e.g., “*attacked*” in Example 8) (Grodner & Gibson, 2005). Grodner and Gibson (2005) provided evidence for this

hypothesis by showing that the SRC-advantage at the main verb disappeared after inserting a prepositional phrase to separate the main and embedded verbs (e.g., inserting “*to the editor*” into the sentence “*The reporter who the senator sent to the editor admitted the error*”). They argued that as long as the distance between the subject and the main verb is the same between the SRC and ORC conditions, the processing cost associated with integrating new input should be the same. Grodner et al. suggested that these results strongly support the hypothesis that WM capacity constrains language processing. However, as just discussed above, the cue-based parsing account would actually predict easier processing of the main verb after the SRC than ORC structure. The finding of an SRC-advantage in some studies but not others might be due to many differences in methodology (e.g., self-paced reading vs. eye-tracking), materials (e.g., in some studies, the main verb is the final word), and so on. So far, none of these previous studies on RC processing have systematically examined the syntactic and semantic features of the intervening NP and taken interference effect into consideration, while controlling for other factors.

However, it is worth noting that Lewis et al.’s model only simulates the syntactic aspect of sentence processing, whereas online sentence parsing is affected by many other factors as well, such as lexical-semantic, contextual, and pragmatic information (Boland, 1997; Tanenhaus & Trueswell, 1995). For example, Traxler et al. (2002) have demonstrated that the robust SRC-advantage in English is greatly reduced when there are semantic cues that help subjects to make prediction about the incoming RC structure (e.g., an inanimate sentential subject strongly implies a ORC structure). It is necessary to

integrate other factors, as well as syntactic prediction in addition to retrieval (e.g., Bartek et al. 2011; Fiebach et al., 2002; Gibson, 1998, 2000) into the sentence parsing models.

1.2.4.3 Explaining the subject-object RC processing differences in Chinese through the cue-based parsing approach

Since the cue-based parsing approach could successfully capture the SRC-advantage in English sentence comprehension, examining the RC processing differences in a language such as Chinese, which has different properties from English (which will be discussed later), would allow us to further test the predictions from cue-based parsing approach that interference is a critical determinant of sentence processing difficulty.

To date, no study has applied the cue-based parsing approach to explain the RC processing asymmetry in Chinese, though the results from many existing studies are actually consistent with the prediction from cue-based parsing approach. If attention is restricted to the matrix verb in Chinese sentences containing either an SRC or an ORC structure, the cue-based parsing approach predicts greater syntactic interference effect in the ORC structure as compared to the SRC structure, because the embedded NP plays a role of grammatical subject in the ORC structure. For instance, taking the sentences from Hsiao and Gibson's study (2003) as an example (as shown in Example 9), while readers trying to link the main verb "*had*" and its subject "*official*", the embedded noun phrase "*tycoon*" would cause greater syntactic interference in the SRC condition (e.g., "*Invite tycoon de official had bad intentions*"; English: "*The official who invited the tycoon had bad intentions*") than in the ORC condition ("*Tycoon Invite de official had bad*

intentions”; English: “*The official who the tycoon invited had bad intentions*”). Because “*tycoon*” satisfies the syntactic retrieval requirement of “a grammatical subject” in the ORC structure and thus, causes more syntactic interference; whereas it plays a role of grammatical object in the SRC structure and does not satisfy the syntactic retrieval cue. Therefore, one would expect to observe longer RTs or greater brain activation on the main verb following an ORC structure than that following an SRC structure. Indeed, this was the result reported in Jäger et al.’s study (2015), as I just summarized. Thus, I suggest that through carefully controlling the context adjacent to the main verb (e.g., blocking the spillover effect from RC processing), there should be an SRC-advantage at the main verb and thus, provide additional support for the critical role of interference in determining sentence processing difficulty.

1.2.5 Evidence from ERP studies of relative clause processing

So far, several studies have examined RC processing in English and Chinese using the ERP technique. In the English studies examining relative clause processing, an LAN effect with a left anterior distribution between 300 and 500 ms has been observed on both the embedded verb and the main verb in sentences containing an ORC structure as compared to sentences with a SRC structure (King & Kutas, 1995). King and Kutas (1995) interpreted this LAN effects in terms of the processing difficulty associated with higher WM load caused by tracking multiple thematic roles for a given noun phrase (i.e., the head noun is the agent of main verb and the patient of embedded verb). In addition, a bilateral fronto-central sustained negativity between 200 and 900 ms has been reported at the linguistically defined “gap” position (e.g., “*The fireman who [speedily rescued the*

cop]/[*the cop speedily rescued*] [*GAP*] *sued the ...*”) in spoken sentences with ORC structure compared with SRC structure (Müller, King, & Kutas, 1997). Müller et al. attributed the slight differences of the ERPs in scalp distribution and time course between their study with spoken sentences and King et al.’s study with written sentences (1995) to the modality difference.

In an ERP study using Chinese sentences, Yang, Perfetti, and Liu (2010) compared processing differences between object-modifying sentences with SRC and ORC, as shown in Example 11.

Example 11. Object-modifying RCs (Yang, Perfetti & Liu, 2010)

- a. SRC
 那个 议员 介绍 GAP 攻击 政客 的 那个
 律师 给 公众 认识。
 The senator introduce GAP attack politician *de* the
 lawyer to public known
 “The senator introduced the lawyer that attacked the politician to the public.”
- b. ORC
 那个 议员 介绍 政客 攻击 GAP 的 那个
 律师 给 公众 认识。
 The senator introduce politician attack GAP *de* the
 lawyer to public known
 “The senator introduced the lawyer that the politician attacked to the public.”

For ERP data, Yang et al. focused on the comparison between different RC types in three regions: 1) first word of RC region (e.g., “*attacked*” in SRC vs. “*politician*” in ORC); 2) second word RC region (e.g., “*politician*” in SRC vs. “*attacked*” in ORC); 3) head noun (e.g., the determiner “*that*” and NP “*lawyer*” in both conditions). At the first word of RC region, Yang et al. (2010) observed an SRC-elicited (i.e., compared SRC condition over ORC condition) P600 between 600 – 800 ms preceded by an N400 effect

at the embedded verb of the SRC (e.g., “*attacked*” in sentence 6a). They suggested that the SRC-elicited P600 effect reflected a phrasal restructuring process of linking the embedded verb (e.g., “*attack*”) to the main-clause verb (e.g., “*introduce*”) in the SRC structure. The reconstruction happened because readers tended to initially anticipate a direct object NP (as what occurs in the ORC condition) after reading the matrix verb in the SRC condition (in their written cloze sentence completion task, 98% of the S-V-fragment were finished with a direct object NP). Thus, readers have to reinterpret the phrase structure. For the second word or RC region, an ORC-elicited N400 effect between 370 – 500 ms was observed. Yang et al. suggested that this N400 effect should be associated with increasing WM demands in maintaining and reactivating unintergrated referents for structural integration in the ORC structure. For the head noun region, no reliable differences were found on the determiner (“*the*”) before “*lawyer*”, but an ORC-elicited sustained AN between 250 - 800 ms was observed on the head noun (“*lawyer*”). In addition, Yang et al. suggested that the ORC-elicited AN is a combined effect of an N400 corresponding to the semantic integration difficulty and a frontal-dominant LAN effect corresponding to the WM demands underlying binding processes (i.e., establishing reference). Furthermore, Yang et al. claimed that the SRC-elicited N400 on the head noun, which was not observed in English studies, might be related to a linguistic-specific property of Chinese. The sharp N400 may reflect the involvement of lexical-semantic processes to resolve syntactic difficulty, because Chinese has a larger number of class-ambiguous words that can be used as both nouns and verbs (e.g., (like “*fish*” in English).

Readers need to identify the word as a possible transitive verb in RC structure first, then proceed to structure reconstruction and semantic processing¹⁹.

In another study, Packard, Ye, and Zhou (2011) avoided comparing words from different classes by making use of the fact that the RC occurs before the head noun in Chinese and focusing on the comparison of the relative marker “*de*” and the head noun, which were at exactly the same position in both SRC and ORC structures. Participants were tested on four conditions crossed by RC modifying position (subject-modifying vs. object-modifying) and RC types (SRC vs. ORC). The examples of the sentences used in Packard et al.’s study are shown as below.

Example 12 (Packard et al., 2011)

a. Subject-modifying SRC

GAP	完成	论文	的	学生	获得了	学位
GAP	complete	thesis	de	student	obtained	degree

“The student that completed her thesis obtained her degree”

b. Subject-modifying ORC

教授	指导	GAP	的	学生	发表了	文章
professor	advise	GAP	de	student	published	article

¹⁹ For the above ERP data, notably, in both English mentioned above (King et al., 1995; Müller et al., 1991) and in this Chinese RC study, researchers compared different words at the same position of the RC region. For example, they compared the noun phrase “*the senator*” in sentence 8a versus verb phrase “*attacked*” in sentence 8b. This was done because researchers tried to control for the sentence position effect caused by the direct comparison of the same words at different position (e.g., “*attacked*” is the 4th word in sentence 8a and the 6th word in sentence 8b). Previous studies have indicated that the amplitude of an ERP effect is affected by the target words’ position in the sentence (Kutas, Federmeier, Coulson, King, & Mute, 2000, cited by Yang et al., 2010). In addition, Yang et al. (2010) suggested that the direct contrast of embedded verbs in different conditions might reflect different processes (revision vs. reanalysis). Yang et al. argued that, as a result, comparing different words at the same position has limited the confounding variables to “one single dimension of lexical differences” (Yang et al., 2010, p89, footnote 3). Yang et al. claimed that the effects they reported are “robust enough to weigh against the lexical-driven alternative”. However, we suggested that comparing different words at the same position might not narrow the confounding variables to lexical factors. Instead, it may actually bring in other confounding variables such as different syntactic and semantic integration processes. For example, processing the verb phrase “*attack*” involves generating retrieval cues to look for its subject, while the processing of the noun phrase “*senator*” does not engender such a process.

“The student that the professor advised published the article”

c. Object-modifying SRC

老师	表扬了	GAP	完成	作业	的	学生
teacher	praised	GAP	complete	assignment	de	student

“The teacher praised the student that completed the assignment”

d. Object-modifying ORC

公司	录用了	老师	推荐	的	GAP	学生
company	hired	teacher	recommend	de	GAP	student

“The company hired the student that the teacher recommended”

For the behavioral measures, no significant differences between conditions were found in accuracy data for the probe sentence (Yes/No answer required) following each sentence. For the ERP data, in the subject-modifying RC condition, a significantly greater P600 effect was found in the SRC over the ORC structure on the relative marker “*de*”; while in the object-modifying RC condition, this SRC-elicited P600 effect was found on the head noun following the “*de*”. Packard et al.’s results were consistent with Gibson et al.’s (2003, 2011) claim that the Chinese SRC is more difficult to process than the ORC, because there is a more distant integration (i.e., filler-gap integrations) in the SRC structure as reflected by the P600 effect. In addition, they suggested that the different locus of the P600 effects in subject-modifying and object-modifying RC condition was caused by the fact that the embedded verb’s integration process happens later in the object-modifying condition. In Chinese, a RC structure without a head noun is very common. For example, as shown in Example 12a, the phrase “*complete thesis de [student] obtained degree*” is also grammatical after leaving out the head noun “*student*”. Thus in the subject-modifying condition, readers may immediately start the integration when “*de*” is presented. Overall, Packard et al. suggested that sentences with subject-modifying SRCs are just more difficult to process overall. Furthermore, Packard et al.

examined the ERP effects on the main verb in the two modifying conditions separately. In the subject-modifying conditions, there was a broadly distributed ORC-elicited N400 effect on the matrix verb, indicating more difficulty in the assignment of thematic roles in ORC condition (i.e., the RC head noun is the subject of the matrix verb and the object of the RC verb). However, in the object-modifying condition, no RC type difference was observed, although there is also a thematic mismatch in the SRC condition. Packard et al. attributed this to the temporary ambiguity in the ORC condition, in which subjects are expecting a direct object rather than a RC structure. Then the ambiguity effect offset the thematic mismatch effect and resulted in approximately equal processing difficulty. To confirm this hypothesis, they examined the ERP effects on the embedded verb in both conditions, and did observed a ORC-elicited P600 effect, which is traditionally associated with temporary syntactic ambiguity (Osterhout & Holcomb, 1992).

To summarize, the investigation into differences in processing difficulty between Chinese SRC and ORC structure is not conclusive yet, with conflicting results from both ERP and behavioral studies. For the ERP results, the differences between Yang et al.'s study (2010) and Packard et al.'s study (2011) may be attributed to differences in the sentence materials they used. The ERP effects on the head noun observed in Packard et al.' study might be confounded with a wrap-up effect at the sentence-final position as the head noun was always the last word in their sentence, whereas there was one more phrase (e.g., "*to the public*" in Example 10) following the head noun in Yang et al.'s study.. It is noteworthy that there were temporary ambiguity problems in both studies (as discussed in section 1.2.4.1.2), which may be orthogonal to the differences of RC processing

difficulty. Therefore, in order to better understand the mechanisms that subserve RC processing, a more careful examination is required.

1.3 Exploring individual differences in sentence processing efficiency

One of the basic approaches to investigate the underlying mechanisms supporting sentence processing is to relate individual differences in other cognitive resources to subjects' performance in sentence comprehension. A significant correlation between a certain cognitive ability and sentence reading implies that this cognitive ability might be a crucial predictor, though such results do not necessarily establish a cause-effect relationship. The individual differences approach has been widely used in behavioral studies, but has been less frequently applied in neuroimaging studies due to its requirement for a large sample size. With respect to the individual differences approach, different accounts make different predictions regarding how brain activity during sentence processing is modulated by subjects' variation in other cognitive factors, such as the WM capacity (as predicted by the capacity-based account), reading experience (as predicted by the experience-based account), and attentional control ability (as predicted by the executive control-based account). In the current project, I aimed to investigate the cognitive resources supporting sentence comprehension through examining how individual differences in brain activity relate to the processing of semantic and syntactic interference. The following section will summarize some relevant studies looking into the correlations between individual differences in basic WM, EF, or verbal knowledge and sentence processing.

The general implication drawn from previous studies investigating the WM-language relation is that WM supports language processing. Behaviorally, subjects with higher WM span generally show less difficulty in language processing, as indexed by shorter RT in self-paced reading or eye-movement measures, and higher accuracy in answering comprehension question. However, researchers have not come to a final conclusion about the nature of the WM-language relation at the neural level. Based on a number of neuroimaging studies, Prat and colleagues (Prat & Just, 2011; Prat, Mason, & Just, 2012) have claimed that individual differences in sentence comprehension could be reflected in three important facets of brain function: neural adaptability, neural efficiency, and neural synchronization. Regarding neural adaptability, Prat et al. suggest that individuals with high capacity may show more neural activation as task demands increase. Prat and colleagues provided supporting evidence for the neural adaptability hypothesis by showing that individuals with higher WM capacity (as indexed by reading span score) showed greater recruitment of both prefrontal cortex and striatum than did individuals with lower WM capacities in the contrast of processing syntactically complex versus simple sentence (Prat et al., 2011) and processing sentences with low-frequency nouns versus sentences with high-frequency nouns (Prat, Keller, & Just, 2007). Regarding neural efficiency, Prat et al. suggested that high capacity subjects could be more efficient and thus, use fewer mental resources than low capacity subjects, as evidenced by less activation in the given brain regions or more focal activation during the task (Prat & Just, 2011; Prat et al., 2007; Reichle, Carpenter, & Just, 2000). For example, Reichle, Carpenter and Just (2000) reported that subjects with better verbal WM capacity showed lower activation in Broca's area when they were guided to use a verbal strategy

in a sentence identification task. Similarly, subjects with better visual-spatial skills showed lower activation in regions involved in visual association (i.e., certain regions in the parietal cortex), when they were encouraged to use a spatial strategy. Reichle et al. suggested that these results demonstrated that high verbal/spatial WM capacity subjects have more available resources to perform verbal/spatial tasks, respectively. In two following studies, Prat et al. (2007) further supported the hypothesis about neural efficiency by showing that during general sentence comprehension (i.e., across experimental conditions), high-WM capacity subjects generally utilized fewer neural resources in the bilateral middle frontal and right lingual gyri than did low-WM capacity subjects (Prat et al., 2007). Last, regarding neural synchronization, high capacity subjects have been reported to show better ability in maintaining synchronization between different brain regions (e.g., Broca's area and Wernicke's area during sentence comprehension), especially when lexical or syntactic processing demands are high (Prat & Just, 2011; Prat et al., 2007).

Although the predictions from the “neural adaptability” and the “neural efficiency” facets seem quite contradictory and predict contrasting results of the WM-language correlations – high capacity subjects' superior performance in behavioral measures (e.g., shorter RTs and higher accuracy) might be related to either less (i.e., neural efficiency) or more (i.e., neural adaptability) brain activation during complex sentence processing, Prat et al. (2007, 2011, 2012) suggested that the individual differences in neural adaptability are related to, but relatively independent from neural

efficiency.²⁰ They argued that the individual differences in neural adaptability might arise from some property of brain function, such as plasticity and control ability, instead of neural efficiency. Moreover, based on the results from a series of studies, Prat et al. (2011) claimed that both neural adaptability and efficiency could be predicted by individuals' WM capacity, depending on the specific task and design, while neural efficiency is primarily predicted by individuals' vocabulary size.

The hypotheses from the “neural adaptability” and the “neural efficiency” facets of brain function are relevant for the purposes of the current project, because they affect the direction of the WM-language relations as well as the exact tasks that potentially have the best ability to explain differences among subjects in ERP studies. In a number of previous ERP studies, it has been demonstrated that individuals with better WM, EF, or verbal knowledge might show either larger or smaller ERP effects, depending on the specific experiment designs and subject populations.

²⁰ Regarding the relationship between “neural adaptability” and “neural efficiency” facets, Prat et al. (2007, 2011) suggested that the individual differences in neural adaptability might arise because of more efficient processing in baseline conditions (thus resulting in more available resources when task demands increase). Nonetheless, the neural adaptability facet cannot be fully explained by subjects' differences in processing efficiency for two reasons. First, results from several studies revealed that individual differences in neural efficiency are primarily a function of variation in word knowledge, whereas the largest differences in neural adaptability are mostly related to individual differences in WM capacity (Prat & Just, 2011; Prat, Mason, & Just, 2012). In Prat et al.'s study (2011), they ran random-effects multiple regression models in which vocabulary percentile and reading span scores were entered simultaneously as predictors of interest. The correlations between WM capacities and brain activity were less straightforward than the correlations between vocabulary scores and brain activity. There were both positive and negative correlations between WM capacity and brain activity. Readers with better WM capacity showed less activation in the right parahippocampus region, and more activation in left orbital and right superior frontal regions than did low-WM capacity readers. In their previous study, Prat et al., (2007) also found that high WM capacity subjects showed less activation in frontal regions. However, Prat et al. attributed the increased efficiency (less activation) in high-WM capacity readers to the likelihood that these subjects also have superior reading experience than low-WM capacity readers. Overall, Prat et al. claimed that vocabulary size was a “better and more consistent” predictor for neural efficiency while WM capacity was primarily related to neural adaptability. Second, brain regions that predicted neural adaptability are different from those that predicted neural efficiency. In Prat et al.'s study (2007, 2011), the greater neural adaptability in high-WM capacity individuals was found in prefrontal cortex and in the striatum, although no neural efficiency differences were detected in these regions. Therefore, they concluded that the individual differences in neural adaptability might arise from some other property of brain function, such as plasticity and control ability.

Some ERP studies have provided evidence supporting Prat et al.'s claim (2011, 2012) about better neural efficiency (i.e., less brain activation in the demanding condition) of the high capacity subjects relative to the low capacity subjects during language processing (Federmeier & Kutas, 2005; Federmeier, McLennan, Ochoa, & Kutas, 2002; Fiebach, Schlesewsky, & Friederici, 2001, 2002; Kazmierski, Blasko, & Dessalegn, 2003; St. George, Mannes, & Hoffman, 1997; Van Petten, Weckerly, McIsaac, & Kutas, 1997; Vos, Gunter, Schriefers, & Friederici, 2001). For example, in Fiebach et al.'s study (2002) as discussed previously, they found the AN effect elicited by ORC structure was modulated by individuals' WM capacity – low-WM span subjects showed stronger and more broadly distributed AN effect than high-WM span subjects. In a series of studies with older adults, Federmeier and colleagues (2002, 2005) examined the relationship between aging, WM capacity, and vocabulary fluency, with subjects' efficiency of making use of rich contextual information during sentence processing. They found that some older adults, whose vocabulary fluency was lower than that of younger adults, did not show facilitation (i.e., a reduced N400 effect²¹) like younger adults when the final word was unexpected but semantically related to the predicted sentence completions (e.g., *“So, along the driveway they planted rows of pines”*). Only a subset of the older adults with higher verbal fluency and larger vocabularies showed an N400 pattern similar to the younger adults. As a result, Federmeier et al. (2002) have argued that the ability to make predictions during sentence comprehension is related to the ability to generate lexical items quickly and appropriately as measured by individuals'

²¹ During sentence processing, highly predictable semantically related or congruous words always elicit a smaller N400 effect than do less predictable or incongruous words. The size of such a reduction in the N400 effect (e.g., smaller amplitude, longer latency, or longer duration) has been related to subjects' processing efficiency, and a larger reduction is related to higher efficiency.

vocabulary fluency. However, in a later study with similar materials, Federmeier and Kutas (2005) found that the peak latency of the N400 effect was positively correlated with subjects' age but negatively correlated with WM capacity, indicating that subjects with lower WM span or of older age had more difficulty in exploiting the predictive information from a strongly constraining sentence context as evidenced by a longer delay of the N400 effect. However, vocabulary fluency did not predict the size or latency of the associated N400 effect. Federmeier et al. did not explicitly discuss the differences in results between this study and their early study (2002), in which the age-related decline in sentence comprehension was attributed to verbal fluency decrease, but not WM-capacity. I suggest that one possible explanation could be that different kinds of inference were measured in these two experiments. St. George and colleagues (1997) have pointed out that all the subjects made basic inferences during sentence comprehension. However, only high-span subjects made elaborative inferences. For example, when presented the sentence "*the tooth was pulled painlessly*", the readers might infer the agent of the sentence "dentist", although such inference is optional but not necessary for comprehension. The high-span subjects showed a smaller amplitude of N400 to the elaborative inference condition as compared to the control condition, while the low-span subjects did not show such reduction in ERP amplitude (St. George et al., 1997). As a result, I suggest that it is possible that subjects' ability to make inferences in the strong contextual constraint condition (e.g., Federmeier et al., 2002) is predicted by their word knowledge, while their ability to make use of available information in the current sentence is predicted by WM capacity. However, further investigation is needed.

On the other hand, many ERP studies have provided evidence supporting Prat et al.'s claim (2011, 2012) about better neural adaptability (i.e., more brain activation in the demanding condition) for high capacity subjects relative to the low capacity subjects during language processing. These ERP studies have shown that subjects with high WM capacity are more sensitive to semantic or syntactic integration difficulty in sentence processing as evidenced by larger amplitudes, shorter latency, or qualitatively different components in their ERP effects (Bornkessel, Fiebach, & Friederici, 2004; Federmeier & Kutas, 2005; Fiebach et al., 2002; Friederici, Steinhauer, Mecklinger, & Meyer, 1998; Gunter, Wagner, & Friederici, 2003; Nieuwland & Van Berkum, 2006; Van Petten et al., 1997; Ye & Zhou, 2008). Most of the studies investigating the role WM in sentence processing employed complex span tasks such as reading span (Daneman & Carpenter, 1980) or operation span (Turner & Engle, 1989) as a measure of subjects' general WM capacity. Then these subjects' WM spans are related to their sentence processing performance as indexed by their ambiguity resolution ability (e.g., recovery from garden-path sentences or resolving of referential ambiguity). Although there are still debates about whether the WM mechanisms underlying ambiguity processing are inhibitory (e.g., keep dominant interpretation only) or activational (e.g., keep multiple interpretations active simultaneously), these studies generally support the view that high-span subjects are more adaptable parsers as they can adjust their sentence comprehension strategy dynamically. For example, Nieuwland et al. (2006) found that the high WM-capacity subjects' sensitivity to referential ambiguity (e.g., "*Jennifer Lopes told Madonna that she had too much money*") is reflected in a frontal negative shift, but no significant effect was observed in the low-span group. They concluded that high-span readers are more sensitive

to the different referential interpretations, thus they are more likely to temporarily keep multiple representations active. However, other studies have found opposite results in which high span subjects show greater commitment to a particular interpretation²². In the analysis with only correctly interpreted sentences, Friederici and colleagues (1998) found that high-span subjects showed a syntactic reanalysis effect (P600) at the sentence final disambiguation position when the correct interpretation is the non-preferred one (Friederici et al., 1998) whereas low span subjects did not show such ERP effect. They argued that committing to a single preferred structure is a more efficient processing strategy. Similar findings have been reported from other studies on comprehending ambiguous sentences, in which high-span subjects elicited a greater N400 (associated with integration or switching difficulty in Gunter et al., 2003) or P600 (associated with syntactic reanalysis in Bornkessel, et al., 2004) response to the non-preferred representation relative to the dominant one. In these studies, the observed ERP responses differed with the nature of the ambiguities (e.g., semantic vs. syntactic ambiguity, early vs. late ambiguity), but high-span subjects generally showed better neural adaptability as the processing demands increased.

In addition, several individual differences studies using ERP techniques have also investigated how other measures like executive control relate to language processing (Ye & Zhou, 2008, 2009b). In a Chinese study, Ye and Zhou (2008) found that ERP

²² In Friederici et al.'s experiment (1998), German is used to examine the processing of subject-object ambiguous and unambiguous structures. For example, in the subject-first sentence "*Das ist die Direktorin, die die Sekretarinnen gesucht hat* (this is the director that the secretaries sought has)" and object-first sentence "*Das ist die Direktorin, die die Sekretarinnen gesucht haben* (this is the director that the secretaries sought have)", the relative pronoun form "die" is ambiguous because it marks feminine nominative and feminine accusative case in both singular and plural. The disambiguation happened in sentence final position (e.g., "hat" vs. "haben") as the number marking information of the auxiliary become available.

components induced by a semantic conflict (e.g., “*The thief kept the policeman in the police station*”) were mediated by individuals’ executive control ability as measured by the verbal Stroop task. Only subjects with high executive control ability showed adaptability as syntactic processing demands changed (e.g., processing passive structures vs. active structures). These subjects showed a sustained positivity (350 - 850 ms) in passive sentences, but an anterior negativity (300 - 600 ms) in active sentences. In contrast, subjects with low executive control ability showed a similar sustained positivity between 350 and 750 ms in both active and passive sentences. Ye et al. suggested that the positivity observed in both groups might be associated with detecting and resolving conflict, while the negativity observed in high-control group only might reflect interference suppression, which was similar to that observed in the verbal Stroop task (Liotti, Woldorff, Perez III, & Mayberg, 2000) or the switching task (Brass, Ullsperger, Knoesche, Von Cramon, & Phillips, 2005). Additionally, there are a large number of bilingual studies providing indirect evidence for the correlation between executive control ability and language processing ability (see a review in Hervais-Adelman, et al., 2011). The existing studies using the ERP technique have revealed several similarities between bilingual language control (e.g., switching between the two languages) and other executive control functions (e.g., as measured by Go/No-go task, Stroop task), although most of these studies examined the differences between bilingual subjects and monolingual subjects on the word level (e.g., in word production tasks), which is not the focus of the current project.

Overall, according to these studies, ERP effects related to sentence processing are sensitive to subjects’ variation in WM, verbal knowledge, and attentional control

abilities. The major finding from these studies is that subjects with better WM (e.g., as measured by complex span tasks) or attentional control (e.g., as measured by verbal Stroop) show better neural adaptability, as evidenced by larger amplitude or shorter latency of their ERP responses when the sentence processing demands increase, while subject with better verbal knowledge (e.g., as measured by verbal fluency) show better neural efficiency. However, most of these studies did not consider all of these potentially relevant cognitive abilities in the statistical analysis simultaneously, though it has been shown that most of these individual differences measures are correlated with each other to some extent (Engle, Nations, & Cantor, 1990; Engle, Tuholski, Laughlin, & Conway, 1999). Therefore, it is hard to interpret the observed correlations to the unique contribution of any single factor. Moreover, the exact cognitive ability reflected by a complex span score (e.g., primary memory capacity, attentional control, and secondary memory retrieval) is still a controversial issue (Unsworth, Fukuda, Awh, & Vogel, 2014). Few of the previous studies distinguished the separate contribution from different WM components to sentence comprehension, where difference would be predicted based on the multi-component model of WM proposed by Martin and colleagues (Martin & He, 2004; Martin & Romani, 1994; Martin, Shelton, & Yaffee, 1994) and also the executive control-based account. Therefore, it is necessary to further investigate the nature of the cognitive mechanisms underlying sentence comprehension.

1.4 Motivation for the present study

The aim of the present work was to investigate the interplay of semantic and syntactic interference effects during the comprehension of well-formed sentences through

using the ERP technique, and to examine the role of WM mechanisms that may underlie semantic and syntactic interference resolution during online sentence comprehension. I was interested in determining whether semantic interference could be observed when the distractor was syntactically unavailable, and how the ERP effects associated with semantic and syntactic interference effects were modulated by subjects' WM or executive control ability, beyond their linguistic experience and general processing speed.

1.4.1 Some special properties of Mandarin Chinese

Chinese provides a particularly interesting test case for addressing my research questions. First, as summarized above, the ORC-advantage during RC processing in Chinese as reported by Gibson and colleagues (2003; 2013) allows us to better examine the syntactic interference effect by isolating it from potential processing differences spilling-over from the RC regions, which were not necessarily related to interference manipulations and were unavoidable to some extent in English studies.

Second, examining the interaction of semantic and syntactic interference effects in Chinese could help us resolve the broader question in psycholinguistic studies about how syntactic and semantic retrieval cues are combined during sentence processing. As discussed earlier, there is a long duration debate about the interplay between semantic and syntactic processes. On the one hand, researchers who support the “syntax-first” approaches have suggested that syntactic processing is an automatic and relatively independent process which is not initially influenced by semantic variables, while subsequent semantic processing depends on the syntactic structure and is constrained by syntactic analysis to a large degree (Ferreira & Clifton, 1986; Frazier, 2002). On the other hand, some researchers have argued that lexical, semantic or conceptual information can

provide immediate constraints on sentence processing and is partially independent of syntactic processing (Boland, 1997; Kim & Osterhout, 2005; Kuperberg, 2007; Trueswell, Tanenhaus, & Garnsey, 2002). With respect to interference effects, the size and time course of interference effects engendered by certain types of distractor information could be taken as an indicator of what cues are used and when they are made use of. If syntactic analysis precedes and is totally independent of semantic processes, then I would expect to observe little or no semantic interference when the intervening NP does not satisfy syntactic constraints in the first place. If semantic analysis happens as early as syntactic analysis or has a strong influence on it, then I would expect to observe a semantic interference effect even if the intervening NP is syntactically unavailable. However, as summarized earlier, previous English studies resulted in mixed findings. Some studies did observe a semantic interference effect even when the distractor was syntactically unavailable (Tan et al., 2011; Van Dyke, 2007), whereas others did not (Van Dyke & McElree, 2011). Van Dyke and McElree (2011) suggested that when an item's syntactic feature was strong enough to eliminate it from the distractor set (e.g., plays the role of grammatical object in a core argument), further semantic processing of this item was blocked.

However, I suggest that this might not be the case in Chinese, because it has been shown that syntactic features play a less dominant role in Chinese than in English (Ye et al., 2006; Zhang & Boland, 2010). Chinese has an impoverished morphosyntactic system with minimal morphosyntactic cues for building sentence constituent relations. The absence of such cues for real-time thematic-role specification (i.e., who-did-what-to-whom) encourages Chinese readers to rely on lexical-semantic relations between adjacent

characters, as well as sequential properties (e.g. word order) to guide their sentence processing. What is more, written Chinese has no word boundaries and Chinese words could consist of a single character or multiple characters. Thus, Chinese readers have to rely on adjacent characters' lexical-semantic relations to figure out word and phrasal groupings for further processing (Yang, Perfetti, & Liu, 2010). Furthermore, some researchers have even argued that compared to structure-dependent sentence processing in English and many other languages, Chinese sentence parsing is a semantically-based, contextually-driven process with a reduced role for syntactic processes (Chu, 1998; Li & Thompson, 1989; Yang et al., 2010; Ye, Luo, Friederici, & Zhou, 2006; Zhang, Yu, & Boland, 2010). In a number of ERP studies in English and other European languages, an N400 effect has failed to be observed in a double-violation condition with combined semantic and syntactic violations (e.g. "*The door lock was in the ****eaten***"). This has been taken as strong evidence supporting the syntax-first parsing approach since the syntactic violation appears to prevent processing of the semantic violation (Friederici, Gunter, Hahne, & Mauth, 2004; Friederici, Steinhauer, & Frisch, 1999; Hahne & Friederici, 2002). However, in studies using Chinese materials, a widely distributed N400 effect indicating semantic processing difficulty was still observed in the double violation condition (Ye, Luo, Friederici, & Zhou, 2006; Yu & Zhang, 2008; Zhang, Yu, & Boland, 2010). These studies have led to the speculation that semantic processing could proceed even when syntactic licensing fails in Chinese. Therefore, I suggest that it is worthwhile to test how semantic and syntactic retrieval cues are combined in a language with different grammatical properties from English, such as Chinese. I expect that semantic

interference has a greater influence in Chinese sentence processing, such as occurring even when the distractor causing semantic interference is in a core argument.

1.4.2 Examining interference resolution using the ERP techniques

As summarized above, most of the previous studies focusing on interference effects during sentence comprehension have mainly relied on behavioral measures, such as recording reading times (RTs) during self-paced reading or during eye-tracking, while a more complete examination using techniques with higher temporal and spatial resolution is needed. It has been shown that the self-paced reading paradigm is sensitive to word-level effects (e.g., RTs vary with the length and frequency of a word) as is the eye-tracking technique, but it may not be sensitive enough for measuring higher-level on-line sentence comprehension (Just, Carpenter, & Woolley, 1982). Moreover, I am interested in processes happening within different stages of interference resolution. Based on Friederici's sentence processing model (1998, 2000, see more discussion later), I predicted that there might be at least two stages of interference resolution: an early stage of interference detection/diagnosis and a later stage of actual interference resolution. None of the previous behavioral or neuroimaging studies were able to dissociate these processes, while the high temporal resolution (on the order of milliseconds) and well-established linguistic effects (as summarized in section 1.1.1.4) in ERPs would allow inferences about the nature of underlying processes during on-line sentence processing in a moment-by-moment fashion as each word is processed.

To summarize, the central goal of current study was to use the ERP technique to further test the cue-based parsing approach, examining the interplay between semantic and syntactic processes, and investigating the nature of the WM mechanisms underlying

language processing, through the examination of semantic and syntactic interference effects and their relations to WM capacity during the reading of Mandarin Chinese sentences. I aimed to answer the following research questions:

(1) If both semantic and syntactic interference impede Chinese sentence comprehension as reported in previous English studies, are the syntactic and semantic retrieval cues combined in the same way? Will semantic interference be observed even for arguments which are syntactically unavailable? How would the time course of the usage of different types of retrieval cues differ from English?

(2) What's the nature of the WM mechanism underlying interference resolution during online Chinese sentence comprehension? If there are specific links between certain WM components and semantic/syntactic interference resolution process as we observed in previous English studies (Tan et al., 2011, 2013), would the links be the same in Chinese? What theoretical accounts could best explain the specificity of the WM-language links we observed?

1.4.3 Predictions

In the current study, I manipulated the degree of both semantic and syntactic interference effects in Chinese sentences. According to the cue-based parsing approach, first, I expected to observe both semantic and syntactic interference during sentence comprehension. Regarding the behavioral measures, subjects were expected to show higher error rates to comprehension questions in the high syntactic or high semantic interference condition relative to the low interference condition. Based on results from

our previous behavioral study (Tan et al., 2011; Van Dyke, 2007), no interaction between the semantic and syntactic interference effects was expected.²³

Regarding the ERP measures, I expected to observe ERP effects associated with both syntactic and semantic interference effects right at the critical verb, where readers were trying to complete the subject-verb dependency. Based on the results from previous ERP studies, I predicted to observe an AN effect that associated with higher WM computation/integration demands in both the high semantic and the high syntactic interference conditions. For the semantic interference manipulation, it is possible that the high semantic interference condition would also elicit an N400 effect, which is associated with semantic integration difficulty, and a late negativity, which is associated with second-pass semantic revision. For the syntactic interference manipulation, the high syntactic interference condition might also elicit a P600 effect following the early AN effect, indicating syntactic integration difficulty.

With respect to the WM mechanisms underlying interference resolutions, different accounts (as summarized in section 1.2) make different predictions (see summary in Table 1). Therefore, I have included a set of ten individual differences measures in order to test the predictions from each account. Based on results from my previous behavioral and neuropsychological studies (Tan et al., 2011, 2013), I expected

²³ It should be noted that a potential problem of the current design is that the foil answer (e.g. “*citizen*”) in the low semantic interference condition will be a semantically related word which never show up in the sentence, whereas the foil answer in the high semantic interference condition is the intervening NP (e.g. “*neighbor*”) from the relative clause. Even if a semantically related word was used to increase difficulty in the low semantic interference condition, it is possible that the question after low semantic interference condition is still easier because the foil NP never appeared in the sentence. In our previous study (Tan et al., 2011), this problem was solved by only presenting the comprehension question and asked subjects to provide a spoken response. However, it is important to control head movement in the ERP studies and thus, a spoken response was avoided. In addition, there might be a ceiling effect in the accuracy data. As a result, I did not put much emphasis on the comprehension question performance, but instead mainly focused on the on-line data.

to observe that the magnitude of the semantic and syntactic interference effects (as quantified by the mean amplitude of corresponding ERP effects) to be modulated by different WM capacities - the ERP effects associated with semantic interference effect should be modulated by individual's semantic STM capacity (as measured by a category probe task), while the ERP effect associated with syntactic interference effect should be modulated by individual's general WM capacity or executive control ability. The predictive power should remain significant even after controlling for individuals' verbal knowledge and processing speed. However, due to the many differences between the current experiment and previous studies, such as the differences in languages investigated and techniques employed, and the potential variation of WM-language relation in different stages of interference resolution, which was investigable in the current ERP study but not previous self-paced reading experiment, the actual relation might differ.

Table 1 Predictions from different accounts about the relations between individual differences measures and interference resolution processes during sentence processing.

Accounts		Predictions
Capacity-based	Multiple capacities	The semantic STM measure (e.g., category probe task) should predict semantic interference resolution only. The phonological STM measure (e.g., digit span task) should be uncorrelated with either types of interference resolution.
	Domain general	The general WM capacity (i.e., composite scores computed from multiple span tasks) should be correlated with the size of both semantic and syntactic interference effects.
	Domain specific	There should be no correlation between any WM span

		tasks and online interference effects.
Non-capacity based	Experience-based	There should be no correlation between interference effects and any WM measures. The linguistic experience measure (e.g., author recognition, vocabulary tasks) should be correlated with the size of both types of interference effect.
	Executive control-based	There should be no correlation between interference effects with the capacity measures (e.g., semantic STM, phonological STM). The executive control measures (e.g., recent negatives task, verbal Stroop task) should be correlated with the size of both types of interference effect.

2 Experiment

2.1 Subjects

40 undergraduate students were recruited from Peking University. All the subjects were native Chinese speakers without a diagnosed reading or learning disability and had normal or corrected-to-normal vision. Subjects received payment for their participation.

2.2 Materials, design, and procedure

2.2.1 Sentence materials

Examples of the experimental sentences are shown in Example 13. There were 80 sets of sentences with four different types of sentences in each set. For each set of sentences, they all began with same introduction region and end with same verb phrase. Syntactic and semantic interference were manipulated in a 2×2 design within the relative clause. When subjects process the main verb “*complain*”, they need to discriminate its real subject (“*resident*”) against the proactive interference from the intervening distractor “*neighbor/wall*”. In the high syntactic interference condition, the distractors were the subject of the relative clause (e.g., sentence 13b), while they were direct objects of the relative clause verb (e.g., Example 13a) in the low syntactic interference conditions. For semantic manipulations, the distractors were semantically impossible in the low interference conditions (e.g., a “*wall*” cannot complain), whereas semantically possible in the high interference conditions (e.g., “*neighbor*” could complain). Repetition of the same nouns and verbs was avoided as much as possible.

Example 13. Experimental sentences (Semantic manipulation in brackets)

a. Low syntactic interference (SRC sentence)

- i. 记者听到那个昨晚撞到危险的围墙的住户天天抱怨房租问题。
- ii. Jizhe tingdao na-ge [_{RC}**GAP** zuowan tuidao weixian-de weiqiang(linju) de]
 zhuhu tiantian baoyuan fangzu wenti.
- iii. Reporter heard DET-CL **GAP** last night hit dangerous wall(neighbor) REL]
 resident everyday complained rent problem.
- iv. The reporter heard that the resident who hit the dangerous wall(neighbor) yesterday complained about the investigation every day.

b. High syntactic interference (ORC sentence)

- i. 记者听到那个昨晚危险的围墙砸伤的住户天天抱怨房租问题。
- ii. Jizhe tingdao na-ge [zuowan weixian-de weiqiang(linju) zashang **GAP** de]
 zhuhu tiantian baoyuan fangzu wenti.
- iii. Reporter heard DET-CL last night dangerous wall(neighbor) hurt **GAP** REL]
 resident everyday complained rent problem.
- iv. The reporter heard that the resident who the dangerous wall(neighbor) hurt yesterday complained about the investigation every day.

Because of the potential problem of local syntactic ambiguities within the Chinese relative clause region (Gibson & Wu, 2013; Lin & Bever, 2006; see Jäger et al., 2015 for a review), following the solution proposed by Jäger et al. (2013), a determiner/classifier (DET-CL) “Na-ge (*the*)”, and an adverbial phrase (ADVP) “tiantian (e.g., *yesterday*)”, which reliably bias a relative clause interpretation were added before a relative clause. In addition, another adverbial phrase (e.g. “*everyday*”) was inserted between the head noun (e.g., “*resident*”) and main verb (e.g., “*complain*”) to separate these two words and thus prevent spillover effects from processing the relative clause. Based on the results from a sentence completion test and an eye-tracking experiment conducted by Jäger et al. (2013; 2015), the local ambiguities in the Chinese RC sentences were eliminated in the experimental materials through providing these syntactic cues, and readers strongly predicted a relative clause after the adverbial phrase (Jäger, Chen, Li, Lin, & Vasishth, 2015). Additionally, two words (i.e., two noun phrases) were added after the critical verb to allow for a continuous processing (or spillover effects) and avoid confounding the interference effects on the critical verb with wrap-up effects at the sentence final position (Just, Carpenter, & Woolley, 1982; King & Just, 1991, Hagoort, 2003). Based on the results from our previous behavioral study (Tan et al., 2011), there might be a spillover effect from processing the main verb on the first word after it. In addition, a wrap-up effect on the last word was predicted by a number of ERP studies, which have shown that processing difficulty somewhere in the sentence also elicits a more global effect (e.g., an increase in the N400 amplitude which is independent of the nature of the violation) at the end of the sentence (Hagoort, 2003; Osterhout & Nicol, 1999).

Stimulus rating. All experimental sentences were matched in length (length = 12 words; mean number of character in each sentence = 26.4 characters; S.D. = 1.3) and the position of the main verb (i.e., the 10th word). After two rounds ($N_1 = 20$; $N_2 = 26$) of online meaningfulness rating from native Chinese speakers on 90 sets of sentences, 80 sets of sentence were revised and chosen as the final materials. To make sure the target sentences were well-formed and matched in meaningfulness, a separate group of 28 native Chinese speakers (mean age = 26.5 years; S.D. = 6.4 years) were recruited to rate the meaningfulness of each sentence on a 7-point scale (1= “Not meaningful at all”; 7 = “Very meaningful”). Two rating lists were constructed with each list only containing one of the semantic manipulations of each syntactic structure within one set. The summary of the meaningfulness rating of the materials is presented in Table 2. Generally, all the sentences were meaningful (mean meaningfulness rating: 4.7, S.D. = 1.2). A two (syntactic interference) $\times$ two (semantic interference) repeated ANOVA were conducted on the meaningfulness rating. As intended, there was no meaningfulness difference between the high and low semantic interference conditions in the by-item analysis (mean meaningfulness rating: 4.6 vs. 4.7), $F_2(1, 79) = 1.88, p = .17, \text{MSE} = .75$, though in the by-subjects analysis, the high semantic interference sentences were rated higher than the low semantic interference sentences, $F_1(1, 27) = 6.36, p = .02, \text{MSE} = .24$. However, subjects rated the sentences containing a SRC structure more meaningful than the sentence containing an ORC structure (mean meaningfulness rating: 5.2 vs. 4.1), $F_1(1, 27) = 34.99, p < .001, \text{MSE} = 30.79$, $F_2(1, 79) = 355.54, p < .001, \text{MSE} = 82.97$. The meaningfulness difference between the SRC and the ORC sentences was unavoidable to some extent, due to the large structural frequency difference in Chinese (82% vs. 18% as

reported by Jäger et al., 2015). However, as all the sentences were grammatical and meaningful (including the fillers) and subjects showed high accuracy in answering comprehension question (as discussed in section 3.1), the general meaningfulness rating differences should not affect the interference effect in online processing that we are most interested in.

Table 2 Meaningfulness rating of the experimental sentences

	Low Syntactic	High Syntactic
Low Semantic	5.24	4.16
High Semantic	5.11	4.02

2.3 Procedures

The experiment was completed in two sessions. In the first session, all the participants were tested on a sentence comprehension task while their brain activity was recorded through electrodes placed on their scalp. The sentence comprehension task lasted about 1.5h whereas the entire the session lasted about 3h in total, including setup time. In the second session, subjects were tested on a set of individual differences measures, including reading span, operation span, digit span, category probe, Stroop, author recognition test (ART), vocabulary test from WAIS-III, and rapid automatized naming (RAN) tasks. The second session lasted about 3h in total.

2.3.1 Sentence comprehension task

In the sentence comprehension task, subjects were seated in an comfortable armchair facing a computer screen at a distance of about 1.5 m. Sentences were presented in a word-by-word, noncumulative fashion at the center of the screen. In addition to the

experimental sentences, 140 filler items (mean number of character in each sentence = 24.4 characters; S.D. = 1.3) were constructed to distract subjects' attention away from the actual target sentences and avoid developing of specific processing strategy: 1) 80 sentences containing a SRC structure with an additional embedded relative clause as Tan et al. (2011) used in previous English studies as the high syntactic interference condition (e.g., *The reporter heard that yesterday said the wall was very dangerous de resident everyday complain the investigation* – English: “*The reporter heard that the resident who said that the wall was dangerous yesterday complained about the investigation every day.*”); 2) 60 sentences in which the first noun phrase (NP₁) was the subject of the matrix verb (e.g., *The manager heard that yesterday waiter beat de customer already left very angry* - English: “*The manager was very angry when he heard that the customer who was beaten by the waiter yesterday has already left.*”)²⁴. Half of the sentences contained a SRC structure while the other half contained an ORC structure; 3) 20 sentences with a time prepositional phrase (e.g., “*President morning criticize the teacher stop student while did not ask about details*” – English: “*The president criticized the teachers who did not ask about details when stopping the student this morning.*”).

To avoid repetition of the sentence content within one participant, each subject only saw one semantic condition from each type of sentence. Eight sentence lists were created and each subject only received one list. Thus, each subject was visually presented with 160 experimental sentences and 140 fillers for a total of 300 sentences. After the presentation of each sentence, there was a comprehension question asking subjects to

²⁴ In the experimental sentences, the grammatical subject to the matrix verb was always the second or the third noun phrase (NP₃) in the sentences. Thus, it was necessary to add the filler sentences in which the first noun phrase was the grammatical subject of the matrix verb, although the word order in the relative clause region was the same between this kind of fillers and the experimental sentences.

choose the correct answer from two alternatives (e.g., “*Who complain? Resident/Neighbor*”) to encourage them to integrate incoming material into a consistent interpretation. Subjects were instructed to indicate the correct answer (yes or no) by pressing a button as quickly and accurately as possible. The position of the correct answer on the screen (left vs. right) was counterbalanced. In the high semantic interference condition, half of the foils answer were the intervening NP in each set (e.g., “*neighbor*”) and the other half foils were the NP₁ (e.g., “*reporter*”). However, in the low semantic interference condition, since the intervening NP in the RC cannot serve as subject to the main verb (e.g., the “*wall*” cannot “*complain*”), half of the foils answer were a semantically related NP to the subject (e.g., “*citizen*”) while the other half was the NP₁.

Before the experiment, there were 12 practice sentences to help participants get familiar with the experiment. Participants were instructed to read each sentence for comprehension and told that there they have to answer a comprehension question following each sentence. All trials began with a fixation point appearing in the center of the screen beginning for 500 ms, and then followed by a 400 ms blank screen. Each word was presented for 400 ms with a 400 ms inter-stimulus interval (ISI). The fixed-rate serial visual presentation (SVP) we adopted here, in which readers were required to fixate on the center of the screen while sentences are presented word by word at a fixed rate, has both advantages and disadvantages (Ditman, Holcomb, & Kuperberg, 2007). A disadvantage is that it is not a natural way to read, and the presentation time is slower than the natural reading speed (200 – 250 ms per word in Chinese, Bai, Yan, Liversedge, Zang, & Rayner, 2008). The slow presentation is used in order to allow the investigation

of later ERP effects (e.g. N400, P600) with effects from one word uncontaminated by the previous word. An important advantage of a fixed SVP is that the researchers can time-lock neural activity to a particular word. Moreover, in Chinese, the basic reading time is about 200 ms per character (Chan & Lee, 2005; L. Tan, Spinks, Eden, Perfetti, & Siok, 2005). In the current experiment, all the critical words (e.g., head noun, main verb, adverbial phrase) consisted of two or more characters. Thus, the presentation rate of 400 ms per word is very close to the natural reading speed, although the 400 ms ISI adds time for additional processes. However, previous studies have shown that as long as the presentation time is not extremely fast or slow, it does not disrupt normal reading processes (Kutas & Van Petten, 1994; Van Berkum, Brown, & Hagoort, 1999). The presentation time of about 400 ms per word (plus about 500 ms ISI) has been demonstrated to be comfortable for Chinese reading (Jiang & Zhou, 2009, 2013; Ye et al., 2007). The comprehension question following each sentence remained on the screen until either the subject responds or 5s elapses. Then the next sentence started after an inter-trial interval of 1000ms.

2.3.2 Cognitive capacity tests

2.3.2.1 Memory tests

Reading span task

A automated version of reading span in Chinese (Cai, Dong, Zhao, & Lin, 2015), which was modified from the English version developed by Unsworth and colleagues (Unsworth, Heitz, Schrock, & Engle, 2005), has been used in the current experiment.

Like the English version, subjects were instructed to judge whether the presented sentence makes sense or not (e.g., “*Andy was stopped by the policeman because he crossed the yellow heaven*”). After each sentence, a letter to be recalled was shown on the screen for 800 msec. At the end of each set of sentences, subjects had to recall all the letters showed in current set in order. At the beginning of the experiment, there were some practice trials helping subjects become familiar with the task. The set size ranged from 3 to 7 items and there were three trials within each set. Thus, there were a total of 75 letters and 75 sentence judgments. The order of set sizes was randomized for each subject. Subjects’ performance was evaluated as the total number of correctly recalled letter. Test duration is about 15 – 20 min.

Operation span task

The automated version of Operation Span (Aospan) developed by Unsworth and colleagues was also included to measure subjects’ WM capacity (Unsworth et al., 2005). Similar to the automated version of reading span task, subjects were required to solve a math operation. Then a digit was presented in the next screen for judging whether it is the correct answer to the math operation or not. After subjects choosing either the “true” or “false” box by comparing this digit to their answer, a letter to be recalled was shown for 800 ms. By the end of each set, subjects have to recall all the letters in the correct order. The experimental trials contain three sets at each set size, with set sizes ranging from 3 to 7 items. This result in a total of 75 sets with 75 letters and 75 math problems. The order of set sizes was random for each participant. Subjects’ performance is evaluated by calculating the total number of correctly recalled letter sets. Test duration is about 15 – 20 min.

Category probe and digit span task

Based on the category probe task in English, a Chinese version was constructed to measure semantic STM. In this task, subjects were presented with an auditory Chinese word list. After a short pause, they heard a probe word and had to judge whether this word is in the same category as any of the words in the list (all the words in one list were drawn from different categories). Before testing, subjects were presented a list of all nine categories (e.g., animals, clothing, fruits, weather, trees, insects, kitchen equipment, and flowers.) that were presented in the experiment as well as all the words belonging to each category. The number of words in each list ranged from 4 to 7 and there were 24 lists at each list length. Subjects' semantic STM capacity was evaluated as the overall accuracy across the list length. Test duration is about 20 min.

The digit span task from the WAIS-III was included to tap subjects' phonological STM. In this task, participants heard a list of digits in Chinese and have to repeat the list back in order. The number of digits in each list ranges from 4 to 11, and there were 2 trials at each level. I expended the list length to 11 digits compare to up to 9 digits in the English WAIS-III, because it has been demonstrated that Chinese speaker obtained a larger digit span than speakers in many other languages due to the shorter articulatory duration of digit names, as digits' names in Chinese have fewer phonemes (Chincotta & Underwood, 1997). As a result, each subject received 16 trials. The overall accuracy for list recall was calculated for each subject. Test duration is about 7 – 10 min.

2.3.2.2 Executive function tests

Classic Verbal Stroop

The classic verbal Stroop task (Stroop, 1935) was adopted in the current experiment to measure subjects' resistance to automatic or prepotent response (Friedman & Miyake, 2004). Subjects were required to name the ink color in all conditions. In the congruent condition, a color word was presented in the congruent color (e.g., the word “blue” in blue ink); while in the incongruent condition, a color word was presented in a different ink color (e.g., the word “blue” in red ink). In the neutral condition, a series of colored asterisks was presented. There were 68 incongruent trials, 72 neutral trials, and 14 congruent trials. Subjects were instructed to press left button when the ink color was red or yellow, and press right button when the ink color was blue or green. Since each button corresponded to two colors, among the 68 incongruent trials, there were 16 trials in which the correct button press was consistent with the word meaning (e.g., the word “red” in yellow ink, the word “blue” in green ink). These trials were excluded from the data analysis, resulted in 52 incongruent trials in the data analysis. Response latencies were recorded from the onset of the stimulus through a button press. In order to take both accuracy and RT into consideration as suggested by Hughes et al. (Hughes, Linck, Bowles, Koeth, & Bunting, 2013), Stroop interference score for each subject was calculated as the inverse score, through dividing the average RT differences (i.e., subtracting the average RT in the neutral condition from the incongruent condition) by the average accuracy differences (i.e., subtracting the average accuracy in the neutral condition from the incongruent condition). Test duration is about 10 min.

Recent negatives task

There is evidence that resistance to proactive interference in memory (e.g., as measured by the reading span task recall, recent negatives, etc.) involves a different type of interference resolution than resistance to a predominant response (e.g., as measured by the Stroop task, antisaccade task, etc.) (Friedman & Miyake, 2004). Given that the current project was interested in proactive interference resolution during sentence comprehension, recent negatives task (Monsell, 1978) was included to measure subjects' resistance to PI. During the experiment, subjects heard a list of Chinese two-character words and had to judge whether the probe word following each list was in the immediately preceding list. In each list, there were three concrete words of comparable frequency (according to Modern Chinese Frequency Dictionary, Beijing Language College, 1986). Ninety-six lists will be constructed. There were 3 types of trials: recent negative, non-recent negative, and positive. Half of the trials were positive trials in which the probe word was presented in the immediately preceding list. The other half of the trials were negative trials in which the probe did not appear in the current list. Furthermore, half of the negative trials were recent negative trials, in which the probe word was present in the immediate previous list; the other half of the negative trials are non-recent negative trials, trials in which the probe word did not appear in the previous two lists (though the probe word may appear in trials prior to the previous trial). Subjects were instructed to press buttons as quickly as possible to make a "yes" or "no" response to the probe word. They received 10 practice trials to familiarize them with the test. Both error rates and reaction times were collected. Interference score for each subject was calculated for each subject as the inverse score, through dividing the average RT differences (i.e., subtracting the average RT in the non-recent negative condition from the recent negative

condition) by the average accuracy differences of these two conditions. Test duration is about 20 – 25 min.

Rapid automatized naming (RAN)

Several previous studies using an individual differences approach have shown that efficiency of speeded naming predicts subjects' language processing efficiency, as subjects with faster naming ability showed more efficient eye-movement in early sentence processing (Gordon et al., 2013; Kuperman & Van Dyke, 2011). Thus, I included rapid automatized naming tests for digits and letters. In this test, two trials in each version were presented to each subject. On each trial, 36 digits or letters were shown in random order in a 4×9 grid. Subjects were instructed to name the Arabic numbers and letters of the alphabet, which consisted of high-frequency lowercase letters (e.g., a, b, d, o, p, s),²⁵ as fast and accurately as they can. Total response time from the pronunciation of first name to the last one was recorded for each trial. Subjects' speeded naming ability was calculated as the average naming speed across the four trials. Test duration for each version is about 5 min.

Lexical decision

Previous studies have shown that subjects with faster word decoding or lexical access speed are generally more efficient in sentence reading (Kuperman & Van Dyke, 2011). To measure individuals' word decoding skill, Kuperman and Van Dyke (2011)

²⁵ The digits and letters selection is following Leong et al.'s study (2008) with Chinese children. Subjects should be able to name the alphabet letters automatically due to overlearning since childhood, and any WM load should be reduced. They have explained that "the alternative of substituting the familiar letters of the alphabet with Chinese characters might approximate the "alpha" aspect but run the risk of changing the structure of the tasks and of introducing the element of unfamiliarity and the need to draw on long-term memory" (Leong, Tse, Loh, & Hau, 2008).

tested subjects on a word identification test, in which subjects were instructed to read aloud sets of words as fast and accurately as possible. Subjects were expected to be able to simply pronounce the words according to letter-sound correspondences without the necessity of knowing the meaning of them. Thus, subjects with high scores in word identification tests were assumed to have mastered the grapheme-to-phoneme correspondences rules better, and could more quickly activate unitary lexical representations for most words. However, in Chinese, there are no simple grapheme-to-phoneme correspondence rules because no unit in the writing system encodes single phonemes (L. Tan & Perfetti, 1998). (See discussion in footnote 26). Due to cross-linguistic differences, I included a written lexical decision task instead of a word identification test to measure subject's semantic processing speed developed by Janssen and colleagues (Janssen, Bi, & Caramazza, 2008). In the lexical decision task, subjects were presented a list of Chinese words and asked to decide as quickly as possible whether each word comprises a real word or not. All of the words consisted of two-characters and no contextual support was provided. The pseudo-words were formed by combining two meaningful Chinese characters together to create a meaningless compound word. Subjects were explicitly instructed that real words are those used in the daily language with fixed meanings while non-words are those not used in daily language and have no fixed or commonly accepted meanings. In addition, in order to control for a phonological priming effect, neither of the two characters within a compound word rhymed with a third character which could form a meaningful word with the other character. For example, although the pseudo-word “掩该 yan[3] gai[1]”²⁶ is meaningless, by

²⁶ In Mandarin Chinese, there are four phonemic tone classes of Chinese words. The tones are marked by 1, 2, 3, or 4 in the brackets.

replacing the second character “该 gai[1]” with a rhyming character “盖 gai[4]” the word “掩盖 yan[3] gai[4]” means “to conceal”. Subjects received 10 practice trials followed by 40 real words and 30 pseudo-words presented in a pseudo-random order. They were instructed to respond by pressing buttons corresponding to “yes” and “no”. Reaction times were recorded for each subject as latencies between onset of the word and their button press. Subjects’ lexical decision performance was calculated as the inverse efficiency score, mean RT and accuracy. Interference score for each subject was calculated for each subject as the inverse score, through dividing the average RT differences (i.e., subtracting the average RT in the non-recent negative condition from the recent negative condition) by the average accuracy differences of these two conditions. Test duration is about 10 min.

2.3.2.3 Reading experience tests

Vocabulary test

Subjects’ reading experience was measured through the vocabulary test from WAIS-RC (Gong, 1992). In the vocabulary test, subjects were required to provide definition to each word. There are 40 words in total in the Chinese version WAIS-RC and I began the test from the 21st item (*invert*) to the 40th item (*plagiarize*) since the words before the 20th were not discriminating enough for undergraduate/graduate students. Subjects received 0, 1, or 2 points on each item based on their answer, as defined by standard scoring criteria. Total scores were calculated. Test duration is about 15 min.

Author recognition test (ART)

A Chinese version of the author recognition test (ART) was constructed following the rules of creating the English ART (Stanovich & West, 1989). In the ART, subjects were instructed to circle the names that they knew to be authors on a list, which consists of 60 authors and 60 name foils. Previous studies have shown that the ART test provides a reliable way to measure individuals' exposure to print, and subjects with higher ART scores could better solve the semantic interference during sentence comprehension (Van Dyke, Johns, & Kukona, 2014). To create the name lists, 30 subjects were recruited to rate 180 Chinese names (or the Chinese translation of some English names, including western authors, e.g., Mark Twain) on a 5-point scale (1 – “Very sure this is NOT an author”; 5 – “Very sure this is an author”). Based on the rating results, 30 famous authors (mean rating = 4.57, S.D. = 0.24), 30 less famous authors (mean rating = 3.54, S.D. = 0.22), 35 appealing foils (mean rating = 2.95, S.D. = 0.26), and 25 less appealing foils (mean rating = 1.90, S.D. = 0.67) were included in the final list. Subjects' performance on ART was calculated as their ability (d') to correctly circle the real authors (z -hit) minus the false alarm rate (z -false alarm). Test duration is about 10 min.

2.4 EEG recording and data analysis

The electroencephalogram was recorded from 61 electrodes mounted in an elastic cap selected from the extended 10-20 system (Sharbrough et al., 1990). The signal on each electrode was referenced to the nose tip during online recording. The vertical eye movements (i.e., vertical electrooculogram, VEOG) were recorded from an electrode placed above the right eye. Horizontal eye movements (i.e., horizontal electrooculogram, HEOG) were recorded from electrodes placed at the outer canthus of right eye. The

electrodes were positioned over the midline (i.e., FPZ, FZ, FCZ, CZ, CPZ, PZ, POZ and OZ), over the left hemisphere (i.e., AF7, AF3, FP1, F7, F5, F3, F1, FT7, FC5, FC3, FC1, T7, C5, C3, C1, TP7, CP5, CP3, CP1, P7, P5, P3, P1, PO7, PO5, PO3, and O1), and over the corresponding locations in the right hemisphere (i.e., AF8, AF4, FP2, F8, F6, F4, F2, FT8, FC6, FC4, FC2, T8, C6, C4, C2, TP8, CP6, CP4, CP2, P8, P6, P4, P2, PO8, PO6, PO2, and O2). The EEG biosignals were amplified with BrainAmps DC amplifiers (BrainProducts, Munchen), filtered with a 0.016-250 Hz bandpass, and sampled at a rate of 500 Hz. Electrode impedance was kept below 5 k Ω during the experiment.

EEG data were analyzed using EEGLAB (Delorme & Makeig, 2004) following general ERP data processing procedures. Only correctly answered sentences were included in the analysis. During offline data analysis, the EEG data were re-referenced to the grand average of all electrodes. A 0.1 Hz high-band pass filter was applied to the data to remove slow drift. Then blinks and eye movements were removed from the data following a procedure based on Independent Component Analysis (ICA) (Jung et al., 2000). After running ICA, artifact detection was further conducted using ERPLAB to remove trials containing artifacts (Lopez-Calderon & Luck, 2014). Any trials with mean voltage exceeding ± 70 μ V or had step-like eye-movement were rejected before application of the EEG averaging procedure. Three participants were excluded due to excessive artifacts (lost more than 20% trials on the matrix verb).

I focused on the ERP responses at the critical verb (e.g., “*complain*” in Example 13), where subjects should start retrieval to complete the subject-verb dependency. Additionally, I have examined four more sentence regions of interest. Two more regions after the critical verb were examined: 1) the spillover region (i.e., the first word after the

critical verb) and 2) the final word (i.e., the second word after the critical verb). These were examined because interference effects might spillover to post-critical verb regions (A. E. Martin et al., 2012, 2014; Tan et al., 2011), and subjects might show a global processing effect in the sentence final region as a result of earlier processing difficulty (Hagoort, 2003; Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992; Osterhout & Nicol, 1999). Two regions before the critical verb were also examined: 1) the head noun (i.e., the first word following the relativizer marker “*de*”) and 4) the adverbial phrase (i.e., the word between head noun and the critical verb). These were examined because many studies have reported processing differences between SRC and ORC sentences at the head noun (Packard et al., 2011; Yang et al., 2010), which is relevant to the debate about the nature of the WM mechanism underlying sentence processing. I also examined ERP effects on the adverbial phrase, which was inserted between head noun and critical verb, in order to make sure that the processing effect on the head noun would not affect the ERP effects on the critical verb.²⁷

After artifact detection, ERPs were segmented for each participant over an epoch from 200 ms before word onset to 800 ms after the onset of the word in each ROI. For all ERP averages, a baseline of 200 ms prior to word onset was used. The mean number of trials free of artifacts was 36.8 in the main verb region (S.D. = 0.31), 36.8 in the spillover region (S.D. = 0.42), and 36.7 in the final region (S.D. = 0.42). For illustration purpose,

²⁷ Some of the previous studies have reported a RC type effect on the relative marker “*de*” in Chinese RC processing. However, due to the limited number of artifact-free trials within the relative clause and on the following “*de*” (average number of trials ≤ 20), I did not examine ERP effects during the relative clause region or at the head nounrelativizer marker “*de*”, although which could be potentially interesting. Because in order to keep as many trials, which were free of eye movements, as possible in the regions of most interests (i.e., main verb and the following region), we instructed the subjects to blink during the first half of the sentences and at the relativizer *de*. As a result, although ICA helped correct subjects’ eye blink, we decided not to run statistic analysis on the “*de*” since very few trials were remained after artificial-rejection without ICA correction or head noun right following it.

ERP waves were smoothed using a low-pass filter with a cutoff frequency of 7 Hz in the figures.

To better find the onset and offset time point of each ERP component, repeated ANOVAs were conducted on the mean amplitude in consecutive 50-ms latency window for each test shifted 50 ms rightward between 0 – 800 ms (i.e., 0 – 50 ms, 50 – 100 ms, etc.). Then based on visual inspection of the grand average waveforms and the consecutive 50-ms latency window analyses, critical time windows were identified for each sentence region of interests. Unless specified, all of the significant ERP effects reported in the results section satisfied the selection criteria that more than two consecutive 50-ms analyses reached significance within the large time window. For statistical analyses, four lateral regions of interests (ROIs) including six electrodes each were identified: left anterior region (i.e., F3, F5, F7, FC3, FC5, FT7), right anterior region (i.e., F4, F6, F8, FC4, FC6, FT8), left posterior region (i.e., CP3, CP5, TP7, P3, P5, P7) and right posterior region (i.e., CP4, CP6, TP8, P4, P6, and P8). Repeated ANOVAs were conducted on the mean amplitude within each time window with the within-subject factor of Semantic interference (high vs. low) $\times$ Syntactic interference (high vs. low) $\times$ Region (anterior vs. posterior) $\times$ Hemisphere (left vs. right)²⁸. Effects along the midline were analyzed on five electrodes (i.e., FZ, FCZ, CZ, CPZ, PZ). Similar to the lateral analysis, repeated ANOVAs were conducted on the mean amplitude within each time window with the factor of Semantic interference (high vs. low) $\times$ Syntactic interference (high vs. low) $\times$ Electrodes (5 levels). Planned comparisons were conducted

²⁸ It should be noted that by-item analyses are usually not performed on ERP data partially because stable ERPs per condition require large number of subjects to reduces the probability that the results affected just by a few odd outliers (Van Berkum, Hagoort, 1999).

separately on the experimental manipulations within each level of variables when there were significant interactions between experimental manipulations (i.e., interference types) and any of the other two topographic variables (i.e., hemisphere and regions). Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) was applied when there were more than one degree of freedom (the original degrees of freedom and corrected *p*-value was reported in the tables).

To access sources of variability that might help understand the WM mechanisms underlying sentence comprehension, the relationships between the different cognitive abilities data obtained from each subject (including reading span, operation span, category probe, digit span, rapid automatized naming, verbal Stroop, recent negatives, lexical decision, author recognition, and WAIS vocabulary tests) and statistically significant ERP effects were examined by using both simple correlation and multivariate linear regression tests. For the simple correlation test, difference scores were calculated for each subject as the index of their interference resolution efficiency through subtracting the mean amplitude of each significant ERP effect in the high semantic/syntactic condition from that in the low semantic/syntactic interference condition, respectively. Then the correlation between the differences scores and each individual difference measure was calculated. For the linear regression test, composite variables of subjects' cognitive ability were calculated if two or more variables highly correlated with each other. Then the ERP responses in the high semantic/syntactic interference conditions (i.e., mean amplitude) were regressed against the mean amplitude in the low semantic/syntactic interference and all the individual differences measures (or composite scores) simultaneously.

3 Results

3.1 Comprehension question accuracy

Accuracy for comprehension questions was calculated for each condition (see Table 3). Participants answered the comprehension question correctly 94% (SD = 5%) of the time. This is consistent with previous findings that subjects are generally very accurate in understanding complex Chinese sentences with relative clauses (Jäger, Chen, Li, Lin, & Vasishth, 2015). A two-way ANOVA crossing syntactic interference (high vs. low) and semantic interference (high vs. low) was conducted on the accuracy. The results revealed a main effect of the semantic interference manipulation, $F(1, 36) = 6.43$, $p = .016$, $MSE = .008$, indicating that subjects tended to make more errors to comprehension question in the high relative to the low semantic interference conditions (HiSem; 6% vs. LoSem: 4%). However, neither the main effect of syntactic interference manipulation (HiSyn; 6% vs. LoSyn: 5%) nor the interaction of semantic $\times$ syntactic interference was significant ($F_s < 1$). These results are consistent with previous English studies showing that the semantic interference effect size is numerically larger than syntactic interference effect size during comprehension question answering (Tan et al., 2011; Van Dyke, 2007). In addition, the fact that this study examined proactive interference, which is much less detrimental than retroactive interference (Van Dyke & McElree, 2011), made it less likely to observe big interference effect. However, the lack of a strong syntactic interference in the error rates should not affect the interpretation of the online effects,

given that subjects might have already resolved the syntactic interference by the end of sentence reading.

Table 3 Descriptive data of mean accuracy (proportion correct) in sentence comprehension task

	Low semantic	High semantic
Low syntactic	0.95 (.05)	0.94 (.05)
High syntactic	0.96 (.04)	0.93 (.05)

Note. The numbers in the parentheses were the standard deviation.

3.2 Event-related potentials (ERPs)

I examined electrophysiological results in five sentence regions: the critical verb region (e.g., “*complain*”), the spillover region (e.g., “*rent*”), the sentence final region (e.g., “*problem*”), the head noun (e.g., “*resident*”), and the adverbial phrase (e.g., “*everyday*”). The ERPs elicited at the critical verb region were of most interest. The ERPs on the two following sentence regions were reported as well to assess spillover effects of semantic/syntactic interference and additional global processing at the sentence final position. In addition, the ERPs on the head noun and the following adverbial phrase were examined for the RC processing differences in Chinese. As can be seen from Figs. 1 – 11, in all of the regions, each individual word elicited an N1-P2 complex in the first 250 ms after the onset of each word, and a P1 component precedes the N1-P2 complex at occipital sites. These are typical ERP components for visually presented materials. In the following analyses, I focused on ERPs in the later time windows (i.e., 250 – 800 ms), which have been demonstrated to be sensitive to sentence processing.

3.2.1 Critical verb

In the critical verb region, based on visual inspection, the consecutive 50-ms latency window analyses, and the results from previous studies investigating the long-distance dependency completion on the matrix verb (Fiebach, Schlesewsky, & Friederici, 2002; King & Kutas, 1995; Kluender et al., 1998) or interference resolution during sentence processing (Martin, Nieuwland, & Carreiras, 2012, 2014), two main time windows were identified: 1) 300 – 500 ms time window for N400 or LAN effect; 2) 650 – 800 ms time window for P600 effect.

Fig. 1 and 2 (with selective electrodes) display the grand average waveform of each condition on the critical verb. Statistical analyses (as shown in Table 4) showed that for syntactic interference manipulation, there was a positive shift in the typical time window of P600 component (i.e., 650 – 800 ms) along the midline, $F(1, 36) = 6.68$, $MSE = 19.84$, $p = .014$. High syntactic interference conditions elicited a greater late positivity compared to the low syntactic interference conditions (HiSyn: .44 μV vs. LoSyn: .11 μV). This effect was robust as it was significant in each of the consecutive 50-ms time window between 650 – 800 ms ($ps < .05$). Although Fig. 1 seems to show that the greater positivity for high syntactic interference conditions may be more prominent in the posterior regions and the effect even seems negative on Fz, there was no reliable interaction between syntactic interference effects and electrodes due to substantial variability between subjects.

For the semantic interference manipulation, no main effect was observed in either 300 – 500 ms or 650 – 800 ms time window. However, in the lateral electrodes, there was a significant interaction of semantic interference $\times$ region in the 650 – 800 ms time

window, $F(1, 36) = 4.77$, $MSE = 12.29$, $p = .036$. Planned comparisons in the anterior and posterior region revealed that high semantic interference conditions elicited a late negativity in the anterior region only (HiSem: $-0.13 \mu V$ vs. LoSem: $0.17 \mu V$), $F(1, 36) = 6.68$, $MSE = 1.62$, $p = .014$, but not in the posterior regions (HiSem: $0.10 \mu V$ vs. LoSem: $-0.18 \mu V$), $F(1, 36) = 2.50$, $MSE = 1.46$, $p = .123$. This late anterior negativity was a reliable effect as it was significant in each of the consecutive 50-ms time windows between 650 – 800 ms ($ps < .05$).

Regarding the interaction of semantic and syntactic interference effects, it was not significant in either pre-identified time window. However, based on the 50-ms consecutive time window analyses, there was a short-lived three-way interaction of semantic $\times$ syntactic $\times$ hemisphere between 450 – 550 ms, $F(1, 36) = 9.45$, $MSE = 6.20$, $p = .004$. Further comparisons within each hemisphere showed that the interaction of semantic $\times$ syntactic was significant in both the left hemisphere, $F(1, 36) = 6.51$, $MSE = 1.25$, $p = .015$, and the right hemisphere, $F(1, 36) = 6.58$, $MSE = 1.89$, $p = .015$. Following up on each interaction, in the left hemisphere, under the high semantic interference condition only, the high syntactic interference condition was more negative relative to the low syntactic interference condition (HiSem/HiSyn: $0.40 \mu V$ vs. HiSem/LoSyn: $0.64 \mu V$), $F(1, 36) = 7.44$, $MSE = 1.09$, $p = .01$. However, in the right hemisphere, under the high semantic interference condition only, the high syntactic interference condition was more positive relative to the low syntactic interference condition (HiSem/HiSyn: $0.26 \mu V$ vs. HiSem/LoSyn: $-0.14 \mu V$), $F(1, 36) = 6.43$, $MSE = 2.91$, $p = .016$. I do not have a good explanation about the hemispheric discrepancy of the short-lived ERP effects elicited by the high relative to the low syntactic interference

conditions in the high semantic interference conditions, as it was not reported in earlier studies investigating long-distance dependency integration or memory retrieval and future replication is necessary. However, I suggest that the hemisphere discrepancy observed between the 450 – 550 ms might be caused by the fact that more than one ERP component occurred and overlapped with each other to some extent within this time window. For example, the N400 effect typically occurs between 300 – 600 ms with a central-posterior distribution, while P600 effect typically occurs between 400 – 800 ms with a slightly right-lateralized central-posterior distribution. Although many studies have shown that the neural generators of the N400 and P600 are non-overlapping (Osterhout & Nicol, 1999), the cognitive processes of semantic integration and syntactic assignment are not fully independent (see Hagoort, 2003, for a review). For example, in the current experiment, the resolution of semantic interference in the high semantic interference condition was actually dependent on the noun phrase's syntactic features, since both NPs (e.g., “neighbor” & “resident” as in Example 13) were semantically plausible subjects of the verb (e.g., “complain”). Therefore, it was possible that the interaction of more than one ERP component during 450 – 550 ms time window resulted in the mixed results observed here.

ROI analysis. A potential concern of investigating interference effects using the current experimental sentences was that the proactive interference effects might be very weak and subtle, and thus hard to observe. In a previous English study investigating both proactive and retroactive semantic interference, Van Dyke and McElree (2011) have shown that although retrieval speed was the same across the PI and RI conditions (as there was no differences in SAT intercept or rate from SAT studies), there was a main

effect of interference type with PI being significant less detrimental than RI in the eye-tracking measures. Subjects showed less regressive eye movements and relative faster reading times in the PI condition. For example, on the critical verb, the total reading times for the PI conditions was 220 ms shorter than for the RI conditions (PI: 431 ms vs. RI: 651 ms). Moreover, the asymptotic accuracy was higher in the PI than RI conditions. Van Dyke and McElree suggested that this is because the distractors in the RI condition are more recent than distractors in the PI conditions and thus, provide a stronger source of interference, or the presentation of subsequent similar items have a detrimental effect on the quality of the target's representation via a process of feature overwriting in the RI condition but not in the PI condition (Nairne, 1990; Oberauer, 2009; Oberauer & Lange, 2008). However, their argument may be up for debate. Van Dyke and McElree did not further discuss the nature of the recency effect, such as how the decay of information affects the degree of interference. However, a general conclusion drawn from this study and many basic research studies on memory (Nairne, 2002; Oberauer & Kliegl, 2006) is that RI is much stronger than PI. In the current experiment, I have only examined PI and the distance over which retrieval happened was relatively shorter than in previous English studies - there were about eight phrases intervening between the matrix verb and its grammatical subject in our recent English studies (Tan et al., 2011, 2013), while there was only one phrase (i.e., an adverbial phrase) in the current Chinese study. Thus, I expected the interference effects to be weaker in the current experiment than in previous studies (e.g., Tan et al., 2011; A.E. Martin et al., 2012, 2014; Van Dyke, 2007) because of the less detrimental nature of PI than RI and the relatively short distance between the subject-verb dependencies. Based on the strong prediction of observing an anterior

negativity effect, which is associated with maintaining or integrating multiple memory representation over long-distance interval, and visual inspection (see Fig. 1), I ran an ROI analysis in the anterior electrodes specifically. Anterior negativity effects are usually maximal at anterior sites, and might fail to show statistically reliable differences across regions due to the weak interference effects that were investigated here. I therefore felt justified to proceed with a ROI analysis on anterior sites only to obtain a more sensitive test of anterior negativities.

In analogy to previous studies in which an anterior negativity was reported in the high WM demand sentences (Fiebach et al., 2002; King & Kutas, 1995; Kluender et al., 1998; Van Berkum, et al., 2007), I examined the anterior electrodes specifically including F3/F4, F5/F6, F7/F8, FC3/FC4, FC5/FC6, and FT7/FT8. Although most previous studies observed an AN that was usually maximal at a left hemisphere site, some studies have observed this effect to be bilateral distributed (King & Kutas, 1995; Nieuwland, Otten, & Van Berkum, 2007; Nieuwland & Van Berkum, 2006; Nieuwland & Van Berkum, 2008; Van Berkum, Brown, & Hagoort, 1999). As a result, repeated ANOVAs were conducted on the mean amplitude with the within-subject factor of Semantic interference (high vs. low) $\times$ Syntactic interference (high vs. low) $\times$ Hemisphere (left vs. right). The results were presented in Table 4 (see also Fig. 3). Based on results from the 50-ms time window analyses, first, the ROI results confirmed previous finding from whole brain analyses that there was a late anterior negativity elicited by the high semantic interference conditions as compared to the low semantic interference condition between 650 – 800 ms (HiSem: -0.13 μ V vs. LoSem: 0.17 μ V), $F(1, 36) = 6.68$, $MSE = 6.47$, $p = .014$. However, different from the whole brain analyses, there was a main effect of semantic interference in the 300

– 500 ms time window with high semantic interference conditions being more negative than the low semantic interference conditions (HiSem: -0.83 μ V vs. LoSem: -0.61 μ V), $F(1, 36) = 4.55$, $MSE = 3.43$, $p = .040$. In addition, there was a significant interaction of syntactic interference $\times$ hemisphere in the 300 – 500 ms time window, $F(1, 36) = 4.86$, $MSE = 1.37$, $p = .034$. Planned comparisons showed that the high syntactic interference conditions was more negative than the low syntactic interference conditions in the left hemisphere (HiSyn: -0.68 μ V vs. LoSyn: -0.43 μ V), $F(1, 36) = 6.15$, $MSE = 1.16$, $p = .018$, but not in the right hemisphere (HiSyn: -0.87 μ V vs. LoSyn: -0.89 μ V), $F < 1$.

Table 4 Results of repeated ANOVAs for the mean amplitude of the critical verb region.

	Time windows					
	300 – 500 ms			650 – 800 ms		
	<i>F</i>	<i>MES</i>	<i>p</i>	<i>F</i>	<i>MES</i>	<i>p</i>
<i>Midline</i>						
Semantic	0.16	0.41	0.695	1.26	6.43	0.270
Syntactic	1.25	4.37	0.270	6.68	19.84	0.014*
Semantic $\times$ Syntactic	0.25	0.81	0.624	0.10	0.34	0.755
Semantic $\times$ Region	1.85	3.72	0.158	1.56	7.12	0.209
Syntactic $\times$ Region	0.30	0.42	0.841	0.71	3.17	0.497
Sem $\times$ Syn $\times$ Reg	0.31	1.05	0.751	1.36	6.55	0.262
	<i>F</i>	<i>MES</i>	<i>p</i>	<i>F</i>	<i>MES</i>	<i>p</i>
<i>Lateral</i>						
Semantic	0.58	0.24	0.450	0.01	0.01	0.914
Syntactic	1.33	0.34	0.256	0.32	0.12	0.578
Semantic $\times$ Syntactic	0.66	0.14	0.423	0.11	0.04	0.743
Semantic $\times$ Region	2.40	4.54	0.130	4.77	12.29	0.036*
Syntactic $\times$ Region	2.87	3.91	0.099	0.00	0.00	0.965

Semantic $\times$ Hemisphere	0.00	0.00	0.996	0.18	0.19	0.674
Syntactic $\times$ Hemisphere	5.15	2.41	0.029* ²⁹	1.12	0.92	0.298
Sem $\times$ Reg $\times$ Hemis	0.00	0.00	0.987	0.00	0.00	0.968
Syn $\times$ Reg $\times$ Hemis	0.08	0.01	0.773	0.17	0.05	0.680
Sem $\times$ Syn $\times$ Reg	2.50	2.85	0.122	2.85	6.01	0.100
Sem $\times$ Syn $\times$ Hemis	2.80	1.63	0.103	0.03	0.01	0.873
Sem $\times$ Syn $\times$ Reg $\times$ Hemis	0.99	0.24	0.326	1.30	0.53	0.261

	<i>F</i>	<i>MES</i>	<i>p</i>	<i>F</i>	<i>MES</i>	<i>p</i>
<u>ROI analysis</u>						
Semantic	4.56	3.43	0.040*	6.68	6.68	0.014*
Syntactic	1.69	0.97	0.201	0.06	0.06	0.805
Semantic $\times$ Syntactic	1.14	0.87	0.294	2.43	2.43	0.128
Semantic $\times$ Hemisphere	0.00	0.00	0.990	0.10	0.10	0.750
Syntactic $\times$ Hemisphere	4.85	1.37	0.034*	0.55	0.55	0.463
Sem $\times$ Syn $\times$ Hemisphere	0.76	0.31	0.389	0.66	0.66	0.423

²⁹ In the lateral electrodes analyses, there was a significant interaction of Syntactic $\times$ Hemisphere in the 300 – 500 ms time window. However, in the 50-ms time window analyses within this large time window, such interaction was not significant in either 350 – 400 ms or 450 – 500 ms time window ($ps > .15$). Given the possibility that certain simple main effect might be more distributed/focal in certain time windows, we still took a close look into the Syntactic $\times$ Hemisphere interaction. Planned comparisons revealed that in the right hemisphere, there was a syntactic interference effect in both 300 – 350 ms and 400 – 450 ms time windows, but not in either 350 – 400 ms or 450 – 500 ms time window ($ps > .15$). There was no significant syntactic interference effect in the left hemisphere. Due to the lack of continuation to meet our selection criteria (i.e., significant in more than two consecutive 50-ms analyses), we did not take this interaction into consideration.

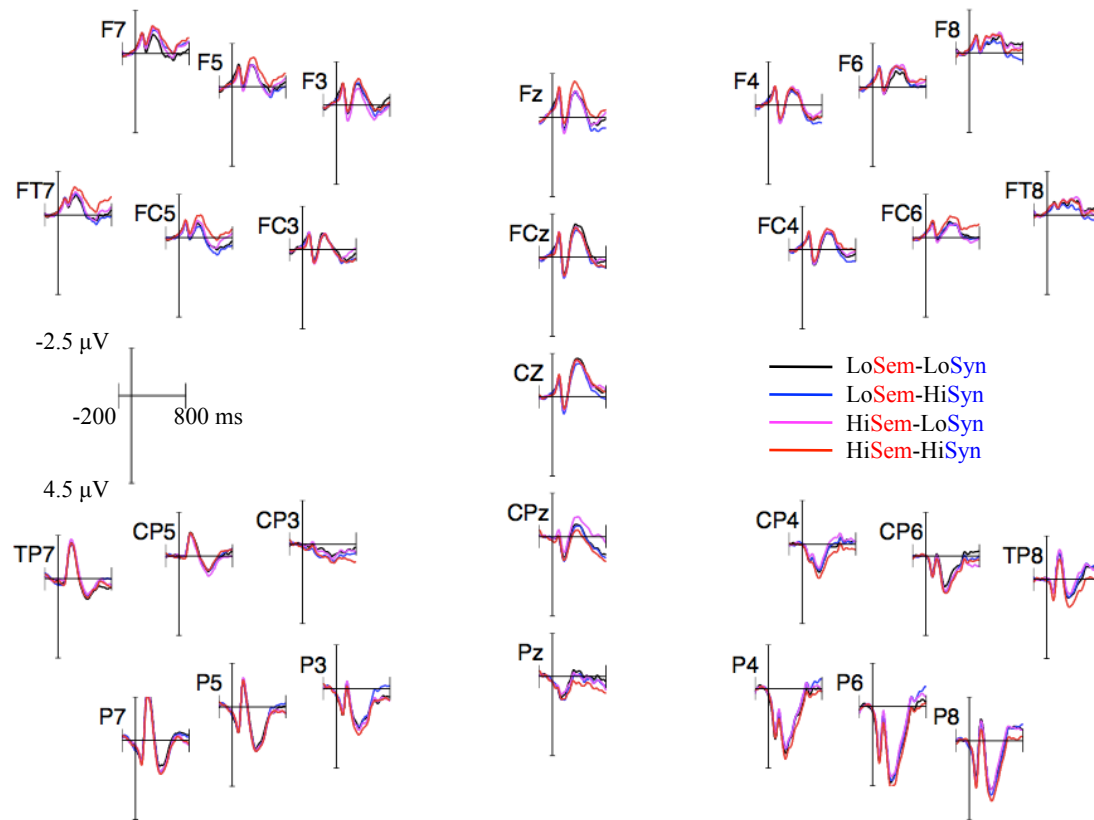


Figure 1 Grand average ERPs for each experimental condition time-locked to the onset of critical verb from -200 to 800 ms, at all electrodes included in the statistical analyses (negativity is plotted up).

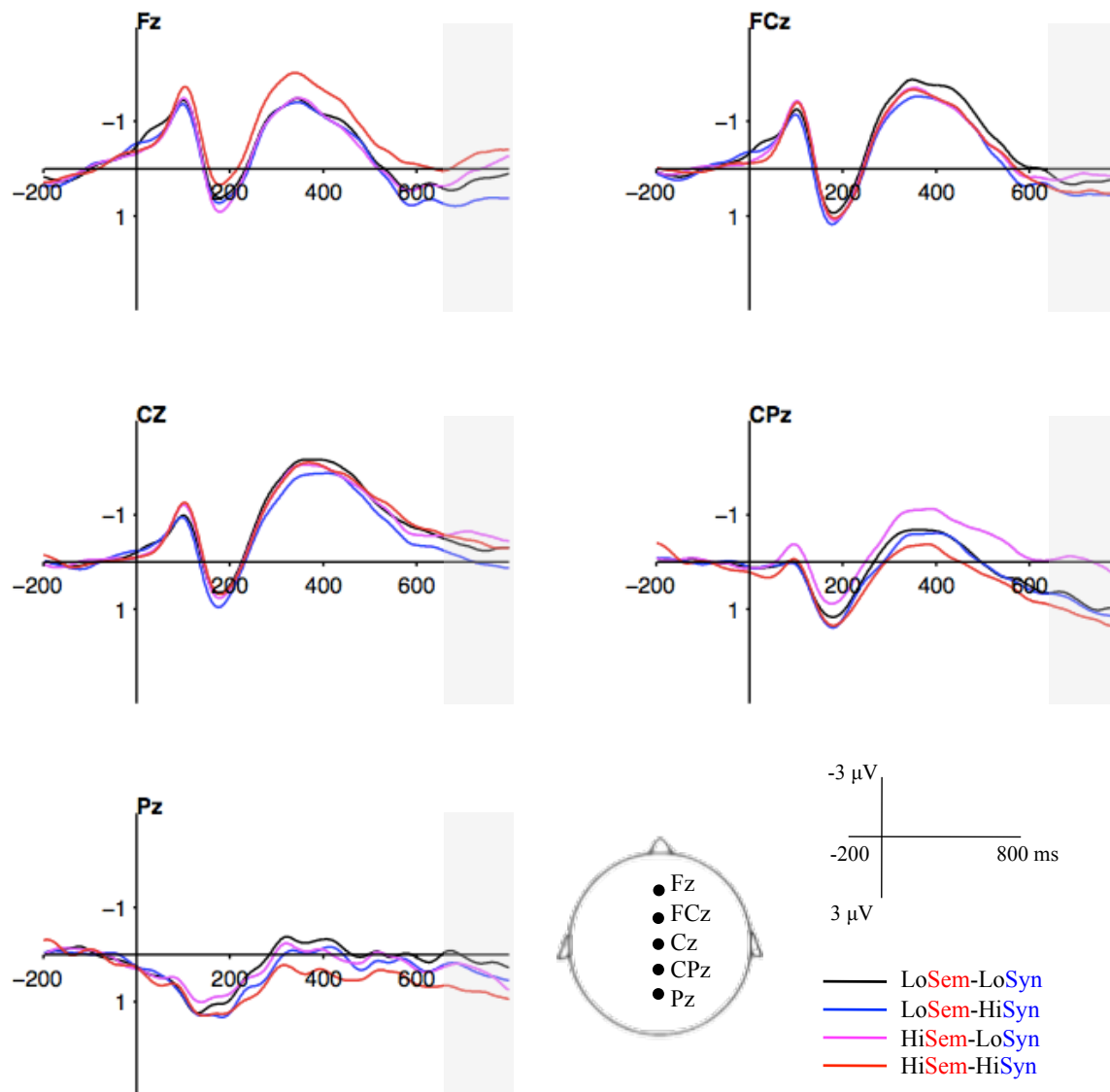


Figure 2 Grand average ERPs for each experimental condition time-locked to the onset of critical verb from -200 to 800 ms, at midline electrode position Fz, FCz, CZ, CPz, and Pz (negativity is plotted up).

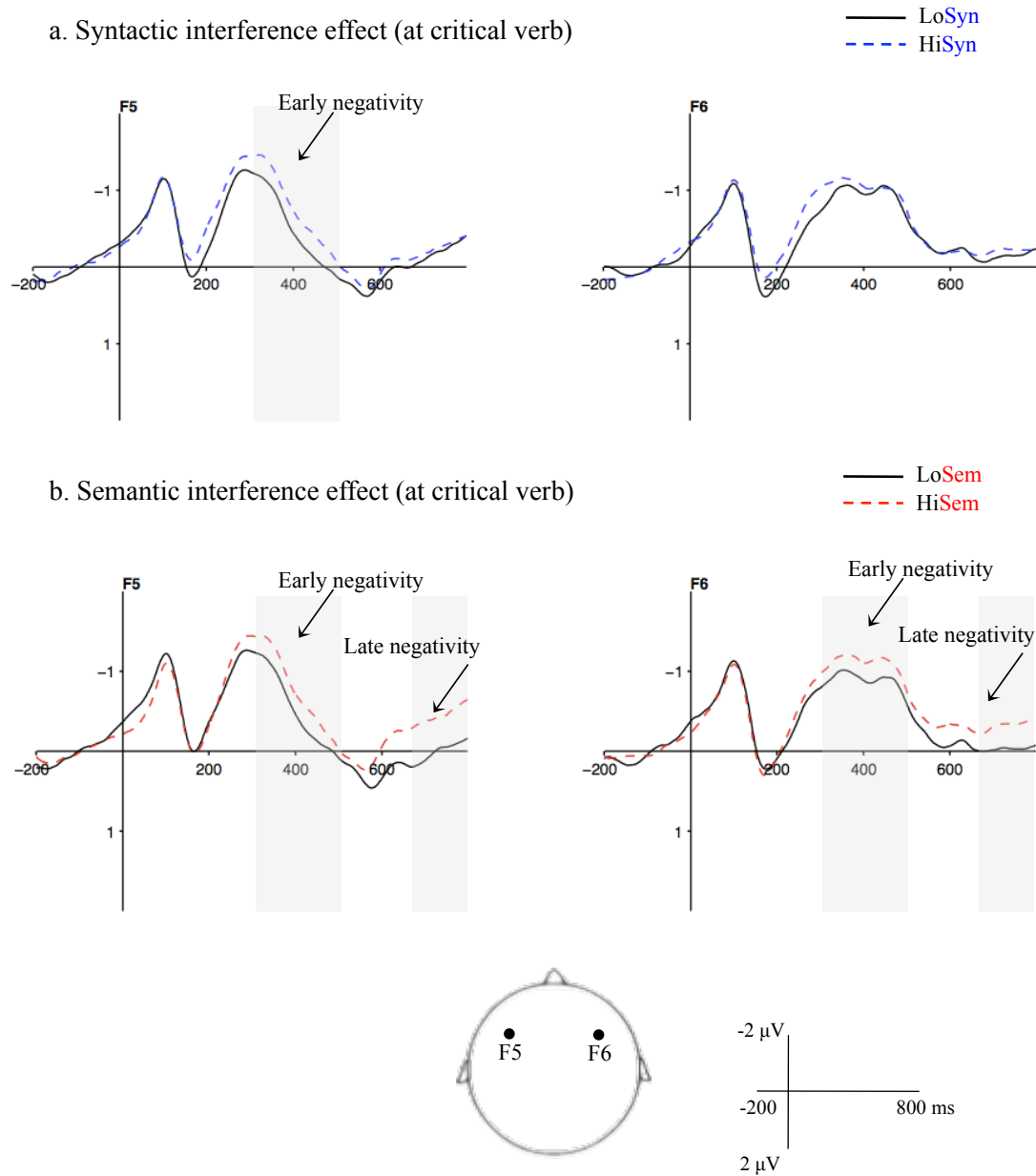


Figure 3 Grand average ERPs for semantic (averaged across low and high syntactic interference conditions) and syntactic (averaged across low and high semantic interference conditions) interference effects time-locked to the onset of critical verb from -200 to 800 ms, at electrode position F5 and F6 (negativity is plotted up).

To summarize, on the critical verb, there were ERP effects elicited by both high semantic and high syntactic interference sentences relative to low interference sentences. Sentence with high syntactic interference elicited a left anterior negativity between 300 –

500 ms larger, followed by a P600 effect along the midline between 650 – 800 ms, relative to the low syntactic interference sentences. Sentences with high semantic interference elicited an early bilateral distributed anterior negativity between 300 – 500 ms, followed by a late negativity with the same scalp distribution between 650 – 800 ms. However, as to whether the negative shift elicited by the high compared to the low semantic interference conditions was two separate negativities, or one sustained component in the current study, it may be up for debate. In the current analysis, although there was no significant semantic interference effect in any 50-ms time window analysis between 500 – 650 ms ($ps > .07$), there was an overall semantic interference effect between 300 – 800 ms with the HiSem condition being more negative than the LoSem condition (HiSem: $-0.37 \mu\text{V}$ vs. LoSem: $-0.14 \mu\text{V}$), $F(1, 36) = 4.94$, $MSE = 3.77$, $p = .033$. It was possible that the interaction of semantic and syntactic interference during 450 – 550 ms or the large individual differences in the time point of starting to resolve semantic interference (A. E. Martin et al., 2014; Tan et al., 2011) capped the sustained nature of the anterior negative effect associated with semantic interference effect. This will be further discussed in the discussion section.

3.2.2 Spillover region

In addition to the critical verb region, I have also examined the ERP effects in the post-critical verb region (i.e., the first word after critical verb, such as “*rent*” in Example 13) as the interference effect might spillover over to the following words (A. E. Martin, et al., 2012, 2014; Tan et al., 2011). Statistical analyses were conducted in the same manner as in the critical verb region. Since I did not have a strong prediction about the effects

that might observe in the spillover region, no ROI analysis was conducted and one time window (i.e., 250 – 800 ms) was identified based on visual inspection and the consecutive 50-ms latency window analysis.

As shown in Table 5 (see also Fig. 4 and 5), in the lateral electrodes, there was no main effect of either semantic or syntactic interference, or any interaction between either semantic or syntactic interference and brain region or hemisphere. However, in the midline electrodes, there was an interaction of semantic and syntactic interference between 250 – 800 ms, $F(1, 36) = 9.01$, $MSE = 17.34$, $p = .005$. Given the early onset and the same scalp distribution of this interaction as the P600 on the critical verb, I suggest that it is a spillover effect elicited by the syntactic interference effect in the critical region, though semantic interference came in and interacted with syntactic interference resolution. Therefore, planned comparisons were conducted on different levels of semantic interference to examine how syntactic interference effect was affected by the semantic features of the distractor. The results revealed that in the low semantic interference condition, the high syntactic interference sentences elicited a positive shift in the 250 – 800 ms time window (HiSyn: $-0.004 \mu V$ vs. LoSem: $-0.46 \mu V$), $F(1, 36) = 8.55$, $MSE = 3.79$, $p = .006$, while there was no syntactic interference effect in the high semantic interference conditions, $F(1, 36) = 1.02$, $MSE = 0.47$, $p = .32$. This finding of a positive effect in the post-critical word was consistent with A. E. Martin et al.'s recent results (2014), in which they found a sustained positivity in the spillover region and argued that it reflected continuous syntactic revising.

Table 5 Results of repeated ANOVAs for the mean amplitude of the spillover region.

Time window	250 – 800 ms
-------------	--------------

	<i>F</i>	<i>MSE</i>	<i>p</i>
<u>Midline</u>			
Semantic	2.49	7.67	0.123
Syntactic	1.53	3.98	0.225
Semantic × Syntactic	9.01	17.34	0.005*
Semantic × Region	1.29	3.33	0.281
Syntactic × Region	1.94	4.02	0.135
Sem × Syn × Reg	1.50	3.48	0.230
	<i>F</i>	<i>MSE</i>	<i>p</i>
<u>Lateral</u>			
Semantic	0.03	0.01	0.870
Syntactic	0.09	0.03	0.770
Semantic × Syntactic	0.11	0.04	0.743
Semantic × Region	2.56	4.23	0.118
Syntactic × Region	1.93	2.61	0.174
Semantic × Hemisphere	2.32	1.18	0.137
Syntactic × Hemisphere	0.33	0.12	0.569
Sem × Reg × Hemis	0.00	0.00	0.948
Syn × Reg × Hemis	0.73	0.26	0.399
Sem × Syn × Reg	0.09	0.07	0.769
Sem × Syn × Hemis	1.87	1.46	0.180
Sem × Syn × Reg × Hemis	1.00	0.33	0.323

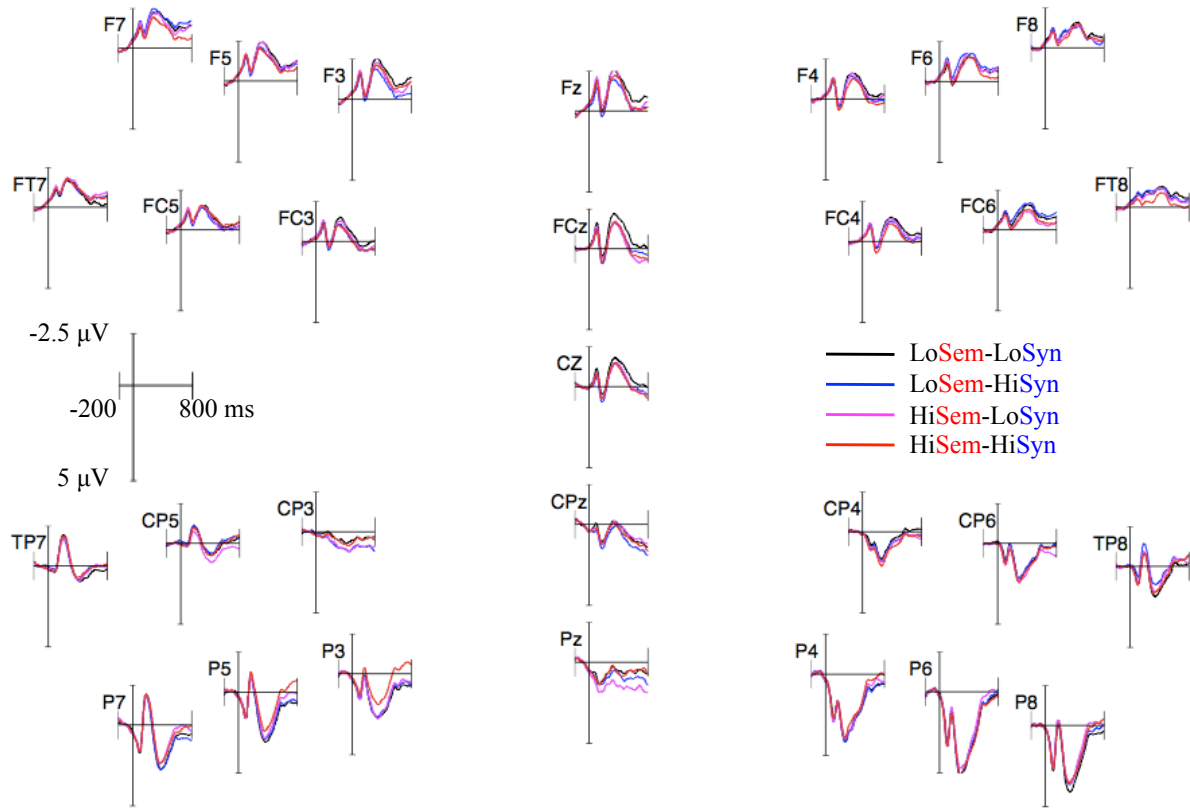


Figure 4 Grand average ERPs for each experimental condition time-locked to the onset of the first word following the critical verb (i.e., spillover region) from -200 to 800 ms, at all electrodes included in the statistical analyses (negativity is plotted up).

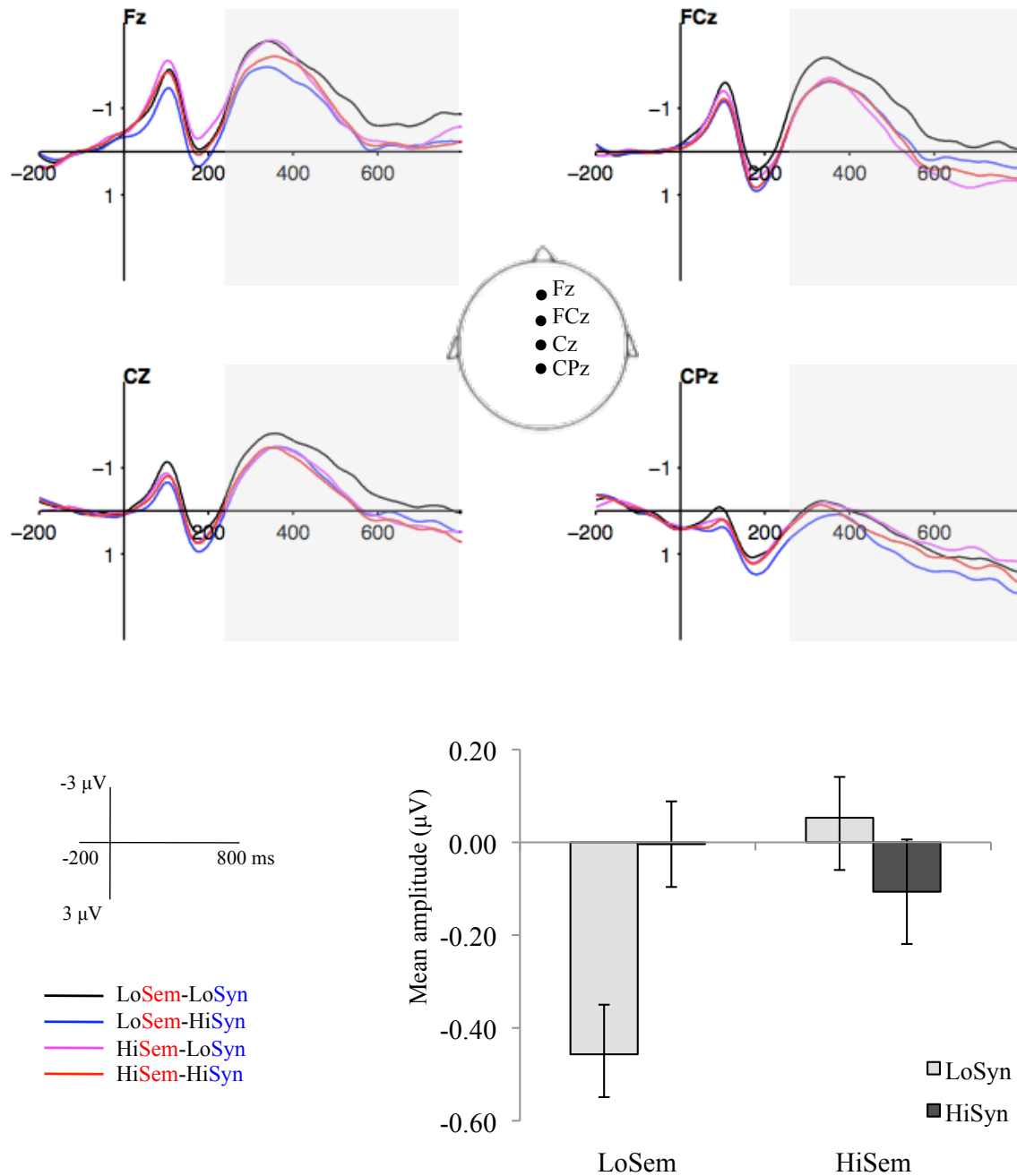


Figure 5 Grand average ERPs for each experimental condition time-locked to the onset of the first word following the critical verb (i.e., spillover region) from -200 to 800 ms, at electrode position Fz, FCz, CZ, and CPz (negativity is plotted up). As there was no interaction of interference effects with electrodes, a bar graph averaged across the five midline electrodes was presented in the bottom right corner for the mean amplitude between 250 – 800 ms. The error bars representing corrected standard error of the mean (Cousineau, 2005).

To summarize, in the spillover region, the semantic interference and syntactic interference interacted with each other along the midline. Further analyses demonstrated that in the low semantic interference condition only, the high syntactic interference condition elicited a positive deflection in the 250 – 800 ms time window. It was less expected that the syntactic interference effect would only be evident in the low but not the high semantic interference condition. Although previous English studies demonstrated that semantic and syntactic interference did not interact (Tan et al., 2011, 2013; Van Dyke, 2007; Van Dyke & McElree, 2011), if there were any interaction between semantic and syntactic interference, one might expect the syntactic interference to be more predominant in the high semantic interference condition due to more processing difficulty in the double interference conditions. In the current experiment, there was the possibility that in the HiSem/HiSyn interference condition, subjects temporarily mistook the distractor NP as the grammatical subject of the critical verb due to its partially match with the semantic retrieval cues and, thus, did not experience interference effect immediately until they revised their initial false interpretation at a later point (e.g., in the sentence final region or during question answering). Previous studies have shown that a cue-based retrieval mechanism is subject to retrieval errors, at least temporarily, and subjects do not always resolve the interference effect immediately (A. E. Martin et al., 2012; Wagers, Lau, & Philips, 2009). However, it should be noted that this speculation might not fit the our results very well, as such retrieval errors are assumed to be more likely to occur at an early point during sentence processing before controlled analysis happens, e.g., at the critical verb, but this was not the case in our current experiment. Nonetheless, it is also possible that this kind of retrieval error only

occurs when subjects start resolving both semantic and syntactic interference, with semantic interference resolution not being initiated until the spillover region. Although subjects detected semantic interference as early as syntactic interference (as indexed by the anterior negativities for both interference effects around 300 ms on the critical verb), they did not or were not able to resolve semantic interference until a later point of the sentence when more information had accumulated. Based on the results from previous English studies, I suggest that syntactic interference resolution could be faster than semantic interference resolution as it involves analyzing a finite set of grammatical features to determine whether each noun phrase matches the critical verb. In contrast, determining semantic fit is more complex due to the possibility of degrees of fit (Glaser et al., 2013; Van Dyke, 2007). Thus, the temporary retrieval error might occur only when subjects were trying to resolve both semantic and syntactic interference. These results will be further discussed in the discussion section. Generally, I suggest that the sustained positivity reflected subjects' continuing effort or difficulty in resolving semantic and syntactic interference.

3.2.3 Exploratory analysis: sentence final region

I have also examined ERP effects on the sentence final position (i.e., the second word after critical verb, e.g., “*problem*” in Example 13) for the global effect to sentence endings caused by processing difficulty in earlier part of the sentences. This analysis was exploratory because so far, it is unclear what the nature of the sentence final “wrap-up” effect is and what cognitive processes happen in the final region (Hagoort, 2003; Van Dyke, 2007). Some researchers have argued that the sentence final effect is related to the

task response requirements, such as preparing for answering comprehension questions (Hagoort, 2003; Schriefers, Friederici, Kuhn, 1995). In an ERP study, Hagoort (2003) found that three different types violation (i.e., semantic, syntactic, and semantic-syntactic double violations), which elicited different ERP effects when the violation happened in the middle part of the sentences, all resulted in a similar enlarged and long-lasting negativity in the sentence-final position. Therefore, Hagoort has suggested that it is hard to disentangle specific effects other than a N400 effect from the global processing costs at the sentence final region. In my current experiment, there should be little semantic or syntactic processing difficulty caused by the final word itself, because all the words were high frequency nouns without ambiguity or violation. In addition, the final words were always the same across the four conditions within each set of sentences. However, more than one potential processing difficulty appeared in the earlier part of the sentences – semantic and syntactic interference, the relative low frequency of the ORC structures relative to SRC structures, the longer distance of filler-gap dependency in the SRC structures relative ORC structures, and the experimental sentences themselves were generally long and complicated (based on the meaningfulness rating in our pilot studies). Therefore, the puzzling nature of the sentence final wrap-up effect and the multiple processing difficulties in the current experiment make it hard to predict and interpret the ERP effects in the sentence final region. Therefore, I only exploratorily examined the ERP data on the final word. Based on visual inspections, the consecutive 50-ms latency window analyses, and the results from previous studies investigating the sentence final wrap-up effect (Hagoort, 2003; Hagoort, Brown, & Groothusen, 1993; Osterhout &

Nicol, 1999; Schriefers, Friederici, & Kuhn, 1995), two time windows were identified as in the critical verb region: 1) 300 – 500 ms time window; 2) 650 – 800 ms time window.

The results are presented in Table 6 (see also Fig. 6 and 7). Although the processing differences in the final region might not be directly caused by the interference manipulation, I still refer to the observed differences between different types of sentences as “semantic interference manipulation” and “syntactic interference manipulation” consistent with previous analyses. In both midline and lateral electrodes, there was no main effect of semantic or syntactic interference. None of the interactions reached significance in the midline (e.g., see FCz in Fig. 7). In the lateral electrodes, there was an interaction of Syntactic $\times$ Hemisphere between 300 – 500 ms, $F(1, 36) = 2.79$, $MSE = 5.54$, $p = .024$, and this interaction was modulated by the brain regions since the three way interaction of Syntactic $\times$ Hemisphere $\times$ Regions was significant as well, $F(1, 36) = 9.48$, $MSE = 7.33$, $p = .004$. In further planned comparisons between 300 – 500 ms, I examined the interaction of Syntactic $\times$ Hemisphere in anterior and posterior regions separately. The results showed that this interaction was only significant in the anterior site, $F(1, 36) = 16.91$, $MSE = 2.49$, $p < .001$, but not in the posterior site ($F < 1$). In the left anterior region, the low syntactic interference condition elicited a more negative shift relative to the high syntactic interference condition (HiSyn: $-1.20 \mu V$ vs. LoSyn: $-1.44 \mu V$), $F(1, 36) = 5.43$, $MSE = 1.07$, $p = .025$, whereas in the right anterior site, the high syntactic interference condition elicited a more negative shift relative to the low syntactic interference condition (HiSyn: $-1.47 \mu V$ vs. LoSyn: $-1.19 \mu V$), $F(1, 36) = 7.72$, $MSE = 1.44$, $p = .009$. In addition, in a later time window (i.e., 650 – 800 ms), the interaction of Syntactic $\times$ Hemisphere was significant in the lateral electrodes, $F(1, 36) = 9.00$, $MSE =$

2.22, $p = .005$. Planned comparison showed that in the right hemisphere only, the high syntactic interference condition elicited a more negative shift relative to the low syntactic interference condition (HiSyn: $-0.40 \mu\text{V}$ vs. LoSyn: $-0.11 \mu\text{V}$), $F(1, 36) = 7.26$, $MSE = 1.52$, $p = .011$, while no syntactic interference effect was observed in the left hemisphere ($p = .205$).

Table 6 Results of repeated ANOVAs for the mean amplitude of the sentence final region.

	Time windows					
	300 – 500 ms			650 – 800 ms		
	<i>F</i>	<i>MES</i>	<i>p</i>	<i>F</i>	<i>MES</i>	<i>p</i>
<i>Midline</i>						
Semantic	0.16	0.41	0.695	0.17	1.03	0.685
Syntactic	1.25	4.37	0.270	0.79	4.80	0.381
Semantic $\times$ Syntactic	0.25	0.81	0.624	0.12	0.84	0.734
Semantic $\times$ Region	1.85	3.72	0.158	0.19	0.91	0.857
Syntactic $\times$ Region	0.30	0.42	0.841	1.51	3.67	0.212
Sem $\times$ Syn $\times$ Reg	0.31	1.05	0.751	0.57	4.59	0.576
<i>Lateral</i>						
Semantic	0.13	0.03	0.725	0.09	0.04	0.772
Syntactic	0.01	0.00	0.938	0.53	0.62	0.473
Semantic $\times$ Syntactic	4.26	1.22	0.046* ³⁰	2.01	1.31	0.165
Semantic $\times$ Region	2.61	2.66	0.115	0.12	0.26	0.736
Syntactic $\times$ Region	0.04	0.03	0.853	1.29	2.85	0.264
Semantic $\times$ Hemisphere	0.03	0.02	0.859	0.12	0.07	0.733
Syntactic $\times$ Hemisphere	5.54	2.79	0.024*	9.48	7.33	0.004*
Sem $\times$ Reg $\times$ Hemis	1.58	0.29	0.216	0.05	0.02	0.828
Syn $\times$ Reg $\times$ Hemis	9.00	2.22	0.005*	3.67	2.17	0.063

³⁰ In the lateral electrodes analyses between 300 – 500 ms, the interaction of Semantic $\times$ Syntactic was only significant in the 350 – 400 ms time window ($p = .038$), but not in any of the following three 50-ms time windows ($ps > .08$). Due to the lack of continuation to meet our selection criteria (i.e., significant in more than two consecutive 50-ms analyses), we did not take this interaction into consideration.

Sem $\times$ Syn $\times$ Reg	2.22	3.80	0.145	0.49	1.50	0.488
Sem $\times$ Syn $\times$ Hemis	0.13	0.03	0.725	0.39	0.31	0.539
Sem $\times$ Syn $\times$ Reg $\times$ Hemis	0.01	0.00	0.938	1.00	0.45	0.323

Note.

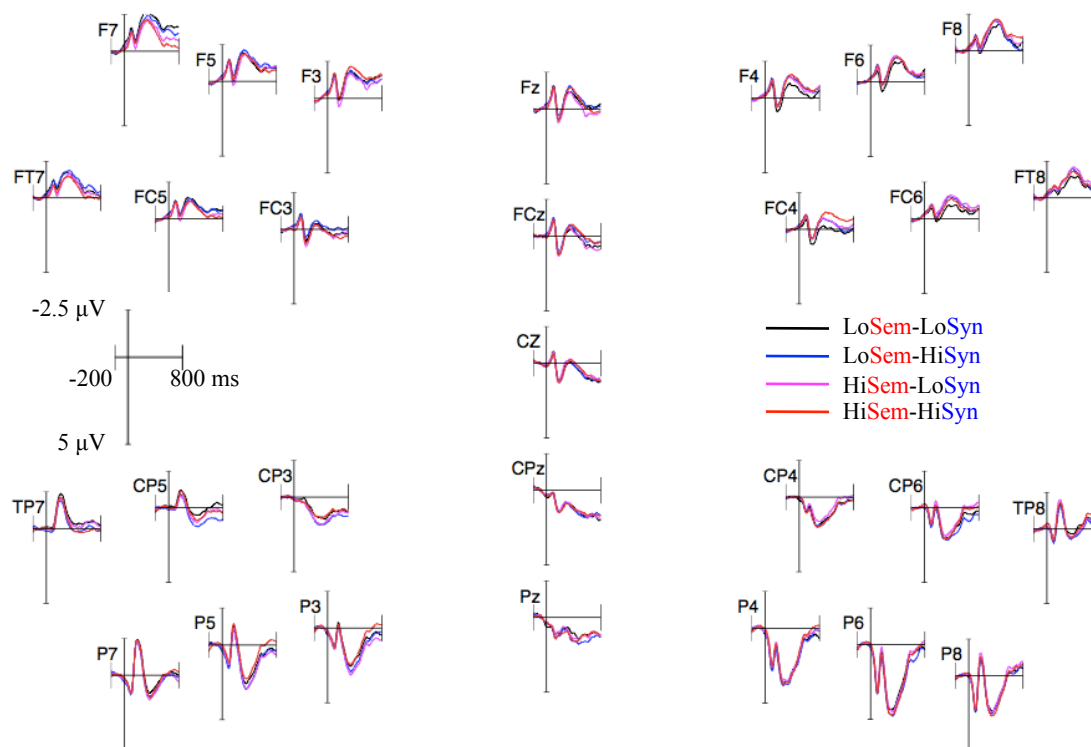


Figure 6 Grand average ERPs for each experimental condition time-locked to the onset of the final word (i.e., the second word after the critical verb) from -200 to 800 ms, at all electrodes included in the statistical analyses (negativity is plotted up).

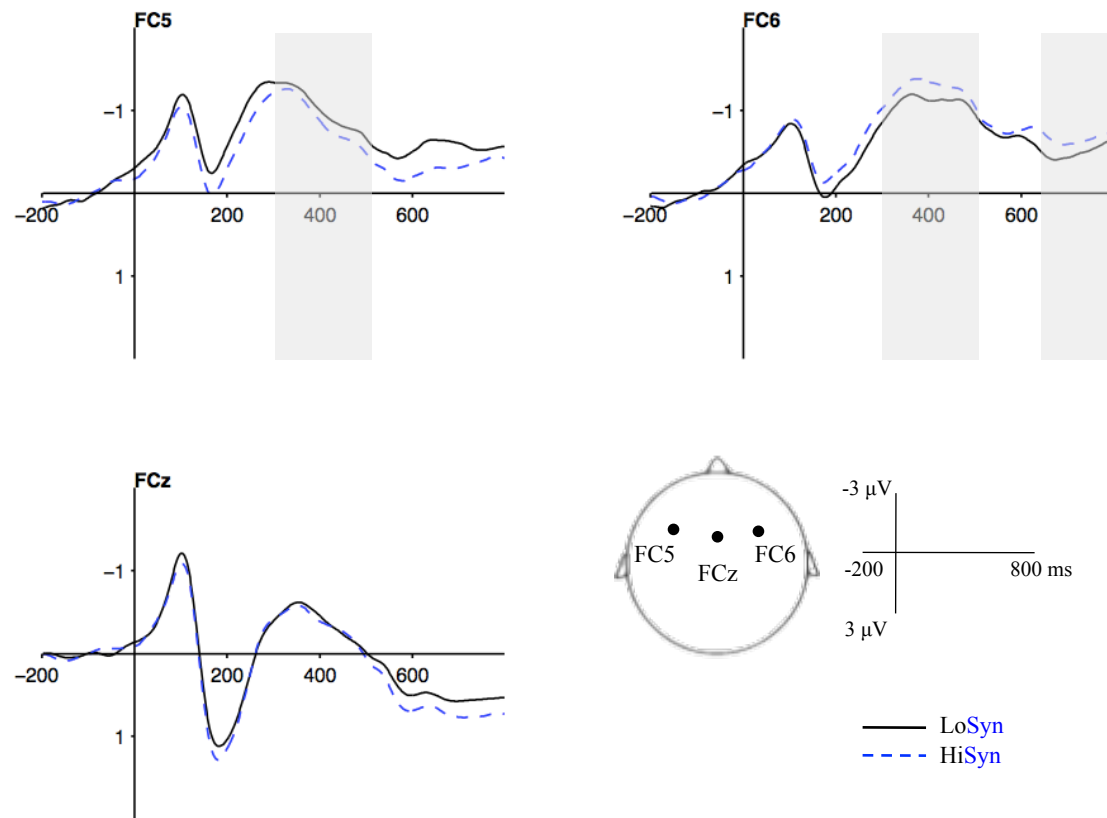


Figure 7 Grand average ERPs for syntactic interference effect (averaged across semantic interference conditions) time-locked to the onset of the final word from -200 to 800 ms, at electrode position F5, F6, and FCz (negativity is plotted up).

(* Although it looks like that there might be a positive syntactic interference effect on FC5, it was not statistically significant in any of the 50-ms time window analyses).

To sum up, in the sentence final position, I did not observe any semantic interference effect. The syntactic interference effect interacted with scalp position. Low syntactic interference sentences elicited a negative shift in the left anterior region between 300 – 500 ms, whereas an opposite effect was observed in the right anterior site with high syntactic interference sentences elicited a negative shift between 300 – 500 ms. The right anterior negativity elicited by the high syntactic interference condition merged into a late positive effect widely distributed over the entire right hemisphere between 650 – 800 ms. I did not have a very adequate explanation about the hemisphere asymmetric

observed here, due to the puzzling nature of the sentence final wrap-up effect and the multiple processing difficulty during the sentences parsing, as I have discussed earlier. One possible explanation for the left anterior negativity elicited by the LoSyn condition was that these ERP effects were related to the ORC-advantage (i.e., processing difficulty in the LoSyn compared to HiSyn sentences) in Chinese. In a self-paced reading experiment investigating Chinese RC processing, Gibson and Wu (2013) have reported and ORC –advantage at the sentence final region. They suggested that this effect should be interpreted as a sentence final wrap-up effect, instead of a spillover effect from the RC processing, because in the “Head noun + Word1 + Word2 + Final word” structure, this effect only occurred at the head noun, where filler-gap integration took place, and the final region, but not at the two words intervening between the head noun and the final word. Therefore, I suggested that the left AN effects (300 – 500 ms and 650 – 800 ms) elicited by the LoSyn conditions might be related to more difficulty in processing SRC structure in earlier part of the sentence, while the right AN effect (300 – 500 ms) elicited by the HiSyn conditions might be related to difficulty in resolving high syntactic interference. However, this speculation is quite tentative and results from the current experiment or previous relevant studies do not provide a clear explanation for it. Most of the previous ERP studies investigating RC processing or interference resolution did not report the effects on the final word (e.g., A. E. Martin, 2012, 2014; Yang et al., 2012) or the critical word was also the final word itself (e.g., head noun, in Packard et al., 2011). Therefore, it is hard to explain the sentence-final effect by referring to previous relevant studies. In addition, I have suggested that individual differences in processing strategies (e.g., when to start resolving semantic interference), WM capacity, executive control, or

other cognitive abilities might play an important role during the wrap-up process (Tan et al., 2011). Previous ERP studies have shown that low- and high-capacity subjects did not only differ in the magnitude of certain ERP component, but also in showing different ERP patterns during complex sentence processing (Nieuwland & Van Berkum, 2006; Ye, Luo, Friederici, & Zhou, 2006). Thus, the sentence-final ERP effects presented an interesting puzzle, and I will not focus on this region in further discussion.

3.2.4 ERP effects at the RC head noun and adverbial phrase

In addition, I examined ERP effects on the RC head noun (e.g., “*resident*” in Example 13) and the adverbial phrase following the head noun (i.e., the phrase between the head noun and the critical verb, e.g., “*everyday*” in Example 13) for the processing differences associated with different RC types. Although there is a general robust SRC-advantage in most western languages (see Traxler et al., 2005 for a review), in previous Chinese studies, several self-paced reading experiments reported an ORC-advantage on the RC head noun, where filler-gap integration is supposed to happen, with the head noun in the ORC structure being processed faster (Chen et al., 2008; Gibson & Wu, 2013; Hsiao & Gibson, 2003). However, there have also been studies that failed to observe any RC type differences on the head noun and some researchers suggested that the processing demand only differed during the RC region (Jäger et al., 2015; Lin et al., 2006). In two recent Chinese ERP studies, researchers observed ERP effects associated with RC type differences at the head noun, though the results differed across the two studies (Yang et al., 2010; Packard et al., 2011). On the RC head noun of sentence containing an object-modifying RC, Yang et al. (2010) reported an ORC-elicited sustained AN effect (250 –

800 ms), while Packard et al. (2011) reported an SRC-elicited P600 effect. I suggest that such discrepancy might be caused by the problem of temporary syntactic ambiguity in both studies, and the confounding problem caused by the sentence final wrap-up effect in Packard et al.'s study as the head noun was also the final word. Therefore, it is worth examining the ERP effects on the RC head noun in the current study due to the inconclusive results from previous studies. Additionally, I examined ERP effects on the adverbial phrase inserting between the head noun and the critical verb, to make sure no matter what kinds of ERP effect observed on the head noun, such effects would not extend beyond the adverbial phrase and confound with interference effects on the critical verb.

3.2.4.1 ERP effects at the head noun

Based on visual inspection and the consecutive 50-ms latency window analyses, two time windows were identified: 1) 300 – 650 ms time window; 2) 650 – 800 ms time window. The statistical results are shown in Table 7 (see also Fig. 8 and 9). Although the processing differences at the head noun might be caused by many factors (e.g., RC structure frequency, memory load of maintaining/integration syntactic predictions) other than the interference manipulation, I still refer to the observed differences between different conditions as “semantic interference” and “syntactic interference” manipulations to be consistent with previous analyses. In the 300 – 650 ms time window, none of the effects reached significance along the midline, while there was an interaction of Syntactic $\times$ Regions in the lateral electrodes, $F(1, 36) = 7.21$, $MSE = 9.04$, $p = .011$. Planned comparison revealed that the high syntactic interference condition (i.e., ORC sentences) elicited a more positive shift compared to the low syntactic condition (i.e., SRC

sentences) in the posterior region only, peaking around 300 ms (HiSyn: 1.45 μ V vs. LoSyn: 1.23 μ V), $F(1, 36) = 4.57$, $MSE = 0.93$, $p = .039$, but not in the anterior regions ($p = .18$). In the 650 – 800 ms time window, along the midline, there was a main effect of semantic interference, with the high semantic interference condition being more negative than the low semantic interference condition (HiSem: -1.03 μ V vs. LoSyn: -0.73 μ V), $F(1, 36) = 6.06$, $MSE = 17.36$, $p = .019$. None of the other effects were significant at the RC head noun in either midline or lateral analysis.

Table 7 Results of repeated ANOVAs for the mean amplitude of the RC head noun.

	Time windows					
	300 – 650 ms			650 – 800 ms		
	<i>F</i>	<i>MES</i>	<i>p</i>	<i>F</i>	<i>MES</i>	<i>p</i>
<u>Midline</u>						
Semantic	1.83	4.61	0.185	6.06	17.36	0.019*
Syntactic	3.41	11.29	0.073	2.83	14.59	0.101
Semantic $\times$ Syntactic	1.08	2.76	0.306	0.49	1.99	0.487
Semantic $\times$ Region	0.22	0.73	0.830	0.71	4.41	0.502
Syntactic $\times$ Region	1.47	5.39	0.237	0.64	5.14	0.508
Sem $\times$ Syn $\times$ Reg	0.48	1.54	0.639	0.85	4.87	0.437
	F	MES	p	F	MES	p
<u>Lateral</u>						
Semantic	0.18	0.06	0.671	1.20	0.53	0.280
Syntactic	0.30	0.08	0.589	0.28	0.08	0.600
Semantic $\times$ Syntactic	1.05	0.29	0.312	0.23	0.11	0.638
Semantic $\times$ Region	0.77	0.81	0.385	4.98	8.92	0.032* ³¹
Syntactic $\times$ Region	7.21	9.04	0.011*	3.21	7.87	0.082
Semantic $\times$ Hemisphere	1.06	0.97	0.311	0.03	0.03	0.864
Syntactic $\times$ Hemisphere	0.60	0.42	0.444	0.29	0.40	0.596

³¹ In the lateral electrodes analyses between 650 – 800 ms, the interaction of Semantic $\times$ Region was only significant in the 650 – 700 ms time window ($p = .011$), but not in any of the following two 50-ms time windows. Due to the lack of continuation to meet the selection criteria (i.e., significant in more than two consecutive 50-ms analyses), I did not take this interaction into consideration.

Sem $\times$ Reg $\times$ Hemis	2.40	0.59	0.130	3.88	1.22	0.056
Syn $\times$ Reg $\times$ Hemis	0.37	0.18	0.549	0.03	0.02	0.854
Sem $\times$ Syn $\times$ Reg	0.87	1.16	0.358	1.14	2.32	0.293
Sem $\times$ Syn $\times$ Hemis	2.36	1.40	0.133	0.01	0.01	0.931
Sem $\times$ Syn $\times$ Reg $\times$ Hemis	3.46	0.78	0.071	0.14	0.07	0.713

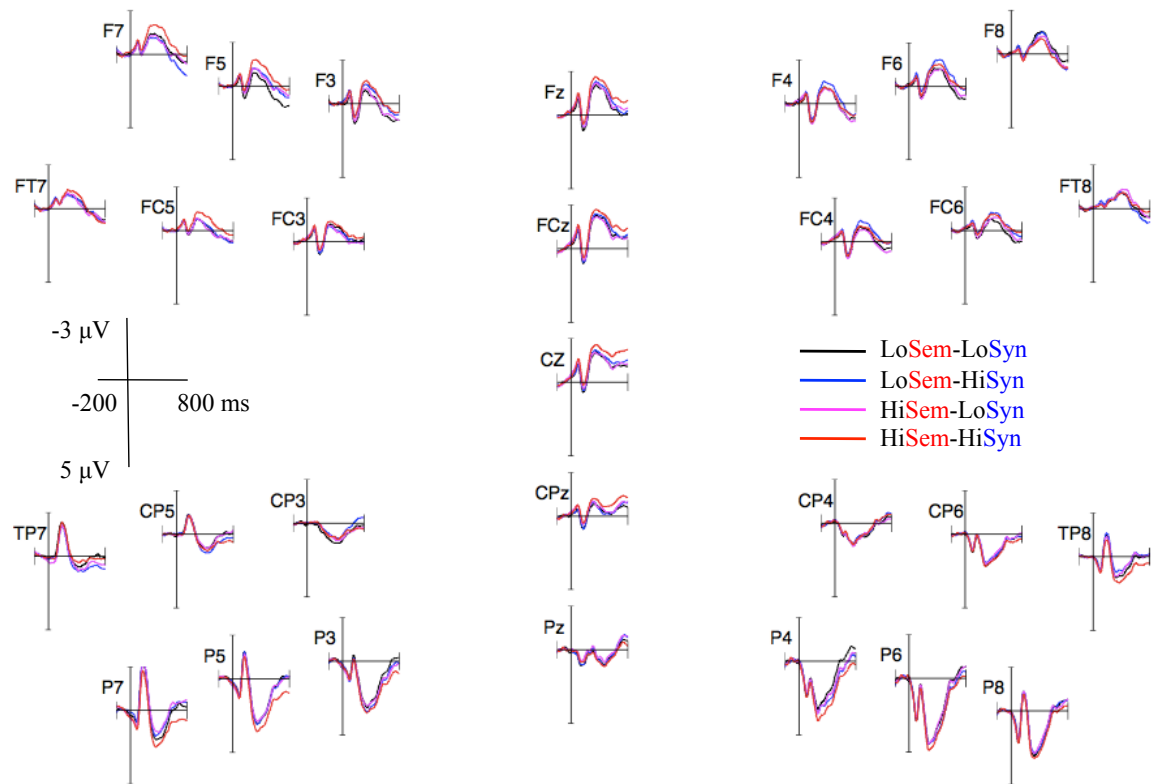


Figure 8 Grand average ERPs for each experimental condition time-locked to the onset of the head noun of the RC (e.g., “resident” in the Example 13) from -200 to 800 ms, at all electrodes included in the statistical analyses (negativity is plotted up).

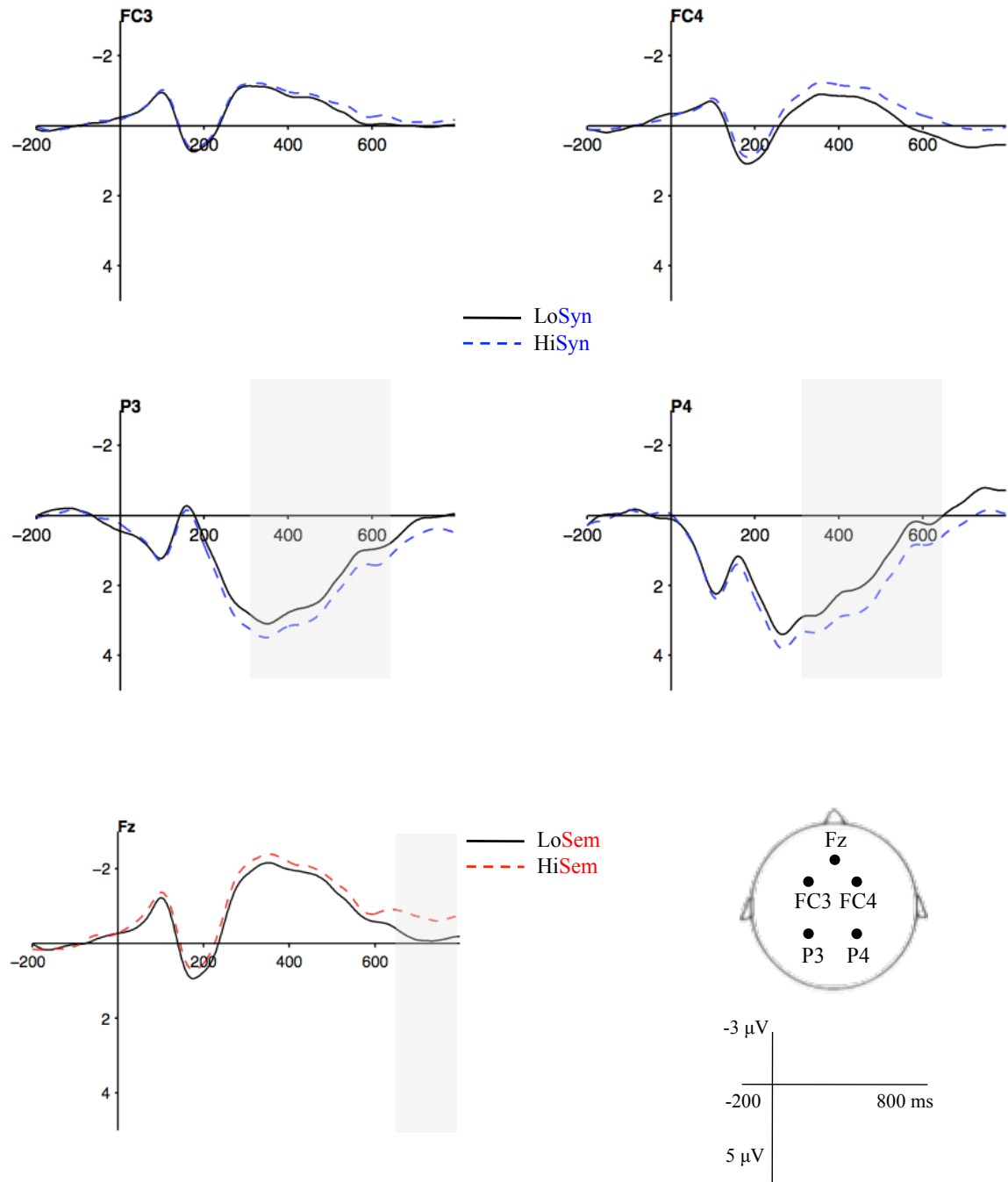


Figure 9 Grand average ERPs for syntactic interference effect (averaged across semantic interference conditions) and semantic interference effect (averaged across syntactic interference conditions) time-locked to the onset of the RC head noun from -200 to 800 ms, at electrode position FC3, FC4, Fz, P3, and P4 (negativity is plotted up).

3.2.4.2 ERP effects at the adverbial phrase

Based on visual inspection and the consecutive 50-ms latency window analyses, only one time window was identified in the adverbial phrase region: 150 – 400 ms. The results are shown in Table 8 (see also Fig. 10 and 11). The ERP differences, if there were any, observed on the adverbial phrase should be a function of the incomplete processing from the head noun since the words in this region were the same across all the four conditions and little retrieval happened in this region according to the cue-based parsing approach (though subjects might make predictions about the incoming verb which is modified by the adverbial phrase). In consistence with previous analyses, we referred to the observed differences between different types of sentences as “semantic interference” and “syntactic interference”. In the 150 – 400 ms time window, there was a significant interaction of Syntactic $\times$ Region $\times$ Hemisphere in the lateral electrodes, $F(1, 36) = 9.54$, $MSE = 1.53$, $p = .004$. Then the Syntactic $\times$ Hemisphere interaction was examined in the anterior and posterior regions separately. The interaction was not significant in the anterior region ($F < 1$). However, there was a marginally significant interaction of Syntactic $\times$ Hemisphere in the posterior region, $F(1, 36) = 3.39$, $MSE = 0.71$, $p = .074$. As shown in Fig. 11, there seemed to be a trend of the HiSyn condition being more positive than the LoSyn condition in the left posterior region (HiSyn: 0.65 μV vs. LoSyn: 0.52 μV ; $F(1, 36) = 1.84$, $MSE = 0.30$, $p = .18$), but being less positive in the right posterior region (HiSyn: 1.54 μV vs. LoSyn: 1.69 μV ; $F(1, 36) = 1.50$, $MSE = 0.42$, $p = .23$). In sum, on the adverbial phrase, there was a three-way interaction of Syntactic $\times$ Regions $\times$ Hemisphere between 150 – 400 ms. There seemed to be a HiSyn-elicited

positivity in the left posterior region and a LoSyn-elicited positivity in the right posterior region, though none of the comparisons reached significant.

Table 8 Results of repeated ANOVAs for the mean amplitude of the adverbial phrase.

Time window	150 – 400 ms		
	<i>F</i>	<i>MSE</i>	<i>p</i>
<u>Midline</u>			
Semantic	0.00	0.01	0.956
Syntactic	0.38	0.45	0.542
Semantic × Syntactic	2.19	3.64	0.148
Semantic × Region	1.73	2.42	0.179
Syntactic × Region	0.10	0.15	0.921
Sem × Syn × Reg	0.19	0.25	0.829
	<i>F</i>	<i>MSE</i>	<i>p</i>
<u>Lateral</u>			
Semantic	0.22	0.04	0.643
Syntactic	1.71	0.21	0.199
Semantic × Syntactic	0.68	0.10	0.416
Semantic × Region	4.29	4.51	0.045* ³²
Syntactic × Region	0.41	0.35	0.528
Semantic × Hemisphere	1.22	0.61	0.277
Syntactic × Hemisphere	0.31	0.20	0.579
Sem × Reg × Hemis	0.51	0.06	0.478
Syn × Reg × Hemis	9.54	1.53	0.004*
Sem × Syn × Reg	1.06	1.03	0.310
Sem × Syn × Hemis	0.06	0.02	0.810
Sem × Syn × Reg × Hemis	0.09	0.02	0.762

³² In the lateral electrodes analyses between 150 – 400 ms, the interaction of Semantic × Region was only significant in the 200 – 250 ms ($p = .035$) and 250 ms -300 ms ($p = .04$) time window, but not in the other two time windows ($p > .1$). Due to the lack of continuation to meet the selection criteria (i.e., significant in more than two consecutive 50-ms analyses), I did not take this interaction into consideration.

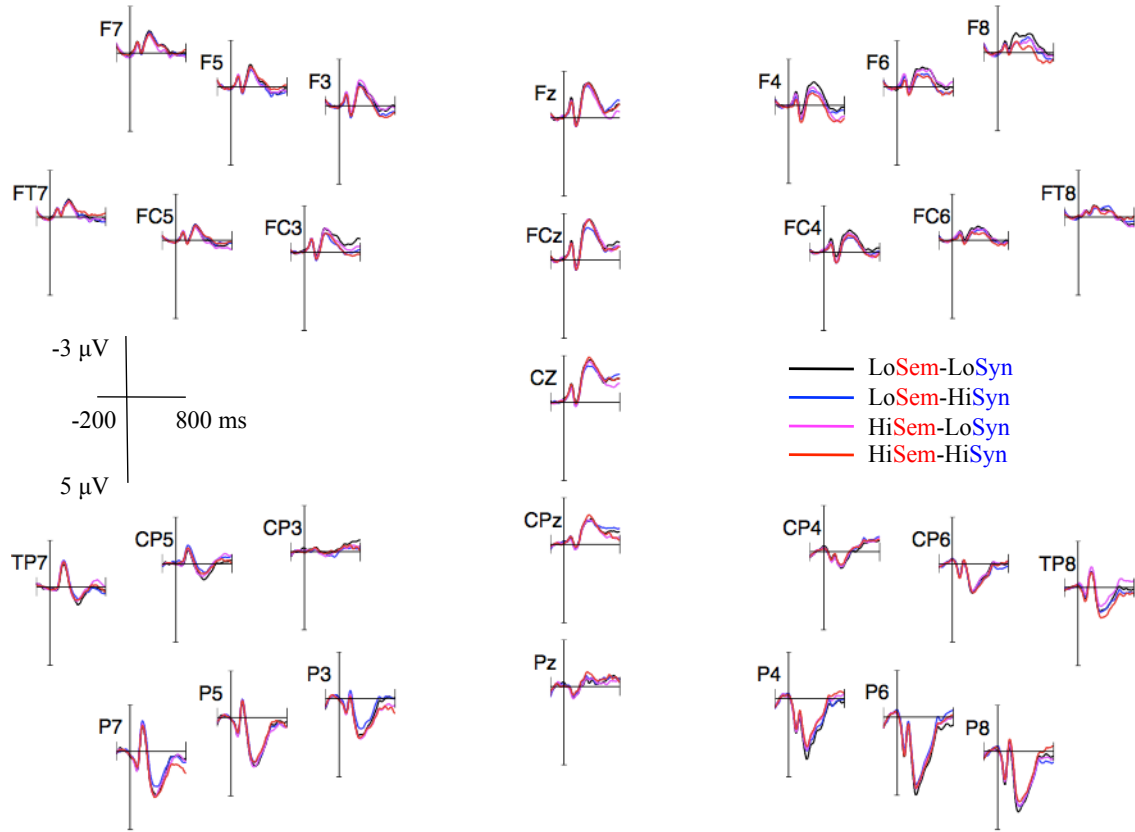


Figure 10 Grand average ERPs for each experimental condition time-locked to the onset of the adverbial phrase before the critical verb (i.e., "everyday" in the Example 13) from -200 to 800 ms, at all electrodes included in the statistical analyses (negativity is plotted up).

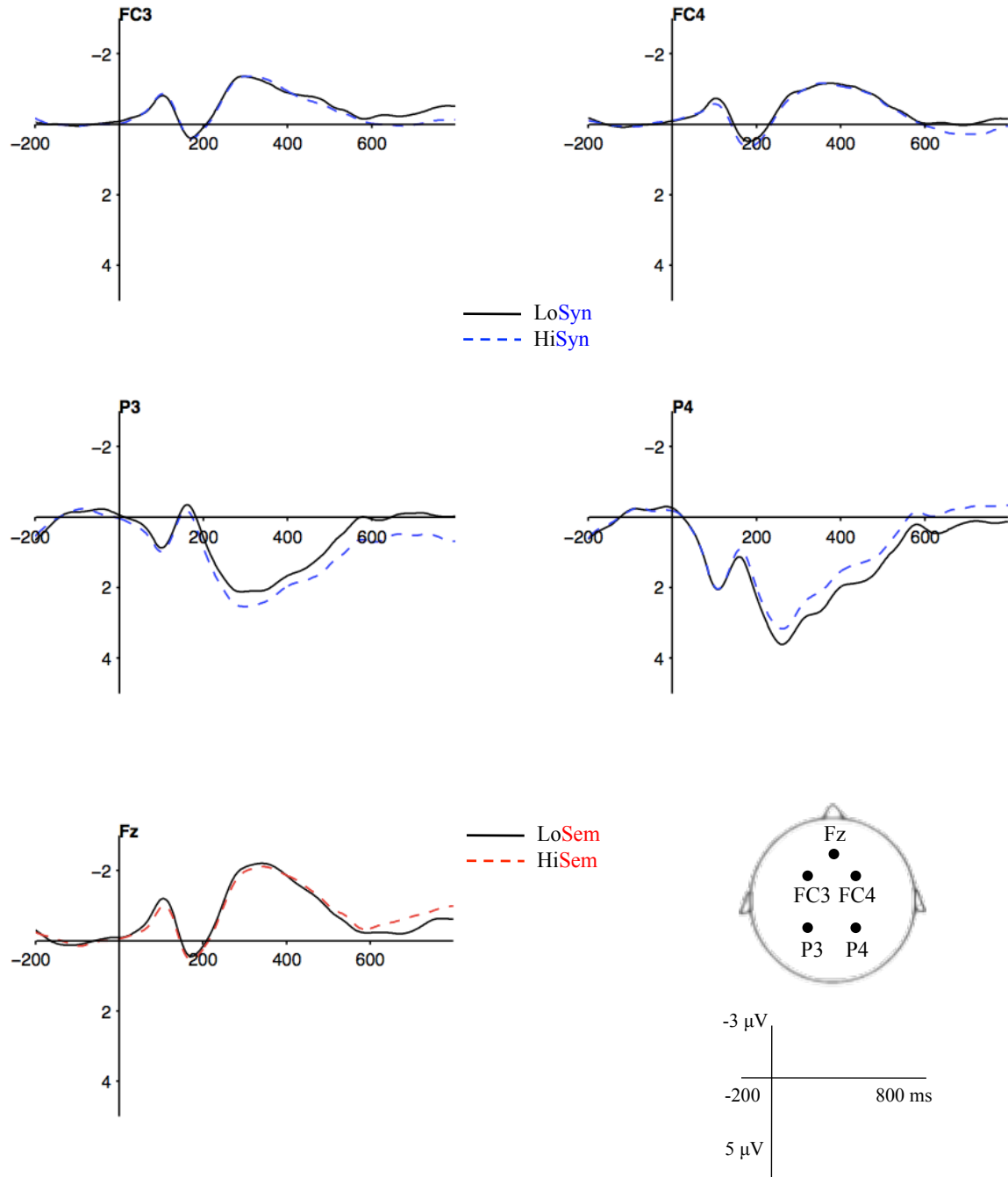


Figure 11 Grand average ERPs for syntactic interference effect (averaged across semantic interference conditions) time-locked to the onset of the adverbial phrase before critical verb from -200 to 800 ms, at electrode position Fz, FC3, FC4, P3, and P4 (negativity is plotted up).

To sum up, for the RC-related processing, on the head noun, I observed an ORC-elicited positivity in the posterior regions (300 – 650 ms), and a high semantic

interference-elicited late negativity along the midline (650 – 800 ms). This posterior positivity was not consistent with Yang et al.'s study (2010), in which they observed an ORC-elicited sustained negativity (250 – 800 ms) at four frontal and central sites (i.e., F3, F4, Fz, Cz), nor with Pakard et al.'s study (2011), in which they observed a SRC-elicited P600. It is hard to reconcile the current study with these two studies, as there were many differences in the experimental materials. However, the current results are in line with the finding from Jäger et al.'s self-paced reading experiment (2015) using similar materials. Jäger et al. observed a marginally significant SRC-advantage on the head noun ($t = 1.97$ in the mixed-effects model analysis, with $t \geq 2$ as a significance criteria). However, Jäger et al. did not further discuss this SRC-advantage on the head noun. They suggested that this marginal effect might not be reliable as it was only observed in the first-pass reading time using the eye-tracking technique but not in the self-paced reading experiment. In addition, they argued that the SRC-advantage was only found in the object-modifying sentences but not in the subject-modifying sentences posed a problem to the interpretation as well. The interaction of modifying types (object- vs. subject-modifying) and RC types (SRC vs. ORC) could be not explained by either WM-based (an ORC-advantage is predicted regardless of modifying types), or experience-based (an SRC-advantage is predicted regardless of modifying types) account. However, I suggest that this SRC-advantage could be potentially interesting for the purpose of testing the debates between WM- and experience-based accounts of sentence processing. First, the eye-tracking measure is more sensitive than the self-paced measure in reflecting moment-to-moment cognitive process (Rayner, 1998). The SRC-advantage might only show up in the early stage of sentence processing. Second, the interaction of modifying types and RC

types has been reported in many Chinese studies (Hsiao, 2003; Lin, 2006; Packard et al., 2011), indicating that the processing of subject- and object-modification sentences differed at least in time course. In two self-paced reading experiments, Lin (2006) and Hsiao (2003) found an SRC-advantage on the relativizer “*de*” and the head noun following an object-modifying RC, whereas no SRC-advantage was found on either relativizer “*de*” or the head noun following the subject-modifying RCs. In an ERP study, Packard et al. found that the SRC-elicited P600 occurred on the relativizer “*de*” in the subject-modifying sentences, whereas on the head noun in the object-modifying sentences. Therefore, Packard et al. suggested that the filler-gap integration happened earlier in the subject-modifying sentences, as in Chinese it is very common to leave out the head noun in the subject-modifying sentences without affecting the grammaticality. As a result, subjects might start the integration immediately when “*de*” appeared. In the current experiment, by using the ERP technique with high temporal sensitivity compared to self-paced reading, I observed this SRC-advantage on the head noun as in Jäger et al.’s eye-tracking experiment, and Lin’s and Hsiao’s self-paced reading experiment. I suggest that this result is worthy of further discussion.

Given the time course and its posterior scalp distribution, I suggest that this might be a P300 effect, which has been associated with detecting temporary ambiguity in garden-path sentences³³ (Friederici et al., 2001) or referential binding ambiguity (Heine et al., 2006; Li & Zhou, 2010), and was sensitive to word/structure frequency in previous studies (Heine, Tamm, Hofmann, Hutzler, & Jacobs, 2006; Polich & Donchin, 1988).

The SRC-advantage seems more consistent with the experience-based account, which

³³ Friederici et al., (2001) investigated the ERPs related to subject-object ambiguous in German relative and complement clauses. In German, both SRC and ORC was analyzed as subject-first structures. Thus, ORC structure requires a revision at later part of the sentence.

predicted more integration difficulty in the ORC sentences on the head noun due to low frequency of the ORC structure. Additionally, the late negativity elicited by the high semantic interference condition confirmed previous findings from both eye-tracking and ERP experiments showing that subjects were sensitivity to the semantic cues during RC processing (Traxler, Morris, & Seely, 2002; Traxler, Williams, Blozis, & Morris, 2005; Weckerly & Kutas, 1999). When subjects need to complete the filler-gap dependency at the head noun, they had more difficulty in the high semantic interference conditions (i.e., both head noun and embedded noun were human beings) compared to the low semantic interference condition when the embedded noun was an inanimate NP, because the inanimate NP reduced the similarity-based confusion. In the high semantic interference conditions, both noun phrases are a plausible agent (or patient) of the embedded verb (e.g., as shown in Example 13, both “*neighbor*” and “*resident*” can be the agent of the embedded verb “*hit*” in the low syntactic interference condition, or “*hurt*” in the high syntactic interference condition). However, in the low semantic interference conditions, in most of the experimental sentences, the thematic roles of the head noun and embedded noun were not reversible. For instance, as shown in Example 13, only the “*resident*” could “*hit*” the “*wall*”, while only the “*wall*” could “*hurt*” the “*resident*”, but not the reverse situation³⁴. Therefore, the semantic interference effects observed on the head noun might reflect RC integration difficulty in the conditions when there was semantic confusion. However, an alternative explanation could be that the semantic interference effect might not be caused by the difficulty in assigning thematic roles related to the

³⁴ Although after translating the sentences into English, the “*the wall hit the resident*” might be a possible situation, in the original Chinese sentence, the word correspond to “*hit*” is a compound word and means “*move and hit*”. The “*wall*” cannot move. Thus, the reserved situation was not a possible one. For the word “*hurt*”, in Chinese, it could only take a human being as the object.

embedded verb, instead, it might be caused by the difficulty in encoding the head noun (e.g., “*resident*” in Example 13) when there was proactive interference from item(s) sharing semantic features (e.g., “*neighbor*”). Some researchers have proposed that extended overwriting happens among items that have to be held concurrently in WM (Oberauer, 2009; Oberauer & Kliegl, 2006). However, the current experiment was not design to distinguish these two possibilities.

Importantly, although these ERP effects associated with RC type differences observed on the head noun spilled-over to the following adverbial phrase, as a three-way interaction of Syntactic $\times$ Hemisphere $\times$ Region between 150 – 400 ms, there was neither a main effect nor an interaction of semantic and syntactic interference from 400 ms prior to the onset of the critical word (or 400 ms after the onset of the adverbial phrase). Thus, the ERP effects observed on the critical verb should not be simply accounted for as a spillover effect from RC processing differences.

3.2.5 Summary of ERP effects during sentence processing

The ERP effects observed in the five sentence regions (i.e., critical verb, spillover region, sentence final position, RC head noun, and adverbial phrase) are summarized in Table 9. Regarding semantic and syntactic interference, the results aligned with the predictions from cue-based parsing approach that both semantic and syntactic interference impair sentence processing through impeding retrieval. Importantly, the current experiment demonstrated that both semantic and syntactic interference occur immediately at the critical verb, where retrieval of earlier information (i.e., grammatical subject) is supposed to be triggered in order to complete the long distance subject-verb

dependency. I suggest that the ANs elicited by both semantic and syntactic interference reflected greater WM demands associated with detecting multiple representations, while the late negativity elicited by semantic interference and the P600-like effect elicited by syntactic interference reflected later revision processes unique to semantics and syntax. The processing difficulty caused by the high syntactic interference condition spilled-over to the first word following the critical verb. In this post-critical verb region, a P600-like effect was elicited by the syntactic interference effect in the low semantic interference condition. This might reflect processing costs associated with continued syntactic interference resolution. The lack of any syntactic interference effect in the high semantic interference condition might be attributed to the fact that subjects temporarily consider the animate distractor in the HiSem/HiSyn condition as the grammatical subject of the main verb and thus, did not experience syntactic interference until revising their initial false interpretation at a later point. In the sentence final region, a complicated pattern of ERP effects was observed with hemispheric asymmetries, which might reflect multiple processing difficulties during the wrap-up process. The ERPs in the sentence final region will not be further discussed due to the puzzling nature of the wrap-up effect.

Regarding the RC processing, on the RC head noun, an ORC-elicited positivity was observed in the 300 – 650 ms time window, which might reflect binding difficulty due to low frequency of the ORC structure. In addition, a late negativity elicited by the semantic interference effect was observed along the midline between 650 – 800 ms, indicating that subjects had more difficulty in parsing the RC head noun when there was proactive semantic interference from an animate NP in the RC region. In the adverbial phrase region immediately following the head noun, there was an interaction of Syntactic

× Hemisphere × Region in an early time window (150 – 400 ms), though none of the further comparisons was close to significance. The fact that no ERP effect was observed during the 400 – 800 ms time window ensured that the ERP effects elicited by RC processing differences did not spillover to the critical verb and confound the interpretation of the interference effect observed in that region.

Table 9 Summary of the time course and scalp distribution of each reliable ERP component on the critical verb, spillover region, and sentence final position.

Region	Interference	Time window	
		300 – 500 ms	650 – 800 ms
Critical verb	Semantic	Negativity (anterior)	Negativity (anterior)
	Syntactic	Negativity (left anterior)	Positivity (midline)
250 – 800 ms			
Spillover region	Syntactic (LoSem)	Positivity (midline)	
Final	Semantic	N/A	N/A
	Syntactic	Negativity (right anterior) Negativity (elicited by LoSyn condition; left anterior)	Negativity (right hemisphere)
300 – 650 ms			
RC head noun	Semantic	N/A	Negativity (midline)
	Syntactic	Positivity (posterior)	
150 – 400 ms			
Adverbial phrase	Semantic	N/A	
	Syntactic	Syntactic × Hemisphere × Region	

3.3 Relations between individual differences measures and interference effect in ERPs

One of the main goals of the current project was to explore the nature of the underlying WM mechanisms involved in sentence comprehension through examining how the different components of WM or EF capacities modulate the mean amplitude of the ERP effects elicited by semantic or syntactic interference effects. Therefore, I tested subjects on a set of ten individual differences measures, then related their performance to each significant ERP effects as summarized in Table 9, except for the sentence final region. Based on the results from our previous studies (Tan et al., 2011, 2013), I predicted a specific relation between the magnitude of the semantic interference effect (as indexed by the mean amplitude difference of the corresponding ERPs) and semantic STM, and between the magnitude of the syntactic interference effect and general WM or attentional control ability, even after controlling for individuals' linguistic experience and general processing speed.

3.3.1 Relations between the individual differences measures

Each subject was tested on ten individual differences measures, including reading span, operation span, digit span, category probe, author recognition test (ART), vocabulary, recent negatives, verbal Stroop, rapid automatic naming (RAN), and lexical decision. As introduced in the procedure section, wherever possible, I chose standardized measurements and scoring methods that are well established in previous psychometric studies to obtain high validity and reliability. Range, mean, and standard deviation for each individual differences measure are presented in Table 10. Although subjects generally performed well in most tasks, their scores were distributed widely on each scale.

Reliabilities of each task are reported as the extent of relation between two variables is limited by the reliability of the measures involved (Schmitt, 1996). Internal reliability was calculated as the split-half correlation adjusted with the Spearman-Brown prophecy formula (Cronbach, 1951). For operation span and reading span, the internal reliability was obtained from previous studies (Redick et al., 2012). As shown in Table 10, all the individual differences measures had high reliability and showed a close to normal distribution. It should be noted that, however, as most of the previous ERP studies investigating the individual differences in ERP effects, I did not calculate the split-half reliability of the ERP signals based on the single-trial activity because of the low signal-to-noise ratio in the single-trial EEG data. Recently, researchers are encouraged to address the validity and reliability of the EEG data, especially when special analytic procedures are employed (e.g., special temporal or spatial filters) (Keil et al., 2014). Nonetheless, it has been demonstrated that the averaged EEG data, which was analyzed following standard procedures, is generally reliable (Lopez-Calderon & Luck, 2014; Picton et al., 2000) and has high test-retest reliability for accessing changing in cognitive states such as WM demands (McEvoy, Smith, & Gevins, 2000). Therefore, as the current study has followed the standard procedure for analyzing EEG data, I proceeded to investigate the correlation between the ERP effects and the ten reliable individual differences measures.

Table 10 Descriptive data and reliability estimates for all the individual differences measurements

Individual differences Measures	Index	Mean	Range	SD	Skewness	Kurtosis	Reliability
Operation span	Total score	65/75	43 – 75	7.7	-0.91	0.83	0.84 ^{b****} (Cronbach's α)
Reading span	Total score	61/75	38 – 73	8.2	-0.62	0.35	0.86 ^{b****}

							(Cronbach's α)
Digit span	Accuracy	0.80	0.44 – 1.00	0.02	-0.95	1.31	0.68**
Category probe	Accuracy	0.83	0.57 – 0.97	0.01	-0.51	0.75	0.84**
Stroop	RT/Accuracy	62.5	(-28.3) – 257.7	60.5	0.97	1.71	0.56*
Recent Negatives	RT/Accuracy	223.0	(-193.26) – 751.9	189.6	0.63	1.03	0.68**
Lexical decision	RT (ms)	725	590 - 1078	115	1.32	1.61	0.99**
Vocabulary	Score	52/58	34 - 58	4.4	-1.88	5.72	0.61**
ART	d'	1.94	0.6 – 3.5	0.66	0.02	0.01	0.91**
RAN	RT (ms)	10.07	7.64 – 14.01	1.48	0.59	0.00	0.88**

Note: a) * $p < .05$. ** $p < .01$. *** $p < .001$.

b) The internal reliability for operation span and reading span were reported by Redick et al. (2012) (Redick et al., 2012). The reading span we used was a direct translation into Chinese. Some sentences have been slightly modified but this should not affect the reliability very much.

The correlations among the individual differences measures are displayed in Table 11. As shown in the table, reading span, operation span, and digit span were all significantly related to each other. This was consistent with previous findings that the two complex span tasks have a high correlation (Redick et al., 2012; Nash Unsworth, Redick, Heitz, Broadway, & Engle, 2009), and there is a phonological component to these working memory measures (Camos, Mora, & Barrouillet, 2013; Camos, Mora, & Oberauer, 2011). The high correlations between RAN with the three memory tasks were in line with many findings that processing speed is a general characteristic that influences other abilities (Fry & Hale, 1996; Salthouse, 1996). For example, there are studies showing that phonological/orthographic decoding speed as measured by the RAN task predicts individuals' eye movement efficiency during sentence reading (Gordon et al., 2013; Van Dyke & Kuperman, 2011). Some researchers have even argued that processing speed differences can account for the correlation between WM and other capacities (Evans et al., 2014; Salthouse, 1996). However, it is worth noting that although there are high correlations among processing speed and WM tasks, there have been numerous

previous studies showing that general processing speed does not fully account for the relation between WM and other higher cognitive abilities, such as intelligence and reading (Baddeley, Logie, Nimmo-Smith, & Brereton, 1985; Conway, Cowan, Bunting, Theriault, & Minkoff, 2002; Engle, Cantor, & Carullo, 1992). As expected, there was a significant correlation between RAN and lexical decision performance, as both tasks tap processing speed. Although one might have expected that RAN would be related to the linguistic experience measures, such as vocabulary, the fact that subjects were naming the same items repeatedly and all were of high frequency, would tend to diminish any lexical effect such as word frequency, to performance on the RAN task.

Table 11 Full correlation matrix for the individual differences measures

Individual differences Measures	Operation Span	Reading span	Category probe	Digit span	Stroop	Recent Negatives	Lexical decision	Vocab	ART
Operation span									
Reading span	0.761**								
Category probe	0.224	0.244							
Digit span	0.585**	0.649**	0.208						
Stroop	0.007	0.218	-0.084	0.258					
Recent Negatives	-0.185	-0.078	-0.087	-0.139	-0.051				
Lexical decision	-0.269	-0.193	-0.082	-0.125	0.308	0.307			
Vocabulary	0.238	0.206	0.311	0.174	-0.019	0.144	-0.063		
ART	0.525**	0.430**	0.047**	0.280	-0.106	-0.230	-0.299	0.416*	
RAN	-0.378*	-0.358*	0.031	-0.390*	0.053	0.039	0.387*	0.004	0.022

Note. a) * $p < .05$.

** $p < .01$.

b) The WM composite variable for WM capacity was calculated by combining z -scores for reading span and operation span.

It was less expected that neither the Stroop nor the recent negatives task was related to the two complex span tasks, since some memory studies have found that an attentional control component (e.g., goal maintenance ability in the Stroop task and proactive interference resolution ability in recent negatives task) plays a critical role in complex span measures (Engle, 2002; Kane, Bleckley, Conway, & Engle, 2001; Unsworth, Fukuda, Awh, & Vogel, 2014). The lack of correlation between general executive control ability and WM capacities might be attributed to the limited sample size I had ($N = 37$), and the relatively high proportion of incongruent trials (50%) in the Stroop task used here. Kane and Engle (2001) found that high- and low- WM subjects did not differ on the Stroop task in the 50%- or 100%- incongruent conditions, but only in a 25%-incongruent condition, where the context makes it difficult to maintain the task goal. The lack of a correlation between Stroop and recent negatives was consistent with Friedman and Miyake's claim that prepotent response inhibition (e.g., as measured by Stroop) is different from resistance to proactive interference (e.g., as measured by recent negatives) (Friedman & Miyake, 2004). In a recent study with both young and aging subjects, Crowther and Martin (2014) also found no correlation between subjects' performance in the verbal Stroop and recent negatives task ($r = .024, p = .83$), and neither task was related to WM capacity ($p > .20$). Moreover, by relating subjects' executive control to a semantically blocked cyclic naming task³⁵, they found that subjects' ability to inhibit a prepotent response (i.e., Stroop effect) predicted their performance in the related

³⁵ In the semantically blocked cyclic naming task, subjects are instructed to repeatedly name pictures in sets of either semantically related or unrelated pictures in different blocks. Typically, naming latencies are longer for the pictures from the semantically related sets than unrelated sets. Crowther and Martin found that they found that subjects with better ability to inhibit a response (i.e., smaller Stroop effect) showed less interference in the related conditions, while subjects with better resistance to proactive interference (i.e., smaller recent negatives effect) showed more change in naming latencies (i.e., smaller repetition priming) for the unrelated conditions

conditions, while their resistance to proactive interference predicted their performance for the unrelated conditions (Crowther & Martin, 2014). Therefore, the dissociation between Stroop and recent negatives tasks observed in the current experiment was in line with this previous study, as the two executive function tasks had different predictive power for different kinds of interference effects.

With respect to the ART and category probe tasks, both of which were created for the purpose of the current experiment and have not been used in previous Chinese studies, it was interesting that the ART was highly correlated with several other tasks (i.e., operation span, reading span, category probe, and vocabulary), whereas category probe was only related to ART. In a previous study with a larger sample size (e.g., $N = 112$; Tan et al., 2011), we found significant correlations between category probe span and reading span ($r = .33$), operation span ($r = .20$), and vocabulary ($r = .21$). As shown in Table 11, there was a trend of showing all these three correlations in the current experiment with similar sizes of correlations, though the limited sample size most likely constrained the power to detect significant relations. It should be noted that there was a potential problem with the Chinese version of the category probe task – the majority of the words we used in this experiment were compound words, in which two or three meaningful characters were melded together, while a large portion of the words under same category shared the same character representing that category. For example, under the category of “trees”, among the 24 names of different trees, 19 out of them contained the exact morpheme “tree” in the words, such as “苹果树 [apple-tree]”, “柳树 [willow-tree]”, etc. It was the similar issue for the category of “flowers”, “insects” and “weather”. This was unavoidable to some extent, as Chinese has been described as a language of

compound words with more than 70% of words are compound words (Zhou, Marslen-Wilson, Taft, & Shu, 1999). The shared morpheme within each category provided additional cues to subjects' probe judgment. However, the relative high r -value between vocabulary and category probe performance ($r = .31$; though not significant) and the relatively low r -value between digit span and category probe ($r = .21$) observed here, indicated that linguistic experience, rather than phonological cues might play a larger role in this Chinese version category probe task. However, it is hard to make a clear conclusion here due to the limited sample size. Overall, although the category probe task had a very high split-half reliability, it may not have a high validity of specifically measuring subjects' semantic STM as expected. Regarding the ART, most of the previous studies have focused on its relation with linguistic experience (e.g., as measured by self-report questionnaires) and general reading abilities (e.g., verbal ACT in Acheson et al., 2008, and eye-movement efficiency during sentence processing in Moore & Gordon, 2014), while very few studies have investigated its relation to other cognitive abilities. The correlation between ART and operation span, reading span, category probe, and vocabulary, but not the tasks which measured processing speed (i.e., RAN and lexical decision) or executive control (i.e., Stroop and recent negatives tasks), might be explained by the fact that the secondary memory retrieval component as tapped by the ART task plays an important role complex memory span measures (Shipstead et al., 2014; Unsworth et al., 2014) and long-term linguistic experience measures. The high collinearity between ART and many other tasks might cause difficulty in interpreting the results from multiple regression models including all the individual differences measures simultaneously. This will be discussed in section 3.3.4.

3.3.2 Compute composite scores

The observation that several of the individual differences measures were moderately or highly related to each other indicated that although some of the individual differences measures included in this experiment captured different aspects of individuals' cognitive abilities, some tapped similar underlying constructs. Thus, in order to reduce the problems of multicollinearity and overfitting caused by including too many correlated measures in regression model and obtain a more reliable measures for a given construct (Nunnally, Bernstein, & Berge, 1967; Saffran & Martin, 1997), I computed composite scores for general WM capacity, linguistic experience, and processing speed by averaging *z*-scores of certain individual difference measures, based on hypotheses from prior studies and the correlation matrix as shown in Table 11. The composite score for general WM was calculated by averaging across the *z*-scores of reading span, operation span, and digit span tasks. Although there were correlations between rapid naming task and the three memory tasks, the correlations between RAN and the memory span tasks ($r_s < .39$) were weaker than the correlations among the memory span tasks ($r_s > .59$). Unsworth et al. (2009) have shown that processing efficiency (e.g., sentence processing speed in the reading span task) has a significant influence on but does not fully mediate the relation between memory capacity and higher cognition. As I did not wish to emphasize the processing speed component of WM, I included RAN and lexical decision as a separate factor tapping processing speed in further analyses, and expected that the correlation between WM capacity and interference resolution efficiency, if there were any, would persist even after partialling out the variance shared with processing speed in the regression model. The composite score for linguistic experience and lexical

representation quality was computed by averaging across the z-score of ART, vocabulary, and category probe tasks. The z-scores of the Stroop and the recent negatives tasks were included by themselves as each of them addressed a relative unique construct. The correlations between the individual differences factors, including both composite scores and individual measures, are shown in Table 12. After computing the composite scores, there was a significant correlation between WM composite and linguistic experience composite ($r = .41, p = .011$). As discussed earlier, this could be explained by the fact that there is a LTM component of the WM measures. For example, the secondary memory retrieval is affected by the quality of long-term knowledge representations (Shipstead et al., 2014; Unsworth et al., 2014). In addition, subjects' processing speed composite was negatively related to their WM composite score ($r = -.39, p = .017$), consistent with previous findings that faster processing speed relates to higher WM capacity.

Table 12 Full correlation matrix of the correlation tests between factors

	WM comp.	Linguistic experience comp.	Processing speed comp.	Recent negatives
WM comp.				
Linguistic experience comp.	.412*			
Processing speed comp.	-.389*	-.100		
Recent negatives	-.152	-.075	.161	
Stroop	-.183	.090	.216	-.051

Note. * $p < .05$. ** $p < .01$.

3.3.3 Simple correlations between individual differences measures and semantic and syntactic interference resolution

Before running multiple regression models to examine the unique contribution of each individual differences factor to the interference resolution, simple correlations were conducted to examine the general correlations between each task included in the present project and subjects' semantic and syntactic interference resolution efficiency. Given that these analyses were intended to determine if collinearity had hidden the contribution of some of the variables in later regression models, a liberal criterion for significance was used without correction for multiple comparisons.

For correlational analyses, the mean amplitude of each significant ERP effect (as summarized in Table 9) elicited by either semantic or syntactic interference effect was related to the each of the ten individual differences measures. That is, for each subject, I calculated a difference score of mean amplitude between the high and low semantic interference conditions, and between the high and low syntactic interference conditions. Then these difference scores were correlated with each of the individual differences measures. The results of the correlation matrix are shown in Table 13, and the scatter plots of the significant correlations are shown in Fig. 12.

Table 13 Correlations between interference effect size on the critical verb and spillover region (or SRC-advantage effect on the head noun) and composite score of individual difference measures.

IDs	Critical				Spillover	Head noun	
	Semantic		Syntactic		Syntactic	Semantic	Syntactic
	Negativity (Anterior) 300 – 500 ms	Negativity (Anterior) 650 – 800 ms	Negativity (Left anterior) 300 – 500 ms	Positivity (Midline) 650 – 800 ms	Positivity (Midline) 250 – 800 ms	Negativity (Midline) 650 - 800 ms	Positivity (Posterior) 300 – 650 ms
Category probe	-0.081	0.138	-0.172	-0.033	0.265	0.23	-0.208
Digit span	-0.174	0.232	-0.126	0.025	-0.007	0.109	-0.015
Reading span	0.113	0.296	0.124	-0.077	0.112	0.275	0.018
Operation span	0.039	0.272	0.126	0.029	-0.012	0.149	0.196
Stroop	-0.042	0.096	-0.041	-0.188	-0.352*	-0.051	0.229
Recent negatives	0.189	0.061	0.268	0.349*	0.305	0.007	-0.117
Vocabulary	-0.035	0.187	0.014	0.024	0.21	0.027	0.032
ART	0.001	0.087	-0.095	-0.143	0.200	0.258	0.165
RAN	0.221	-0.041	0.031	-0.074	-0.198	-0.088	-0.052
Lexical decision	0.100	0.148	0.107	0.013	-0.164	-0.036	0.094

Note. a) * $p < .05$.

b) The correlations between ERP effects on the head noun (i.e., SRC-advantage effect and semantic interference effect) and individual differences factors are also presented in this table, but will be discussed in later section 3.3.5.

As shown in Table 13, on the critical verb, there was a positive correlation between the recent negatives effect size and the magnitude of the P600 effect elicited by the high syntactic interference condition ($r = .349, p = .034$), even after controlling for individuals' linguistic experience and processing speed ($r = .357, p = .035$)³⁶. Subjects with better resistance to proactive interference as indexed by the recent negatives task (i.e., smaller effect size) had less difficulty in resolving syntactic interference on the critical verb (i.e., smaller mean amplitude for the P600 effect). In the spillover region, there was a similar correlation between the recent negatives effect size and the P600-like sustained positivity elicited by the syntactic interference effect in the low semantic interference conditions. This correlation was marginally significant ($r = .305, p = .066$) but became significant after controlling for linguistic experience and processing speed ($r = .385, p = .022$). In addition, there was a negative correlation between the Stroop effect and this P600-like positivity ($r = -.352, p = .033$), indicating that subjects who were more tolerant of interference from a prepotent response (i.e., had a smaller Stroop effect) actually had more difficulty in syntactic interference resolution, though this correlation became marginally significant after controlling for linguistic experience and processing speed ($r = -.313, p = .067$).

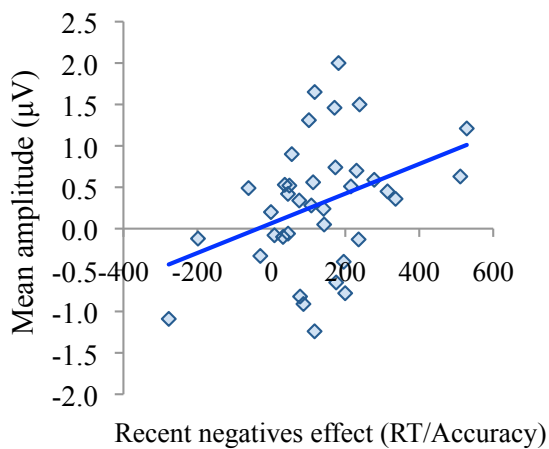
³⁶ For all the correlations between WM measures (i.e., reading span, operation span, digit span) and interference effect size and between EF measures (i.e., Stroop, recent negatives) and interference effect size, I have run partial correlation with linguistic experience composite score (i.e., computed from ART, category probe, and vocabulary) and processing speed composite score (i.e., computed from RAN and lexical decision) as control variables. The pattern of the partial correlation results was similar to the pattern of the simple correlation results. All the non-significant correlations in the simple correlation analyses remained non-significant. Therefore, I only reported the partial correlation results for the selective correlations, which were significant in the simple correlation analyses.

The current results were generally in line with the prediction from accounts that claim that attentional control plays an important role in sentence comprehension (Novick et al, 2005; Vuong & Martin, 2014; Ye & Zhou, 2009). However, these results were not consistent with those from previous studies in our lab, in which Tan et al. (2011, 2013) consistently observed a robust positive link between individuals' semantic STM capacity (as measured by category probe task) and semantic interference resolution ability. The power to detect such a correlation might be constrained by the potential validity problem of Chinese version category probe task and the limited sample size. In addition, the current results did not replicate previous findings of a negative link between general WM capacity and syntactic interference effect size obtained for the healthy young subjects (Tan et al., 2011), or a positive link between Stroop effect size and syntactic interference effect size in the aphasic patients (Tan et al., 2013), both of which indicated that the better the performance (i.e., higher general WM capacity or smaller Stroop effect size) the less difficulty in resolving syntactic interference. However, due to the many differences in the methodology (e.g., self-paced reading vs. ERP technique) and experiment materials (e.g., English vs. Chinese), I suggest that the discrepancies in the correlations obtained from different studies might reflect different WM mechanisms underlying different stages or aspects of interference resolution. The observation of the opposite directions of the correlation of the Stroop-syntactic interference effect and the correlation of recent negatives–syntactic interference effect was consistent with Crowther & Martin's finding (2014) that Stroop effect and recent negatives effect related to different indices of interference resolution in the picture naming task – the Stroop effect was positively related to the change in naming latencies across semantically related

cycles, while the recent negative effect was negatively related with the change in naming latencies across semantically unrelated cycles. However, given that the correlation between Stroop effect and syntactic interference effect became non-significant after controlling for individual's linguistic experience and processing speed, this correlation will not be further discussed.

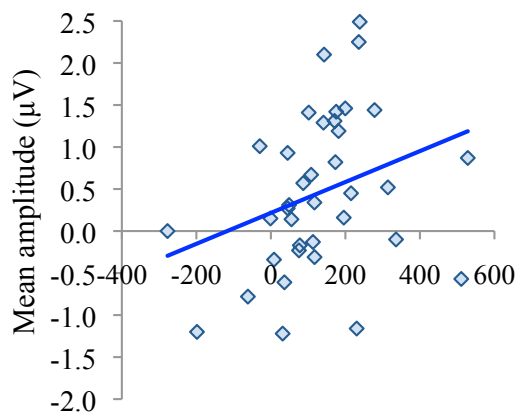
a. Critical verb

Syntactic interference effect

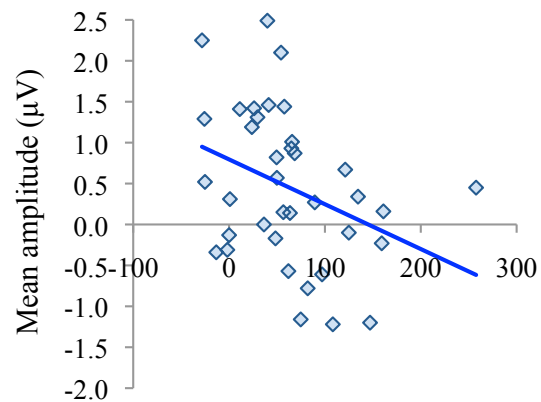


b. Spillover region

Syntactic interference effect (in the low semantic interference condition)



Recent negatives effect (RT/Accuracy)



Stroop effect (RT/Accuracy)

Figure 12 Scatter plots of the significant correlations between the mean amplitude of the ERP effects and individual differences measures. Fig. a shows the relation of recent negatives effect and syntactic interference-elicited P600 in mean amplitude (µV) in the critical verb region (e.g., “complain” in Example 13), collapsed across semantic interference conditions. Fig. b shows in the low semantic interference

conditions, the relation of recent negatives effect/Stroop effect and syntactic interference-elicited sustained positivity in mean amplitude (μV) in the spillover region (e.g., “*rent*” in Example 13).

3.3.4 Multiple regression approach

In order to test the unique contribution of each cognitive factor to subjects’ semantic and syntactic interference resolution ability and avoid a potential problem of using differences scores³⁷, multiple regressions were performed on each ERP component with the mean amplitude in the more difficult condition (e.g., the high syntactic/semantic interference condition) as the dependent measure and mean amplitude in the easier condition (e.g., the low syntactic/semantic interference condition) as well as the five relatively independent individual differences factors as computed in section 3.3.2 (i.e., WM composite, recent negatives effect, Stroop effect, linguistic experience composite, and processing speed composite) as predictors. The regression results are shown in Table 14.

³⁷ There have been a long lasting debate about the appropriateness of using difference scores (or “raw change”, “raw gain”) versus residuals in the analysis of covariance in the study of individual differences (Cronbach & Furby, 1970). The analyses reported above implemented a difference score approach. However, some researchers (Cronbach & Furby, 1970; Lord, 1956) have pointed out that that difference scores tend to be much more unreliable than the scores themselves. Cronbach and Furby (1970) claimed that it is more straightforward to ask about the regression of performance in condition Y (e.g., the high interference condition) on performance in condition X (e.g., the low interference condition) and other predictors of interest. The significance of the regression weights for the predictors of interest indicates how well these measures predict performance beyond that accounted for by the baseline condition. However, some researchers have also argued that difference scores can be reliable under certain conditions (Zimmerman & Williams, 1982; Rogosa, Brandt, & Zimowski, 1982; Rogosa & Willett, 1983; Rogosa, Brandt, and Zimowski, 1982). Because of these debates, I ran multiple regressions on the data to determine whether these two analyses would provide similar results.

Table 14 Multiple regressions on the interference effect and individual differences measures

IDs	Critical								Spillover		Head noun			
	Semantic				Syntactic				Syntactic		Semantic		Syntactic	
	Negativity		Negativity		Negativity		Positivity		Positivity		Negativity		Positivity	
	(Anterior)		(Anterior)		(Left anterior)		(Midline)		(Midline)		(Midline)		(Posterior)	
	300 – 500 ms		650 – 800 ms		300 – 500 ms		650 – 800 ms		250 – 800 ms		650 - 800 ms		300 – 650 ms	
	<i>B</i> -value	<i>p</i> -value	<i>B</i> -value	<i>p</i> -value	<i>B</i> -value	<i>p</i> -value	<i>B</i> -value	<i>p</i> -value	<i>B</i> -value	<i>p</i> -value	<i>B</i> -value	<i>p</i> -value	<i>B</i> -value	<i>p</i> -value
Constant	-0.28	0.024	-0.23	0.042	-0.26	0.056	0.34	0.012	0.30	0.018	0.05	0.782	-0.32	0.048
Low Interf condition	0.90	0.000	0.61	0.000	0.65	0.000	0.92	0.000	0.66	0.000	1.14	0.000	0.98	0.000
WM comp.	0.10	0.500	0.24	0.140	-0.01	0.960	0.15	0.428	0.07	0.687	0.11	0.489	0.03	0.876
Recent negatives	0.07	0.500	0.13	0.259	0.02	0.798	0.29	0.034*	0.36	0.005*	0.04	0.754	-0.06	0.706
Stroop	-0.06	0.616	-0.07	0.544	-0.02	0.825	-0.12	0.404	-0.18	0.176	-0.02	0.855	0.17	0.267
Linguistic exp comp.	-0.05	0.746	-0.18	0.254	0.10	0.433	-0.13	0.482	0.28	0.110	0.17	0.303	-0.01	0.961
Processing speed	0.26	0.254	0.19	0.402	0.00	0.998	0.00	0.997	-0.15	0.535	-0.01	0.964	0.00	0.989

Note. a) * $p < .05$. b) The correlations between ERP effects on the head noun (i.e., SRC-advantage effect and semantic interference effect) and individual differences factors are also presented in this table, but will be discussed in later section 3.3.

As expected, subjects' performance in the low syntactic/semantic interference conditions strongly predicted their performance in the high syntactic/semantic interference conditions ($p < .001$). Generally, the results of the regression models confirmed the results obtained from differences score measures. The only significant weight was for the recent negative effect on the magnitude of the syntactic interference effect during sentence processing. On the critical verb, subjects who showed a larger recent negatives effect tended to elicit a greater P600 effect in the high relative to the low syntactic interference condition, $B = .29, p = .034$. In the spillover region, subjects who showed a larger recent negatives effect tended to elicit a greater sustained P600-like effect (250 – 800 ms) along the midline associated with the syntactic interference effect in the low semantic interference condition, $B = .36, p = .005$. However, the relation between Stroop effect and P600-like effect to syntactic interference in the spillover region was no longer significant, $B = -.18, p = .176$. This multiple regression finding is due to the same factors as in the correlational analysis after controlling for subjects' processing speed and linguistic experience. Thus, the executive control ability as indexed by Stroop effect was not a unique and robust predictor for syntactic interference resolution in the spillover region, unlike the executive control ability measured by recent negatives effect. None of the other predictors were even close to significance in any of the models. In sum, resistance to proactive (as measured by the recent negatives task) was a unique predictor of individuals' online syntactic interference resolution ability. Subjects who were more resistant to proactive interference had less difficulty in syntactic interference resolution during online sentence processing.

3.3.5 Relations between composite measures and RC processing

Last, I have examined the relations between the individual measures and ERP components elicited by RC processing on the head noun. Regarding the ORC-elicited posterior positive effect (300 – 650 ms), since it was more consistent with the experience-based account, I predicted that the mean amplitude of this positivity should be modulated by individuals' linguistic experience - subjects with richer lexical semantic representation or linguistic experience should show smaller amplitude. Regarding the negative ERP effect (650 – 800 ms) elicited by the high semantic interference condition, based on the previous finding that high WM span subjects were more sensitive to semantic cues (Traxler et al., 2002, 2005), I predicted that the mean amplitude of this negativity should be modulated by certain components of individuals' WM capacity. However, few of the previous studies have examined the nature of the relation between certain components of WM capacity and sensitivity to semantic cues and thus, it is hard to make a precise prediction about the specific task and direction of the relations.

Correlational analyses and multiple regression analyses were performed in the same manner as above on the mean amplitude of the posterior positivity elicited by ORC structure (300 – 650 ms) and negativity along the midline (650 – 800 ms) elicited by semantic interference. The results are shown in Table 13 and 14. As shown in the tables, subjects' reading performance on the head noun was not related to any of the individual differences measures. The lack of relations with any individual differences measures was not predicted by either WM-based or experience-based account. It was interesting that the relation between syntactic processing, specifically the relation to interference resolution efficiency, and to resistance to proactive interference found in the critical verb and the

spillover regions, was not obtained here. This indicated that the underlying mechanism involved in solving ORC processing difficulty was at least partially different from that involved in solving syntactic interference - subjects have to select the target NP against the distractor(s) at the main verb, whereas there is little interference at the head noun following the RC structure because only one NP and one embedded verb were presented in the RC region could be integrated with the head noun. Regarding the lack of relations between the size of the semantic interference effect and any individual differences measures, such might be explained by the lack of an appropriate measure for semantic STM in the current experiment. In previous English studies, semantic STM as measured by the category probe was the only significant predictor for semantic interference resolution (Tan et al., 2011, 2013). However, the category probe task in Chinese did not uniquely tap subjects' semantic STM, although the limited sample size of the present study might also constrain the power to detect these subtle WM-language relations.

3.4 Summary of major results

In the current experiment, as predicted by the cue-based parsing approach, evidence for both semantic and syntactic interference effects was obtained, demonstrating that subjects had more difficulty in comprehending sentences containing high semantic or syntactic interference. Importantly, both high semantic and syntactic interference sentences elicited an early anterior negativity starting around 300 ms after the onset of the critical verb. This early AN, which has typically been associated with greater WM computation or integration demands (King & Kutas, 1995; Fiebach et al., 2002), indicated that Chinese readers made use of semantic retrieval cues as early as syntactic retrieval cues. Additionally, high semantic interference sentences also elicited a late AN

(650 – 800 ms), while high syntactic interference sentences also elicited a P600 effect (650 – 800 ms), reflecting semantic and syntactic revision, respectively. In the spillover region, there was an interaction between semantic and syntactic interference effects between 250 – 800 ms, with high syntactic interference condition elicited a sustained positivity along the midline in the low semantic interference condition only. I suggest that there was the possibility that readers temporarily consider the distractor in the HiSem/HiSyn condition, which partially matched both semantic and syntactic retrieval cues, as the grammatical subject of the main verb. Therefore, they did not experience more processing difficulty in the HiSem/HiSyn condition as compared to other conditions. However, this speculation has difficulty in explaining why the illusion of grammaticality did not occur in the critical region, where sentence processing might rely more on familiarity information during early stage of processing and thus be more vulnerable to interference effects. Instead, the interaction was observed in a later stage of interference resolution, when subjects already started processing the following phrase and controlled attention was assumed to be involved. I suggest that it is possible that the interaction only occurs when subjects actually started resolving both types of interference, and semantic interference resolution might not be initiated until the spillover region, though subjects detected the semantic interference as early as syntactic interference. Future studies are needed to address the nature of the interaction of semantic and syntactic interference effects.

Importantly, the mean amplitudes of the P600 effect to the high syntactic interference condition on the critical verb and a similar effect (in the low semantic interference conditions only) in the spillover region were predicted by subjects'

resistance to proactive interference. Subjects who were more resistant to proactive interference (i.e., smaller recent negative effect) showed a smaller magnitude of these two syntactic interference-elicited ERP effects. There was a negative link between subjects' resistance to a prepotent response (e.g., as measured by the Stroop task) and their syntactic interference resolution ability in the spillover region. However, this relation disappeared after controlling for subjects' linguistic experience and general processing speed. Thus, the executive control indexed by the Stroop effect might not be a unique and robust predictor for syntactic interference resolution like that indexed by the recent negatives effect. Regarding the WM mechanisms underlying semantic interference resolution, I did not replicate the robust relation between the semantic interference effect size and semantic STM (as measured by category probe) found in previous English studies (Tan et al., 2011, 2013). The semantic interference effect elicited ERPs were not related to any individual differences factors. These results might be caused by the lack of validity of the Chinese category probe task and the limited sample size.

Moreover, I have examined the RC processing differences on the RC head noun and following adverbial phrase. On the head noun, there was an ORC-elicited P300-like posterior positive effect, which might be associated with processing low compared to high frequency structure in Chinese (i.e., ORC vs. SRC), and a semantic interference effect-elicited late negativity along the midline. However, neither effect on the head noun was correlated with the individual difference factors. On the adverbial phrase, which was inserted between the head noun and the critical verb, there was a three way interaction of Syntactic $\times$ Hemisphere $\times$ Region between 150 - 400 ms. However, there was no significant ERP effect from 400 ms prior to the onset of the critical verb. This result

ensured that the interference effects observed in the critical verb region should not be simply attributed to a spillover effect from processing RC structure.

4 Discussion

The present study aimed to investigate semantic and syntactic interference effects in sentence comprehension, and the role of WM capacity in interference resolution. Through analysis of ERPs during the reading of well-formed Mandarin Chinese sentences, the results confirmed predictions from the cue-based parsing approach that interference from non-target constituents is an important determinant of sentence comprehension difficulty (Lewis & Vasishth, 2005; Lewis et al., 2006; Van Dyke & McElree, 2006; Van Dyke, 2007). These results converge with those of previous English studies (Glaser et al., 2013; Tan et al., 2011, 2013; Van Dyke, 2007; Van Dyke & McElree, 2006, 2011) and extend those findings to show that in Chinese, the semantic interference effect arises even when the distractor's syntactic features strongly eliminated it from the distractor set. In addition, there was no time course difference in detecting the semantic and syntactic interference effects as both effects elicited an anterior negative ERP effect starting as early as 300 ms after the onset of the critical verb. However, the actual resolution process for the two types of interference might differ, as different ERPs were observed in a later time window on the critical verb (650 – 800ms). I suggest that interference resolution might involve two processes – an early diagnosis based on familiarity information and a later revision/resolution process during which executive control is involved. Importantly, regarding the role of WM capacity plays in interference resolution, the current results demonstrated that subjects with better resistance to proactive interference, as measured by the recent negatives task could better resolve syntactic interference as they showed a smaller P600 effect on the critical verb and the

spillover region, even after controlling for their linguistic experience and processing speed. Taken together with our previous studies (Tan et al., 2011, 2013) these results are in line with the proposal that attentional control ability supports syntactic revision in sentence processing (Novick, Trueswell, & Thompson-Schill, 2005; Vuong & Martin, 2011; Ye & Zhou, 2009b). In the following, different ERP effects for different sentence regions and implications for WM mechanism underlying sentence processing will be discussed separately.

4.1 Interference resolution during sentence comprehension

The first objective of this study was to examine interference effects during Chinese sentence comprehension. It is noteworthy that all the sentences used in the current experiment were grammatically well-formed and differed only in the degree of semantic or syntactic interference, which allowed us to assess natural sentence processing without structural ambiguities or violations.

4.1.1 Critical verb

As predicted, on the critical verb, anterior negative effects were elicited by both high semantic and high syntactic interference conditions relative to the low interference conditions, starting around 300 ms after the onset of the critical verb and lasting for about 200 ms (i.e., 300 – 500 ms). The early occurrence of both semantic and syntactic interference effects in Chinese adds to the growing body of cross-linguistic evidence demonstrating that interference effects play an immediate (i.e., within 300 ms) and important role in determining sentence processing difficulty. The finding of ANs is in line with previous data that such ERP effects reflect high WM demand. In the previous memory studies, ANs has been observed when subjects were required to retain verbal

information in WM, as compared to a control condition of completing a searching task without retention demand (Ruchkin, Johnson, Canoune, & Ritter, 1990). In psycholinguistic studies, ANs have been associated with difficulty in establishing long-distance dependencies (Fiebach et al., 2002; King & Kutas, 1995; Kluender & Kutas, 1993; Mullet et al., 1997; Weckerly & Kutas, 1999; Yang & Perfetti, 2010), ambiguity-related referential binding (e.g., when a particular NP could refer to more than one equally suitable referents) (Van Berkum, Brown, & Hagoort, 1999; Van Berkum, Brown, Hagoort, & Zwitterlood, 2003), and comprehending sentences that presented events out of chronological order (Münte, Schiltz, & Kutas, 1998)³⁸. Thus, the observation of ANs is in agreement with my prediction that subjects have more difficulty in the high semantic or syntactic interference condition when they detected more than one memory representation that partially matched the retrieval cues.

However, as mentioned earlier, a major issue in the discussion of ANs concerns the variations in their time course (e.g., short lived vs. sustained) and scalp distribution (e.g., left lateralized vs. bilaterally distributed). Therefore, it is hard to compare the ANs observed across different studies, and then to relate each one to a very specific memory process. Friederici and Weissenborn (2007) have pointed out that "... the ultimate solution for the variance across different studies in the literature is not simple as the studies have used different violations types, languages and modalities. All these parameters may influence the appearance of the E/LAN effects" (Friederici &

³⁸ In Münte et al.'s study (1998), they reported a left anterior negativity when additional memory operations are required. For example, while reading the sentences "*After/Before the scientist submitted the paper, the journal changed its policy*", readers showed a LAN effect in the "*before*" conditions as they have to compute the actual order of occurrence since the two events were mentioned in the reverse temporal order. Moreover, the amplitude differences between "*before*" and "*after*" conditions was positively related to subjects' reading span – subjects with better WM capacity showed greater differences. Therefore, they suggested that the LAN reflects WM computation demand.

Weissenborn, 2007). As a result, although researchers have generally agreed that at least part of the neural generators of the anterior negativity family are overlapping and reflecting increased WM load (see a review in Fiebach et al., 2002), the multifaceted view of WM (Shipstead et al., 2014; Unsworth et al., 2014) and the many-to-one mapping from memory operations to ANs make it hard to unequivocally relate the ANs to an exact component(s) of WM (i.e., primary memory capacity, attention control, and secondary memory retrieval according to Shipstead et al., 2014 and Unsworth et al.'s, 2014).

It should be noted that although the ANs elicited by the high syntactic and high semantic interference had the same time-course (i.e., 300 – 500 ms), the two negativities differed in scalp distribution. The AN associated with high semantic interference conditions was bilaterally distributed, while the AN associated with high syntactic interference conditions was left hemispheric lateralized. The same time-course and partially overlapping topographical profile of the two ANs implied that at least part of the processing consequences of high semantic and syntactic interference involved a similar underlying processing mechanism, such as an increased demand on WM operations. However, the more widely distributed ERPs in the high semantic interference relative to the high syntactic interference conditions indicated that the two processes differed to some extent and did not reflect the exact same underlying WM process. Based on the current ERP results, it is hard to tell to what extent semantic and syntactic interference resolution can be separated, as I did not further conduct a spatial analysis (e.g., the moving source approach model or the stationary source approach) to locate the neuroanatomical source of the ERP effects. The different scalp distribution of the

semantic interference- and syntactic interference-elicited anterior negativities might be attributed to the neural activation from different brain regions, or different levels of activation from the same regions. In previous studies, attempts have been made to localize the neural generators of different ERP components. Through intra-cranial depth recording of ERPs, magnetoencephalography (MEG), or through fMRI, researchers proposed that most ERP components arise from more than one functionally and spatially distinct neural generator (Hagoort, 2003; Osterhout & Nicol, 1999). Friederici et al. (2000, 2003) reported that the early LAN (ELAN), which was elicited by word category violations (e.g., “*The fish was in caught*”), was caused by greater activation in the inferior frontal and anterior temporal cortex with a clear left hemisphere dominance (Friederici, Wang, Herrmann, Maess, & Oertel, 2000; Friederici, Rüschemeyer, Hahne, & Fiebach, 2003). However, so far, no study has systematically examined the differences in the neural generators of the different ANs. It is unclear whether the LAN and the bilateral AN have distinct or the same neural basis. Nonetheless, I found an interesting phenomenon reported in both ERP and some fMRI studies that syntactically related processes tend to elicit a left lateralized activation, whereas semantically or conceptually related processes, or high WM load process (e.g., long-distance integration) tends to elicit bilateral activation (Friederici, Rüschemeyer, et al., 2003; Kluender et al., 1998; Service, Helenius, Maury, & Salmelin, 2007; Van Berkum, Brown, & Hagoort, 1999; Van Berkum, Brown, Hagoort, & Zwitterlood, 2003).

For example, based on the results from both German and English ERP studies, Kluender et al. (1998) have suggested that the local and global anterior negativities are experimentally dissociable – the local AN is associated with (morpho)syntactic

processing, while the global AN is associated with increased WM load (Kluender et al., 1998). This claim is consistent with numerous findings that local LAN is associated with demanding syntactic processing, such as the detection of word category violation or morphosyntactic violation (e.g., number/gender/person agreement errors in subject-verb agreement) (Friederici & Weissenborn, 2007; Gunter, Friederici, & Schriefers, 2000; Osterhout & Mobley, 1995), while a sustained AN with a bilateral distribution is more likely to be observed when demanding conceptual-semantic processing was involved, such as when two or more referents satisfied the semantic constraints of a pronoun (Nieuwland & Van Berkum, 2008; Van Berkum et al., 1999; Van Berkum et al., 2003; Van Berkum, Koornneef, Otten, & Nieuwland, 2007; Van Berkum, Zwitserlood, Bastiaansen, Brown, & Hagoort, 2004). For example, Van Berkum et al. (1999) presented subjects sentences in the context of short introductory stories, such as “*David had told the two girls (vs. the boy and the girl) to clean up their room before lunch time. But one of the girls had stayed in bed all morning, and the other had been on the phone all the time. David told the girl that ...*”. They observed a bilateral AN emerging 300 – 400 ms after the onset of “*the girl*” in the last sentence “*David told the girl that ...*”, which last for several hundreds of milliseconds. Van Berkum et al. suggested that the important implication from this result was that subjects could make immediate use of referential information to parse a subsequent local structural ambiguity. Although Van Berkum et al. examined referential ambiguity at the context level, the observation of a bilateral AN elicited by high semantic interference condition in the current study aligned with their finding as both reflected a processing difficulty in semantic representation integration. In a recent study, Service et al. (2007) used the MEG technique to locate the neural basis of

the LAN and N400, both of which have the same time course but different topographical distribution. They found that although both the LAN and N400 arise from activation in the STG region, semantic processing (i.e., the N400) showed a bilateral activation pattern as reported in previous studies (Marinkovic et al., 2003; Service, Helenius, Maury, & Salmelin, 2007), whereas syntactic processing (i.e., LAN) showed a left lateralized activation pattern.

With respect to the fMRI evidence, the combined data from some relevant studies revealed a potentially interesting finding that although both semantic and syntactic processes rely on a temporal-frontal network, lexical-semantic processing tends to activate the relevant brain regions bilaterally (e.g., middle STG bilaterally, insular cortex bilaterally, as in Friederici et al., 2003), whereas syntactic processing tends to activate the relevant brain regions in the left hemisphere (e.g., anterior left STG, left posterior frontal operculum, left basal ganglia). Moreover, Friederici et al. (2003) directly compared the brain regions underlying the processing of semantically and syntactically incorrect sentences and found that semantic processing caused higher levels of brain activation than did syntactic processing in the middle STG bilaterally, while syntactic processing caused higher levels of activation in the left basal ganglia (Friederici, Rüschemeyer, Hahne, & Fiebach, 2003). However, it should be noted that Friederici et al. and many other researchers who found lateralization differences between semantic and syntactic processing did not draw conclusions about the hemispheric lateralization of semantic versus syntactic processing. The results from numerous studies showed that some brain regions are bilaterally activated for both semantic and syntactic processes (e.g., STG), some semantic processing more strongly activates the left hemisphere (e.g. in Glaser et

al.'s study investigating semantic and syntactic interference in English sentences, 2013), and in cases, semantic and syntactic information are demonstrated to be processed by the same neural system but in different ways (Kuperberg et al., 2003).³⁹ More importantly, the scalp distributions of the ERP effects may not necessarily correspond to the real brain location of the source because the ERPs reflect several synchronously active sources (Picton et al., 2000). Thus, in the current experiment, the finding that the AN elicited by the syntactic interference effect was more left lateralized and the AN elicited by the semantic interference effect bilaterally distributed does not necessarily imply that semantic interference resolution is more bilateral in nature than syntactic interference resolution. Future studies investigating the neural basis of LAN and AN are needed.

To sum up, I suggest that the early ANs elicited by the high semantic and the high syntactic interference conditions reflect the demanding syntactic/semantic processes due to the detection of multiple memory representations that at least partially match the retrieval cues, and the resistance to multiple semantic and syntactic representations might be supported by overlapping but partially distinct underlying neural mechanisms.

Regarding the two late effects, the late positivity elicited by the high syntactic interference condition resembles a P600 effect. This positivity with a midline maximum started about 650 ms after the onset of the critical verb and did not return to baseline at the onset of the next word (i.e., 650 – 800 ms). Traditionally, the P600 component has generally been associated with the syntactic reanalysis, such as when reading a

³⁹ Through comparing the results from an ERP and a fMRI study using the exact same sentence materials, Kuperberg et al. (2003) found that both semantic and syntactic information are processed by the same neural system (i.e., left temporal and inferior frontal region) but in different ways with an increased activation in the left temporal and inferior frontal regions for semantic violations and a decreased activation for morphosyntactic violations.

word/phrase that is ungrammatical given the preceding sentence context (Friederici, Mecklinger, Spencer, Steinhauer, & Donchin, 2001; Hagoort, 2003; Hagoort, Wassenaar, & Brown, 2003; Hahne & Friederici, 1999; Kuperberg, 2007; Osterhout & Holcomb, 1992), Later studies, however, demonstrated that the P600 reflects syntactic integration difficulty in general that is not restricted to repair or reanalysis of outright syntactic violations (Fiebach, Schlesewsky, Lohmann, Von Cramon, & Friederici, 2005; Hagoort, 2003; Kaan, Harris, Gibson, & Holcomb, 2000; van de Meerendonk, Kolk, Vissers, & Chwilla, 2010). With respect to the overall syntactic interference resolution process, a similar LAN-P600 compound has been reported in several previous studies (Friederici, von Cramon, & Kotz, 1999; Friederici, Hahne, & Mecklinger, 1996; Friederici, Rüschemeyer, et al., 2003; Hahne & Friederici, 1999), though most of these experiments observed the early negativity when there was an explicit (morpho)syntactic violation. Friederici and colleagues suggested that the two syntactic processing related ERPs should be attributed to two functionally different stages of syntax processing, i.e., the LAN is related to an initial, automatic structure-building process, and the late P600 that is related to a controlled second-pass parsing process. As a result, based on the finding that the P600 component was preceded by a LAN-like component in the current experiment, I suggest that the P600 effect reflects a second-pass syntactic integration or revision after the detection of multiple memory representations that are syntactically partially suitable.

With respect to the late negativity elicited by the high versus low semantic interference conditions, however, I suggest that it could be either interpreted as a separate ERP component from the earlier anterior negativity (300 – 500 ms), or a continuation of the early effect (i.e. one sustained anterior negativity between 300 – 800 ms). Although

the semantic interference effect did not reach significance between 500 – 650 ms, it was overall significant in the 300 – 800 ms time window with the high semantic interference condition being more negative than the low semantic interference condition. As I have discussed earlier, there were two possible reasons that might cause the discontinuation of the anterior negativity: 1) The interaction of semantic and syntactic interference between 450 and 550 ms. During this time window, as reported in the results section 3.2.1, in the high semantic interference conditions, the syntactic interference effect elicited a negative deflection in the left hemisphere but a positive deflection in the right hemisphere. This puzzling hemisphere asymmetry might be caused by the fact that more than one ERP component presented and overlapped with each other to some extent and thus, cancelled out certain effects, such as the semantic interference elicited anterior negativity. 2) The substantial inter-subject variability might obscure the sustained nature of the observed negativity. A similar issue has been reported by A. E. Martin et al. (2014) in the study manipulating both grammaticality and morphosyntactic interference effects in Spanish (see more details in section 1.1.1.4.2). They observed an anterior negativity starting as early as 100 ms but which stopped at 400 ms, though they predicted a sustained negativity (e.g., 100 – 1000 ms) based on their earlier study using similar materials (A. E. Martin, 2012). A. E. Martin et al. suggested that the lack of continuity of this effect might be caused by the fact that a subset of the participants showed positive deflections in the 400 – 1000 ms time window. Their speculation was supported by the posthoc analyses, in which subjects were divided into two groups based on whether they showed a negative ERP deflection to the grammaticality manipulation. Then a sustained negativity induced by the high morphosyntactic interference condition was observed in the 400 – 1000 ms

time window in the group of subjects who showed a grammaticality-related ERP effect, but not in the group who did not. However, A. E. Martin et al. admitted that such posthoc analysis was quite exploratory, being unmotivated by any prior hypothesis. They did not discuss anything about the potential differences in cognitive abilities or processing skills between the two groups. Instead, they just suggested that since the group-split rendered the interference effects trivial, it should reflect group differences in the “overall response pattern”. Though it was still unclear what kind of response pattern each group showed and why they showed different response patterns.

Moreover, there was a third possibility that the individual differences in onset latency caused the lack of continuity of the AN induced by semantic interference. In our previous self-paced reading experiment investigating semantic and syntactic interference effects with English materials (Tan et al., 2011), we found considerable variation in when subjects detected and attempted resolution of semantic interference. In Tan et al.’s study (2011), although the main effect of the semantic interference was significant in the spillover region following the critical verb, only 59% of the 112 subjects showed a semantic interference effect in this region. However, a large majority of the subjects showed a positive semantic interference effect in offline question answering RT (84%). In the current experiment, for the semantic interference ERP effects, 65% of the subjects (N = 24) showed the early AN, 65% showed the late AN, and 57% (N = 21) showed both the early and the late ANs. As a result, the variation in semantic interference ERP effects during online sentence processing may have prevented us from observing a sustained negativity associated with semantic interference resolution. Although there might be a cross-linguistic difference in the relative time-course of semantic and syntactic

interference effects between English and Chinese, and the variation in the starting point of semantic interference might not be as great as in English, future analyses might profitably examine possible individual differences in the onset of semantic interference.

For the current results, if the late anterior negativity (650 – 800 ms) is a relatively independent ERP component from the early anterior negativity (300 – 500 ms), I suggest that the late negative deflection reflects a second-pass process of updating or replacing the existing semantic representations (Baggio, Van Lambalgen, & Hagoort, 2008; Friederici, Steinhauer, & Frisch, 1999; Jiang, Li, & Zhou, 2013; Politzer-Ahles, Fiorentino, Jiang, & Zhou, 2013; Zhang, Yu, & Boland, 2010). Taken together with the LAN-P600 pattern observed for syntactic interference resolution, these results are generally consistent with Friederici's two-stage assumption (1998, 1999, 2003) that there are two different processing aspects during integration/revision: a first stage of diagnosis and a second stage of actual reanalysis. In a previous study, Baggio et al. (2008) observed a similar late anterior negativity emerged at about 400 ms following the onset of sentence final word (e.g., "*paper*"), in the sentence such as "*The girl was writing letters when here friend spilled coffee on the paper (vs. tablecloth)*". They suggested that in the "*tablecloth*" condition, readers would automatically make inference such as "the girl has written a letter" based on the assumption that spilling coffee on the tablecloth is usually neutral and would not affect the writing activity. However, readers have to suppress or recompute their initial inference in the "*paper*" condition for the reason that spilling coffee on the paper is sufficient to terminate the writing activity. Thus they suggested that the late negativity on the final word "*paper*" was associated with recomputation of the existing representations to arrive at a new mental representation. In a Chinese study,

Jiang et al. (2013) also observed a late anterior negativity between 550 – 800 ms on the critical verb when the sentence continuation was incongruent with the subjects' expectation (or pragmatic inference). For example, when read the sentence "*Zhanghong can hear **even** such loud (vs. tiny) sounds clearly. He had sharp hearing*" (Note: in Chinese, the verb "*hear*" come after the "*even*" phrase, and the sentence was presented in the order as "*Even so loud (vs. tiny) sound Zhanghong can hear very clearly ...*"), subjects had more difficulty in processing the "*loud*" condition because as cued by the word "*even*", they were expecting an event with low likelihood of happening, such as hearing a tiny sound. Jiang et al. suggested that the late anterior negativity reflected that subjects were doing a second-pass revision according to the likelihood of the events to satisfy the pragmatic constraints of the existing representation. Based on these results, I suggest that the two semantic interference-related anterior negativities might reflect two different stages of semantic processing, similar to the syntactic interference LAN-P600 compound. The early anterior negativity reflects initial automatized semantic representation, and the later anterior negativity reflects subjects' effort at updating/recomputing the semantic representations after detecting the semantic interference from earlier distractors. Subjects might have constructed one interpretation first, and then reanalyzed or recomputed the initial representation when they detected that there was semantic or syntactic interference.

On the other hand, if the late AN (650 – 800 ms) is a continuation of the early AN (300 – 500 ms), and the high semantic interference sentences in fact elicited a sustained AN (300 – 800 ms), I suggest that the sustained AN (or *Nref* effect) might reflect a deep level rather than a shallow level of binding difficulty (Martin, Nieuwland, & Carreiras,

2012, 2014; Nieuwland & Van Berkum, 2006; Van Berkum, Brown, & Hagoort, 1999; Van Berkum, Brown, Hagoort, & Zwitterlood, 2003; Van Berkum, Koornneef, Otten, & Nieuwland, 2007; Van Berkum, Zwitterlood, Bastiaansen, Brown, & Hagoort, 2004).

This means that at least in the correctly-answered sentences, the processing difficulty did not just arise at the level of memory for text (i.e., shallow level binding), e.g., subjects remembered that there are more than one animate NPs in the sentences. Instead, processing difficulty arises because subjects at least temporarily think about the possibility of integrating the distractor with the critical verb. My assumption is based on the finding from Nieuwland and Van Berkum's study (2007), in which they observed a sustained AN only in the condition when there was genuine referential ambiguity from two eligible referents, but not in the condition when two referents were mentioned but only one was eligible. As just discussed, in an earlier study, Van Berkum et al. (1999) observed a referentially induced AN effect due to having two eligible antecedents. In a following study, Nieuwland and Van Berkum (2007) further demonstrated that such referentially induced sustained AN reflects deep level processing, rather than a superficial "resonance-type memory access mechanism", for the reason that such an ERP effect disappeared if one of the eligible candidates left the scene (e.g., *David had told the two girls to clean up their room before lunch time. But one of the girls had been sunbathing in the front yard all morning, and the other had actually just driven off in his car for some serious downtown shopping. As he gazed at the empty driveway, David told the girl that ...*). Assuming that the sustained AN reflects deep level processing, I suggest that subjects in the current experiment did not just realize or remember that there were

another semantically plausible noun phrase in the sentence, but actually at least temporarily considered its possibility or eligibility of being the agent to the verb.

The fact that the sustained AN was only observed in the high semantic interference condition but not in the high syntactic interference condition implies that: First, as I have discussed above, at least part of the neural mechanisms underlying semantic and syntactic interference resolution are relative independent; Second, the semantic interference might not be fully resolved by the end of the critical verb. As discussed earlier, the resolution of semantic interference actually depends on the analysis of the NP's syntactic features, given that both nouns are equally compatible as the subject of a verb on semantic grounds (e.g., as in Example 13, both "*neighbor*" and "*resident*" are possible agent of the critical verb "*complain*"), but no P600 effect was observed in the high semantic interference condition. However, it should be noted that there are some potential concerns in relating the current findings to the findings from the series of studies conducted by Van Berkum and colleague (1999, 2003, 2004, 2007), in which an AN was obtained. First, Van Berkum et al. examined the referential binding at the discourse level and did not dissociate semantic, syntactic, and pragmatic binding processes. Although most of these studies focused on the subjects' recomputation of semantic or pragmatic representations, it is possible that syntactic revision also contributed to the observed sustained anterior negativity. Therefore, there is the possibility that when subjects made use of syntactic information to resolve semantic interference, a sustained anterior negativity rather than a typical P600 will be obtained. Second, although it is possible that the sustained AN reported in many of these studies was actually composed of more than one component (e.g., an early negativity and a late

negativity) and was visible as one continuous effect due to the stronger violation or ambiguity than in the current experiment, so far, no study has systematically tested such a hypothesis. Therefore, I consider the sustained AN as a different possibility from the two-stage processing model, and reflected a general deep level difficulty in semantic integration.

4.1.2 Spillover region

At the first noun phrase after the critical verb (e.g., “*rent*” in Example 13), an interaction of syntactic and semantic interference was observed. During the 250 – 800 ms time window, the LoSem/HiSyn condition elicited a sustained positive deflection peaking around 600 ms along the midline as opposed to the LoSem/LoSyn condition, indicating that the syntactic interference effect continued to impede sentence processing during the spillover region though only in the low semantic interference condition. Although such positivity was only marginally significant in the 50-ms time window analysis between 350 – 450 ms ($.05 < ps < .10$), I suggest that this should not be a P300-P600 compound as reported in some previous studies for three reasons: 1) there was no clear pattern of two peaks in which a first significant peak between 300 – 400 ms was followed by a second less pronounced peak around 600 ms (Friederici, Mecklinger, Spencer, Steinhauer, & Donchin, 2001; Li & Zhou, 2010); 2) the P300 and the P600 effect were supposed to have different spatial factors as demonstrated by Friederici et al. (2001). Though I did not conduct a principle factor analysis as Friederici et al. did, the scalp distribution of the positivity varied little between 250 – 800 ms; 3) the P300-P600 compound has been interpreted as reflecting an early diagnosis and an actual reanalysis of syntactic processing difficulty, such as in garden-path sentences. However, in the current

experiment, the observation of both syntactic and semantic interference starting in the critical region indicated that the interference diagnosis or actual resolution should have already started at the preceding critical verb. Therefore, I suggested that it was an early onset P600 effect in this post-critical region. This result was in line with A. E. Martin et al.'s recent study (2014), in which they reported a P600 effect in the post-critical region associated with high morphosyntactic interference conditions. The early onset of this P600-like effect might be caused by the fact that the P600 effect in the critical region, which was elicited by the syntactic interference effect as well, spilled-over to this following region. As shown in Fig. 2, it is obvious that on the critical verb, the large positive deflection was cut off at the 800-ms epoch boundary without returning to the baseline (each epoch was defined as the 0 – 800 ms time window after onset of each word). In the spillover region, the continuing P600 effect, which spilled-over from the critical region, was not visible in the 0 – 250 ms time window due to the 200 ms pre-word baseline correction. The sustained nature of this positivity might reflect the continuous effort in recovering from subtle interference effects, as earlier ERP studies have demonstrated that the onset latency of P600-like positivity was correlated with diagnosis of syntactic revision difficulty, whereas the duration of the positivity was related to the reanalysis proper (see Friederici et al., 2001, for a review).

However, it should be noted that the syntactic interference elicited P600-like effect observed on the post-critical noun phrase might not a simple continuation of the early P600 effect on the critical verb. Although both positive ERP effects have the same scalp distribution (i.e., along the midline electrodes), it was only observed in the low semantic interference conditions in the spillover region. It was unexpected that the

HiSem/HiSyn condition, which was supposed to be the most difficulty condition, did not significantly differ from any of the other three conditions in the simple pairwise comparisons ($F_s < 1$). Regarding the lack of syntactic interference effect in the high semantic interference condition, it is possible that this is due to the fact that subjects temporarily licensed the animate distractor as the subject of the critical verb and thus, they did not experience interference until they realized that their initial interpretation was wrong at a later point. Such speculation may not fit our current results very well, as it is generally assumed that the illusion of grammaticality of distractor should occur in the early stage of syntactic revision when subjects relied more on familiarity information, instead of during the later stage of revision when attention control becomes involved (A.E. Martin, 2012). For instance, the illusion of grammaticality of the distractor has been reported in A.E. Martin et al.'s ERP study (2012), in which they obtained an enlarged ERP effect in the low morphosyntactic interference condition relative to the high interference condition⁴⁰. There was even an observation that syntactic retrieval interference sometimes yields a facilitation effect, as in cases where an incorrectly retrieved NP can make an *ungrammatical* sentence appear more acceptable. For example,

⁴⁰ Although most of the previous studies observed the illusion of grammaticality of distractor in retroactive interference conditions, in which the target was more distant from the critical word (e.g., matrix verb) that triggered retrieval than the distractor, I suggested that the illusion could happen even in the proactive interference condition, when the correct target was less distant from the critical verb than the distractor (e.g., in the current experiment, target was 1-word far from the main verb, while distractor: was about 3-4 word far from the main verb). According to the cue-based parsing approach, both the target and the distractor were accessed through a direct, content-addressable, parallel retrieval. Thus, the distance differences between target-verb and distractor-verb should not affect the retrieval speed (McElree, 2006; Van Dyke & McElree, 2011), though the target word might have higher activation level due to its recency (Lewis & Vasishth, 2005). However, the activation differences between the target and distractor might not be strong enough for the subjects to distinguish between target and distractor immediately (Engelmann et al., 2015). The calculation of exact activation value of the distractor and target was beyond the scope of the current study. In general, I suggest there is the possibility that subjects did not experience more processing difficulty in the HiSemHiSyn condition for the reason that they temporarily considered the distractor as the grammatical subject.

it has been shown that subjects read the sentence “*The key to the cabinets/cabinets unsurprisingly *were rusty*” faster when the second noun phrase is plural (e.g., “*cabinets*”) than singular, though the second condition was actually ungrammatical (Wagers, Lau, & Phillips, 2009). Wager et al. argued that agreement attraction in ungrammatical sentences reflects the fact that a cue-based retrieval mechanism is subject to retrieval errors. Some researchers have tried to explain interference effects using computational modeling (Engelmann, Jäger, and Vasishth, 2015). Engelmann et al. (2015) suggested that whether interference could cause an inhibition or facilitation effect⁴¹ depends on the activation difference between target and distractor. However, it is not very clear how to estimate an item’s activation level according to these current computational models, since most models do not accurately simulate every aspect of interference effect. Further studies will be required to fully address how the activation differences between target and distractor(s) decide whether the observed interference is inhibition, facilitation, or absent, which would be needed to explain the interaction of semantic and syntactic interference observed in our study.

⁴¹ Regarding the facilitation effect, one might expect that besides showing shorter RTs on the critical word in the high interference condition, subjects should still make more errors to the comprehension questions in the high interference condition. Therefore, for the current data, if there was a facilitation effect in the HiSem/HiSyn condition, subjects should make more errors in such condition. However, the current results showed that there was a semantic interference effect in the expected direction (i.e., more errors in the HiSem condition), but there was no syntactic interference effect. The lack of a syntactic interference might be explained by several reasons. First, as I have discussed, there seemed to be a ceiling effect in the accuracy data that prevent the observation of some effects. Second, the processes that occurred during the question answering are very complicated. Although some studies have found reliable attractor effects in error rate data, some did not (see Wagers et al. 2009 for a review). Wagers et al. suggested that the offline tasks are more complicated than the online measures as the former one might involve semantic integration and “engender some sort of late reprocessing or regeneration of the surface form of the sentence”. Thus, although I did not observe a syntactic interference effect or interaction of syntactic and semantic interference effects in the accuracy data, this does not contradict the finding that there might be a facilitation effect in the online RT data.

Summary of the interference effects. Overall, the current ERP results were consistent with the predictions from the cue-based parsing approach that both semantic and syntactic interference impede subjects' sentence comprehension. More importantly, the high temporal resolution of the ERP technique reveals the moment-by-moment interference resolution, which was not visible in previous behavioral, neuropsychological or fMRI studies (Glaser et al., 2013; Tan et al., 2011, 2013; Van Dyke, 2007; Van Dyke & McElree, 2011). The relatively early LAN-like effects elicited by semantic and syntactic interference effects on the critical verb demonstrated that at least part of interference detection or resolution occurs very quickly. Chinese readers are sensitive to both semantic and syntactic interference as early as within 300 ms after the critical word onset, though the underlying cognitive mechanism supporting the diagnosis or resolution process might be different between syntactic and semantic interference conditions given the different topographical distribution (semantic interference: bilateral anterior vs. syntactic interference: left anterior). In the later time window on the critical verb, there was a late AN elicited by the high semantic interference condition only, indicating a second-pass revision of semantic processing due to semantic interference (or maybe consider it as part of a sustained anterior negativity between 300 – 800 ms, which reflect deep level semantic integration difficulty). In addition, readers also showed a P600 effect elicited by the high syntactic interference condition between 650 – 800 ms along the midline, indicating a syntactic integration/revision difficulty. The syntactic interference effect continued on the post-critical word, as an early onset P600-like effect but presented in the low semantic interference condition only, reflecting subjects' continuous effort in resolving syntactic interference. I speculated that the interaction of semantic and syntactic

interference might be caused by the fact that subjects temporarily mistook the distractor as the target of retrieval and thus, did not experience very much interference effect in the HiSem/HiSyn condition. However, this speculation might not explain the time course of the interaction in the current experiment, as the illusion of grammaticality has been assumed to happen at early stage of processing (A. E. Martin et al., 2012; Van Dyke, 2007). Further investigation is required.

4.2 Time course of semantic and syntactic interference

As mentioned in the introduction, a second objective of the current project was to investigate how semantic and syntactic interference interact to determine sentence processing efficiency. In psycholinguistic studies, different models of sentence processing try to account for how and when different types of information, especially semantic and syntactic information, are combined into a sentence-level representation during on-line sentence processing. The high temporal resolution of the ERP technique allowed us to look at the time course of semantic and syntactic interference resolution as a way to infer the time course of semantic and syntactic processing. I expected the current data to provide some new evidence about whether Chinese sentence processing better fits the predictions from the syntax-first model (Ferreira & Clifton, 1986; Frazier, 2002; Hahne & Friederici, 2002) or the constraint-based model (Boland, 1997; Kim & Osterhout, 2005; Kuperberg, 2007; Trueswell, Tanenhaus, & Garnsey, 2002).

Although researchers generally agreed that interaction between syntactic and lexical-semantic information takes place during a later processing stage, there are still

debates about whether the initial processing state only take syntactic information into consideration (Friederici, 2002; Friederici, Gunter, Hahne, & Mauth, 2004; Friederici, Steinhauer, & Frisch, 1999; Friederici & Weissenborn, 2007; Hagoort, 2003; Hahne & Friederici, 2002). Based on a number of lines of evidence from ERP and fMRI studies, Friederici et al., (1999, 2002, 2007) proposed that there are three fixed time windows for language processing: 1) Phase 1 (< 300 ms) for building local structure, which is word category based and usually happens within 300 ms after word onset; 2) Phase 2 (300 – 500 ms) for further syntactic and semantic processes taking place in parallel to construct the syntactic, semantic, and thematic relations between words; and 3) Phase 3 ($\pm$ 600 ms) for final stage reanalysis and repair, where the different kinds of information are integrated together to reach a final interpretation. Friederici and Weissenborn (2007) suggested that the time window for each stage could vary as a function of many factors, such as language-specific factors, particular features of experimental materials, and presentation mode. In the current experiment, given that both types of interference manipulation should have little influence on Phase 1 processing of the critical verb (i.e., there was no word category violation or ambiguity), I was interested in how semantic and syntactic interference interacted during Phase 2 and 3, after 300 ms of the word onset.

In most Indo-European languages (e.g., German, Dutch, French, English), although there is considerable supporting evidence for the assumption of independent syntactic processing during Phase 1 and interactive semantic and syntactic processing during Phase 3 (Gunter, Friederici, & Schriefers, 2000; Kim & Osterhout, 2005; Kuperberg, Kreher, Blais, Caplan, & Holcomb, 2005; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003). However, it was still controversial whether the two processes were

independent or interactive in Phase 2 (see Friederici & Weissenborn, 2007, for a review).

⁴² Some evidence supports independent processing of semantics and syntax during Phase

2. For example, Gunter and colleagues (2000) manipulated article-noun gender agreement (e.g., “*She travels das_{Neuter}/*den_{Masculine} (the) land_{Neuter}*”) and the cloze probability of the nouns (“*She travels the land on a strong camel*” vs. “*She travels the land with an old Warburg car*”). They found that LAN effects induced by gender violations was not affected by the cloze probability manipulation, and the N400 effect induced by low cloze probability was not affected by the gender violation, though an interaction of the two manipulations was found in the P600. Therefore, Gunter et al. suggested that semantic and syntactic processes are relatively independent during Phase 2. On the other hand, Hagoort (2003) suggested that semantic and syntactic processes are interactive in Phase 2. He also manipulated gender agreement (i.e., determiner-noun gender agreement) and semantic match (i.e., adjective-noun mismatch). The result showed that when there was a double violation on the noun phrase (e.g., “*The_{com} broken umbrella_{com} ...*” vs. “**The_{neut} *honest umbrella_{com} ...*”), the size of the N400 induced by the semantic violation was boosted by an additional syntactic violation as compared to the single semantic violation condition. However, the size of the P600 effect induced by the syntactic violation was not affected by the additional semantic violation. Therefore, Hagoort (2003) argued that syntactic processing is relatively independent of semantic

⁴² According to Friederici et al.’s sentence processing model, syntactic process differs between Phase 1 and Phase 2. Phase 1 only involves local phrase structure building, such as using local syntactic information to predict the incoming word’s category. Phase 2 involves structural dependency computation, such as processing information of agreement. Phase 1 process may be similar across languages, but the phase 2 may differ as different languages put different weights on the morphosyntactic information (Friederici, et al., 2006). In a German study, Rossi et al. (2006) reported a LAN-P600 compound associated with the agreement violation, and an ELAN-P600 compound associated with word category violation or double violation conditions (i.e., both word category and agreement violations). They suggested that the absence of the LAN effect in the double violation condition demonstrated the relative independency and primacy of local structure building in Phase 1 process as compared to dependency computation in Phase 2 process.

context, whereas semantic integration is influenced by syntactic processing. Friederici et al. (2007) explained the discrepancies between these studies as the different types of violation they investigated. For example, in Hagoort's study (2003), the gender information they manipulated was lexically based. Thus, the processing of gender information process was not independent of lexical-semantic processes, but instead had an influence on semantic processes.

The evidence for the interplay between syntactic and semantic processes is more complicated in Chinese, as there are studies showing that even word category processing in Phase 1 interacts with semantic processing (Ye, Luo, Friederici, & Zhou, 2006; Zhang, Yu, & Boland, 2010), ERP results in most other languages have demonstrated that semantic integration does not happen for words that are not syntactically licensed (e.g., as in a word category violation) (Friederici et al., 2004; Friederici et al., 1999; Hahne & Friederici, 2002; Isel, Hahne, Maess, & Friederici, 2007). In contrast, in a Chinese study, Ye et al. (2006) found that although word category information processing happened prior to semantic processing, word category information interacted with semantic information in a very early time window following onset of the critical word (250 ms – 400 ms), which was earlier than observed in most other languages. In a more recent study, Zhang et al. (2010) found that both of the ERP effects, which have been associated with semantic and syntactic violations, respectively, showed up in the double violated condition (e.g., local phrase and selectional restriction violation, “*The girl ate a very **skit and gloves.*”). Therefore, they concluded that semantic processing is not prevented by word category information in Phase 1. These findings are consistent with the notion that semantic information plays an earlier and more important role in online Chinese

processing. Researchers have pointed out that the large number of ambiguous words (i.e., many words can be a verb or a noun) and the lack of morphosyntactic cues make Chinese readers rely more heavily on lexical-semantic information, which is the only information available for disambiguation (Chu, 1998; Li & Thompson, 1989; Yang, Perfetti, & Liu, 2010; Ye, Luo, Friederici, & Zhou, 2006; Zhang, Yu, & Boland, 2010).

Based on these studies, I hypothesized that the semantic interference effect should occur early (e.g., as early as syntactic interference effect) and be relatively independent from the syntactic interference effect during Chinese sentence comprehension, which is different from interference effects observed in most of the previous English studies. As summarized earlier, in English, although evidence for semantic interference effect exists (Kush et al., 2015; Tan et al., 2011; Van Dyke, 2007; Van Dyke & McElree, 2011; Van Dyke et al., 2014), the issue is more controversial than syntactic interference effect. Researchers found that the semantic interference effect occurred later than the syntactic interference effect in both eye-tracking (Van Dyke, 2007) and fMRI (Glaser et al., 2013) studies, and it was even blocked by syntactic processing in certain conditions as observed in eye-tracking and speed-and-accuracy paradigms (Van Dyke & McElree, 2011). These findings are consistent with the notion that syntactic processing generally precedes semantic processing in most languages (Boland & Blodgett, 2001; Friederici, 2002; Friederici, Gunter, Hahne, & Mauth, 2004; Hagoort, 2003). However, since Chinese has been argued to be a more semantically based language, a more robust and earlier semantic interference effect was expected. According to Friederici's sentence comprehension model (1999, 2002, 2007), I hypothesized that both semantic and syntactic interference should occur beginning in Phase 2, after subjects had successfully

processed the word category information of the critical verb. Importantly, ERP effects related to semantic interference should not be fully blocked by strong syntactic features of the distractor (e.g., if the distractor was an object in the SRC structure, which is a core argument) and should be present as early as the syntactic interference effect.

Generally, the current ERP results bore out my hypotheses about interference resolution in Chinese sentence processing and also implied several differences in comparison to interference resolution in English. First, the results confirmed the hypothesis that both semantic and syntactic interference would occur during Phase 2 and 3 as defined in Friedrici et al.'s language processing model (2002, 2007), by showing that high semantic/syntactic interference sentences elicited ERP effects during 300 – 500 ms and 650 – 800 ms time windows after the critical verb onset⁴³. Second, I observed a semantic interference effect even when the distractor was in an SRC structure, when the distractor had strong syntactic features (i.e., object in a core argument) to eliminate it from the distractor set, whereas in previous English studies, the semantic interference effect disappeared in such conditions (Van Dyke & McElree, 2011). Third, I did not observe a significant time course difference for semantic versus syntactic effects during online sentence processing, as both types of interference elicited a LAN-like effect starting around 300 ms after the onset of critical verb, while previous English studies found syntactic interference at an earlier point in the sentence than semantic interference when both interference effects presented (Glaser et al., 2013; Van Dyke, 2007). Fourth, although there was a puzzling interaction of semantic $\times$ syntactic $\times$ hemisphere between

⁴³ The lack of clear main effects associated with either semantic or syntactic interference manipulation in the pre-critical region (i.e., the adverbial phrase region, such as “*everyday*” in Example 13) indicated that the ERP effects observed at the critical verb could not be simply explained as a spillover effect from processing different types of RC (see more discussion in later section).

450 – 550 ms, the semantic and the syntactic interference elicited two relative independent early anterior negativities with different topographical distribution between 300 – 500 ms.

The current findings are of interest as they provide additional evidence relevant to the assumption that during online Chinese sentence comprehension, semantic processing is a relatively independent process that happens as early as syntactic processing. These results seemed to be more consistent with the constraint-based model, which proposes that lexical, semantic or conceptual sources of information can provide immediate constraints on sentence processing and are partially independent of syntactic processing (Boland, 1997; Kim & Osterhout, 2005; Kuperberg, 2007; Trueswell, Tanenhaus, & Garnsey, 2002). Although it is hard to make a parallel comparison to previous English behavioral or fMRI studies due to the large differences in technique, the fact that I observed an early occurring semantic interference effect together with a syntactic interference effect, even when the distractor was syntactically unequivocally unlicensed indicates that semantic processing plays an early and important role in Chinese sentence processing. However, it should be noted that there are at least two speculations as to why I observed such an early semantic interference-elicited ERP effect: 1) during Phase 2 and 3 processing in Chinese, semantic and syntactic processes operate in parallel. Thus, the distractor's syntactic features, which were strong enough to eliminate it from distractor set in English, did not affect the parallel semantic retrieval/integration in Chinese; 2) there might still be an asymmetry in interplay between syntactic and semantics processes, and syntactic processes have a certain priority. However, syntactic features were given less weight in Chinese than in English and thus, did not fully block further semantic

processing of the distractors. However, evidence from the current experiment could not unambiguously distinguish these two possibilities. The observation of both semantic and syntactic interference starting around the same time (i.e., within 300 ms) seems to favor the first explanation. Nevertheless, since the two early negativities associated with semantic and syntactic interference effect differed in scalp distribution and perhaps other aspects that were not examined in the current analysis (e.g., peak latency, source generator), further studies are needed to tell apart the two potential explanations.

Additionally, it is noteworthy that there might still be a time course difference as to when people started resolving syntactic and semantic interference, though the detection or diagnosis of semantic and syntactic interference happened almost simultaneously. As I have pointed out earlier, the resolution of both semantic and syntactic interference actually depends on the analysis of items' syntactic features. However, on the critical verb, the P600 effect induced by review of syntactic decisions was only observed for the syntactic interference effect, but not for the semantic interference effect. Moreover, although there was no interaction of semantic and syntactic interference effects in the 650 – 800 ms time window, due to the subtle nature of the interference effects, I have specifically examined whether a P600 effect for the semantic interference effect could be observed in the HiSyn condition. Such a posthoc analysis is motivated by the fact when subjects have to refer to syntactic features to resolve semantic interference, there should be little checking of syntactic decisions in the LoSyn condition as compared to the HiSyn condition. In the LoSyn condition, subjects could easily look up the relevant syntactic features to help resolve the semantic interference effect and thus, no significant P600 effect is expected. However, no P600 effect was induced by the semantic interference

effect in the HiSyn condition in either midline or lateral electrodes ($F_s < 1$). Therefore, I suggest that subjects may not make use of the items' syntactic features to help resolve semantic interference as immediately as for syntactic interference.

For both Chinese and English, we have suggested that syntactic interference resolution may be faster as it involves analyzing a finite set of grammatical features to determine whether each noun phrase matches the critical verb (Tan et al., 2011). In contrast, determining semantic fit is more complex due to the possibility of degrees of fit. For example, consider an example sentence from our English experiment “*The critic who had enjoyed the memorable play at the new theatre will visit the director*”. In this sentence, the intervening noun phrase “*play*” causes little syntactic interference as it is unambiguously assigned an object role, which does not match the subject role that is required when retrieving the subject NP of “*will visit*”. However, in terms of semantic fit to “*will visit*”, it is the case that non-concrete subjects can appear with a movement verb used in a certain circumstances (e.g., “*The play will visit Houston*”). It is the same issue in some of the current Chinese experimental sentences. As a result, I suggest that although the semantic interference effect occurred as early as the syntactic interference effect in Chinese, the actual resolution of semantic interference might happen later than the resolution of syntactic interference. However, there is the possibility that although the resolution of semantic interference was based on the analysis of items' syntactic features, it might not elicit a P600 effect as the resolving of syntactic interference and the late or sustained anterior negativity reflected the semantic resolving process. These speculations needed to be examined in future studies, possibility with a more strict manipulation of the inanimate NP's semantic feature.

In conclusion, the present study provides evidence for an early and relatively independent semantic interference effect during Chinese sentence comprehension. In addition to the ERP effects elicited by syntactic interference, ANs (or a sustained AN) were elicited by high semantic interference even when the distractor's syntactic features could strongly eliminate it from the distractor set. I suggest that at least for interference resolution, which happens after local phrase structure has been built based on word category information, semantic information plays an immediate and relatively independent role in Chinese. The outputs of parallel semantic and syntactic processing serve as the input to a final stage of integration to achieve a final interpretation. The current results differed from observations from previous studies in English, which showed that the failure of syntactic licensing prevented further semantic processing, but were in line with previous finding that semantic processing in Chinese can happen before a syntactic commitment has been made, and proceeds even in the syntactically inappropriate condition (Boland, 1997; Hagoort, 2003a; Kuperberg, 2007; Ye et al., 2006; Zhang et al., 2013; Zhang et al., 2010). However, it should be noted that the generalizability of our conclusion was restricted by the fact that there is no comparable ERP study in English at this time, and most of the previous ERP studies investigating the interplay of syntactic and semantic process adopted an ambiguity or violation paradigm (e.g., garden-path sentences, subject-verb agreement violation, phrase structure violation). Thus, the different time course we observed in the current study from other experiments might be caused by the differences in methodology and types of processing difficulty examined. Overall, the implication from our current experiment for the language processing model was that semantic processing might play a more immediate

and important role in Chinese than in some Indo-European languages. A general language comprehension model for cross-linguistic evidence must take the different weights of semantic information in different languages into consideration. Future research using ERP technique to investigate the interplay of semantic and syntactic interference in other languages is needed to confirm my speculation.

4.3 Relations of semantic and syntactic interference effects to WM capacity

To date, there are still debates about the WM mechanisms underlying language comprehension (see Caplan & Waters, 1999, 2014 for a review). One of the goals of the present study was to examine the WM mechanism underlying semantic/syntactic interference resolution in Chinese sentence comprehension, which could potentially inform us about the general role of WM mechanism supporting online sentence comprehension. By using ERP technique, the well-established linguistic and memory effects and the high temporal resolution of ERPs allows us to investigate whether and how certain component(s) of WM capacity play a role in interference resolution.

As I have discussed in section 4.1, the ANs elicited by both high semantic and high syntactic interference conditions at the critical verb between 300 – 500 ms imply that certain WM operations are involved in interference resolution immediately, because that family of ANs has been commonly related to increased WM demands during sentence comprehension in previous studies (Felser, Clahsen, & Münte, 2003; Fiebach, Schlesewsky, & Friederici, 2001, 2002; King & Kutas, 1995; Kluender & Kutas, 1993; Matzke, Mai, Nager, Rüsseler, & Münte, 2002; Münte, Schiltz, & Kutas, 1998).

However, most of these studies just related the observed ANs to general WM process based on the author's theoretical predictions about WM-language relation, instead of systematically examining relations to measures of different aspects of WM capacity (e.g., Kluender & Kutas, 1993; Münte et al., 1998). The many-to-many mapping between hypothesized language/memory processes and ERP components make it hard to draw a firm conclusion. In addition, among the few studies in which subjects' WM capacity was measured, most of them only employed a single WM task – specifically the reading span task (e.g., Fiebach et al., 2002; Friederici et al., 2004; Gunter et al., 2003; King & Kutas, 1995; Van Patten et al., 1997; Vos et al., 2001). Although the internal consistency of the reading span task is high (Conway et al., 2005), its validity and test-retest reliability has been questioned (MacDonald & Christiansen, 2002; Waters & Caplan, 2003). Moreover, recent memory models have argued for a multiple sources view of the variance within WM system (i.e., capacity, attention control, and secondary memory retrieval), and suggested that the relation between WM capacity and higher cognition is multifaceted in that all of the three processes are important (Shipstead & Engle, 2013; Unsworth & Engle, 2007; Unsworth, Fukuda, Awh, & Vogel, 2014). Furthermore, as I have summarized, some researchers have questioned whether the WM-language relation actually derives from other factors, such as linguistic experience (MacDonald & Christiansen, 2002; Perfetti, 2007; Reali & Christiansen, 2007; Van Dyke et al., 2014). As a result, in the current experiment, I included ten individual differences measures tapping different aspects of WM capacity (i.e., phonological STM, semantic STM, and general WM), executive function (i.e., resistance to proactive interference and resistance to prepotent response), processing speed, and linguistic experience, in order to investigate

the unique contribution of certain WM components to specific aspects of language comprehension. Based on our recent results from both a behavioral study and a neuropsychological study in English (Tan et al., 2011, 2013), I expected to observe specific relations between certain WM capacities and interference resolution – the magnitude of individuals’ syntactic interference effect should be predicted by a general WM capacity or executive control ability, and the magnitude of individuals’ semantic interference effect should be predicted by a measure of semantic STM capacity. However, the current experiment did not observe any relation between the semantic interference effect size (as indexed by the mean amplitude of the significant ERPs) and the category probe span, or between the syntactic interference effect size and the general WM capacity, in both correlational analyses and multiple regression models. Subjects’ Stroop performance did not predict their syntactic interference resolution as observed in the previous neuropsychological study (Tan et al., 2013). Instead, in the present study there was a clear effect of resistance to proactive interference on the magnitude of P600 effects elicited by syntactic interference on both the critical verb region (650 – 800 ms) and the following noun phrase (250 – 800 ms). Subjects who were more resistant to proactive interference in the recent negatives tasks, which did not contain sentence level parsing, showed a reduction in the P600 amplitude. In addition, none of the other individual differences measures were predictive of the size of either semantic or syntactic interference related ERP effects in any region.

The finding that P600 effects induced by high syntactic interference in both the critical verb region and the following region were correlated with the recent negative effect provide additional evidence to the speculation that these two positivities might be

one sustained effect reflecting continuous difficulty in resolving syntactic interference. These results are most consistent with the executive control-based account that participants' attention control ability plays a critical role in language processing (Mason, Just, Keller, & Carpenter, 2003; Novick, Trueswell, & Thompson-Schill, 2005; Thompson-Schill, Bedny, & Goldberg, 2005; Vuong & Martin, 2011; Ye & Zhou, 2009). It is interesting that although there were two distinct ERPs (i.e., an LAN and a P600) at the critical verb, both of which were associated with the syntactic interference effect, the recent negatives effect size was only related to the sustained P600-like effect (observed at the critical word and the following word), but not to the LAN effect. These results seemed to be consistent with the assumption that during memory or online sentence processing, executive control ability is exploited to recover from processing difficulty or misinterpretation in the later stage, whereas early processing proceeds via a rapid, parallel, familiarity-based direct retrieval of recent presentations (Fiebach, Schlesewsky, & Friederici, 2001; Hahne & Friederici, 1999; Lewis, Vasishth, & Van Dyke, 2006; Li & Zhou, 2010; Öztekin & McElree, 2010; Van Dyke & McElree, 2006). In an ERP study, Hahne and Friederici (1999) observed a similar ELAN-P600 compound. They found that the occurrence of the P600 effect was modulated by the proportion of incorrect trials (i.e., 20% vs. 80%) - P600 effect was only observed in the low proportion condition, whereas the ELAN effect was consistently observed in both low and high proportion conditions. Hahne and Friederici suggested that the finding that P600 was only observed in the 20% incorrect sentences condition confirmed the assumption that early and late ERP components could be characterized as a fast, automatic process and a slow, controlled processes, respectively. In line with Hahne and Friederici's study, I suggest that the

possible two-stage interference resolution process observed in the current experiment - the early process of detecting/diagnosing syntactic interference and the later process of actual resolving syntactic interference - might be supported by different underlying mechanisms, with executive control only involved in the second-stage process, while subjects with poor and good attentional control ability (as least as indexed by the individual differences measurements I have included) were equally sensitive to proactive interference. The assumption about a link between attentional control ability and late stage sentence processing is supported by neuropsychological findings showing that lesions in the left basal ganglia, which has been demonstrated to be involved in attentional control, only affected late syntactic integration processes (Friederici, Kotz, Werheid, Hein, & von Cramon, 2003; Moro et al., 2001; Ullman, 2001). For example, Friederici et al. (2003) reported that Parkinson's disease patients with focal basal ganglia lesions showed relatively normal early automatic syntactic processes (as indexed by anterior negativity) but poor late syntactic integration processes (as indexed by P600) as compared to the age-matched controls.

More relevant to the purpose of our current study, further evidence for the dissociation of early and late stages of interference resolution has been reported in a recent memory study using a SAT paradigm (Oztekin & McElree, 2010). Oztekin and McElree (2010) related individuals' general WM capacity (as measured by reading span and operation span tasks) to retrieval dynamics in a short-term recognition task, in which there were both recent and non-recent negative trials as I have manipulated in the recent negatives task. The results demonstrated that individuals' WM capacity did not predict their recognition performance at early time points when judgments were mainly based on

familiarity, indicating that both low- and high-WM capacity subjects were equally susceptible to proactive interference. However, for the recent negative trials only, individuals' WM capacity predict their performance at later time points where specific episodic information was retrieved and combined with familiarity in making a judgment, with the low span group taking longer to show evidence of this controlled retrieval of episodic information. Oztekin and McElree hypothesized that the low WM capacity subjects' delayed use of episodic information was either the result of later deployment of controlled episodic retrieval processes or a delay in resolving the interference between familiarity and episodic information. In the current experiment, I did not examine the variation of onset latency or peak latency for each ERP component, which might be more directly correspond to the processing speed parameters in the SAT paradigm. In a previous study, Federmeier and Kutas (2005) found that older adults generally showed a delayed N400 compared with young adults, and subjects with higher WM span (as measured by reading span task) showed shorter peak latency of the N400 effect induced by unexpected sentence continuation. Therefore, it might be worth examining the relations between general WM capacity and onset/peak latency of relevant ERP effects. However, the finding that there was no relation between the size of early ERP effect and any individual differences measures, but subjects who were more sensitive to proactive interference showed increased amplitude of the P600 effect was consistent with Oztekin and McElree's assumption that low capacity subjects had more difficulty in the late, controlled interference resolution process.

The assumption that the controlled process is only involved in the later stage of interference resolution also helps reconcile the results from several experiments

conducted in our lab (Tan et al., 2011, 2013). According to this line of reasoning, I suggested the early interference effects may be a direct result of degradation of semantic and syntactic memory representation due to a loss of the fidelity of representations – resulting from, for example, a loss of specific features due to decay over time (Barrouillet, Bernardin, & Camos, 2004; Barrouillet, Portrat, Vergauwe, Diependaele, & Camos, 2011) or to feature overwriting among the memory items themselves (Lewandowsky & Oberauer, 2009; Lewandowsky, Oberauer, & Brown, 2009; Nairne, 2002). In addition, I suggest that variation in the degradation rates might also affect the interference resolution at later points, as some of the previous studies demonstrated that the quality of lexical representation constrains the success of controlled retrieval (Perfetti, 2007; Van Dyke et al., 2014). Therefore, one would expect to observe a relation between the overall semantic/syntactic interference effect size and the ability to actively maintain semantic/syntactic information, respectively. This assumption is confirmed by our previous finding that subjects' category probe span, which reflects the different rates of degradation of semantic representation between subjects, predicted their overall semantic interference effect size in both healthy young population and aphasic patients (Tan et al., 2011, 2013). Although unfortunately, in the current ERP experiment, due to the lack of a valid measurement for semantic and syntactic STM in Chinese, I was not able to examine the relation between semantic/syntactic STM and the magnitude of semantic/syntactic interference induced ERP components.

With respect to the later stage of interference resolution, during which controlled processes are involved, one would expect to observe a relation between individuals' attentional control ability and the interference effect size, which does not occur in the

early stage. However, few of the previous relevant studies (were able to) dissociate the late stage process from the overall process, as most of these studies used a self-paced reading paradigm (e.g., Tan et al., 2011, 2013; Van Dyke et al., 2014), which only reflects overall processing. But the observation of a link between attentional control and an interference effect in dependent measures reflecting overall processing could support the role of attentional control to some extent. Indeed, in a series of studies conducted in our lab (Tan et al., 2011, 2013; Vuong & Martin, 2014), we confirmed the important role of attentional control in interference resolution during sentence processing. Although there was variation in what specific measure of attentional control ability has the greatest predictive power due to the many differences in sentence materials and experimental paradigms across studies, I suggest that the different WM-interference relations observed from different experiments could be reconciled in a way that they all implied a general link between attentional control and interference resolution. A tentative account for the specificity of the relations can be derived from the interaction of two factors: first, the different stages of sentence processing were investigated in different experiment as I have just discussed, such as the overall online processing as indexed by the phrase-by-phrase self-paced RT data (Tan et al., 2011), the overall offline process as indexed by the accuracy data in the neuropsychological study (Tan et al., 2013), and the separable stages of online process as indexed by the ERP effect in the current experiment. Different cognitive mechanisms might become more predominant in different stages. The second factor that determines the specificity of the WM-language relation might be the weights each WM/EF task allocated to the each of the three factors. Based on the recent WM models (Shipstead et al., 2014; Unsworth et al., 2014), I suggest that the predictive power

of all these WM/EF tasks we used is accounted for by subjects' variation in capacity, attentional control, and secondary memory retrieval, though the way these three factors combined differed. For example, the recent negatives task might be more strongly related to the attentional control factor (specifically the resistance to proactive interference, which is different from the resistance to prepotent response interference as measured by Stroop (Crowther & Martin, 2014; Friedman & Miyake, 2004), although it also requires maintenance and retrieval constructs. The Stroop task also relies heavily on attentional control, though it might be less related to the capacity and retrieval components than the recent negatives task, for the reason that subjects are not required to temporarily maintain a list of items as in the Stroop task. Last, complex span performance is strongly related to all of the three constructs because both storage (e.g., maintain the letter for later recall) and processing components (e.g., read the sentences) as tapped by the span tasks are supported by all of the three factors (Unsworth et al., 2014). Thus, subjects' performance on these WM/EF tasks could be either correlated or relative independent, depending on the specific experimental design. For instance, a significant correlation between Stroop effect and WM capacity as operationalized by complex span tasks was reported in some experiment (Engle, 2002; Kane, Bleckley, Conway, & Engle, 2001; Kane & Engle, 2003) but not others (Engle, 2002; Shipstead et al., 2014), and a dissociation between resistance to prepotent interference (e.g., as measured by Stroop) and to proactive interference (e.g., as measured by recent negatives task) has been reported in some studies (Crowther & Martin, 2014; Friedman & Miyake, 2004) but not others (Pettigrew & Martin, 2014). The investigation of the independent and jointly contribution of each of the three factors to the

WM-language relation is beyond the scope of our current project, given the very limited sample size I had and the limited sample of WM/EF tasks I employed.

Moreover, it is very likely that the attentional control ability underlying the later stage of interference resolution supports both semantic and syntactic interference resolution. This assumption is supported by our previous finding of links between the attentional control component of WM capacity and both syntactic interference and semantic interference (Tan et al., 2011). We observed a link between individuals' attentional control ability (as measured by reading span and operation span) and both semantic interference (e.g., in the comprehension question accuracy) and syntactic interference (e.g., in the self-paced reading times) resolution ability in different regions. We suggested that the WM-semantic and -syntactic interference relation was observed at different time points possibly because the attention control ability predominately relates to the type of interference being resolved at the moment at that point (see Tan et al., 2011 for more discussion). The absence of any link between semantic interference and WM/EF capacity might be a result of differences in task and material presentation between the current study and previous studies. In sum, a general conclusion derived from all these studies is that there is a specific link between attentional control and the resolution of interference effect during sentence processing, though the exact nature of this relation requires future investigation. I argue that the finding of a relation between WM and interference resolution measures would be consistent with the cue-based parsing approach, if such relations could be appropriately attributed to the attentional control component or perhaps the secondary retrieval aspect of WM measures. Future work should include specific measures of long-term memory retrieval ability in order to

determine if these account for the relation between WM and sentence processing measures.

Conclusion regarding the WM mechanisms underlying interference resolution. By relating the magnitude of each significant ERP effect to individual difference measures, I found that subjects' performance in the recent negatives task predicted the magnitude of the P600 effect induced by the syntactic interference effect in both the critical verb region and the spillover region. Subjects who were more resistant to proactive interference have less difficulty in syntactic interference resolution, as they showed a reduced mean amplitude difference between high and low interference conditions. Taken together with our previous findings that the size of syntactic interference effect could be predicted by the attentional control ability as indexed by WM composite score (Tan et al., 2011) and the Stroop effect (Tan et al., 2013), I suggest that there is a general link between attentional control and interference resolution, which is evident in the late stage controlled process of interference resolution aligning with what Oztekin and McElree found in a memory recognition task (2010). However, the nature of the relation is inconclusive, because of the multifaceted nature of WM (i.e., capacity, attention control, and secondary memory constructs), the different stages of interference resolution I investigated, the different languages examined, and so on. Regarding semantic interference resolution, my results did not replicate the previous findings of a very specific relation between semantic STM and semantic interference resolution in online RTs, or a relation between general WM capacity and semantic interference resolution in offline comprehension question accuracy. I suggest that the underlying mechanisms for

semantic interference resolution could not be addressed here adequately due to the lack of a good measure of semantic STM capacity.

4.4 Chinese relative clause processing

One of the original motivations for us to investigate interference effects in Mandarin Chinese was that an ORC-advantage during filler-gap integration in RC processing has been reported in both behavioral studies (Chen, Ning, Bi, & Dunlap, 2008; Gibson & Wu, 2013; Hsiao & Gibson, 2003; Vasishth, Chen, Li, & Guo, 2013) and an ERP study (Packard et al., 2011). Although there has been contradictory evidence arguing for an ORC-advantage in Chinese as in many other languages (Lin, 2006; Lin & Bever, 2006; Vasishth et al., 2013; Yang, Perfetti, & Liu, 2010), for the purposes of the current experiment, the potential ORC-advantage would have further ensured (in addition to the manipulation of inserting an adverbial phrase between the head noun and the critical verb) that the greater difficulty in the processing critical verb following an ORC structure (i.e., high syntactic interference condition) was not just a spillover effect from the RC processing differences, because the ORC structure itself was assumed to cause less processing difficulty. Therefore, in addition to the critical verb and following words, I have also examined ERP effects in the head noun region and the adverbial phrase region. As expected, at the adverbial phrase, although there was a significant interaction of Syntactic $\times$ Regions $\times$ Hemisphere between 150 – 400 ms, there was no effects in the (-400) – 0 ms time window prior to the critical verb onset. Therefore, the ERPs we observed on the main verb could not be simply accounted for as a continuous effect

which spilled over from the RC region. On the RC head noun, there was an ORC-elicited positive effect (300 – 650 ms) in the posterior regions that peaked around 300 ms.⁴⁴

This ORC-elicited posterior positivity was different from either the ORC-elicited frontal-central sustained negativity (250 – 800 ms) reported in Yang et al.'s study (2010) or the SRC-elicited P600 reported in Packard et al.'s study (2011). There are many differences in the sentence materials between the present study and the two earlier studies - for instance, the sentences used in both Yang et al. and Packard et al.'s studies had an issue of temporary structural ambiguity⁴⁵ to some extent and the critical verb was the final region in Packard et al.'s study. Thus, I will not focus on how to reconcile these results. Based on the time course and scalp distribution, I suggest that the posterior distributed positivity I observed to the ORC structure was likely to be a P300 effect, which has been traditionally been associated with the occurrence of rare events embedded in a sequence of events (i.e., oddball-paradigm, such as go/no go task) (Enriquez-Geppert, Konrad, Pantev, & Huster, 2010; Polich, 2007)⁴⁶. In psycholinguistic studies,

⁴⁴ On the RC head noun, there was also a semantic interference-elicited late negativity along the midline between 650 – 800 ms. As discussed in section 3.2.4.1, I suggest that it reflected a second-pass semantic integration difficulty as observed on the later critical verb region in the current experiment, and reported in many previous studies (Baggio, Van Lambalgen, & Hagoort, 2008; Friederici, Steinhauer, & Frisch, 1999; Jiang, Li, & Zhou, 2013; Politzer-Ahles, Fiorentino, Jiang, & Zhou, 2013; Zhang, Yu, & Boland, 2010). The semantic interference effect did not interact with RC types. The following discuss will not focus on this ERP effect, because only the syntactic processing differences between the two RC types were relevant to the WM-language relations, of which I am interested in.

⁴⁵ In addition, it should be noted that a number of previous studies demonstrated the processing differences between object modification and subject modification. For example, most of the studies reported an ORC advantage employed subject modification sentences (e.g., Gibson & Wu, 2011; Hsiao & Gibson, 2003); while most of the studies reported an SRC advantage employed object modification sentences (e.g., Lin & Bever, 2006). However, the sentence structure we used here was more complicated. The relative clause was embedded in a sentential complement. Thus, it is hard to make a direct comparison to most of the previous studies.

⁴⁶ Previous studies have shown that P300 is not a monolithic component, but a composite waveform made up of at least three distinct components: the frontal-central P3a, the centro-parietal P3b, and a longer duration late positivity slow wave (see Osterhout & Hagoort, 1999 for a review). Given the posterior distribution, we suggested that the one we observed here is more like a P3b, which is the most classical

although there has been a long-lasting debate about whether the P300 and the P600 are two independent effects or not (see Friederici et al., 2001 and Osterhout & Hagoort, 1999 for a review), the P300 effect has been associated with the early diagnosis/detection of temporary ambiguity (Friederici, Mecklinger, Spencer, Steinhauer, & Donchin, 2001; Hahne & Friederici, 1999; Li & Zhou, 2010) and sensitivity to word/structure frequency (Heine, Tamm, Hofmann, Hutzler, & Jacobs, 2006; Polich & Donchin, 1988). On the head noun of the current experimental sentences, there should be no structural ambiguity in any sentence type and little difference in the degree of interference across conditions⁴⁷. The capacity-based model of sentence processing such as the DLT (Gibson, 1998, 2000) would predict that one would observe an ORC-advantage due to greater cost due to the longer distance of integration, while the experience-based account would predict an SRC-advantage due to higher frequency of the preceding SRC structure. The observation of a larger amplitude of ERP component associated with the ORC structure seems to favor the experience-based account, indicating that subjects had more processing difficulty on the head noun when they needed to integrate it with the gap position in a low frequency structure (ORC) than in a high frequency structure (SRC). According to the experience-based account, syntactic processing is constrained by a wide variety of probabilistic factors at the lexical, syntactic, semantic, and context levels (MacDonald & Christiansen, 2002; Misyak & Christiansen, 2012; Reali & Christiansen, 2007; Wells, Christiansen, Race, Acheson, & MacDonald, 2009). Although the role of probabilistic factors has been

P300 with largest amplitude. However, the sub-category of the P300 effect did not affect our conclusion here.

⁴⁷ According to the cue-based parsing approach, when encountered the head noun, readers need to retrieve earlier information (e.g., embedded verb, embedded noun) to integrate with it. As there was only one word satisfy each retrieval cue (the influence from the first NP, e.g., “*reporter*” in Example 13, and its verb, e.g., “*heard*” in Example 13, should be balanced across conditions), I suggest that there should be little interference effect on the head noun.

mostly studied in the context of ambiguity resolution, a recent study conducted by Wells et al. (2009) demonstrated that increasing exposure to equal amount of ORC and SRC structures increased reading speeds for ORC more than for SRC. They suggested that such results were consistent with the experience-based account, which predicts that increasing experience has more benefit on less regular sentence types (i.e., ORC structure, which did not adhere closely to the overwhelmingly subject-verb-object word order in English), and that experience is a powerful factor in determining sentence processing efficiency.

The speculation that the SRC-advantage observed here might support the experience-based account was further strengthened by the fact that the ORC-elicited P300 effect has been shown to be a function of the probability of the stimulus, with the less plausible stimulus associated with larger amplitude (Polich & Donchin, 1988; Ruchkin et al., 1990). Therefore, at least in the current experiment, when the problem of temporary structural ambiguity has been controlled to a large extent, structural frequency seemed to have a stronger influence on the RC processing than distance differences between the linguistic dependencies. However, I would not generalize the current results to argue against the WM-based account, as many other studies have observed a robust locality effect during sentence processing as predicted by Gibson's theory (e.g., Bartek et al., 2010), and Packard et al (2011) did observe an SRC-elicited P600 revealing greater syntactic revision difficulty in the SRC structure. Instead, I suggest that both low structural frequency and distal integration combine to increase sentence processing difficulty. In the current experiment, it is possible that the overwhelming frequency advantage of SRC structures (SRC: 82% vs. ORC: 18%, Jäger et al., 2015) concealed the

relative small distance advantage of ORC structures (e.g., the distance between head noun and embedded verb, SRC: 3 words vs. ORC: 1 words). A potential problem with this explanation is that it has difficulty in accounting for the lack of relation between subjects' linguistic experience (e.g., measured by the vocabulary, category probe, and ART) and the size of the P300 effect. One possible explanation is that the measurements I included for tapping subjects' linguistic experience are not sensitive predictor for the moment-to-moment sentence comprehension, as previous studies mainly reported a relation between the vocabulary or ART and overall offline sentence processing ability, such as accuracy in answering comprehension question or verbal SAT/ACT (Acheson, Wells, & MacDonald, 2008; Moore & Gordon, 2014; Stanovich & West, 1989).

Moreover, Misyak and Christiansen (2012) have shown that after controlling for many other factors (e.g., verbal WM, STM, motivation, IQ), vocabulary and ART were no significant longer predictors for comprehending relevant types of sentence. So far, very few studies directly examined individual differences in linguistic experience and sentence comprehension. In a recent study, Misyak and Christiansen (2012) investigated the relationship between linguistic experience and language processing through an individual differences approach. They tested subjects on a variety of individual differences measures, including verbal WM (i.e., reading span), digit span, vocabulary, ART, fluid intelligence, cognitive motivation, and statistical learning ability as measured by the artificial grammar learning tasks.⁴⁸ Regarding the prediction from experience-based

⁴⁸ In a typical artificial grammar learning task, there are usually two phases. In the study phase, subjects were passively exposed to a set of brief strings generated by an artificial grammar (e.g., continuous sequences of non words from an artificial lexicon). Then in the testing phase, subjects were required to generalize the appropriate regularities to new materials, such as judging the grammaticality of some new strings.

account,⁴⁹ they did not find a correlation between vocabulary and sentence comprehension, while the relation between ART and sentence comprehension disappeared after controlling for all the factors in a regression model. Instead, Misyak and Christiansen found that statistical learning ability was the best predictor for online sentence comprehension, which remained significant after partialling out all the other actors. Although the limited sample size ($N = 30$) greatly constrained their statistic power, the implication from Misyak and Christiansen's results for my current study is that it might be that statistical learning ability (which was not measured in my current experiment), but not receptive vocabulary or ART performance (which I have measured) reliably predicts subjects' sentence processing performance efficiency. Therefore, no relation between vocabulary/ART and sentence processing was observed in the current study. However, Misyak and Christiansen did not further discuss the differences in the cognitive constructs as measured by the vocabulary, ART, and artificial grammar learning tasks. What and how linguistic experience interacts with WM capacity is an issue that cannot be concluded from the limited evidence of Misyak and Christiansen's and my current experiment, and future large-scale studies that simultaneously manipulate structural frequency and integration distance, and measure individual differences from more reliable tests (or using latent variables) are required.

⁴⁹ Misyak and Christiansen (2012) did observe a link between verbal WM and language comprehension, though it became marginal significant after controlling for other factors in the regression models. However, I suggest that the problem of including only reading span task as the verbal WM measures made their conclusion about WM-language relation questionable. Misyak and Christiansen also admitted that "any comparisons between statistical learning and verbal WM/STM measures are limited by the tasks used to assess them (p 320)".

4.5 Limitations of the study and future directions

Although the current study observed some reliable and interesting results concerning how interference effects impede Chinese sentence comprehension, as well as the nature of the WM mechanisms underlying language processing, there are some limitations of this study that need to be controlled for or further examined in future studies. First, there are some concerns about the statistic analysis conducted in the current project. Regarding the correlational and regression results, the reliability of these tests is greatly constrained by the limited sample size and the reliability of the ERP effects. Ideally, a larger sample size like that used in our previous behavioral study ($N = 120$, Tan et al., 2011) is necessary for obtaining more a robust and generalizable conclusions about the nature of the WM-language relation. Moreover, although it has been widely demonstrated that averaged EEG data, which was analyzed following standard procedures, is highly reliable (Lopez-Calderon & Luck, 2014; Picton et al., 2000), the low signal-to-noise ratio of individual EEG trials prevents researchers from conducting mixed-effects model analyses on the EEG data. In recent years, psycholinguists have emphasized the importance of using mixed-effects models as an alternative to the repeated ANOVA measures, because the former method allows simultaneous adjusting for random effects caused by subjects and item variations and has been demonstrated to be more powerful than traditional techniques (Baayen, 2008; Baayen, Davidson, & Bates, 2008). Many methods have been proposed to improve the reliability of single-trial EEG recoding, such as using independent component analysis (ICA; I have only used it to identify and remove eye-movement in the current study) to blindly separate stimulus- or response-locked EEG activities into separate components (Jung et al., 2001). It is

necessary to combine these advanced EEG analysis techniques and statistical methods in future EEG studies. Last, a general criticism of using an individual differences method in psychological studies is the multiple comparisons problem (Ryan, 1959; Wilkinson, 1999). The multiple comparisons problem occurs when a set of statistical analysis is conducted simultaneously, which causes Type I errors to increase (i.e., the chance of obtaining at least one statistically significant result is greater than 5%). Many methods have been proposed to correct for multiple comparisons, such as the Bonferroni correction of the p -value and empirical Bayes methods. However, it is arguable how to balance the multiple comparisons problem and motivated analyses. Many other researchers have argued that the interpretation of well-intentioned, theoretically motivated comparisons should be less subject to multiple comparisons problems and further correction is not necessary. For example, in my current experiment, it was necessary to include all the individual difference measures, which were motivated by different theories, because I am interested in which account provides the best fit to my current data. Since I have run 35 simple correlation analyses in section 3.3.3, the corrected p -value according to the Bonferroni correct would be .001, which is a very stringent correction. None of the correlations would reach significance with the corrected p -value. Therefore, although the Bonferroni correction may have controlled for Type I error, it might, at the same time, conceal the potential theoretically important link between attentional control and language processing, which was confirmed in the later multiple regression analysis which included several predictors simultaneously, reducing the multiple comparisons problem (see section 3.3.4). In the future, employing a larger sample size and more valid individual differences might help researchers observe more

robust and reliable WM-languages relations, which could survive the multiple comparison correction.

Second, it should be noted that any comparisons between the different cognitive abilities or skills are limited by the reliability and validity of the tasks used to assess them. Although I have used well-established individual differences measures as possible and computed latent variables for some cognitive constructs in the regression analysis, the reliability and validity of each task is always a concern in individual differences studies. Though the reliability of each task was high (see Table 10), the results that some tasks have higher reliability than others make the beta weights for the later predictors in regression analysis misleadingly attenuated (Misyak & Christiansen, 2012). Specifically, a problem of the current study is that I did not have valid measures for subjects' semantic and syntactic STM, which are critical for testing the hypothesis derived from the multiple capacities account (Martin et al., 1996, 1999, 2004). It is a cross-linguistic problem that there is no appropriate empirical test for syntactic STM at this time. For semantic STM, a modified Chinese version of the category probe task was shown to lack validity. In Chinese, the category probe span might reflect more of subjects' long-term knowledge, instead of the semantic STM capacity as it was designed for, as suggested by its greater correlation with subjects' linguistic experience measures than with other WM measures. Further investigation with a valid semantic STM measure is needed to determine whether semantic STM could predict subjects' semantic interference resolution efficiency in Chinese, as we have found in previous English studies (Tan et al., 2011, 2013). In addition, the findings that even a small change in the experimental materials (e.g., the ratio of congruent vs. incongruent trials in the Stroop task) affects the underlying

mechanisms tapped by the task pose a problem in the comparison between different studies. It is necessary to pin down the exact cognitive capacity measured by each individual difference task.

Third, based on the results from previous studies, there are several other analyses worth trying with the current EEG data, such as examining relations between different individual differences measures and the onset/peak latency of the significant ERPs, which were demonstrated to be sensitive to verbal WM and vocabulary capacity in some studies (Federmeier & Kutas, 2005; Federmeier, McLennan, Ochoa, & Kutas, 2002).

Additionally, although examining ERPs time-locked to each individual word could provide evidence for moment-to-moment processes, some researchers have argued that multiple words analysis (e.g., analyzing ERPs spanning a whole clause as one epoch) could better detect the WM operations supporting temporary maintenance of unintegrated structural information before completing the linguistic dependency (Fiebach et al., 2002; King & Kutas, 1995). For the current project, multiple words analysis might help test my speculation that the P600-like effect in the spillover region is a spillover effect from the critical region.

Last, as mentioned earlier, in order to make a direct comparison about the time course difference of semantic and syntactic processes between Chinese and English, a comparable ERP study in English is required. Most of my discussion about the cross-linguistic differences in time course between semantic and syntactic interference resolutions was based on the ERP results from my current study and behavioral/fMRI results from previous English studies. So far, no English study has examined interference effects using a technique with high temporal resolution such as ERP. The conclusion

about the interaction of semantic and syntactic processes in English was mainly drawn from studies examining semantic and syntactic processes using a violation or ambiguity paradigm. The WM/EF mechanisms underlying the violation or ambiguity conditions might differ from that underlying the natural sentence parsing. Therefore, an English ERP study investigating the exact time course of semantic and syntactic interference resolution is necessary to address the cross-linguistic differences.

4.6 Conclusion

The results of this study confirmed findings from previous English studies demonstrating syntactic and semantic interference from unavailable constituents during sentence processing, and provided cross-linguistic evidence supporting the cue-based parsing approach. The novel aspect of the present study was the investigation of the interplay and time course of the two types of interference effects using the high-temporal resolution ERP technique. The current results differ from previous English studies in showing that the failure of syntactic licensing did not prevent further semantic processing, as the semantic interference effect was observed even when the distractor's syntactic features strongly eliminated it from the distractor set. More importantly, both syntactic and semantic interference effects elicited an early AN effect starting around 300 ms at the critical verb without a time course difference. These results imply that semantic processing might play a more immediate and important role in Chinese than in English, because the semantic interference effect was delayed and even blocked by the distractor's unsuitable syntactic features in previous English studies using eye-tracking with temporal precision as well. In addition, for the syntactic interference effect, the left lateralized AN (i.e., LAN) effect was followed by a P600 effect, while for semantic interference effect,

the bilateral distributed AN effect was followed by a late AN (or maybe just one sustained anterior effect). I suggest that the observation of a LAN-P600 and an AN-late AN ERP patterns for different interference effects, respectively, is consistent with a two stage model of sentence processing where the first stage consists of diagnosis and the second stage consists of actual reanalysis.

Regarding the nature of the WM-language relations, this study has demonstrated the predictive value of attentional control for individual differences in syntactic interference effects during sentence comprehension. Subjects who were more sensitive to proactive interference (as measured by the recent negatives task), had more difficulty in the late stage syntactic interference resolution, as indexed by an increased mean amplitude of the P600 effect. Taken together with the previous findings that the magnitude of syntactic interference effects could be predicted by subjects' general WM capacity or the Stroop effect, even after controlling for their verbal knowledge, I suggest that there is a general link between attentional control and syntactic interference resolution, which is evident in the late stage controlled process of interference resolution. However, the nature of this link remains speculative to the extent that the nature of general WM is speculative.

Last, with respect to the RC processing differences in Chinese, a P300-like effect induced by the ORC-disadvantage was observed. This positivity may reflect integration difficulty due to the overwhelming frequency differences of SRC structure over ORC structure. However, as the experience-based account would predict processing difficulty for the less frequent structure (i.e., ORC) as soon as readers recognized this structure, such as during the early part of the RC region (which was not examined in the current

experiment due to the limited number of artifact-free trials), future research specifically examining the RC processing differences in the RC regions, and the underlying cognitive abilities/skills (e.g., linguistic experience, statistical learning ability, WM capacity) are necessary. However, the current results are more consistent with the experience-based account than the WM-based account by showing an SRC-advantage during head noun integration.

References

- Acheson, D. J., Wells, J. B., & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods*, 40, 278-289.
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390-412.
- Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual review of psychology*, 63, 1-29.
- Baddeley, A., Logie, R., Nimmo-Smith, I., & Brereton, N. (1985). *Components of fluent reading*. *Journal of memory and language*, 24, 119-131.
- Baddeley, A. D., & Hitch, G.J. (1977). Recency re-examined. In S. Dornic (Ed.), *Attention and performance VI* (pp. 647-667). Hillsdale, NJ: Erlbaum.
- Bader, M., & Lasser, I. (1994). German verb-final clauses and sentence processing: Evidence for immediate attachment. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 225–242). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Baggio, G., Van Lambalgen, M., & Hagoort, P. (2008). Computing and recomputing discourse models: An ERP study. *Journal of Memory and Language*, 59, 36-53.
- Bai, X., Yan, G., Liversedge, S. P., Zang, C., & Rayner, K. (2008). Reading spaced and unspaced Chinese text: evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 1277-1287.
- Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, 133, 83-100.
- Barrouillet, P., Portrat, S., Vergauwe, E., Diependaele, K., & Camos, V. (2011). Further evidence for temporal decay in working memory: reply to Lewandowsky and Oberauer (2009). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 1302-17.
- Bartek, B., Lewis, R. L., Vasishth, S., & Smith, M. R. (2011). In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 1178-1198.
- Boland, J. E. (1997). The relationship between syntactic and semantic processes in sentence comprehension. *Language and Cognitive Processes*, 12, 423-484.
- Boland, J. E., & Blodgett, A. (2001). Understanding the constraints on syntactic generation: Lexical bias and discourse congruency effects on eye movements. *Journal of Memory and Language*, 45, 391-411.

- Bornkessel, I. D., Fiebach, C. J., & Friederici, A. D. (2004). On the cost of syntactic ambiguity in human language comprehension: an individual differences approach. *Cognitive Brain Research*, 21, 11-21.
- Brass, M., Ullsperger, M., Knoesche, T. R., Von Cramon, D. Y., & Phillips, N. A. (2005). Who comes first? The role of the prefrontal and parietal cortex in cognitive control. *Journal of Cognitive Neuroscience*, 17, 1367-1375.
- Brown, C., Brown, P., & Hagoort. (1993). The Processing Nature of the N400: Evidence from Masked Priming. *Journal of Cognitive Neuroscience*, 5, 34-44.
- Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 10, 12-21.
- Butterworth, B., Campbell, R., & Howard, D. (1986). The uses of short-term memory: A case study. *The Quarterly Journal of Experimental Psychology*, 38, 705-737.
- Cai, R., Dong, Y., Zhao, N., & Lin, J. (2015). Factors contributing to individual differences in the development of consecutive interpreting competence for beginner student interpreters. *The Interpreter and Translator Trainer*, 9, 1-17.
- Camos, V., Mora, G., & Barrouillet, P. (2013). Phonological similarity effect in complex span task. *The Quarterly Journal of Experimental Psychology*, 66, 1927-1950.
- Camos, V., Mora, G., & Oberauer, K. (2011). Adaptive choice between articulatory rehearsal and attentional refreshing in verbal working memory. *Memory & Cognition*, 39, 231-244.

- Caplan, D., & Futter, C. (1986). Assignment of thematic roles to nouns in sentence comprehension by an agrammatic patient. *Brain and Language*, 27, 117-134.
- Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, 22, 77-94.
- Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language*, 3, 572-582.
- Chan, A., & Lee, P. (2005). Effect of display factors on Chinese reading times, comprehension scores and preferences. *Behaviour & Information Technology*, 24, 81-91.
- Chen, B., Ning, A., Bi, H., & Dunlap, S. (2008). Chinese subject-relative clauses are more difficult to process than the object-relative clauses. *Acta Psychologica*, 129, 61-65.
- Chen, E., West, W. C., Waters, G., & Caplan, D. (2006). Determinants of BOLD signal correlates of processing object-extracted relative clauses. *Cortex*, 42, 591-604.
- Chincotta, D., & Underwood, G. (1997). Digit span and articulatory suppression: A cross-linguistic comparison. *European Journal of Cognitive Psychology*, 9, 89-96.
- Chu, C. C. (1998). *A discourse grammar of Mandarin Chinese*. New York, NJ: Peter Lang Publishing Inc.

- Colman, K. S., Koerts, J., Stowe, L. A., Leenders, K. L., & Bastiaanse, R. (2011). Sentence comprehension and its association with executive functions in patients with Parkinson's disease. *Parkinson's Disease*, 2011, 1-15.
- Conrad, R., & Hille, B. A. (1958). The decay theory of immediate memory and paced recall. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 12, 1-6.
- Conway, A. R. A., Cowan, N., Bunting, M. F., Theriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, 30, 163-183.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12, 769-786.
- Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, 1, 42-45.
- Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87-114.
- Cowan, N., Elliott, E. M., Scott Saults, J., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, 51, 42-100.

- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests.
Psychometrika, 16, 297-334.
- Cronbach, L. J., & Furby, L. (1970). How we should measure "change": Or should we?
Psychological Bulletin, 74, 68-80.
- Crowther, J. E., & Martin, R. C. (2014). Lexical selection in the semantically blocked cyclic naming task: the role of cognitive control and learning. *Frontiers in Human Neuroscience*, 8, 1-20.
- Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450-466.
- Daneman, M., & Hannon, B. (2007). What do working memory span tasks like reading span really measure? In N. Osaka, R. H. Logie, & M. D'Esposito (Eds). *The Cognitive Neuroscience of Working Memory* (pp. 21-42), Oxford, England: Oxford University Press.
- Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134, 9-21.
- Ditman, T., Holcomb, P. J., & Kuperberg, G. R. (2007). An investigation of concurrent ERP and self-paced reading methodologies. *Psychophysiology*, 44, 927-935.
- Engelmann, F., Jäger, L., & Vasishth, S. (2015) Cue confusion and distractor prominence explain inconsistent effects of retrieval interference in human sentence processing.

28th CUNY conference on human sentence processing, University of Southern California, LA.

- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11, 19-23.
- Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: a test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 972-992.
- Engle, R. W., Nations, J. K., & Cantor, J. (1990). Is "working memory capacity" just another name for word knowledge? *Journal of Educational Psychology*, 82, 799-804.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General*, 128, 309-331.
- Enriquez-Geppert, S., Konrad, C., Pantev, C., & Huster, R. J. (2010). Conflict and inhibition differentially affect the N200/P300 complex in a combined go/nogo and stop-signal task. *Neuroimage*, 51, 877-887.
- Federmeier, K. D., & Kutas, M. (2005). Aging in context: Age-related changes in context use during language comprehension. *Psychophysiology*, 42, 133-141.
- Federmeier, K. D., McLennan, D. B., Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken

language processing by younger and older adults: An ERP study.

Psychophysiology, 39, 133-146.

Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory and Language*, 54, 541-553.

Fedorenko, E., Gibson, E., & Rohde, D. (2007). The nature of working memory in linguistic, arithmetic and spatial integration processes. *Journal of Memory and Language*, 56, 246-269.

Felser, C., Clahsen, H., & Münte, T. F. (2003). Storage and integration in the processing of filler-gap dependencies: An ERP study of topicalization and *wh*-movement in German. *Brain and Language*, 87, 345-354.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348-368.

Fiebach, C. J., Schleewsky, M., & Friederici, A. D. (2001). Syntactic working memory and the establishment of filler-gap dependencies: Insights from ERPs and fMRI. *Journal of psycholinguistic research*, 30, 321-338.

Fiebach, C. J., Schleewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language*, 47, 250-272.

- Fiebach, C. J., Schlesewsky, M., Lohmann, G., Von Cramon, D. Y., & Friederici, A. D. (2005). Revisiting the role of Broca's area in sentence processing: syntactic integration versus syntactic working memory. *Human brain mapping, 24*, 79-91.
- Fodor, J. D., & Frazier, L. (1980). Is the human sentence parsing mechanism an ATN? *Cognition, 8*, 417-459.
- Frazier, L., 2002. Sentence processing. In: Altmann, G. (Ed.), *Psycholinguistics: Critical Concepts in Psychology, vol. 3*. Routledge, London.
- Friederici, A. D. (1998). Diagnosis and reanalysis: Two processing steps the brain may differentiate, in: J.D. Fodor, F. Ferreira (Eds.), *Reanalysis in Sentence Processing* (pp. 177 – 200). Kluwer Academic Publishers, Dordrecht.
- Friederici, A., D., Von Cramon, D. Y., & Kotz, S. A. (1999). Language related brain potentials in patients with cortical and subcortical left hemisphere lesions. *Brain, 122*, 1033-1047.
- Friederici, A., D., Wang, Y., Herrmann, C. S., Maess, B., & Oertel, U. (2000). Localization of early syntactic processes in frontal and temporal cortical areas: a magnetoencephalographic study. *Human brain mapping, 11*, 1-11.
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Science, 6*, 78-84.
- Friederici, A. D., & Frisch, S. (2000). Verb argument structure processing: The role of verb-specific and argument-specific information. *Journal of Memory and Language, 43*, 476-507.

- Friederici, A. D., Gunter, T. C., Hahne, A., & Mauth, K. (2004). The relative timing of syntactic and semantic processes in sentence comprehension. *Neuroreport*, *15*, 165-169.
- Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). The temporal structure of syntactic parsing: Early and late effects elicited by syntactic anomalies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1219-1248.
- Friederici, A. D., Kotz, S. A., Werheid, K., Hein, G., & von Cramon, D. Y. (2003). Syntactic processes in Parkinson's disease: a dissociation between early automatic and late integrational processes. *Neuropsychology*, *17*, 133-142.
- Friederici, A. D., Mecklinger, A., Spencer, K. M., Steinhauer, K., & Donchin, E. (2001). Syntactic parsing preferences and their on-line revisions: a spatio-temporal analysis of event-related brain potentials. *Cognitive Brain Research*, *11*, 305-323.
- Friederici, A. D., Rüschemeyer, S., Hahne, A., & Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cerebral Cortex*, *13*, 170-177.
- Friederici, A. D., Steinhauer, K., & Frisch, S. (1999). Lexical integration: Sequential effects of syntactic and semantic information. *Memory & Cognition*, *27*, 438-453.
- Friederici, A. D., Steinhauer, K., Mecklinger, A., & Meyer, M. (1998). Working memory constraints on syntactic ambiguity resolution as revealed by electrical brain responses. *Biological Psychology*, *47*, 193-221.

- Friederici, A. D., & Weissenborn, J. (2007). Mapping sentence form onto meaning: The syntax–semantic interface. *Brain Research, 1146*, 50-58.
- Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. *Journal of experimental psychology: General, 133*, 101-135.
- Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science, 7*, 237-241.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*, 1-76.
- Gibson, E. A. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In Y. Miyashita, A. Mirantz, & W. O’Neil (Eds.), *Image, language, brain* (pp. 95– 126). Cambridge, MA: MIT Press.
- Gibson, E., & Wu, H. H. I. (2013). Processing Chinese relative clauses in context. *Language and Cognitive Processes, 28*, 125-155.
- Glaser, Y. G., Martin, R. C., Van Dyke, J. A., Hamilton, A. C., & Tan, Y. (2013). Neural basis of semantic and syntactic interference in sentence comprehension. *Brain and Language, 126*, 314-326.
- Gong, Y. X. (1992). Manual of Wechsler Adult Intelligence Scale-Chinese version: Changsha: Chinese Map Press.

- Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1411.
- Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, 51, 97-114.
- Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 1304-1321.
- Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-Load Interference in Syntactic Processing. *Psychol Science*, 13, 425-430.
- Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95-112.
- Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science*, 29, 261-290.
- Gunter, T. C., Friederici, A. D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience*, 12, 556-568.
- Gunter, T. C., Wagner, S., & Friederici, A. D. (2003). Working memory and lexical ambiguity resolution as revealed by ERPs: A difficult case for activation theories. *Journal of Cognitive Neuroscience*, 15, 643-657.

- Hagoort, P., & Brown, C. M. (1994). Brain responses to lexical-ambiguity resolution and parsing. In C. Clifton Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 45– 80). Hillsdale, NJ: Erlbaum.
- Hagoort, P. (2003). Interplay between Syntax and Semantics during Sentence Comprehension: ERP Effects of Combining Syntactic and Semantic Violations. *Journal of Cognitive Neuroscience*, 15, 883-899.
- Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in Cognitive Science*, 9, 416-423.
- Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and cognitive processes*, 8, 439-483.
- Hagoort, P., Wassenaar, M., & Brown, C. M. (2003). Syntax-related ERP-effects in Dutch. *Cognitive Brain Research*, 16, 38-50.
- Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, 11, 194-205.
- Hahne, A., & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research*, 13, 339-356.
- Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. Paper presented at the Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies.

- Hambrick, D. Z., Kane, M. J., & Engle, R. W. (2005). The role of working memory in higher-level cognition. In R. Sternberg & J. E. Pretz (Eds.), *Cognition and intelligence: Identifying the mechanisms of the mind* (pp. 104–121). New York: Cambridge University Press.
- Hamilton, A. C., Martin, R. C., & Burton, P. C. (2009). Converging functional magnetic resonance imaging evidence for a role of the left inferior frontal lobe in semantic retention during language comprehension. *Cognitive Neuropsychology*, 26, 685-704.
- Heine, A., Tamm, S., Hofmann, M., Hutzler, F., & Jacobs, A. M. (2006). Does the frequency of the antecedent noun affect the resolution of pronominal anaphors?: An ERP study. *Neuroscience letters*, 400, 7-12.
- Hickok, G., Zurif, E., & Cansecogonzalez, E. (1993). Structural description of agrammatic comprehension. *Brain and Language*, 45, 371-395.
- Hochstadt, J. (2009). Set-shifting and the on-line processing of relative clauses in Parkinson's disease: results from a novel eye-tracking method. *Cortex*, 45, 991-1011.
- Hochstadt, J., Nakano, H., Lieberman, P., & Friedman, J. (2006). The roles of sequencing and verbal working memory in sentence comprehension deficits in Parkinson's disease. *Brain and Language*, 97, 243-257.
- Hsiao, F., & Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition*, 90, 3-27.

- Hsiao, F (2003). The syntax and processing of relative clauses in Mandarin Chinese. PhD dissertation, Department of Linguistics, MIT, Cambridge, MA. Available through MIT working papers in linguistics (MITWPL).
- Hughes, M. M., Linck, J. A., Bowles, A. R., Koeth, J. T., & Bunting, M. F. (2013). Alternatives to switch-cost scoring in the task-switching paradigm: Their reliability and increased validity. *Behavior Research Methods*, 46, 702-721.
- Isel, F., Hahne, A., Maess, B., & Friederici, A. D. (2007). Neurodynamics of sentence interpretation: ERP evidence from French. *Biological Psychology*, 74, 337-346.
- Ishizuka, T., Nakatani, K., & Gibson, E. (2006). Processing Japanese relative clauses in context. Paper presented at the 19th Annual CUNY Conference on Human Sentence Processing. New York: City University of New York.
- Jäger, L., Chen, Z., Li, Q., Lin, C.-J. C., & Vasishth, S. (2015). The subject-relative advantage in Chinese: Evidence for expectation-based processing. *Journal of Memory and Language*, 79, 97-120.
- Janssen, N., Bi, Y., & Caramazza, A. (2008). A tale of two frequencies: Determining the speed of lexical access for Mandarin Chinese and English compounds. *Language and Cognitive Processes*, 23, 1191-1223.
- Jiang, X., Li, Y., & Zhou, X. (2013). Even a rich man can afford that expensive house: ERP responses to construction-based pragmatic constraints during sentence comprehension. *Neuropsychologia*, 51, 1857-1866.

- Jung, T., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2001). Analysis and visualization of single-trial event-related potentials. *Human brain mapping, 14*, 166-185.
- Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., Mckeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology, 37*, 163-178.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122-149.
- Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111*, 228-238.
- Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15*, 159-201.
- Kan, I. P., & Thompson-Schill, S. L. (2004a). Effect of name agreement on prefrontal activity during overt and covert picture naming. *Cognitive, Affective, & Behavioral Neuroscience, 4*, 43-57.
- Kan, I. P., & Thompson-Schill, S. L. (2004b). Selection from perceptual and conceptual representations. *Cognitive, Affective, & Behavioral Neuroscience, 4*, 466-482.
- Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130*, 169-183.

- Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, 132, 47-70.
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, 133, 189-217.
- Kazmerski, V. A., Blasko, D. G., & Dessalegn, B. G. (2003). ERP and behavioral evidence of individual differences in metaphor comprehension. *Memory & Cognition*, 31, 673-689.
- Keenan, E. L., & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic inquiry*, 8, 63-99.
- Keil, A., Debener, S., Gratton, G., Junghöfer, M., Kappenman, E.S., Luck, S.J., Luu, P., Miller, G.A., Yee, C.M., 2013. Committee report: publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. *Psychophysiology*, 51, 1-21.
- Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52, 205-225.
- King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, 30, 580-602.

- King, J. W., & Kutas, M. (1995). Who did what and when? Using word-and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, 7, 376-395.
- Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, 5, 196-214.
- Kluender, R., Münte, T., Cowles, H. W., Szentkuti, A., Walenski, M., & Wieringa, B. M. (1998). Brain potentials to English and German questions. Paper presented at the Journal of Cognitive Neuroscience.
- Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, 29, 627-645.
- Krott, A., Baayen, R. H., & Hagoort, P. (2006). The nature of anterior negativities caused by misapplications of morphological rules. *Journal of Cognitive Neuroscience*, 18, 1616-1630.
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146, 23-49.
- Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., & Caplan, D. (2003). Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *Journal of Cognitive Neuroscience*, 15, 272-293.

- Kuperberg, G. R., Kreher, D., Blais, K., Caplan, D., & Holcomb, P. J. (2005). Semantic influences on syntactic processing: evidence from event-related potentials. *Journal of Cognitive Neuroscience*, 29.
- Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, 17, 117-129.
- Kuperman, V., & Van Dyke, J. A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language*, 65, 42-73.
- Kutas, M., & Hillyard, S. A. (1980). Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science*, 207, 203-205.
- Kutas, M., & Van Petten, C. K. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 83–143). New York: Academic Press.
- Law, S.-P., & Leung, M.-T. (1998). Sentence comprehension in Cantonese Chinese aphasic patients. *Aphasiology*, 12, 49-63.
- Law, S.-P., & Leung, M.-T. (2000). Sentence processing deficits in two Cantonese aphasic patients. *Brain and Language*, 72, 310-342.
- Leong, C. K., Tse, S. K., Loh, K. Y., & Hau, K. T. (2008). Text comprehension in Chinese children: Relative contribution of verbal working memory, pseudoword

reading, rapid automatized naming, and onset-rime phonological segmentation.

Journal of Educational Psychology, 100, 135-149.

Lewandowsky, S., & Oberauer, K. (2009). No evidence for temporal decay in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 1545-1551.

Lewandowsky, S., Oberauer, K., & Brown, G. D. A. (2009). No temporal decay in verbal short-term memory. *Trends in Cognitive Sciences*, 13, 120-126.

Lewis, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research*, 25, 93-115.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29, 375-419.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science*, 10, 447-454.

Li, C. N., & Thompson, S. A. (1989). Mandarin Chinese: A functional reference grammar. Berkeley: University of California Press.

Li, X., & Zhou, X. (2010). Who is *ziji*? ERP responses to the Chinese reflexive pronoun during sentence comprehension. *Brain Research*, 1331, 96-104.

- Lin, C.-J. C. (2006). Grammar and Parsing: A Typological Investigation of Relative-Clause Processing. PhD dissertation, Department of Linguistics and Anthropology, The University of Arizona.
- Lin, C.-J. C., & Bever, T. G. (2006). Subject preference in the processing of relative clauses in Chinese. Paper presented at the Proceedings of the 25th west coast conference on formal linguistics.
- Lin, Y., & Garnsey, S. (2011). Animacy and the resolution of temporary ambiguity in relative clause comprehension in Mandarin. In H. Yamashita, Y. Hirose, & J. Packard (Eds.), *Processing and producing head-final structures* (pp. 241–275). Dordrecht, Netherlands: Springer.
- Liotti, M., Woldorff, M. G., Perez III, R., & Mayberg, H. S. (2000). An ERP study of the temporal course of the Stroop color-word interference effect. *Neuropsychologia*, 38, 701-711.
- Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, 8, 1-14.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, 109, 35-54.
- MacWhinney, B. (2005). The emergence of grammar from perspective. In D. Pecher & R. A. Swann (Eds.), *Grounding cognition: The role of perception and action in*

memory, language, and thinking (pp. 198–223). New York, NY: Cambridge University Press.

Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language*, 47, 50-68.

Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., & Halgren, E. (2003). Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron*, 38, 487-497.

Martin, A. E., Nieuwland, M. S., & Carreiras, M. (2012). Event-related brain potentials index cue-based retrieval interference during sentence comprehension. *Neuroimage*, 59, 1859-1869.

Martin, A. E., Nieuwland, M. S., & Carreiras, M. (2014). Agreement attraction during comprehension of grammatical sentences: ERP evidence from ellipsis. *Brain and Language*, 135, 42-51.

Martin, R. C. (1987). Articulatory and phonological deficits in short-term memory and their relation to syntactic processing. *Brain and Language*, 32, 159-192.

Martin, R. C., & He, T. (2004). Semantic short-term memory and its role in sentence processing: A replication. *Brain and Language*, 89, 76-82.

Martin, R. C., & Romani, C. (1994). Verbal working memory and sentence comprehension: A multiple-components view. *Neuropsychology*, 8, 506-523.

- Martin, R. C., Shelton, J. R., & Yaffee, L. S. (1994). Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language*, 33, 83-111.
- Martin, R. C., Wu, D. H., Freedman, M. L., Jackson, E. F., & Lesch, M. (2003). An event-related fMRI investigation of phonological versus semantic short-term memory. *Journal of Neurolinguistics*, 16, 341-360.
- Mason, R. A., Just, M. A., Keller, T. A., & Carpenter, P. A. (2003). Ambiguity in the brain: what brain imaging reveals about the processing of syntactically ambiguous sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 1319-1338.
- Matzke, M., Mai, H., Nager, W., Rüsseler, J., & Münte, T. (2002). The costs of freedom: An ERP-study of non-canonical sentences. *Clinical Neurophysiology*, 113, 844-852.
- McBride-Chang, C., & Ho, C. S.-H. (2000). Developmental issues in Chinese children's character acquisition. *Journal of Educational Psychology*, 92, 50-55.
- McElree, B. (2006). Accessing recent events. *Psychology of Learning and Motivation*, 46, 155-200.
- McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, 48, 67-91.
- McEvoy, L. K., Smith, M. E., & Gevins, A. (2000). Test-retest reliability of cognitive EEG. *Clinical Neurophysiology*, 111, 457-463.

- McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General*, *141*, 302-320.
- Mecklinger, A., Schriefers, H., Steinhauer, K., & Friederici, A. D. (1995). Processing relative clauses varying on syntactic and semantic dimensions: An analysis with event-related potentials. *Memory & Cognition*, *23*, 477-494.
- Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: an individual differences study. *Language Learning*, *62*, 302-331.
- Monsell, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology*, *10*, 465-501.
- Moore, M., & Gordon, P. C. (2014). Reading ability and print exposure: item response theory analysis of the author recognition test. *Behavior Research Methods*, DOI 10.3758/s13428-014-0534-3.
- Moro, A., Tettamanti, M., Perani, D., Donati, C., Cappa, S. F., & Fazio, F. (2001). Syntax and the brain: disentangling grammar by selective anomalies. *Neuroimage*, *13*, 110-118.
- Müller, H. M., King, J. W., & Kutas, M. (1997). Event-related potentials elicited by spoken relative clauses. *Cognitive Brain Research*, *5*, 193-203.
- Münté, T., Matzke, M., & Johannes, S. (1997). Brain activity associated with syntactic incongruencies in words and pseudo-words. *Journal of Cognitive Neuroscience*, *9*, 318-329.

- Münter, T. F., Heinze, H., & Mangun, G. R. (1993). Dissociation of brain activity related to syntactic and semantic aspects of language. *Journal of Cognitive Neuroscience*, 5, 335-344.
- Münter, T. F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, 395, 71-73.
- Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, 18, 251-269.
- Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology*, 53, 53-81.
- Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, 3, 151-165.
- Nieuwland, M. S., Otten, M., & Van Berkum, J. J. (2007). Who are you talking about? Tracking discourse-level referential processing with event-related brain potentials. *Journal of Cognitive Neuroscience*, 19, 228-236.
- Nieuwland, M. S., & Van Berkum, J. J. (2006). Individual differences and contextual bias in pronoun resolution: evidence from ERPs. *Brain Research*, 1118, 155-167.
- Nieuwland, M. S., & Van Berkum, J. J. A. (2008). The neurocognition of referential ambiguity in language comprehension. *Language and Linguistics Compass*, 2, 603-630.

- Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, 5, 263-281.
- Nunnally, J. C., Bernstein, I. H., & Berge, J. M. F. t. (1967). Psychometric theory (Vol. 226): McGraw-Hill New York.
- Oberauer, K. (2002). Access to information in working memory: exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 411-421.
- Oberauer, K. (2009). Interference between storage and processing in working memory: Feature overwriting, not similarity-based competition. *Memory & Cognition*, 37, 346-357.
- Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, 55, 601-626.
- Oberauer, K., & Lange, E. B. (2008). Interference in verbal working memory: Distinguishing similarity-based confusion, feature overwriting, and feature migration. *Journal of Memory and Language*, 58, 730-745.
- Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785-806.
- Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, 34, 739-773.

- Osterhout, L., & Nicol, J. (1999). On the Distinctiveness, Independence, and Time Course of the Brain Responses to Syntactic and Semantic Anomalies. *Language and Cognitive Processes*, 14, 283-317.
- Ozeki, H., & Shirai, Y. (2007). Does the noun phrase accessibility hierarchy predict the difficulty order in the acquisition of Japanese relative clauses? *Studies in Second Language Acquisition*, 29, 169-196.
- Öztekin, L., & McElree, B. (2010). Relationship between measures of working memory capacity and the time course of short-term memory retrieval and interference resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 383-397.
- Packard, J. L., Ye, Z., & Zhou, X. (2011). Filler-gap processing in Mandarin relative clauses: Evidence from event-related potentials Processing and producing head-final structures (pp. 219-240): Springer.
- Packard, Jerome L., Zheng Ye, and Xiaolin Zhou. 2011. Filler-gap processing in Mandarin relative clauses: Evidence from event-related potentials. In H. Yamashita, Y. Hirose, & J. Packard (Eds.), *Processing and producing head-final structures* (pp. 219–240). Dordrecht, Netherlands: Springer.
- Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11, 357-383.

- Phillips, C., Wagers, M., & Lau, E. (2010). Grammatical illusions and selective fallibility in real-time language comprehension. In J. Runner (Ed.), *Experiments at the interfaces. Syntax & Semantics* (Vol. 37). Bingley, UK: Emerald Publications.
- Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., . . . Rugg, M. D. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*, 127-152.
- Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*, 2128-2148.
- Polich, J., & Donchin, E. (1988). P300 and the word frequency effect. *Electroencephalography and Clinical Neurophysiology*, *70*, 33-45.
- Politzer-Ahles, S., Fiorentino, R., Jiang, X., & Zhou, X. (2013). Distinct neural correlates for pragmatic and semantic meaning processing: An event-related potential investigation of scalar implicature processing using picture-sentence verification. *Brain research*, *1490*, 134-152.
- Prat, C. S., & Just, M. A. (2011). Exploring the neural dynamics underpinning individual differences in sentence comprehension. *Cerebral Cortex*, *21*, 1747-1760.
- Prat, C. S., Keller, T. A., & Just, M. A. (2007). Individual Differences in Sentence Comprehension: A Functional Magnetic Resonance Imaging Investigation of Syntactic and Lexical Processing Demands. *Journal of Cognitive Neuroscience*, *19*, 1950-1963.

- Prat, C. S., Mason, R. A., & Just, M. A. (2012). An fMRI investigation of analogical mapping in metaphor comprehension: The influence of context and individual cognitive capacities on processing demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 282-294.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372-422.
- Real, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, 57, 1-23.
- Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, 28, 164-171.
- Reichle, E. D., Carpenter, P. A., & Just, M. A. (2000). The neural basis of strategy and skill in sentence-picture verification. *Cognitive Psychology*, 40, 261-295.
- Rodd, J. M., Longe, O. A., Randall, B., & Tyler, L. K. (2010). The functional organisation of the fronto-temporal language system: evidence from syntactic and semantic ambiguity. *Neuropsychologia*, 48, 1324-1335.
- Ruchkin, D. S., Johnson, R., Canoune, H., & Ritter, W. (1990). Short-term memory storage and retention: An event-related brain potential study. *Electroencephalography and Clinical Neurophysiology*, 76, 419-439.
- Ryan, T. A. (1959). Multiple comparison in psychological research. *Psychological Bulletin*, 56, 26-47.

- Saffran, E. M., & Martin, N. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology*, 14, 641-682.
- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, 103, 403-428.
- Saults, J. S., & Cowan, N. (2007). A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *Journal of Experimental Psychology: General*, 136, 663-684.
- Schriefers, H., Friederici, A. D., & Kuhn, K. (1995). The processing of locally ambiguous relative clauses in German. *Journal of Memory and Language*, 34, 499-520.
- Service, Helenius, P., Maury, S., & Salmelin, R. (2007). Localization of syntactic and semantic brain responses using magnetoencephalography. *Journal of Cognitive Neuroscience*, 19, 1193-1205.
- Sharbrough, F., Chatrian, G., Lesser, R. P., Lüders, H., Nuwer, M., & Picton, T. W. (1990). Guidelines for standard electrode position nomenclature American EEG Society: Bloomfield.
- Shipstead, Z., & Engle, R. W. (2013). Interference within the focus of attention: Working memory tasks reflect more than temporary maintenance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 277-289.

- Shipstead, Z., Lindsey, D. R. B., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language*, 72, 116-141.
- Shivde, G., & Thompson-Schill, S. L. (2004). Dissociating semantic and phonological maintenance using fMRI. *Cognitive, Affective, & Behavioral Neuroscience*, 4, 10-19.
- St. George, M., Mannes, S., & Hoffman, J. E. (1997). Individual differences in inference generation: An ERP analysis. *Journal of Cognitive Neuroscience*, 9, 776-787.
- Stanovich, K. E., & West, R. F. (1989). Exposure to Print and Orthographic Processing. *Reading Research Quarterly*, 24, 402-433.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, 18, 643-662.
- Su, Y. C., Lee, S. E., & Chung, Y. M. (2007). Asyntactic thematic role assignment by Mandarin aphasics: a test of the Trace-Deletion Hypothesis and the Double Dependency Hypothesis. *Brain and Language*, 101, 1-18.
- Tan, L., Spinks, J. A., Eden, G. F., Perfetti, C. A., & Siok, W. T. (2005). Reading depends on writing, in Chinese. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 8781-8785.
- Tan, L., & Perfetti, C. A. (1998). Phonological codes as early sources of constraint in Chinese word identification: A review of current discoveries and theoretical

accounts Cognitive processing of the Chinese and the Japanese languages. *Reading and Writing*, 10, 165-200.

Tanenhaus, M. K., & Trueswell, John C. (1995). Sentence comprehension. In Miller, J. L. & Eimas, P. D. (Ed). *Speech, language, and communication. Handbook of perception and cognition* (2nd ed.), Vol. 11, (pp. 217-262). San Diego, CA, US: Academic Press.

Thompson-Schill, S. L., Bedny, M., & Goldberg, R. F. (2005). The frontal lobes and the regulation of mental activity. *Current opinion in neurobiology*, 15, 219-224.

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, 47, 69-90.

Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, S. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language*, 53, 204-224.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (2002). Semantic influences on parsing. In Altmann, G. T., M. (Ed). *Psycholinguistics: critical concepts in psychology*. Vol. 3, (pp. 172-216). Routledge, New York.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127-154.

Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience*, 2, 717-726.

- Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*, 104-132.
- Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, *71*, 1-26.
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498-505.
- Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, *17*, 635-654.
- Van Berkum, J. J., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, *41*, 147-182.
- Van Berkum, J. J., Brown, C. M., Hagoort, P., & Zwitserlood, P. (2003). Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology*, *40*, 235-248.
- Van Berkum, J.J.A., Zwitserlood, P., Bastiaansen, M.C.M., Brown, C.M., Hagoort, P., 2004. So who's "he" anyway? Differential ERP and ERSP effects of referential

success, ambiguity and failure during spoken language comprehension. Annual meeting of the Cognitive Neuroscience Society (CNS-2004), San Francisco, April 18–20.

Van Berkum, J. J., Koornneef, A. W., Otten, M., & Nieuwland, M. S. (2007).

Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research*, 1146, 158-171.

Van Berkum, J. J. A., Zwitserlood, P., Bastiaansen, M. C. M., Brown, C. M., & Hagoort, P. (2004). So who's "he" anyway? Differential ERP and ERSP effects of referential success, ambiguity and failure during spoken language comprehension. Supplement to the Journal of Cognitive Neuroscience, 16, 70.

van de Meerendonk, N., Kolk, H. H. J., Vissers, C. T. W. M., & Chwilla, D. J. (2010).

Monitoring in language perception: mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, 22, 67-82.

Van Dyke, J. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 407-430.

Van Dyke, J., & Johns, C. L. (2012). Memory interference as a determinant of language comprehension. *Language and linguistics compass*, 6, 193-211.

Van Dyke, J., Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition*, 131, 373-403.

- Van Dyke, J., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, 49, 285-316.
- Van Dyke, J., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, 55, 157-166.
- Van Dyke, J., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, 65, 247-263.
- Van Petten, C., Weckerly, J., McIsaac, H. K., & Kutas, M. (1997). Working memory capacity dissociates lexical and sentential context effects. *Psychological Science*, 8, 238-242.
- Vasishth, S., Chen, Z., Li, Q., & Guo, G. (2013). Processing Chinese Relative Clauses: Evidence for the Subject-Relative Advantage. *PloS one*, 8, e77006.
- Vasishth, S., & Lewis, R. L. (2006). Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language*, 82, 767-794.
- Vos, S. H., Gunter, T. C., Schriefers, H., & Friederici, A. D. (2001). Syntactic parsing and working memory: The effects of syntactic complexity, reading span, and concurrent load. *Language and Cognitive Processes*, 16, 65-103.
- Vuong, L. C., & Martin, R. C. (2011). LIFG-based attentional control and the resolution of lexical ambiguities in sentence context. *Brain and Language*, 116, 22-32.

- Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61, 206-237.
- Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, 85, 79-112.
- Waters, G. S., & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, & Computers*, 35, 550-564.
- Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, 1, 442-452.
- Weckerly, J., & Kutas, M. (1999). An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology*, 36, 559-570.
- Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, 58, 250-271.
- Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594-604.
- Yang, C. L., Perfetti, C. A., & Liu, Y. (2010). Sentence integration processes: an ERP study of Chinese sentence comprehension with relative clauses. *Brain and Language*, 112, 85-100.

- Ye, Z., Luo, Y. J., Friederici, A. D., & Zhou, X. (2006). Semantic and syntactic processing in Chinese sentence comprehension: evidence from event-related potentials. *Brain Research, 1071*, 186-196.
- Ye, Z., & Zhou, X. (2008). Involvement of cognitive control in sentence comprehension: evidence from ERPs. *Brain Research, 1203*, 103-115.
- Ye, Z., & Zhou, X. (2009a). Conflict control during sentence comprehension: fMRI evidence. *Neuroimage, 48*, 280-290.
- Ye, Z., & Zhou, X. (2009b). Executive control in language processing. *Neuroscience & Biobehavioral Reviews, 33*, 1168-1177.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of memory and language, 46*, 441-517.
- Yu, J., & Zhang, Y. (2008). When Chinese semantics meets failed syntax. *Neuroreport, 19*, 745-749.
- Zhang, Y., Yu, J., & Boland, J. E. (2010). Semantics does not need a processing license from syntax in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 765-781.
- Zhou, X., Marslen-Wilson, W., Taft, M., & Shu, H. (1999). Morphology, orthography, and phonology reading Chinese compound words. *Language and Cognitive Processes, 14*, 525-565.